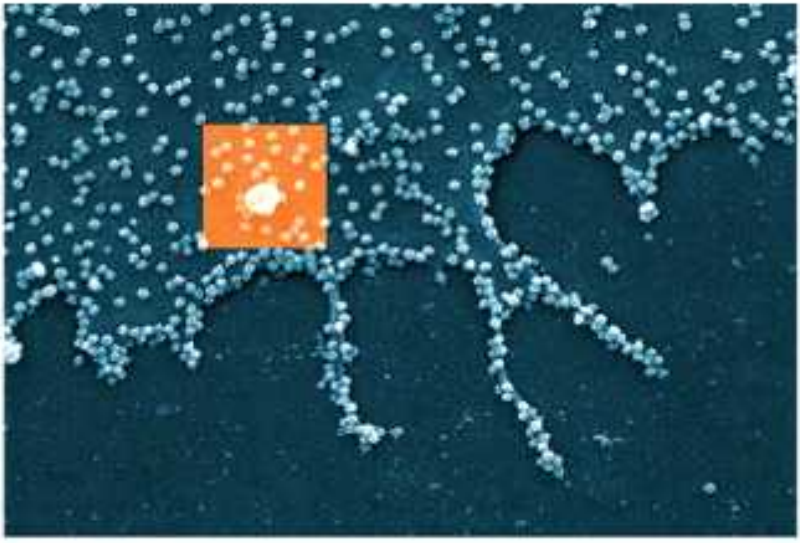
A CLINICAL APPROACH TO



MEDICINE

Second Edition

editors

Yong Yau ONG, Keng Thye WOO
Han Seong NG, Patrick TAN &
Ong Teng TANG

A CLINICAL APPROACH TO MEDICINE

Second Edition

This page intentionally left blank

A CLINICAL APPROACH TO MEDICINE

Second Edition

editors

Yong Yau ONG, Keng Thye WOO
Han Seong NG, Patrick TAN &
Ong Teng TANG

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

A CLINICAL APPROACH TO MEDICINE
Second Edition

Copyright © 2005 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-238-926-1
ISBN 981-256-073-4 (pbk)

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore.

Contents

List of Contributors

A Vathsala
Senior Consultant
Department of Renal Medicine
Singapore General Hospital
Singapore

Alexander P Auchus
Senior Consultant
Department of Neurology
SGH-NNI
Singapore

Amy Ng Suah Hoon
Senior Consultant
Department of Cardiology
National Heart Centre
Singapore

Andrew Beng-Siong Pan
Consultant
Department of Neurology
SGH-NNI
Singapore

Anne Hsu
Senior Consultant
Department of Respiratory & CCM
Singapore General Hospital
Singapore

Bernard Kwok Wing Kuin
Consultant
Department of Cardiology
National Heart Centre
Singapore

Bernard Thong Yu Hor
Consultant
Department of Rheumatology, Allergy & Immunology
Tan Tock Seng Hospital
Singapore

C Rajasoorya
Chairman, Medical Board
Senior Consultant
Department of Medicine
Alexandra Hospital
Singapore

Chang Hui Meng
Senior Consultant
Department of Neurology
SGH-NNI
Singapore

Charles Chuah Thuan Heng
Associate Consultant
Department of Haematology
Singapore General Hospital
Singapore

Charles Chuah
Associate Consultant
Department of Haematology
Singapore General Hospital
Singapore

Cheng Yew Kuang
Associate Consultant
Department of Rheumatology, Allergy & Immunology
Tan Tock Seng Hospital
Singapore

Chng Hiok Hee
Senior Consultant & Head
Department of Rheumatology, Allergy & Immunology
Tan Tock Seng Hospital
Singapore

Choong Hui Lin
Senior Consultant
Department of Renal Medicine
Singapore General Hospital
Singapore

Chow Wan Cheng
Senior Consultant & Head
Department of Gastroenterology
Singapore General Hospital
Singapore

Christopher Chen
Senior Consultant
Department of Neurology
SGH-NNI
Singapore

Colleen Kim Thomas
Consultant
Department of Internal Medicine
Singapore General Hospital
Singapore

Constance Lo
Consultant
Department of Respiratory & Critical Care Medicine
Singapore General Hospital
Singapore

Daphne Khoo
Senior Consultant and
Acting Head
Department of Endocrinology
Singapore General Hospital
Singapore

Dato BA Johan
Senior Consultant
Department of Cardiology
National Heart Centre
Singapore

Ding Yew Yoong
Senior Consultant & Head
Department of Geriatric Medicine
Tan Tock Seng Hospital
Singapore

Ding Zee Pin
Senior Consultant
Department of Cardiology
National Heart Centre
Singapore

Dr Khoo Kei Siong
Visiting Senior Consultant and
Former Head
Department of Medical Oncology
National Cancer Centre
Singapore

Wong Nan Soon
Associate Consultant
Department of Medical
Oncology
National Cancer Centre
Singapore

Einar Wilder–Smith
Senior Consultant, Director
Research Department of
Medicine
Coordinator, Neurology
Diagnostic Laboratory
Clinical Associate Professor NUS
Singapore

Evelyn Wong
Consultant
Department of Emergency
Medicine
Singapore General Hospital
Clinical Lecturer, NUS
Singapore

Fatimah Lateef
Consultant
Department of Emergency
Medicine
Singapore General Hospital
Singapore

Feng Pao Hsii
Emeritus Consultant
Department of Rheumatology,
Allergy & Immunology
Tan Tock Seng Hospital
Singapore

Fong Kok Yong
Senior Consultant & Head
Department of Rheumatology and
Immunology
Singapore General Hospital
Clinical Associate Professor, NUS
Singapore

Gerald Teoh Kheng Hoe
Consultant & Clinical Scientist
Department of Haematology
Singapore General Hospital
SingHealth Research Division
Principal Investigator, Multiple Myeloma Research Laboratory
Singapore

Goh Su-Yen
Registrar
Department of Endocrinology
Singapore General Hospital
Singapore

Goh Yeow Tee
Senior Consultant
Department of Haematology
Singapore General Hospital
Singapore

Ho Lai Yun
Senior Consultant
Department of Neonatal & Developmental Medicine
Singapore General Hospital
Clinical Associate Professor, NUS
Singapore

Ho Su Chin
Consultant & Clinician Scientist
Department of Endocrinology
Singapore General Hospital
SingHealth Research Division
Singapore

Hsu Li Yang
Registrar
Department of Internal Medicine
Infectious Diseases Unit
Singapore General Hospital
Singapore

Humeira Badsha
Consultant
Department of Rheumatology, Allergy & Immunology
Tan Tock Seng Hospital
Singapore

J Raghuram
Consultant
Department of Respiratory & Critical Care Medicine
Singapore General Hospital
Singapore

John Chiam
Registrar
Department of Internal Medicine
Singapore General Hospital
Singapore

John Thomas
Senior Consultant & Head
Department of Neurosurgery
SGH-NNI
Singapore

Julian Thumboo
Consultant
Department of Rheumatology and Immunology
Singapore General Hospital
Associate Professor
Department of Medicine
National University of Singapore
Singapore

Julian Wei-Tze Tang
Clinical Virology Specialist
Registrar/Lecturer
Centre for Virology
Division of Infection and Immunity
Royal Free and University College Medical School
Windeyer Institute for Medical Sciences
London

K Puvanendran
Senior Consultant
Department of Neurology
SGH-NNI
Singapore

Koh Tian Hai
Medical Director
National Heart Centre
Clinical Professor, NUS
Singapore

Koh Wei Howe
Consultant Rheumatologist and Physician
Koh Wei Howe Arthritis & Rheumatism Medical Clinic
Mount Elizabeth Medical Centre
Singapore

Lau Kean Wah
Senior Consultant
Department of Cardiology
National Heart Centre
Clinical Associate Professor, NUS
Singapore

Lee Lai Heng
Senior Consultant & Head
Department of Haematology
Singapore General Hospital
Singapore

Lee Siew Luan, Grace
Visiting Consultant
Department of Renal Medicine
Singapore General Hospital
Singapore

Leng Poh Hock
Associate Consultant
Department of Respiratory and Critical Care Medicine
Singapore General Hospital
Singapore

Leonard Koh
Senior Consultant
Department of Endocrinology
Singapore General Hospital
Singapore

Leong Hoe Nam
Registrar
Department of Internal Medicine
Infectious Diseases Unit
Singapore General Hospital
Singapore

Leong Khai Pang
Senior Consultant
Department of Rheumatology, Allergy & Immunology
Tan Tock Seng Hospital
Singapore

Lim Heok Seng
Senior Consultant
Department of Endocrinology
Singapore General Hospital
Singapore

Lim Lay Cheng
Senior Consultant
Department of Haematology
Singapore General Hospital
Singapore

Lim Ling Choo
Registrar
Department of Medicine
Alexandra Hospital
Singapore

Lim Shih-Hui
Senior Consultant
Department of Neurology
SGH-NNI
Singapore

Lim Soo Teik
Senior Consultant
Department of Cardiology
National Heart Centre
Singapore

Lim Soon Thye
Associate Consultant
Department of Medical Oncology
National Cancer Centre
Singapore

Lim Swee Han
Senior Consultant & Head
Department of Emergency Medicine
Singapore General Hospital
Singapore

Ling Moi Lin
Director, Quality Management
Infection Control Officer
Senior Consultant
Department of Pathology
Singapore General Hospital
Singapore

Linn Yeh Ching
Consultant
Department of Haematology
Singapore General Hospital
Singapore

Lo Yew Long
Consultant
Department of Neurology
SGH-NNI
Singapore

Loo Chian Min
Consultant
Department of Respiratory and
Critical Care Medicine
Singapore General Hospital
Singapore

Lui Hock Foong
Consultant
Department of Gastroenterology
Singapore General Hospital
Singapore

Mak Koon Hou
Senior Consultant
Department of Cardiology

National Heart Centre
Singapore

Ng Han Seong
Chairman, Division of Medicine
Senior Consultant
Department of Gastroenterology
Singapore General Hospital
Singapore

Ng Heng Joo
Associate Consultant
Department of Haematology
Singapore General Hospital
Singapore

Nicholas Paton
Consultant
Director, Clinical Research Unit
Tan Tock Seng Hospital
Clinical Associate Professor NUS
Singapore

Ooi Choon Jin
Consultant
Department of Gastroenterology
Singapore General Hospital
Singapore

Pang Shiu Ming
Consultant
Dermatology Unit
Singapore General Hospital
Singapore

Patrick Tan
Senior Consultant
Haematology & Stem Cell
Transplant Centre
Mount Elizabeth Hospital
Singapore

Pavanni Ratnagopal
Head & Senior Consultant
Department of Neurology
SGH-NNI
Singapore

Peter AC Lim
Clinical Associate Professor
Senior Consultant & Head
Department of Rehabilitation
Medicine
Singapore General Hospital
Singapore

Department of Physical Medicine
& Rehabilitation
Baylor College of Medicine
Houston, Texas

Peter Eng
Consultant
Department of Endocrinology
Singapore General Hospital
Singapore

Philip Choo
Chairman, Medical Board
Senior Consultant
Department of Geriatric Medicine
Tan Tock Seng Hospital
Singapore

Philip Eng
Senior Consultant & Head
Department of Respiratory &
Critical Care Medicine

Clinical Associate Professor, NUS
Singapore General Hospital
Singapore

R Ponampalam
Consultant
Department of Emergency Medicine
Singapore General Hospital
Singapore

Richard Epstein
Department of Medicine
The University of Hong Kong
Queen Mary Hospital
Clinical Professor
Hong Kong

Ruth Kam
Senior Consultant
Department of Cardiology
National Health Centre
Singapore

Sandeep Kumar Rajan
Senior Consultant
Department of Medical Oncology
National Cancer Centre
Singapore

See Hui Ti
Associate Consultant
Department of Medical Oncology
National Cancer Centre
Singapore

Soh Lay Tin
Senior Consultant
Department of Medical Oncology
National Cancer Centre
Singapore

Steven J Mesenas
Consultant
Department of Gastroenterology
Singapore General Hospital
Singapore

T Thirumoorthy
Director & Senior Consultant
Dermatology Unit
Singapore General Hospital
Singapore

Tai E Shyong
Consultant & Clinician Scientist
Department of Endocrinology
Singapore General Hospital
SingHealth Research Division
Singapore

Tan Ban Hock
Head & Consultant
Department of Internal Medicine
Infectious Diseases Unit
Singapore General Hospital
Singapore

Tan Chee Eng
Senior Consultant and Head
Department of Endocrinology
Singapore General Hospital
Singapore

Tan Chee Kiat
Senior Consultant
Department of Gastroenterology
Singapore General Hospital
Singapore

Tan Eng King
Consultant & Clinician Scientist

Department of Neurology
SingHealth Research Division
Singapore

Tan Hak Koon
Consultant
Department of Obstretrics & Gynaecology
Singapore General Hospital
Singapore

Tan Hwee Huan
Visiting Consultant
Department of Endocrinology
Singapore General Hospital
Director of Clinical Projects
SingHealth Polyclinics
Singapore

Tan Keng Leong
Consultant
Department of Respiratory and Critical Care Medicine
Singapore General Hospital
Singapore

Tan Ru-San
Consultant
Department of Cardiology
National Heart Centre
Singapore

Tang Ong Teng
Senior Consultant
Department of Internal Medicine
Singapore General Hospital
Singapore

Teo Wee Siong
Senior Consultant
Department of Cardiology
National Health Centre
Singapore

Udaya Seneviratne
Registrar
Department of Neurology
SGH-NNI
Singapore

V Anantharaman
Chairman
Division of Ambulatory and Critical Care Medicine
Senior Consultant
Department of Emergency Medicine
Singapore General Hospital
Clinical Associate Professor, NUS
Singapore

Widjaja Luman
Senior Consultant
Department of Gastroenterology
Singapore General Hospital
Singapore

William Hwang Ying Kee
Consultant & Clinical Scientist
Department of Haematology
Singapore General Hospital
SingHealth Research Division
Medical Director, Singapore Cord Blood Bank
Singapore

Wong Gee Chuan
Consultant
Department of Haematology

Singapore General Hospital
Singapore

Wong Kok Seng
Senior Consultant & Head
Department of Renal Medicine
Singapore General Hospital
Singapore

Wong Meng Cheong
Senior Consultant
Department of Neurology
Associate Clinical Professor
SGH-NNI
Singapore

Woo Keng Thye
Emeritus Consultant and
Advisor
Department of Renal Medicine
Singapore General Hospital
Clinical Professor, NUS
Singapore

Yap Chin Kong
Senior Consultant
Gastroenterologist
Gastroenterology and Medical
Clinic
Mt Elizabeth Medical Centre
Singapore

Yean Leng Lim
Professor and Director of
Cardiovascular Therapeutics
Western Health
University of Melbourne
Visiting Consultant, NUH
Clinical Professor, NUS
Singapore

Preface

This 2nd edition follows a successful publication of our first, which was well received.

As was previously the case with the previous edition, this book was written with the purpose of providing a basic text for practising clinicians, general practitioners, physicians in training and medical students in the Asia Pacific region. We have added new chapters and recent information to the book and have welcomed new contributors.

This edition has been extensively revised and has grown in size. For example, it now contains a new chapter on S.A.R.S.

We have tried to provide more local information. The publication is not exhaustive, however, and for a more comprehensive review, the reader is referred to other larger standard texts available.

We wish to express our thanks to the many individuals, authors, editors, publishers and the administrative staff of Division of Medicine, Singapore General Hospital who have made it possible to complete this 2nd edition.

We acknowledge the late Professor Seah Cheng Siang a brilliant physician and a man of vision who taught many generations of doctors, including the many who had contributed to this present edition.

The Editors
2004

Cardiology

This page intentionally left blank

1

The Approach to the Patient with Chest Pain

Lim Soo Teik and Koh Tian Hai

INTRODUCTION

Chest discomfort is a common symptom , most often caused by benign conditions, but occasionally may be due to life-threatening medical emergencies. The approach to chest pain, therefore, is to exclude the benign conditions, and to rapidly identify and treat the potentially fatal and serious conditions in a "fast track" at every level of healthcare.

Chest pain is a common presentation of cardiac disease, but it can also be caused by conditions affecting organs throughout the thorax and abdomen.

MEDICAL CONDITIONS PRESENTING WITH CHEST PAIN

Serious Conditions	Less Serious Conditions
• acute coronary syndrome	• cardiac — pericarditis
• aortic dissection	• pleuritis
• pulmonary embolism	• musculoskeletal

Serious Conditions	Less Serious Conditions
• spontaneous pneumothorax	• GI — reflux esophagitis, esophageal spasm, cholecystitis, peptic ulcer, pancreatitis • early herpes zoster • psychiatric

In evaluating a patient presenting with chest discomfort, the following sequence of questions are helpful:

Q1: Could the chest discomfort be due to an acute, potentially life-threatening condition that warrants immediate hospitalization and aggressive evaluation and treatment? These conditions include acute coronary syndrome, aortic dissection, pulmonary embolism and spontaneous pneumothorax.

Q2: If not, could the discomfort be due to a chronic condition likely to lead to serious complications? Examples include stable angina, pulmonary hypertension, aortic stenosis.

Q3: If not, could the discomfort be due to an acute condition that warrants specific treatment? These include pericarditis, pneumonitis/pleuritis, herpes zoster.

Q4: If not, could the discomfort be due to another treatable chronic condition? Conditions that are not immediately life-threatening include esophageal reflux, esophageal spasm, peptic ulcer disease, gallbladder disease, costochondritis and other musculoskeletal disorders.

SYMPTOM EVALUATION

The history should include questions about the *quality, location, nature of onset, duration,* and *associated features* of the discomfort. Typical features of various types of chest pain are summarized below.

1) Ischemic cardiac pain
   - Pain often diffuses over a wide area of the anterior chest wall and is not localized. Pain that is experienced only at a peripheral site in the chest is rarely of cardiac origin.
   - Pain may radiate to left +/− right arm, jaw, neck and back.
   - Often described as "dull," "constricting," "pressing," "heavy feeling," "tearing," "tightness," "terrifying," and sometimes

"intolerable." However, the severity of the symptom is highly variable, while symptom is poorly correlated with risk of developing serious complications.

- Stable anginal pain is precipitated by exertion and is relieved by resting. It is frequently made worse by large meals or a cold wind. With unstable angina, similar pain may be brought on by minimal exertion and may also occur at rest. In contrast, pain associated with a specific movement (turning, stretching, bending, coughing) is likely to be musculoskeletal in origin.
- Myocardial infarction pain usually takes several minutes, or longer, to develop. In contrast, the pain of aortic dissection, massive pulmonary embolism or of pneumothorax is usually very sudden in onset. Patient may have nausea and vomiting, and often appears pale, diaphoretic, and is cold to the touch due to associated autonomic nervous system stimulation. Patient shows no response to posture, movement, food.

2) Reflux esophagitis and esophageal spasm
   - A common cause of chest pain, presenting with heart burn. The pain can mimic that of angina very closely. It is sometimes precipitated by exercise and may be relieved by nitrates.
   - Worse in recumbent position, also during strain. Sometimes pain is related to food or drink intake.
   - No ECG changes.
3) Pulmonary embolism
   - Most common symptom is dyspnea. Patient may have visceral chest pain believed to be due to distension of the pulmonary artery or pulmonary infarction. Onset is usually sudden if massive.
   - Tachypnea, hypoxemia, hypercarbia; with no pulmonary congestion on chest X-ray. Tachycardia is usually present.
   - May resemble inferior wall infarct: ST elevation in II, III, aVF. Sinus tachycardia common. Rare to have $S_I Q_{III} T_{III}$.
4) Pneumothorax
   - Dyspnea is the most common symptom.
   - One-sided chest pain and related to respiratory movement.
   - Auscultation and chest X-ray often clinch the diagnosis.
5) Aortic dissection
   - Pain is severe, sharp and tearing, with changing localization.

- Type A dissection may cause unequal upper limb pulses, new aortic regurgitation, and may involve coronary ostium, usually the right coronary artery, causing inferoposterior infarction.
- Sometimes widened mediastinum on chest X-ray.

6) Musculoskeletal and costochondral pain
   - Costochondral and chondrosternal syndromes are the most common causes of anterior chest musculoskeletal discomfort.
   - Very variable in site and intensity, usually accompanied by local tenderness.
   - Posture or movement of chest influences the pain.
7) Pleuritis
   - "Jabbing," "sharp," and "catching" pain on breathing, coughing or movement. Pain is usually unilateral, and often localized.
8) Peptic ulcer, cholecystitis, pancreatitis
   - Clinical examination.
   - Inferior wall ischemia sometimes may resemble acute abdomen.

Clinicians frequently employ "therapeutic trials" with sublingual nitroglycerin or antacids, and a common error is to assume that a response to either of these interventions clarifies the diagnosis. While such information is helpful, the patient's response may be due to placebo effect. Response to antacid, or failure of nitroglycerin to relieve chest pain, do not exclude the diagnosis of coronary disease.

Figure 1 presents an algorithm for the diagnosis of chest pain.

PHYSICAL EXAMINATION

Careful history taking is essential before proceeding to physical examination, as the physical examination can be apparently normal even with underlying serious condition.

1) Patient's respiratory and hemodynamic status must first be assessed. If either of these is compromised, initial management should focus on stabilizing the patient before pursuing diagnostic evaluation.
2) Physical examination should also include:
   a) evaluation of BP in both arms and pulses in both legs;
   b) cardiac murmurs; third and fourth heart sounds; pericardial rub;
   c) intensity of breath sounds; pleural rub; evidence of pneumothorax, pulmonary embolism, pneumonia, or pleurisy.

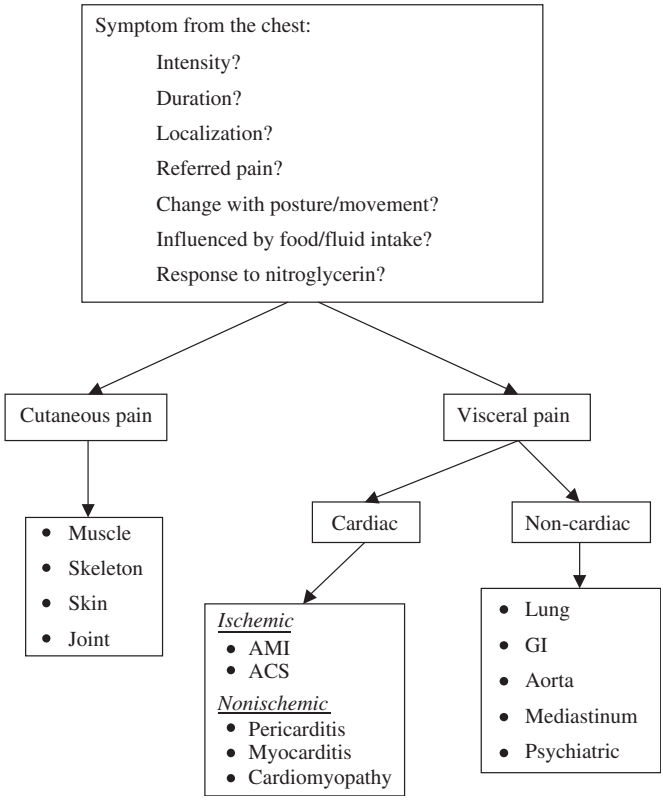


Fig. 1 Algorithm for diagnosis of chest pain.

The physical features of myocardial infarction are those of the accompanying autonomic disturbances (sweating, pallor) or those of heart failure (pulmonary congestion or edema, fourth heart sound), or diminished output (cold periphery). The patient looks systemically "ill," although the absence of such signs does not exclude the diagnosis. Similar autonomic disturbance may accompany any cause of severe pain.

DIAGNOSTIC TESTS IN ACUTE CHEST PAIN

The tests serve: 1) to determine the diagnosis; 2) to quickly identify the high risk patients for the fast track, and 3) to delineate patients with little or no risk of having life-threatening conditions. It must be recognized that

although cardiac investigations may be specific when abnormal, lack of sensitivity means that a normal or nonspecific result does not exclude the diagnosis.

Electrocardiogram

- Essential first screening test for adults with chest pain, aimed mainly to identify patient with myocardial ischemia.
- New ST elevation is sensitive and specific for myocardial infarction. Usually appears within minutes after symptom onset. However, this is present on initial admission ECG in only ~30–40% of hospitalized patients with AMI.
- ST depression indicates ischemia, but has poor power to identify ongoing myocardial infarction (only ~50% of patients with ST depression develop MI).
- T-wave inversion is nonspecific. Multiple differential diagnoses. Only about one-third of patients with chest pain and T-wave inversion on admission ECG develops MI.
- About one-third of patients admitted to emergency department with acute chest pain have normal ECG. Of these, 5–40% may have evolving MI. Serial ECGs, aided by other tests (biochemical markers, stress test, etc), are essential in the chest pain evaluation if the initial ECG is not diagnostic.

Biochemical Markers

Biochemical markers in serum are measured to detect or exclude myocardial necrosis. This identifies the subgroup of patients at higher risk of developing major adverse cardiac event. Troponin T and troponin I, myoglobin and creatine kinase (CK) MB, are the most frequently used. Their specificity and release profile are summarized in Table 2 and Fig. 2.

- Single values of these biochemical markers do not have high sensitivity for AMI or for prediction of complications. Also, because of the time-frame constraints, the use of a single necrosis marker determination is not generally advised in GP/primary care settings.
- For ruling out of myocardial necrosis, myoglobin is a better marker from 3 to 6 hrs after the onset of symptoms compared with CK-MB mass and troponin. Maximal negative predictive value of myoglobin

Table 2

Marker	Cardiac Specificity?	First Rise after Necrosis (h)	Mean Time to Peak Elevation (h)	Time to Return to Normal Range
Myoglobin	No	1–3	6–7	12–24 hrs
CK total	No	4–8	24	36–48 hrs
CK-MB	++	3–4	24	24–36 hrs
Troponin T	++++	3–4	12–48	10–14 days
Troponin I	++++	4–6	24	4–7 days

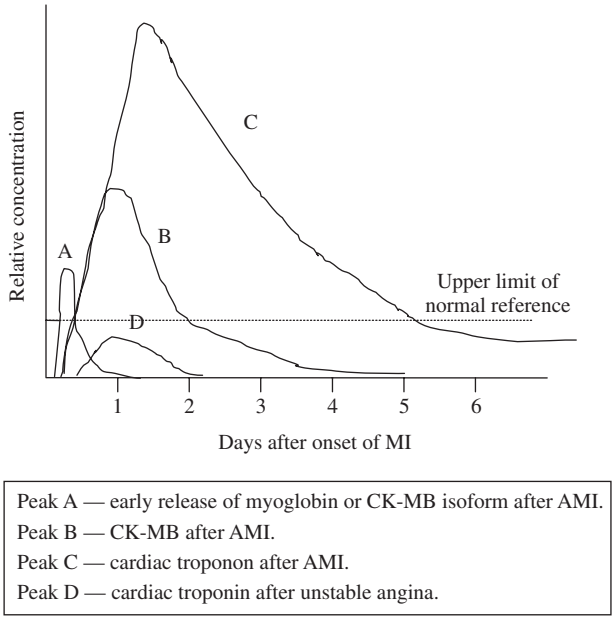


Fig. 2 Plot of the appearance of cardiac markers in blood vs. time after onset of symptoms.

reaches ~90% during this time frame. Because of its lack of cardiac specificity, an isolated measurement of myoglobin in patients with nondiagnostic ECG should not be relied on to make the diagnosis of AMI, but should be supplemented by a more cardiac-specific marker such as troponin.

- From 7 hr after onset of symptoms, CK–MB and troponin T seem to have a higher negative predictive value than myoglobin.
- In AMI, the magnitude of the rise of troponin is typically more than 20 times above the reference range, much higher than the 5- to 20-fold increase of CK–MB above the upper limit of the reference range.
- Routine diagnosis of AMI can be accomplished within 12 hr of using CK–MB, cTnT, or cTnI by obtaining measurements approximately every 6 to 12 hr. In chest pain unit/emergency department, a combination of assays of myoglobin, troponin, and/or CK–MB at 0 hr, 3–6 hr, 6–12 hr after presentation, will be able to detect >95% of myocardial necrosis after the onset of symptoms.
- Elevated troponin is an independent predictor of adverse cardiac outcome in patients presenting with chest pain, with or without diagnostic ECG changes.

Imaging Techniques

- Chest X-Ray
  - Often performed as a routine in the evaluation of patients suspected of having cardiac symptom, though its value has not been established in patients defined as low risk from history or physical examination.
  - One-quarter of these may show some abnormalities, e.g. cardiomegaly, pneumonia and pulmonary edema.
- If aortic dissection is suspected, imaging studies of the aorta must be pursued promptly. Appropriate tests include a chest computer tomography (CT) scan with contrast, or a magnetic resonance imaging in hemodynamically stable patients, or a transesophageal echocardiogram in patients who are less stable. A chest X-ray is not sufficient to exclude this diagnosis.
- If pulmonary embolism is suspected, initial tests usually include ventilation-perfusion scan, or a spiral chest CT scan, and/or pulmonary arteriography.
- In conditions where the clinical history, ECG and biochemical measurements for myocardial damage are equivocal or unavailable, imaging techniques (with rest myocardial perfusion scans or 2-D echocardiography) may be helpful. If normal, they help identify low-risk patients, who can be eligible for early discharge or undergo early stress testing

and avoiding hospital admission. Early use of treadmill exercise testing for such patients is now an accepted management strategy for low-risk patients. Exercise testing is not appropriate, however, for patients who: 1) report pain that is believed to be ischemic, occurring at rest; or 2) have ECG changes consistent with ischemia not known to be old.

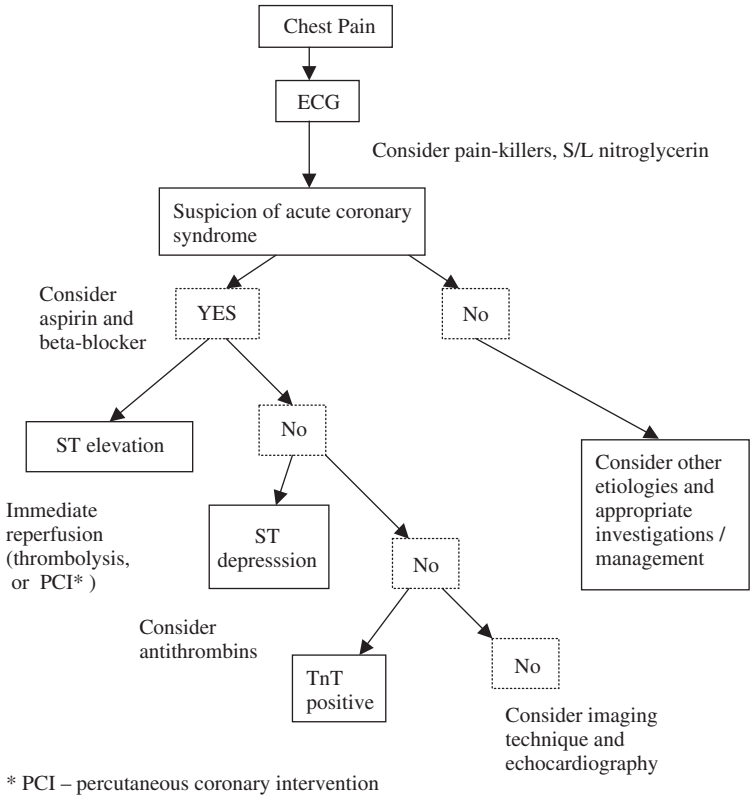


Fig. 3 Evaluation and management of chest pain in emergency department.

CONCLUSION

With careful history taking and physical examination, supplemented by targeted diagnostic tests, chest pain caused by serious conditions can often be quickly identified for early and immediate treatment.

This page intentionally left blank

2

Hypertension

Tan Ru-San and B. A. Johan

INTRODUCTION

Hypertension is a common medical disorder and a major risk factor for the development of stroke, coronary artery disease, cardiac failure and renal dysfunction. As it is frequently asymptomatic, it is often under-diagnosed and under-treated.

Worldwide, hypertension affects an estimated 690 million persons, exacting a tremendous cost in cardiovascular morbidity and mortality.[1] In Singapore, according to the National Health Survey,[2] 27.3% (30.5% males, 24.0% females) of adults aged 30–69 are hypertensive. The age-specific prevalence increases markedly from age 40. Overall, Malays have the highest prevalence (31.5%), followed by Chinese (26.9%) and Indians (24.6%). Of note, 53.0% of hypertensives in the survey were not previously diagnosed.

HYPERTENSION AND CARDIOVASCULAR RISK

The blood pressures of individuals in a population conform to a bell-shaped distribution. Hypertension is thus best understood as the arbitrary

limit above which the epidemiological evidence of benefit of intervention outweighs the risk of inaction.

The 1997 Sixth Joint National Committee[3] (JNC VI) and the 1999 World Health Organization–International Society of Hypertension[4] (WHO–ISH) define hypertension as a systolic blood pressure ⩾ 140 mmHg and/or a diastolic blood pressure ⩾ 90 mmHg. The JNC VI classification of blood pressure levels in adults is summarized in Table 1.

These new practice guidelines incorporate cardiovascular risk assessment as an integral component. Patients are stratified into 3 categories based on the presence of coronary risk factors (the details of specific coronary risk factors are discussed below) and hypertensive target organ damage. Both blood pressure level and cardiovascular risk category become important considerations in the management of patients (Table 2). Note

Table 1 JNC VI Classification of Blood Pressure

Category	SBP (mmHg)		DBP (mmHg)
Optimal	< 120	and	< 80
Normal	< 130	and	< 85
High-Normal	130–139	or	85–90
Hypertension			
Stage I	140–159	or	90–99
Stage II	160–179	or	100–109
Stage III	⩾ 180	or	⩾ 110

DBP-diastolic blood pressure; SBP-systolic blood pressure.

Table 2 Blood Pressure Categories and Cardiovascular Risk in Hypertension Treatment

Blood Pressure	Risk Group A	Risk Group B	Risk Group C
High-normal	Lifelong lifestyle modification	Lifelong lifestyle modification	Drug therapy^{+}
Stage 1	Lifestyle modification up to 12 mth	Lifelong lifestyle modification up to 6 mth*	Drug therapy
Stages 2 & 3	Drug therapy	Drug therapy	Drug therapy

Group A: no risk factor, no target organ damage/clinical cardiovascular disease.
Group B: ⩾ 1 risk factor (no diabetes), no target organ damage/clinical cardiovascular disease.
Group C: target organ damage/clinical cardiovascular disease and/or diabetes.
* When multiple risks are present, treat initially with drugs.
$^{+}$For heart failure, renal insufficiency, diabetes.

Table 3 Optimal Blood Pressure in High-Risk Patients

Condition	Target Blood Pressure
Coronary ischemia	$< 130/85$ mmHg[5]
Proteinuria (< 1 g/day)	$< 130/85$ mmHg[3]
Proteinuria (≥ 1 g/day)	$< 125/75$ mmHg[3]
Diabetes	$< 130/85$ mmHg[3]
Diabetes (with nephropathy)	$< 130/80$ mmHg[6]

that diabetes is accorded special status: Its presence alone automatically places the patient in the highest risk category.

The optimal blood pressure is a function of the overall cardiovascular risk status. As the level of blood pressure is continuously related to the risk of target organ damage, it follows that the target levels of blood pressure in high-risk individuals should be lower than that of low-risk individuals. Table 3 details the recommended target blood pressure levels for some conditions.

EVALUATION OF THE HYPERTENSIVE PATIENT

There are 4 primary aims in the clinical evaluation of a hypertensive patient:

1) Confirm the diagnosis of hypertension;
2) Establish the etiology of hypertension, and rule out secondary causes of hypertension;
3) Assess for target organ damage; and
4) Identify other coronary risk factors, and determine overall cardiovascular risk.

DIAGNOSIS OF HYPERTENSION

In general, the diagnosis of hypertension — and subsequent therapeutic decisions — should be made based on the average of ≥ 2 blood pressure readings taken at each of ≥ 2 office visits after the initial screening visit.[3] The exception to this would be in the event of a hypertensive urgency or emergency, where immediate treatment would have to be instituted.

The importance of accurate blood pressure measurement cannot be overstated. To minimize inter-observer variability, the proper procedure must be adhered to:

1) Ensure at least 5 minutes' rest, and no smoking or caffeine ingestion for 30 minutes prior to measurement.
2) Seat the patient comfortably, with back and bare arm supported.
3) Use a mercury sphygmomanometer (or a recently calibrated aneroid device) and an appropriately sized inflatable cuff.
4) Inflate cuff 20–30 mmHg above the disappearance of the radial pulse; and deflate cuff pressure at a rate of 2–3 mmHg per second.
5) Auscultate over the brachial artery, noting the first and fifth phases of Korotkoff sounds, which correspond to the systolic and diastolic blood pressures, respectively.
6) Record blood pressure to the nearest sphygmomanometer marking, i.e. nearest 2 mmHg.

Some patients have white coat hypertension: The blood pressure is elevated in the clinic, but is within the normal range outside the clinic setting. This diagnosis can only be made by home self-monitoring or ambulatory monitoring of blood pressure. It is important to make this diagnosis, as it carries a much better prognosis than sustained hypertension, and may arguably not require any treatment at all.

Electronic blood pressure monitors are now commonly used. Their ease of use makes them ideal for home self-monitoring and ambulatory blood pressure monitoring. These monitors detect pressure oscillations distal to a deflating cuff — the maximal pressure oscillation corresponds to the mean arterial blood pressure, and the systolic and diastolic blood pressures are empirically derived using algorithms. Agreement with mercury sphygmomanometer blood pressure measurements is generally satisfactory with currently available electronic arm devices, provided they have been subjected to validation tests.[7] Wrist and finger devices should not be used as they are inaccurate.

ETIOLOGY OF HYPERTENSION

Primary Hypertension

More than 90% of hypertensive patients have primary (also called essential or idiopathic) hypertension, i.e. no specific cause can be found. The

exact pathophysiology of primary hypertension is not fully elucidated. Alterations in and the interplay of multiple blood pressure regulatory systems — the central and peripheral autonomic nervous systems, circulating and tissue renin-angiotensin-aldosterone systems, kallikrein-kinin system, renal parenchymal mechanisms and endothelial vasoactive factors — participate in the initiation and maintenance of hypertension.

It has been estimated that 30–60% of the variation in blood pressure between individuals is attributable to the effect of genetic contribution, while environmental factors (excess sodium intake, stressful lifestyle, etc.) account for another 20% of the variation.

Secondary Hypertension

Less than 10% of hypertensive patients have identifiable secondary etiology (Table 4). Among the secondary causes of hypertension, chronic renal

Table 4 Causes of Secondary Hypertension

Renal parenchymal disease
Acute and chronic glomerulonephritis
Interstitial nephritis
Chronic pyelonephritis
Polycystic kidney disease
Hydronephrosis
Diabetic nephropathy
Renovascular disease
Atherosclerotic renal artery stenosis
Fibromuscular dysplasia
Intrarenal vasculitis
Renin-producing tumors
Sleep apnea
Endocrinopathies
Cushing's syndrome
Primary aldosteronism
Pheochromocytoma
Acromegaly
Hypothyroidism and hyperthyroidism
Exogenous hormones
Estrogens, glucocorticoids, mineralocorticoids
Coarctation of aorta
Pregnancy-related hypertension
Miscellaneous
Neurological disorders
Acute stress

parenchymal disease is the most prevalent, followed by renovascular disease. Endocrine conditions occur in less than 1%, the most common being primary aldosteronism. Drug-induced causes must not be forgotten: Patients on glucocorticoid therapy or women on oral contraceptives are susceptible to iatrogenic hypertension.

It is important to diagnose secondary hypertension because some of the causes may be potentially curable by operation or are amenable to specific therapeutic intervention. While it is not feasible to investigate every hypertensive patient for secondary causes, the clinician should select for further specialized testing, patients with features that suggest secondary hypertension (Table 5).

Some of the commoner causes of secondary hypertension are reviewed here.

1) Renal parenchymal disease

Compared with tubular disease (e.g. chronic pyelonephritis, interstitial nephritis), glomerular disease (e.g. chronic glomerulonephritis, diabetic nephropathy) is more commonly associated with hypertension. An exception is polycystic kidney disease, where hypertension is an early manifestation. Abnormal urinalysis and elevated serum creatinine level (late feature) are clues that suggest the diagnosis. Renal ultrasonography is indicated to determine renal size and symmetry.

Table 5 Indications to Screen for Secondary Hypertension

Onset < 30 years or > 50 years
Severe elevation of blood pressure $\geq 180/110$ mmHg
Refractory hypertension despite maximal therapy
Target organ damage
Fundoscopic hypertensive changes $\geq$ grade 2
Serum creatinine > 1.5 mg/dL
Cardiomegaly (chest radiograph)
Left ventricular hypertrophy (electrocardiograph, echocardiography)
Features suggestive of secondary causes
Headaches, palpitations and sweating (pheochromocytoma)
Abdominal bruit (renal artery stenosis)
Radiofemoral delay (coarctation of aorta)
Unprovoked hypokalemia (hyperaldosteronism)
Truncal obesity, abdominal stria (Cushingnoid features)
Obesity, snoring, daytime somnolence (sleep apnea)

2) Renal artery stenosis

Atherosclerotic renal artery stenosis (two-thirds of cases) usually presents after 50 years of age, frequently in association with concomitant coronary, cerebral or peripheral vessel disease and diabetes; and is bilateral in 35–50% of cases. Fibromuscular dysplasia usually occurs unilaterally in young women, and may affect other vessels, such as the carotid artery. The captopril renal scintiscan in which administered angiotensin-converting enzyme (ACE) inhibitor decreases the glomerular filtration rate of the affected kidney (reflected by reduced radioactive tracer uptake) is an effective non-invasive screening test. Duplex renal sonography is also useful but is operator-dependent. Magnetic resonance angiography is a promising new screening test.

3) Sleep apnea

Sleep-disordered breathing is increasingly recognized as a cause of hypertension. The hypertension is driven by excess catecholamine release. The disease occurs frequently in obese men, but may be missed in non-obese women. Symptoms include nocturnal snoring, daytime fatigue and somnolence). Detection is by sleep oximetry, or more elaborate polysomnography. Treatment with continuous positive airway pressure — or oral appliances or uvulopalatoplasty for selected patients — may yield substantial blood pressure reduction.

4) Primary aldosteronism

In a hypertensive patient not on diuretic therapy, a serum potassium level < 3.2 mmol/L (unprovoked hypokalemia) or excessive urinary potassium excretion (> 30 mmol/day) should prompt further investigations to rule out primary aldosteronism. A ratio of plasma aldosterone (ng/dL) to plasma renin activity (ng/ml/h) level > 20 suggests the diagnosis. Unsuppressed supine plasma aldosterone level > 10 ng/dL 4 hour post-saline infusion confirms autonomous aldosterone secretion. Further specialized investigations (postural tests, abdominal imaging and invasive adrenal venous sampling) allow differentiation between the 2 major causes — adrenocortical adenoma (60%) and bilateral adrenocortical hyperplasia (40%).

5) Cushing's syndrome

Cushingnoid features can be subtle and may lack specificity. A 24-hour urinary-free cortisol excretion < 100 μg/day screens out the diagnosis in most patients, although some alcoholics and depressives have modest

elevations. In the dexamethasone suppression test, 1 mg of dexamethasone is taken orally at 11 pm and plasma cortisol level taken at 8 am the next day. A level of $< 5\,\mu g/dL$ excludes Cushing's syndrome in most patients. If these screening tests are abnormal, further specialized investigations are required to define the anatomic lesion — pituitary microadenoma (80%), adrenal adenoma or hyperplasia.

6) Pheochromocytoma

This catecholamine-secreting tumor of the adrenal medulla presents with either persistent or paroxysmal hypertension. The simultaneous occurrence of the classic symptom triad — headaches, palpitations and sweating — is strongly suggestive of the disorder. Elevated ratio of urinary metanephrine 24-hour urinary metanephrine (μg)/creatinine (mg) ratio > 2 and plasma free metanephrine level (where available) establish the diagnosis. Urinary vanillyl mandelic acid assay lacks sufficient sensitivity and specificity and is no longer recommended. In principle, biochemical confirmation should always precede anatomic delineation of the lesion. Indiscriminate adrenal imaging may unmask incidental benign cysts or functionally inconsequential tumors, which are best left alone.

TARGET ORGAN DAMAGE

Hypertension can cause a spectrum of end organ damage (Table 6). The following mechanisms, or a combination of them, are implicated:

1) Direct pressure effect

Increased intraluminal pressure imposes mechanical strain on blood vessels and various organs, resulting in hemodynamic perturbations. The pathologic effects of hypertension on small and large blood vessels are exemplified by its effects on retinal and peripheral arteries (Table 6). Occasionally, an acute severe rise in blood pressure can have serious consequences — acute pulmonary edema, aortic dissection, hemorrhagic stroke and cerebral edema (hypertensive encephalopathy).

2) Neurohormonal activation

Neurohormones (e.g. renin-angiotensin-aldosterone and sympathetic nervous systems) are initially released to physiologically compensate for the effects of elevated blood pressure in the various organs. Their sustained secretion, however, exerts a deleterious effect by perpetuating hypertension and organ damage (see below).

Table 6 Hypertensive Damage to Major Target Organs

Heart	Left ventricular hypertrophy Impaired left ventricular diastolic and systolic function Coronary microvascular and atherosclerotic disease Atrial fibrillation and ventricular arrhythmia
Blood vessels	Arterial dilatation and aneurysm formation Aortic dissection Peripheral artery disease
Brain	Acute and chronic hypertensive encephalopathy Intracerebral hemorrhage, subarachnoid hemorrhage Intraparenchymal stroke, lacunar infarction
Eye	Retinal microaneurysms, hemorrhage and exudates Optic neuropathy
Kidney	Nephrosclerosis Proteinuria, azotemia

3) Accelerated atherosclerosis

Chronic hypertension contributes to atherosclerotic plaque formation, especially in the presence of other risk factors like diabetes, hyperlipidemia and smoking. In addition, it can also cause endothelial and microvascular dysfunction.

In the heart, elevated systemic blood pressure leads to increased peak systolic wall stress. In response, the cardiac muscles undergo gradual hypertrophy and alteration in chamber geometry to normalize the wall stress. Concomitant activation of the vascular renin-angiotensin-aldosterone system leads to further adverse remodeling, enhanced myocardial fibrosis and eventual heart failure. In the coronary circulation, endothelial and microvascular dysfunction occur early. Coronary artery involvement gives rise to coronary ischemia and infarction, leading to further compromise of myocardial function.

In the kidney, chronic blood pressure elevation causes compensatory preglomerular vasoconstriction in an attempt to limit intraglomerular capillary hypertension. Renal blood flow decreases, which activates the renin-angiotensin system leading to postglomerular vasoconstriction in an effort to restore capillary pressure. This neurohormonal-induced intrarenal vasoconstriction leads to eventual cell injury, sclerosis, nephron loss and further aggravation of systemic hypertension.

In the brain, hypertension is associated with both hemorrhagic and ischemic stroke. The direct effect of raised intraluminal pressure leads to

brain hemorrhage. Hypertension-induced fibrinoid necrosis and hyaline damage, as well as atherosclerotic changes, are important in the pathogenesis of ischemic strokes.

CORONARY RISK ASSESSMENT

The risk of cardiovascular disease is dependent on the level of blood pressure as well as the presence or absence of target organ damage or other coronary risk factors (Table 7). These are determined during the routine clinical evaluation of patients with hypertension. A suggested scheme of essential components of history taking, physical examination and laboratory investigations is detailed in Table 8.

MANAGEMENT OF HYPERTENSION

Hypertension is managed by a combination of lifestyle modification and drug therapy, depending on the blood pressure level and cardiovascular risk stratification (Table 2).

Non-drug Therapy

Non-pharmacologic interventions are valuable in the prevention and control of hypertension. Useful lifestyle modifications include:

1) Losing weight, if overweight.
2) Moderation of alcohol intake
   < 30 ml ethanol/day for men; < 15 ml ethanol/day for women.

Table 7 Cardiovascular Risk Stratification in Hypertensive Patients

Major Risk Factors
Smoking
Dyslipidemia
Diabetes mellitus
Age > 60 years
Sex (men and postmenopausal women)
Family history of cardiovascular disease: women < 65 years; men < 55 years
Target Organ Damage/Clinical Cardiovascular Disease
Heart disease (left ventricular hypertrophy, angina, prior infarct, prior coronary revascularization, heart failure)
Stroke or transient ischemic attack
Nephropathy
Peripheral arterial disease
Retinopathy

Table 8 Clinical Evaluation of Hypertension

History
Duration of hypertension
Past cardiovascular disease
Family history
Lifestyle habits (smoking, alcohol consumption, exercise)
Medication (past and present, including oral contraceptive pills in women)
Symptoms to suggest secondary hypertension (Table 5)
Examination
Measure blood pressure accurately (⩾ readings)
Cardiac, peripheral vascular, abdominal, neurologic, fundoscopic examination
Signs of secondary hypertension (Table 5)
Signs of target organ damage (left ventricular hypertrophy or enlargement, stroke, retinopathy, etc.)
Screening Investigations
Full blood count, serum chemistry, fasting lipid profile and glucose, urinalysis, electrocardiogram
Further Investigations (where indicated)
Creatinine clearance, 24-hour urine protein, renal ultrasound (exclude renal etiology of hypertension/assess hypertensive renal damage)
Echocardiography (assess for myocardial hypertrophy and function)
Oral glucose tolerance test, glycosylated hemoglobin (to diagnose diabetes/assess diabetic control)
Specialized tests for secondary hypertension (see above)

3) Aerobic exercise
   30–45 minutes of aerobic activity most days of the week.
4) Reduced consumption of dietary sodium
   Limit sodium chloride intake to 6 g (one teaspoon) a day. Even this is more than 10 times the body's daily sodium chloride requirement of 500 mg.
5) Smoking cessation and reduced saturated fat consumption for overall cardiovascular health.

The effects of potassium and calcium supplements on lowering blood pressure are considered too small (except in the undernourished) to advocate their routine use for blood pressure lowering.

Drug Therapy

Treatment of hypertension by the use of pharmacological agents has been shown to reduce hypertension-induced morbidity and mortality. In

a meta-analysis of 14 major randomized primary prevention trials in hypertension, comprising 37 000 patients with a mean follow-up of 5.1 years, a modest decrease of diastolic blood pressure of 5–6 mmHg was associated with significant reductions in all stroke events (42%), all coronary heart disease events (14%) and cardiovascular mortality (21%).[8]

Early retrospective studies[9,10] of hypertensive patients lend support to the J-shaped curve hypothesis, in which blood pressure lowering beyond a limit (believed to be a diastolic blood pressure level of about 85–90 mmHg) imposes increased risk. However, these studies were methodologically flawed.[11] Several large-scale prospective studies[12–15] have demonstrated that intensive diastolic blood pressure lowering to 80 mmHg or beyond in different patient groups significantly reduces cardiovascular morbidity and mortality.

There are many classes of anti-hypertensive drugs (Table 9). Some of the older drugs such as the central alpha-agonists, peripheral adrenergic inhibitors and direct vasodilators are not frequently used nowadays

Table 9 Classes of Anti-hypertensive Drugs and Common Examples

Diuretics
Hydrochlorothiazide (thiazides)
Frusemide (loop diuretic)
Spironolactone, Amiloride (potassium-sparing diuretics)
Indapamide (diuretic-like compound)
Beta-blockers
Atenolol, Bisoprolol, Metoprolol, Propanolol
Carvedilol, Labetalol (Alpha- and beta-blockers)
Calcium-channel blockers
Amlodipine, Nifedipine (dihydropyridines)
Diltiazem, Verapamil (non-dihydropyridines)
Angiotensin converting enzyme inhibitors
Captopril, Enalapril, Lisinopril, Perindopril, Ramipril
Angiotensin receptor blockers
Candesartan, Irbesartan, Losartan, Valsartan
Alpha-blockers
Prazosin, Doxazosin
Central alpha-blockers
Clonidine, Methyldopa
Peripheral adrenergic inhibitors
Guanethidine, Reserpine
Direct vasodilators
Hydralazine, Minoxidil

because of associated severe or unpleasant side-effects. A detailed discussion of the drug classes is beyond the scope of this review.

Drug selection should be predicated on the following considerations:

1) Blood pressure control
   The overwhelming therapeutic priority is for optimal blood pressure control. The physician must be prepared to use drugs in combination or to change ineffective treatment, if necessary. Combination therapy, especially, has the potential to increase efficacy at lower doses of the individual drugs, thus improving the side-effect profile.
2) Specific indications
   Drug selection should be individualized on the basis of clinical assessment and associated morbid conditions (see Table 10). Specific drug therapy has been found to confer morbidity and mortality benefits in certain conditions.
3) Potential contraindications
   Absolute contraindications include the use of beta-blockers in severe asthmatics, or the use of ACEs in patients with advanced renal insufficiency who are not candidates for renal replacement therapy. The physician should also be alert to potential drug interactions in patients taking multiple drugs.
4) Compliance
   The ideal drug should preferably have a once-a-day dosing schedule and yet provide consistent blood pressure control throughout the day, i.e. a high trough-to-peak activity.

Table 10 Comorbid Conditions and Preferred Anti-hypertensive Drugs

Condition	Drug
Diabetes, with proteinuria	ACE inhibitor, ARB
Renal insufficiency*	ACE inhibitor, ARB
Heart failure	ACE inhibitor, diuretic
	ARB
	Beta-blocker^{+}
Myocardial infarct	Beta-blocker
	ACE inhibitor (with systolic dysfunction)
Chronic angina	Beta-blocker

ACE inhibitor = angiotensin-converting enzyme inhibitor;
ARB = angiotensin receptor blocker.
* Use with caution when creatinine level > 265 mmol/L.
$^{+}$Use carvedilol or bisoprolol with caution in severe heart failure.

5. Cost
   From a public health point of view, the primary aim is to attain the desired blood pressure level using whatever available drug, or combination of drugs, with minimal or an acceptably low level of side-effects, and at a relatively low cost. Cost considerations may sometimes conflict with some of the factors discussed above.

CONCLUSION

Hypertension is a ubiquitous disorder with a well-established role in cardiovascular disease causation. Comprehensive evaluation of the hypertensive patient comprises several components: Exclusion of secondary (potentially treatable) causes of hypertension; assessment for hypertension-induced target organ complications; and overall cardiovascular disease risk assessment. The aim of treatment is to prevent target organ damage and to decrease cardiovascular mortality. Current practice guidelines stress the importance of blood pressure control. Lower blood pressure treatment thresholds are required in high-risk patients (those with diabetes and target organ damage). Evidence-based use of specific drugs to treat hypertension in special patient groups yields morbidity and mortality benefits, and should be advocated where available and where cost permits.

REFERENCES

1. World Hypertension League Yearbook 2000–2001: *Fighting Hypertension into the Next Millennium*, World Hypertension League, Toledo, OH, 2001.

2. *National Health Survey 1998*, Epidemiology and Disease Control Department, Ministry of Health, 1999.

3. Sixth Report of the Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure, *Arch Intern Med* **157**:2413–2446, 1997.

4. World Health Organization-International Society of Hypertension Guidelines for the management of hypertension: guideline subcommittee, *J Hypertens* **17**:151–183, 1999.

5. Ryan TJ, Antman EM, Brooks NH, *et al.*, 1999 update: ACC/AHA guidelines for the management of patients with acute myocardial infarction. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (committee on Management of Acute Myocardial Infarction), *J Am Coll Cardiol* **34**: 890–911, 1999.

6. Bakris GL, Williams M, Dworkin L, *et al.*, Preserving renal function in adults with hypertension and diabetes: a consensus approach. National Kidney Foundation

Hypertension and Diabetes Executive Committees Working Group, *Am J Kidney Dis* **36**:646–661, 2000.

7. O'Brien E, Waeber B, Parati G, Staessen J, Myers MG, Blood pressure measuring devices: recommendations of the European Society of Hypertension, *Brit Med J* **322**:531–536, 2000.
8. Collins R, Peto R, MacMahon, *et al.*, Blood pressure, stroke and coronary heart disease. Part 2, short term reductions in blood pressure: overview of randomized drug trials in their epidemiological context, *Lancet* **335**:827–838, 1990.
9. Cruickshank JM, Thorp JM, Zacharias FJ, Benefits and potential harm of lowering blood pressure, Lancet **1**:581–583, 1987.
10. Alderman MH, Ooi WL, Madhavan S, Cohen H, Treatment-induced blood pressure reduction and the risk of myocardial infarction, *JAMA* **262**:920–924, 1989.
11. Zanchetti A, To what extent should blood pressure be brought by hypertensive treatment? *High Blood Pressure Cardiovasc Prevention* **1**:197–200, 1992.
12. MacMahon SW, The effect of antihypertensive drug treatment on the incidence of stroke and of coronary heart disease, *Clin Exp Hypertens* **A11**:807–823, 1989.
13. SHEP Cooperative Research Group, Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP), *JAMA* **265**:3255–3264, 1991.
14. Yusuf S, Peto R, Lewis J, *et al.,* Beta-blockade during and after myocardial infarction: an overview of the randomized trials, *Prog Cardiovasc Dis* **27**:335–337, 1985.
15. Hansson L, Zanchetti A, Carruthers SG, *et al.,* Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomized trial. HOT Study Group, *Lancet* **351**:1755–1762, 1998.

This page intentionally left blank

3

Ischemic Heart Disease

Lim Yean Leng and Mak Koon Hou

INTRODUCTION

Ischemic heart disease (IHD), once thought to be uncommon at the turn of this century, is now one of the commonest medical conditions. Indeed, in 1997, they accounted for >15 million deaths worldwide, second only to infections. By the next century, cardiovascular diseases will probably emerge as the most common cause of death. IHD is currently a major public health problem in both developed and developing countries. In the Asia-Pacific region, rapid economic achievements of several nations did not only bring about greater affluence and longer life expectancy, but also lifestyle-related diseases such as hypertension, diabetes mellitus, obesity and IHD.

Although the mortality rate of IHD has been steadily increasing over the past decades, a gradual decline has been observed in Singapore since 1990. Indeed, death rates have decreased from 106 per 100 000 residents in 1990 to 99 per 100 000 residents in 1997, an almost 7% relative reduction (Fig. 1), especially among men. However, with an aging population, the absolute number of patients with IHD continues to rise. With limited

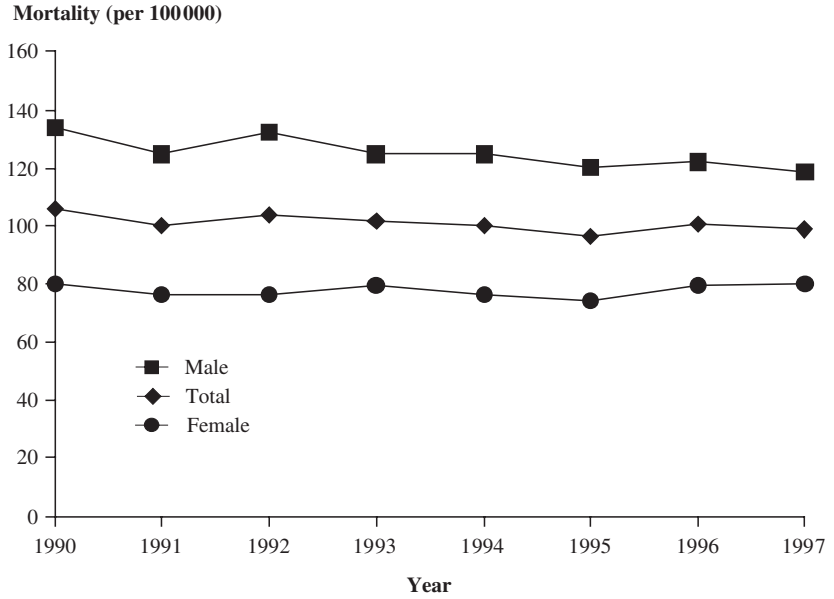


Fig. 1 Age-standardized mortality rates from ischemic heart disease by gender.

Age standardization is based on "OLD" World Standard Population. Data provided by Epidemiology and Disease Control, Ministry of Health, and Registry of Births and Deaths, Singapore.

resources available, the treatment and prevention of IHD have become a complex clinical management issue and a major challenge to healthcare providers in the new millennium.

CORONARY ATHEROSCLEROSIS

The principal underlying mechanism for IHD is coronary atherosclerosis. Currently, disease severity is often equated with angiographic demonstration of the "number and degree of stenoses" or coronary plaques. Recent observations indicate that it is not the severity of stenosis (plaque volume) that determines the outcome, but the type of stenosis (plaque composition) and the extent of collateral circulation. Therefore, there is a changing paradigm in viewing atherosclerosis more as an inflammatory disease than an anatomic disease. Reduction of inflammation ("plaque

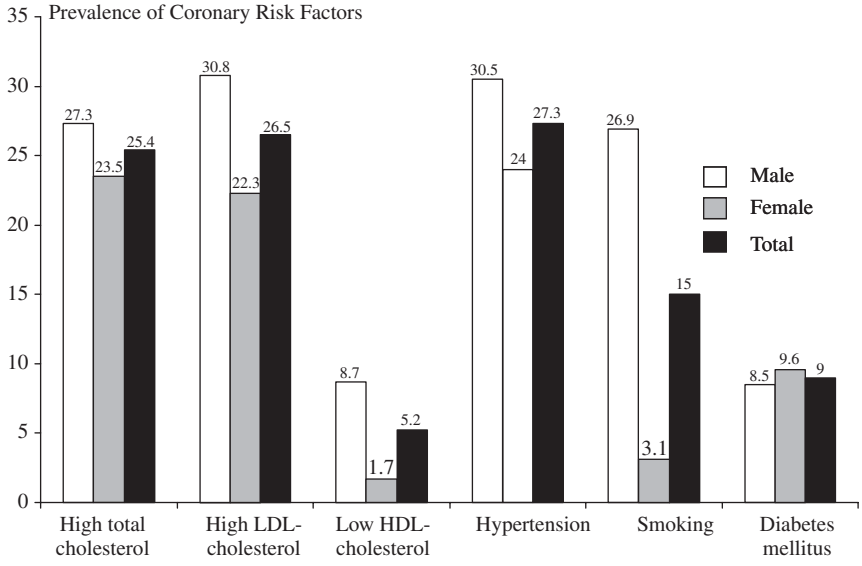


Fig. 2 Prevalence of coronary risk factors among adults (aged 18 to 69 years) in Singapore.

Data obtained from the National Health Survey, 1998.

LDL = low-density lipoprotein; HDL=high-density lipoprotein.

High total cholesterol implies ⩾ 6.2 mmol/L; high LDL-cholesterol implies ⩾ 4.1 mmol/L; low HDL-cholesterol implies < 0.9 mmol/L; hypertension implies a history of hypertension or blood pressure > 140/90 mmHg; diabetes mellitus implies a history of diabetes or positive oral glucose tolerance test; smoking implies smoking of ⩾ 1 stick of cigarette a day.

passivation") is more important than reduction of stenosis in the treatment of acute coronary syndromes (ACS). Stable angina is usually due to stenotic but stable and very slowly progressing lesions, while ACS and sudden death are usually due to sudden and rapid progression of non-critical lesions due to plaque disruption, often complicated by thrombosis, with or without concomitant vasospasm.

Significant luminal diameter stenosis (> 70%) in a major epicardial artery leads to diminution of coronary blood supply to the myocardium. Tonal properties of the vascular wall and the presence of occlusive or non-occlusive thrombus are other mechanisms contributing to the clinical manifestations of coronary insufficiency.

PATHOPHYSIOLOGY OF MYOCARDIAL ISCHEMIA

Myocardial ischemia is present when there is imbalance of oxygen supply and demand to the myocardium. This condition may occur when there is reduction of blood supply to the myocardium, in the presence of normal (at rest) or excessive (during exercise) demand. Obstructive coronary artery disease compromising the lumen is the underlying pathophysiology in this clinical setting.

Myocardial ischemia and angina can also occur in non-obstructive IHD. Endothelial dysfunction with or without the presence of atherosclerotic plaque may produce flow disturbances resulting in myocardial ischemia and angina, the so-called "*normal coronary angina syndromes*." Myocardial ischemia can also result from left ventricular (LV) hypertrophy or hypertrophic obstructive cardiomyopathy, when the myocardial oxygen demand exceeds normal supply.

Obstructive Coronary Artery Disease

Unlike skeletal muscles, the heart depends almost totally on aerobic metabolism. Dissimilar to other tissues, the amount of myocardial oxygen extraction from the blood is substantial, ranging from 70–80%, at resting state. These characteristics limit the capacity of the heart to provide for increased oxygen requirements during exercise or other forms of stress, and the major way to meet the increase in myocardial oxygen demand is to increase coronary blood flow.

The pathophysiology and clinical correlation of myocardial ischemia is summarized in Fig. 3. Significant luminal narrowing of an epicardial coronary artery will reduce coronary blood flow. Apart from blood flow, LV hypertrophy, heart rate, systolic wall tension and myocardial contractility are other key determinants of oxygen requirement of the heart. The clinical spectrum of IHD, ranging from stable angina to ACS, evolves from the interaction of these diverse factors.

Progressive decrease in myocardial perfusion is clinically reflected by progressive ECG changes of myocardial ischemia. This starts with upsloping ST depression (mild ischemia) and progresses to deep symmetrical T-wave inversion (evidence of subendocardial necrosis or the so-called non-ST-segment-elevation myocardial infarction [NSTEMI]). Total occlusion (100% luminal occlusion) of a coronary artery with no antegrade or

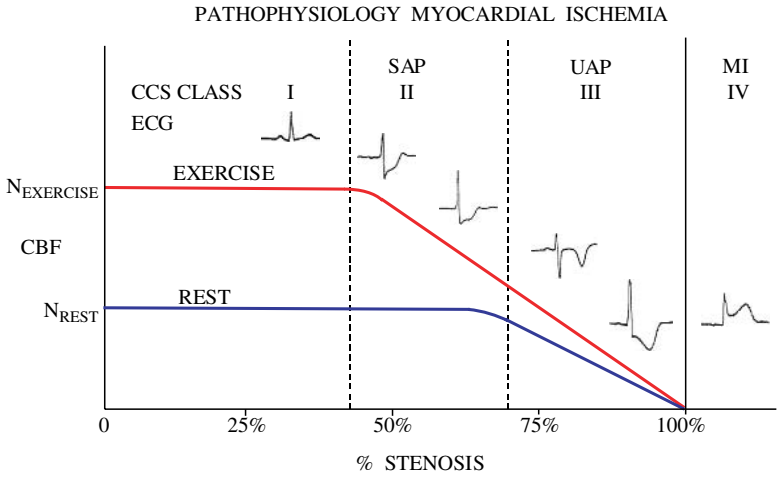


Fig. 3 Schematic diagram showing the relationship between coronary blood flow (CBF) and luminal narrowing.

At rest, coronary blood flow is significantly lower than during exercise to meet the demands of the myocardium. However, after a certain critical degree of narrowing of the lumen, coronary blood flow diminished considerably. Whether the patient develops symptoms of coronary insufficiency depends on a variety of factors, such as metabolic demands and presence of collateral circulation. When the lumen is occluded, ACS ensues. The Romanized numerals indicate the Canadian Cardiovascular Society classification of severity of angina.

SAP = stable angina pectoris; UAP = unstable angina pectoris; MI = myocardial infarction; CCS = Canadian Cardiovascular Society classification of severity of angina (see Table 1); ECG = electrocardiogram.

retrograde blood flow will result in transmural ischemia reflected by a ST-segment elevation in the ECG. Myocardial necrosis and infarction will ensue if there is no reperfusion in >20 minutes.

Non-obstructive Coronary Artery Disease ("Normal Coronary Angina Syndromes")

Coronary endothelial dysfunction and vasospasm, as mechanisms of reduced coronary flow, can occur in the presence of systemic hypertension, diabetes mellitus, raised serum lipid levels, especially low-density lipoprotein (LDL) and lipoprotein (a) (Lp[a]), anxiety and sympathetic overactive states, cigarette smoking and several other factors. Enhanced platelet aggregation due to increased circulating catecholamines and accentuated

vascular reactivity to vasopressor and vasodilator substances, and hormones such as estrogens and androgens are other mechanisms of endothelial dysfunction. Altered coagulative states such as increased serum fibrinogen level and plasminogen activator inhibitor type 1 (PAI-1), and other thrombin dependent factors also predispose to vasoconstriction. *Normal coronary angina syndromes* as a clinical group can be manifested as typical or atypical angina associated with the finding of normal coronary arteries at coronary angiography. The presence of "slow flow" of contrast (clearance of contrast in >3 systolic cycles) after coronary injection confirms the diagnostic group of *Slow Flow Phenomenon*. Angina associated with epicardial coronary artery spasm, demonstrated at angiography and transient ST-segment elevation on the ECG constitutes the *Prinzmetal Angina* group. Finally, spasm of intramyocardial microvasculature produces *microvascular angina*, the other common group of normal coronary angina syndromes.

The demonstration of normal coronary arteries is important as non-obstructive myocardial ischemia is associated with good prognosis. Symptomatic treatment with nitrates, beta-blockers or calcium antagonists and medical reassurance are the principal components of management of this group of patients.

ANGINA PECTORIS

Symptoms

Stable angina

Effort-related *angina pectoris* is described as a discomfort in the chest, which is brought on by exertion, emotion or cold weather, and relieved by rest or nitroglycerin. In addition to effort-related angina, there are 2 other characteristic angina, namely *decubitus angina* and *postprandial angina*. The former denotes angina on lying in the decubitus position, and is due to transient increased venous return to the heart causing LV dilatation. Acute LV dilatation produces increased wall tension and correspondingly raises oxygen demand, in accordance to the Law of Laplace. Decubitus angina typically improves when the patient assumes an upright position. In contrast, postprandial angina is due to diversion of coronary blood flow to the gastric circulation to facilitate the digestive process, the so-called "mesenteric artery steal syndrome." Both

decubitus and postprandial angina occur only when coronary perfusion is critically balanced and imply a significant amount of myocardium is in jeopardy. These are often associated with extensive atherosclerotic disease.

Angina is a result of reversible myocardial ischemia without necrosis. It develops rapidly and plateaus in <1 minute, and rarely lasts >15 minutes. The probability of chest pain being myocardial ischemia can be semi-quantitatively scored with a maximum of 10, allocating 2 units to each of the characteristics: Precipitating factor, location, nature and radiation, associated symptoms and relieving factors. A score close to 10 denotes high probability angina and a score below 5 is atypical of angina. Typical angina pectoris, especially severe and crushing in nature, lasting >20 minutes is considered to be one of the three major diagnostic criteria of acute myocardial infarction (AMI). Raised myocardial creatine kinase (CK) enzyme and ECG ST-segment elevation are the other 2 diagnostic criteria. Although the discomfort is generally located substernally, the site of angina has been described as from the umbilicus to the jaw and neck, including teeth and ear. The description of angina may be atypical in character, more commonly encountered in women, the elderly and those suffering from diabetes. Angina may be associated with symptoms of cold sweat, breathlessness or giddiness. The CCS Grading Scale is most commonly used to classify the clinical severity of anginal symptoms (Table 1).

Table 1 Canadian Cardiovascular Society (CCS) Functional Classification of Angina Pectoris

Class	Ordinary Activity	Description
I	No limitation	Angina from rapid, strenuous or prolonged exertion
II	Slight limitation*	Angina from walking >2 blocks or up >1 flight of stairs
III	Marked limitation	Angina from walking <2 blocks or up <1 flight of stairs
IV	Inability to perform	Angina at rest or minimal exertion

*Angina occurs during walking or climbing of stairs rapidly or under additional stress, such as after meals, in cold or hot weather, against a strong wind or emotional tension, and first few hours after awakening.

Unstable angina and acute coronary syndromes

New or recent onset (< 30 days' duration) angina, increasing frequency and severity (change of > 1 CCS class of angina), and angina less responsive to nitroglycerin are considered as unstable angina. Unstable angina is often associated with abnormal resting ECG changes, ranging from ST-segment depression to deep T-wave inversion, but not ST-segment elevation. Patients with a history of unstable angina and resting ECG changes of ST-segment depression or T-wave inversion (but not ST-segment elevation) and CK enzyme elevation are referred to as having suffered NSTEMI. Unstable angina, NSTEMI and ST-segment elevation MI (STEMI) together constitute ACS.

Silent myocardial ischemia

Myocardial ischemia may also be *asymptomatic*. The adverse outcomes of patients with silent myocardial ischemia are increasingly recognized. Abnormal diastolic function may be the principal feature of coronary insufficiency. Instead of chest discomfort, these patients may present with *effort-related dyspnea*. Functional assessment of dyspnea is categorized into 4 grades of severity, similar to the classification used for angina. O*rthopnea* and *paroxysmal nocturnal dyspnea* (PND) are typical presentations of severe myocardial ischemia and denotes acute LV dysfunction. Decubitus angina is the equivalent symptom to orthopnea and PND.

PHYSICAL EXAMINATION

There are usually no abnormal cardiovascular clinical signs detected on physical examination in patients presenting with IHD, unless the amount of myocardium in jeopardy is significant enough to cause transient acute LV decompensation. Myocardial ischemia due to significant left main and proximal coronary disease, especially affecting the left anterior descending (LAD) artery, is likely to produce acute LV failure. In these patients, the presence of added heart sounds, third or fourth, as well as pulmonary crepitations, confirms LV dysfunction due to myocardial ischemia.

Occasionally, when the etiology of the angina is due to non-coronary conditions such as aortic stenosis, hypertrophic cardiomyopathy and severe pulmonary hypertension. Characteristic physical signs of these conditions will be present. Less commonly, there may be signs of hyperdynamic circulatory states such as anemia, hyperthyroidism or fever. As there is a paucity of physical signs, an excellent clinical history is crucial in making the diagnosis of IHD. Angina associated with systemic hypotension is a grave prognostic sign and indicates severe and critical coronary artery disease. Atrial or ventricular premature beats may be present. When there is transient papillary muscle ischemia, a functional and sometimes fluctuating systolic murmur of mitral incompetence may be heard at the mitral area.

ELECTROCARDIOGRAMS OF MYOCARDIAL ISCHEMIA

ECG changes of myocardial ischemia and infarction are manifested mainly by changes in the ST-segment and the T-wave. ST-segment depression is the hallmark of myocardial ischemia. Mild ischemia is represented by up-sloping ST-segment depression. More severe ischemia is associated with horizontal or down-sloping ST depression of varying degree. The deeper the ST-segment depression, the greater the amount of ischemia. Down-sloping ST-segment depression or deep symmetrical T-wave inversion is indicative of NSTEMI. With MI, there is accompanying cardiac enzyme (CK, troponin I and T, myoglobin) release. When the vessel is occluded and flow is absent, ST-segment elevation is indicative of STEMI. Clinical correlation of ECG changes, symptomatic functional angina classes and coronary blood flow through a progressively narrowing stenotic coronary artery is schemetically represented in Fig. 3. The progression from transmural ischemia to infarction is a function of time. Clinically, myocardial salvage by myocardial reperfusion after total flow cessation is still possible up to 12 hours from vessel occlusion. This time window is also dependent on the presence of collateral circulation to the infarct-related vessel. Often the infarct-related artery can be accurately localized from the surface ECG for STEMI. Changes in the inferior leads indicate right coronary artery (RCA) involvement; anterior precordial leads, V1 to V4, implicate the LAD; and limb leads I and aVL, as well as

lateral chest leads V5 and V6, reflect ischemia in the territories supplied by the left circumflex (LCX) or diagonal artery. Ischemia of the right ventricle and posterior aspect of the left ventricle, localized to the right-sided chest leads (V4R to V6R) and standard left-sided V1 to V2 precordial leads, implicate RCA involvement proximal to the right ventricular branch.

NON-INVASIVE AND FUNCTIONAL TESTING OF MYOCARDIAL ISCHEMIA

Although the key to diagnosis of IHD lies in obtaining a good clinical history, functional testing provides additional information in establishing the extent of the disease and prognosis. The usefulness of functional testing is limited by the fact that pre-test disease prevalence affects post-test likelihood of significant IHD. Figure 4 summarizes a clinical pathway used to evaluate chest pain. Briefly, chest pain that is typical of angina pectoris, with or without ECG evidence of myocardial ischemia, should be attributed to IHD until proven otherwise. Patients with typical angina and resting ECG evidence of ischemia are by definition unstable and suffer from ACS. These patients require urgent admission to a monitored bed in a coronary care unit and should be offered coronary angiography for definitive diagnosis and treatment (Fig. 5). Patients with atypical chest pain but definite ECG evidence of myocardial ischemia must be managed in the same way as those with IHD. In contrast, patients with chest pain which is not typical of angina and without resting ECG evidence of myocardial ischemia may be subjected to functional testing, especially when there is a lingering suspicion that the chest pain could be due to coronary insufficiency.

The principle of functional testing lies in provoking myocardial ischemia by increasing the cardiac workload, and therefore oxygen demand. Gradual increase in workload can be achieved with exercise protocol utilizing physical activity, such as a treadmill or bicycle ergometer, or pharmacological agents, such as dobutamine, adenosine and dipyridamole. Coronary insufficiency is determined by the presence of poor exercise tolerance, and exercise- or pharmacologic-induced angina, hypotension, electrocardiographic changes of ischemia, and arrhythmia, and abnormalities in regional wall motion and perfusion. Cardiac imaging techniques with stress echocardiography and radionuclide

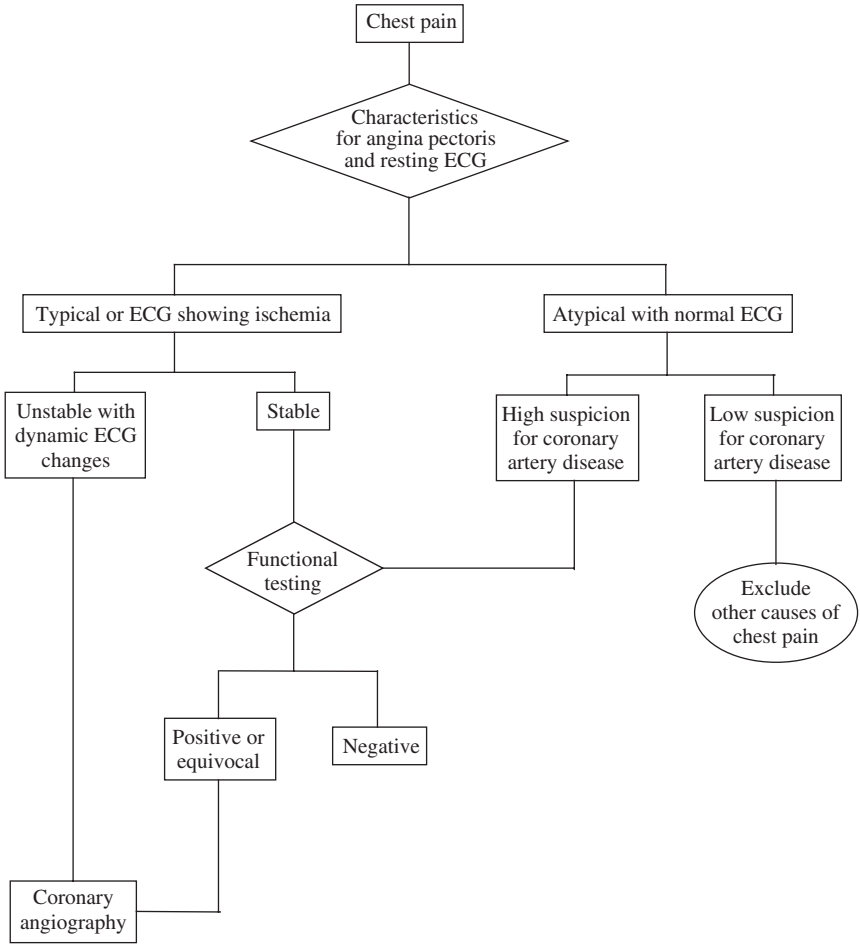


Fig. 4 Flow chart showing a clinical strategy for evaluation of chest pain. ECG = electrocardiogram.

scintigraphy, perfusion scanning or cardiovascular magnetic resonance imaging (CVMRI) are in general superior to ECG stress testing for the detection of IHD. In addition, electron-beam computed tomography (EBCT) can detect coronary artery calcification which correlates with the presence of coronary atheroscleorosis. However, the clinical significance of such findings, especially in asymptomatic subjects, is not clear. Myocardial viability and reversibility of hibernating myocardium due to chronic ischemia is best assessed using CVMRI or positron emission tomography (PET).

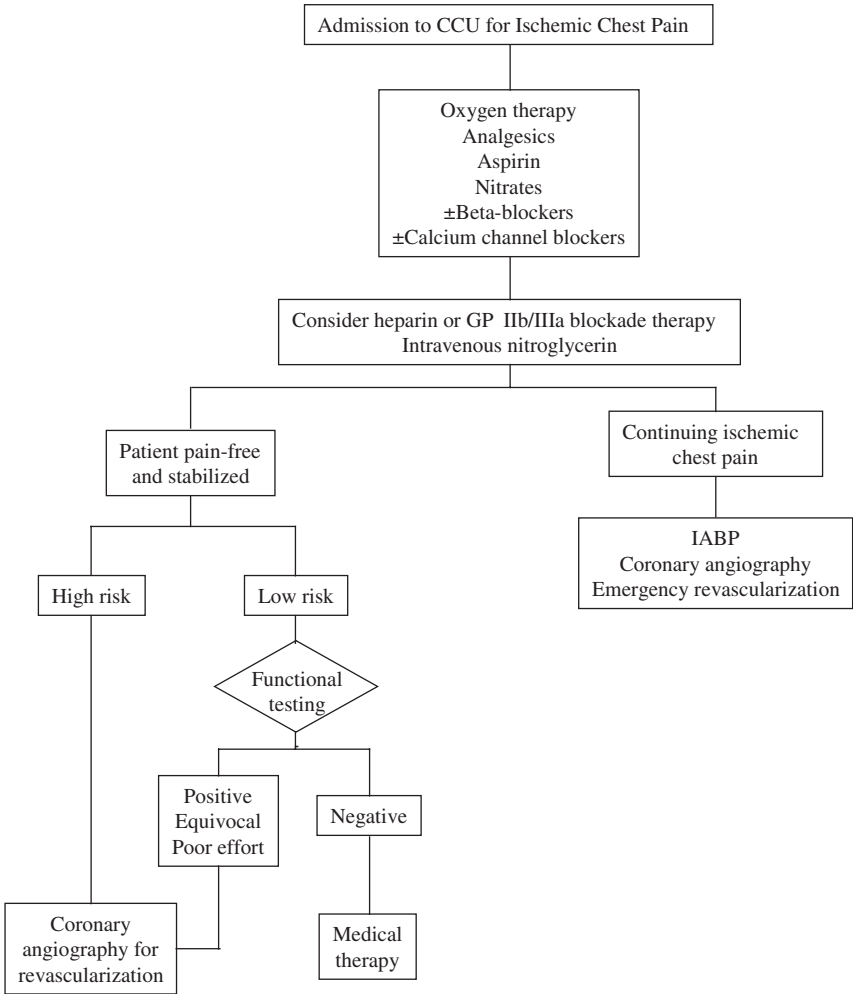


Fig. 5 A strategy for management of patients admitted to the coronary care unit.

CCU = coronary care unit; GP = glycoprotein; IABP = intraaortic balloon counterpulsation.

INVASIVE ASSESSMENT OF ISCHEMIC HEART DISEASE — CORONARY ANGIOGRAPHY

Coronary angiography remains the "gold standard" against which all other non-invasive tests are judged for the detection of IHD. However,

coronary angiography provides only an anatomical definition of the lumen while non-invasive cardiac investigations can provide information on haemodynamic and functional significance of an atherosclerotic lesion. The results of these tests are complementary and provide a sound basis for choosing the most appropriate treatment strategy for a patient.

Anatomically, the extent of coronary artery disease is often described in terms of the number and sites of lesions detected by coronary angiography. Extent of disease is classified as single-, double- or triple-vessel, with or without left main trunk involvement. The 3 major coronary vessels are the LAD and LCX branches of the left coronary artery and RCA. For a lesion to be clinically significant, there should be cross-sectional narrowing of $>50\%$. The presence of a significant left main or proximal LAD has the greatest bearing in determining treatment options. The hemodynamic or functional significance of an anatomic coronary lesion can be objectively assessed by non-invasive functional testing. Absolute CBF and fractional coronary flow reserve (FFR) before and after a stenotic lesion can be measured using Doppler pressure/flow wires introduced into the coronary arteries invasively. These data are helpful in guiding appropriate therapeutic options of IHD. Non-invasive methods of coronary angiography are currently being developed. These include magnetic resonance angiography (MRA) and EBCT coronary angiography. Both techniques are limited by poor image resolution, and thus unable to replace invasive coronary angiography yet. The combined diagnostic and therapeutic applications of coronary angiography help to maintain its prominent role in the current management of IHD. The use of coronary intravascular ultrasound to assess the presence and nature of atherosclerotic plaque is well-established. Because of its invasive nature, it is not routinely used as a diagnostic tool for IHD, but in conjunction with PTCA.

MANAGEMENT OF ISCHEMIC HEART DISEASE

Management of Patients with Chest Pain (Fig. 4)

The resting ECG determines the necessity of hospital admission. Those with ACS should be admitted to the coronary care unit (Fig. 5). A patient with a recent onset chest pain, either typical or atypical of angina, and resting ECG abnormalities consistent with myocardial ischemia, requires hospital admission to a monitored bed. High-risk patients ($\geqslant 2$ major

coronary risk factors) with a typical history of new onset angina should also be warded. Low coronary risk patients with atypical chest pain and low probability of IHD can be managed as an outpatient.

Stable Angina Pectoris

Pharmacologic therapy

Several pharmacological agents have been shown to be effective for the treatment of angina, both stable and unstable. These medications can be conveniently remembered with the alphabets A, B, C, E and G. **A** stands for Aspirin, **B** for Beta-blockers, **C** for Calcium blockers, E for ACE (angiotensin-converting enzyme) inhibitors and G for Glycerine trinitrates. Patients with IHD will benefit from one or more of these drugs. The conventional triple anti-anginal therapy refers to the combined use of nitrates, beta-blockers and calcium antagonists. A fourth class of anti-anginal therapy, the so-called metabolic or cytoprotective drugs which shift cardiac myocytes metabolism from the free fatty acid to the glycolytic pathway, has been introduced into clinical use.

Nitrates (Table 2)

Glycerine trinitrates (GTN), *dinitrates* and the more long-acting *mononitrates* are highly effective for the relief of anginal symptoms. Patients with proven or suspected IHD should be given nitrates; sublingual, topical, oral and intravenous, either alone or in combination. The intravenous form of nitrate administration is usually reserved for patients with unstable angina or ACS. As the major side-effects of nitrates are hypotension and headache, a lower dose may be given initially in susceptible individuals. It is useful to tell patients that headache often subsides after a few days of starting nitrate administration. The concomitant administration of nitrates and sildenafil citrate, used for treating erectile dysfunction, is particularly hazardous and may lead to severe hypotension and death.

Beta-blocking agents (Table 3)

Beta-adrenergic blockers have proven benefits in the treatment of stable chronic IHD, as well as impending MI or post-MI angina. They are also effective anti-hypertensive agents. This group of drugs is relatively

Table 2 Nitrates: Formulation and Dosage

Formulation	Dose	Frequency	Onset of Action (minutes)	Duration of Action
Rapid-acting, Short duration				
Sublingual nitroglycerin tablet	0.3–0.6 mg	When required	2	20–30 minutes
Oral nitroglycerin spray	0.4 mg	When required	2	20–30 minutes
Sublingual isosorbibe dinitrate tablet	2.5–10 mg	When required	5–10	1–2 hours
Rapid-acting, Intermediate duration				
Oral nitroglycerin, slow release	2.6, 5.2 mg	Three times a day	2–5	3–5 hours
Oral isosorbibe dinitrate tablet	10–30 mg	Three times a day	15	4–6 hours
Oral isosorbibe mononitrate tablet	20 mg	Twice a day	30	5–7 hours
Nitroglycerin ointment (2%)	0.5–2″	Twice a day	15	8 hours
Long-acting				
Oral isosorbibe dinitrate, slow release	80–120 mg	Daily	~60	10–12 hours
Oral isosorbibe mononitrate, slow release	60–240 mg	Daily	~60	10–14 hours
Transdermal nitroglycerin patch	0.2–0.8 mg/hr	Daily	30	12–24 hours

Table 3 Common Beta-adrenergic Receptor Blockers

Name (*Proprietary*)	Beta-1-Selectivity	Intrinsic Sympathomimetic Activity	Frequency	Total Daily Dose (mg)
Acebutolol (*Sectral*)	yes	Partial	Twice daily	200–600
Atenolol (*Tenormin*)	yes	No	Daily	50–200
Metoprolol (*Betaloc*)	yes	No	Twice daily	100–400
Slow release (*Betaloc Zok*)			Daily	100–400
Nadolol (*Corgard*)	no	No	Daily	80–240
Pindolol (*Visken*)	no	Yes	Two to three times daily	15–45
Propranolol (*Inderal*)	no	No	Two to three times daily	40–360
Long-acting (*Inderal LA*)			Daily	60–320
Sotalol (*Sotacor*)	no*	No	Twice daily	160–480
Timolol (*Blocadren*)	no	No	Twice daily	15–45

* type 3 anti-arrhythmic effect.

contraindicated in patients with a history of asthma or bronchospasm, severe peripheral vascular disease, heart blocks and insulin-dependent diabetes mellitus. It may elevate serum LDL-cholesterol in some individuals. Unless contraindicated, beta-blocking agents, together with aspirin, should be the first-line therapy for the treatment of IHD.

Calcium antagonists (Table 4)

C*alcium channel blockers*, especially short-acting dihydropyridines popularly used for treating systemic hypertension, may be potentially harmful when administered to patients with unstable angina. Part of this adverse association may be related to the reflex adrenergic activation. Nonetheless, calcium antagonists are effective in alleviating symptoms of stable angina, especially the long-acting formulations such as *amlodipine*. Generally, when a dihydropyridine is used, concomitant beta-blockade is recommended.

The use of different calcium antagonists in patients with IHD must take cognizance of their vasodilatory, negative-inotropic and anti-arrhythmic properties. Dihydropyridines with potent vasodilatory effects are most suitable for treating patients with hypertension, especially those with impaired LV function. Calcium channel blockers with negative-inotropic effects such as verapamil should be avoided in patients with poor LV function. Conversely, they may be useful for patients with IHD associated with tachyarrhythmias.

Metabolic or cytoprotective agents

In the presence of ischemia, cardiac myocytes shift their energy substrate preferentially from the normal free fatty acid to glucose metabolism. An indirect way to increase glucose uptake and decrease circulating fatty acid levels is to use an infusion of glucose and insulin. The more direct approach is to stimulate myocardial glucose oxidation by activating pyruvate dehydrogenase, the rate-limiting enzyme for glucose oxidation. The contribution of glycolysis to adenosine triphosphate (ATP) production increases during mild ischemia of the heart. During severe ischemia, oxidation of carbohydrate and fatty acids effectively ceases, and glycolysis becomes the major source of ATP production. A new class of anti-anginal agents has been developed to shift energy substrate preference in the heart under ischemic condition. *Trimetazidine* and *Ranolazine* act directly

Table 4 Calcium Channel Blockers

Agent (*Proprietary*)	Dose (mg)	Frequency	Heart Rate	Effect on Atrioventricular Node	Blood Pressure
Dihydropyridines					
Amlodipine (*Novasc*)	2.5–10	Daily	±	±	↓↓
Felodipine (*Plendil*)	5–20	Daily	↑	±	↓↓
Isradipine (*Dynacirc*)	2.5–10	Daily	↑	±	↓↓
Lacidipine (*Lacipil*)	2–6	Daily	↑	±	↓↓
Nicardipine (*Cardibloc*)	10–20	Thrice daily	↑	±	↓↓
Nifedipine (*Adalat*)	5–30	Thrice daily	↑	±	↓↓
Long-acting (*Adalat GITS*)	30–180	Daily	±	±	↓↓
Nilvadipine (*Escor*)	4–8	Daily	↑	±	↓↓
Non-dihydropyridines					
Diltiazem (*Herbesser*)	30–90	Three to four times daily	↓↓	↓	↓
Long-acting (*Herbesser Retard*)	60–360	Once or twice daily	↓↓	↓	↓
Verapamil (*Isoptin*)	40–120	Three to four times daily	↓	↓↓	↓
Long-acting (*Isoptin Retard*)	120–240	Once or twice daily	↓	↓↓	↓

± = little or no effect; ↓ = mild reduction; ↓↓ = moderate reduction; ↑ = mild elevation.

by inhibiting fatty acid oxidation and increasing glucose oxidation. These agents are routinely used in conjunction with the other 3 established groups of anti-anginal agents.

Anti-platelet agents

The other important pharmacological group of agents of well-proven benefits in the treatment of IHD is the *anti-platelet agents*, of which *aspirin* is the most time-tested. Newer oral anti-platelet agents such as *ticlopidine* and *clopidogrel* are especially useful in patients who cannot tolerate aspirin.

Statins and other lipid-lowering agents

Recent large primary and secondary prevention studies of coronary artery disease clinical trials have shown marked reduction in cardiac and stroke event rates in patients treated with *statins* or *3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase inhibitors* (refer to chapter 17 on hyperlipidemia). These reductions, up to 40–50%, seem out of proportion to the degree of lesion regression as shown by quantitative angiographic studies. The reasons lie in the anti-inflammatory effect or pleotrophic effects of statins in plaque stabilization. Statins should be routinely prescribed to patients with IHD with serum LDL-cholesterol levels >100 mg/dL (2.56 mmol/L) for secondary prevention.

Angiotensin-converting enzyme (ACE) inhibitors

ACE inhibitors are shown to have anti-ischemic and anti-arrhythmic properties. The use of ACE inhibitors in AMI has been associated with reduction in mortality. However, further studies are necessary to determine if they are indeed efficacious for patients with stable angina pectoris.

Coronary Revascularization

General considerations

Coronary revascularization procedures have become a major therapeutic modality among patients with IHD. Although coronary artery

bypass grafting (CABG) has a much longer history, PTCA is by far the commonest procedure of coronary revascularization in many countries. The number of coronary revascularization procedures, by PTCA and CABG, performed in the Asia-Pacific countries are summarized in Tables 5 and 6, respectively. Undoubtedly, both procedures are effective in improving the quality of lives of patients with symptomatic IHD but not necessarily improving the survival of patients. Indiscriminate use of both therapeutic procedures without appropriate clinical indications will inevitably escalate healthcare costs, and may increase morbidity and mortality.

For improvement in the quality of life (QOL) issues, coronary revascularization is indicated for patients whose symptoms are refractory to medical therapy. Refractory medical therapy is defined as a period of medical therapy with maximum dosage of all 4 classes of anti-anginal therapy which the patient is able to tolerate. Often the patients would be admitted to hospital with administration of intravenous nitroglycerin as

Table 5 Percutaneous Coronary Intervention in the Asia-Pacific Region in 1998

Country	Population (millions)	Number of		Rate (per 100 000)	Year Started
		Procedures	Centers		
Australia	17	15 000	45	88.2	1981
Bangladesh	125	108	1	0.1	1995
China	1300	5000	90	0.4	1984
Hong Kong	6.8	3000	15	44.1	1984
India	970	14 000	62	1.4	1985
Indonesia	200	1260	13	0.6	1987
Japan	125	100 000	1200	80	1981
Korea (South)	45	10 000	32	22.2	1983
Malaysia	20	1200	15	6.0	1984
New Zealand	3.5	2600	7	74.3	1981
Pakistan	130	1150	10	0.9	1995
The Philippines	75	439	4	0.6	1989
Singapore	3.5	3000	4	85.7	1984
Taiwan	21	9500	24	45.2	1981
Thailand	60	2100	16	3.5	1989
Vietnam	80	50	1	0.1	1996
TOTAL	3181.8	168 407	1539	5.3	

Information provided by the Asia-Pacific Society of Interventional Cardiology.

well. The failure of pharmacological treatment may be attributed to the inability of these drugs to control symptoms at their maximum doses, or when the patient is unable to tolerate or suffers from adverse effects from the medications. For improvement in survival outcomes, there are particular angiographic subsets of patients who may benefit more from coronary revascularization. In general, the worse the disease as judged by the number of major vessels involved and the state of the LV function, the greater the benefit from surgical revascularization. Left main and multivessel disease, especially those with poor LV function, will benefit from surgery from the long-term survival point of view. Recent evidence from the Bypass Angioplasty Revascularization Investigation (BARI), a large randomized controlled trial, has shown that diabetic patients with multivessel disease faired poorly from both CABG surgery and PTCA. However, survival was lower for PTCA. A major limitation was that this trial was conducted prior to the widespread use of coronary stenting in the angioplasty treatment arm. Nonetheless, recent studies with the use of coronary stents continued to show the superiority of CABG for patients with diabetes.

Table 6 Coronary Artery Bypass Grafting in the Asia-Pacific Region

Country	Population (millions)	Number of		Rate (per 100 000)	Year Obtained
		Procedures	Centers		
Australia	18.5	20 050	42	108.4	1997
Bangladesh	122	15	2	0.01	1996
Hong Kong	6.4	339	7	0.5	1996
India	970	20 264	?	2.1	1996
Indonesia	204	650	16	0.3	1996
Israel	6	5340	15	89.0	1996
Japan	126	13 688	413	10.9	1996
Korea (South)	46	1083	54	2.4	1996
Malaysia	21	739	12	3.5	1996
Pakistan	138	1487	11	1.1	1996
Singapore	3.3	1939	4	58.8	1997
Taiwan	22	1064	88	4.8	1996
Thailand	63	2657	15	4.2	1997
TOTAL	1746.2	69 315	>679	4.0	

Information provided by the Asia-Pacific Society of Cardiology.
? = unknown

Surgical coronary revascularization vs. medical therapy

Most of the evidences on the comparison of CABG and medical therapy are provided by 3 large landmark prospective randomized trials. They are the European Coronary Surgery Study (ECSS), the Veterans Administration (VA) Coronary Artery Bypass Cooperative Study Group, and the Coronary Artery Surgery Study (CASS). These trials were conducted prior to routine use of arterial graft conduits and represented comparison mainly of saphenous vein bypass surgery to medical therapy. Taken together with other smaller trials and retrospective data from established registries, these studies demonstrated survival benefit for the surgically treated group in patients with left main or left main "equivalent" coronary artery disease and multi-vessel coronary artery disease, especially when proximal LAD is involved and LV dysfunction is impaired. The majority of the patients in these earlier studies were <65 years of age and women were not frequently enrolled. Of these 3 trials, only the CASS utilized the left internal mammary artery as a conduit, and in a relatively small proportion of patients (14%). Furthermore, in the medically treated group, aspirin was not widely used, and ACE inhibitors and statins were not available then. Despite these limitations, the superiority of surgical revascularization in these subsets of coronary artery disease patients was clear. The number of diseased major epicardial arteries involved, the amount of myocardial ischemia and the degree of global LV dysfunction are the major predictors of surgical risk or success.

Percutaneous coronary revascularization vs. medical therapy

PTCA is now a well-established standard treatment modality for symptomatic IHD. Patients without significant symptoms of angina are also candidates for revascularization by PTCA if objective evidence of a large area of reversible myocardial ischemia can be demonstrated.

There is a paucity of clinical trials comparing PTCA and medical therapy. This limitation is further confounded by the rapid advancement of interventional technologies and adjunctive pharmacotherapy. Most comparison studies were almost obsolete by the time the results were published. The ACME (A Comparison of Angioplasty with Medical Therapy in the Treatment of Single Vessel Coronary Artery Disease) study was the

first attempt to compare medical therapy with PTCA. This study randomized 212 patients with stable angina, exercise-induced myocardial ischemia, and the presence of a 70–99% narrowing of a major epicardial artery, to PTCA or medical therapy with a 6-month follow-up period. The investigators found a qualitative but not survival benefits in patients treated by PTCA. More patients in the PTCA group were free of angina and required less medications at the end of the study (64% vs. 46%, $P < 0.01$). PTCA patients also experienced an increase in exercise duration (2.1 vs. 0.5 minutes, $P < 0.0001$). However, more PTCA patients required target vessel revascularization.

A recent study compared aggressive lipid-lowering therapy to PTCA as therapeutic strategies for patients with stable angina pectoris. In the Atorvastatin Versus Revascularization Treatments (AVERT) study, 341 patients with single- or double-vessel coronary disease were randomly allocated to medical therapy, including a potent statin, atorvastatin, at a high dosage of 80 mg/day, or PTCA with usual care, including standard doses of lipid-lowering therapy. At the end of 18 months, the occurrence of the composite endpoint, consisting of death, resuscitated cardiac arrest, non-fatal MI, cerebrovascular accident, non-protocol revascularization procedure, objective worsening of angina or repeat hospitalization, was 13% in the atorvastatin group and 21% in the angioplasty group. This difference amounted to a 36% relative reduction for patients receiving atorvastatin ($P = 0.048$). The majority of the benefit was derived from the excess number of patients with worsening of angina or hospitalization and non-protocol revascularization. However, there was little difference in mortality or non-fatal MI rates.

From these studies, PTCA appears to be effective in relieving angina symptoms and improving exercise tolerance in the short- and intermediate-term. However, the procedure did not improve survival and may increase the need for additional revascularization procedures. Aggressive therapy of the accompanying coronary risk factors, in particular elevated LDL-cholesterol, is associated with improved outcomes.

Surgical vs. percutaneous coronary revascularization

In general, when comparing CABG with PTCA, clinical outcomes were improved among high-risk patients undergoing surgical revascularization. High-risk subsets are those with double- or triple-vessel coronary

atherosclerotic disease, especially where the LAD artery (with $\geqslant 95\%$) narrowing is involved. Patients with low baseline risk fared better with PTCA. For patients with moderate baseline risk, there was a trend towards improved outcomes for patients treated with PTCA.

The issues relating to the comparison between CABG and PTCA among patients with multi-vessel disease are quite similar to those in patients with single-vessel disease. There was no difference in mortality after a 2.7-year period of follow-up from a pooled analysis of 3371 patients from 9 randomized trials comparing these 2 treatment strategies (4.4% CABG; 4.6% PTCA). However, the rate of repeat revascularization procedures at 1-year were considerably higher among patients who underwent PTCA (33.7% vs. 3.3%; $P < 0.0001$). In the more recent BARI study, the 5-year mortality rate was similar among 1829 patients randomly assigned to CABG (10.7%) or PTCA (13.7%) (absolute difference 3.0%; 95% confidence interval -0.2–6.0%; $P = 0.19$). Again, repeat revascularization procedures were performed more frequently among those treated with PTCA (54%) compared to CABG (8%) at 5 years. An important finding in this study was the higher mortality among patients with diabetes mellitus who were treated with PTCA (34.5%) compared with CABG (19.4%; $P = 0.003$).

Tremendous improvement has been achieved in modern day pharmacologic therapy of stable angina. New anti-platelet and lipid-lowering agents and heart failure treatment regimens have greatly improved the option of medical therapy. At the same time, several new devices in PTCA have enabled the procedure to be performed more safely, successfully and with lower restenosis rate. In particular, coronary stenting has been shown to reduce the incidence of abrupt vessel closure and the rate of restenosis following PTCA. Better stent design and deployment techniques have also reduced the occurrence of stent thrombosis and eliminated the need for anti-coagulation after stenting. Correspondingly, substantial improvement in surgical revascularization techniques has also been occurring. The widespread use of internal mammary artery and probably other arterial conduits has enhanced long-term graft patency. In addition, better myocardial preservation techniques, adjunctive pharmacological therapies, and less invasive surgical approaches, the so-called "minimal invasive surgical techniques," have significantly improved the outcome of CABG.

In summary, patients with left main trunk disease or "equivalent" and triple-vessel coronary artery disease are best treated surgically, for optimal long-term clinical outcome. Otherwise, coronary revascularization should only be recommended for those with angina pectoris refractory to medical therapy. The principal limitation of PTCA is restenosis leading to increased rates of repeat revascularization procedures. Patients with diabetes mellitus and multi-vessel diffuse atherosclerotic disease belong to a special subset of patients in whom CABG is likely to be a better therapeutic option in the long-term.

MANAGEMENT OF ACUTE CORONARY SYNDROMES

The strategy for the management of ACS is summarized in Fig. 5. In addition to anti-ischemic therapy described in the section on management of stable angina pectoris (p. 40 of this chapter), several other pharmacological approaches are directed against the thrombus, the principal culprit in ACS.

Anti-platelet Therapy

Aspirin, Ticlopidine, Clopidogrel

As platelet adhesion and aggregation induced by plaque rupture or erosion is the initiating event of ACS, anti-platelet therapy is pivotal in the management of this condition. There is little dispute in the beneficial role of aspirin for the treatment of patients with ACS. Its main action is to inhibit the enzyme, cyclo-oxygenase, which generates thromboxane A_2 that subsequently induces platelet aggregation. Indeed, aspirin has been shown to reduce the relative risk of fatal or non-fatal MI by 71% in the acute phase, 60% at 3 months, and 52% at 2 years. When there are no contraindications, aspirin should be the first-line anti-platelet agent used to treat all patients with ACS (Table 7).

Thienopyridines, such as ticlopidine and clopidogrel, is another group of anti-platelet agents which block adenosine diphosphate-induced

Table 7 Randomized Trial of Aspirin for Patients with Acute Coronary Syndromes

Study/Author	Patients	Aspirin		MI or Death		Relative Reduction (%)	*P*-value
		Duration (months)	Daily Dose (mg)	Control (%)	Aspirin (%)		
Unstable Angina Trials							
Lewis	1266	3	324	10.1	5.0	51	<0.001
Cairns	278	24	1300	14.1	11.5	20	0.008
Théroux	239	<1	650	11.9	3.3	72	0.012
Wallentin	288	3	75	17.6	7.4	58	0.004
Pooled	2071	~6		11.8	6.0	49	<0.001
Acute MI Trials							
Antiplatelet Trialists' Collaboration	20 543	~1	~150–325	14.4	10.6	26	<0.001

MI = myocardial infarction.

platelet aggregation. Ticlopidine has been shown to reduce the relative risk of fatal and non-fatal MI by 46% compared to those receiving placebo. It is recommended as an alternative agent to aspirin. However, platelet and neutrophil counts have to be closely monitored among patients receiving ticlopidine as severe neutropenia may occur. In contrast, clopidogrel, a new longer-acting anti-platelet agent with fewer adverse effects than ticlopidine is effective in preventing vascular events among patients with previous MI, stroke or peripheral vascular disease. A randomized trial, CAPRIE (Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events), demonstrated a relative-risk reduction of 8.7% in favor of clopidogrel. The CURE (Clopidogrel in Unstable angina to prevent Recurrent ischemic Events) trial showed the benefits of the combination of aspirin and clopidogrel in the treatment of patients with ACS compared with aspirin alone. However, the 20% relative reduction of cardiovascular death, MI or stroke was accompanied by an increase in life-threatening bleeding events (2.2% vs. 1.8%).

Glycoprotein (GP) IIb/IIIa receptor blockade

Recently, molecular techniques have identified the platelet surface GP IIb/IIIa receptor as the final common pathway for platelet aggregation. This integrin binds fibrinogen which cross-links adjacent platelets to form the platelet plug. The anti-platelet aggregation action of aspirin, ticlopidine and clopidogrel is, at best, modest because they inhibit only one of the several pathways for platelet aggregation and therefore do not completely prevent platelet aggregation. Currently, >100 agonists have been identified which are capable of activating the GP IIb/IIIa receptor directly or indirectly. A new family of potent anti-platelet agents that block the GP IIb/IIIa receptor has been developed.

The "4P" trials are 4 different clinical studies which evaluated the peptide and non-peptide GP IIb/IIIa inhibitors, *lamifiban, tirofiban* and *eptifibatide* in ACS. In all these trials, aspirin was also given. These trials clearly demonstrated the efficacy and safety of the agents (Table 8). Contrary to what was expected, oral GP IIb/IIIa inhibitors were not shown to be effective, and may be related to doses administered and possibly agonistic actions.

Table 8 Use of Glycoprotein IIb/IIIa Inhibitors in Acute Coronary Syndromes: Clinical Outcomes of the 4P Trials

Characteristic	PARAGON			PRISM			PRISM-PLUS			PURSUIT	
Study Group	*Placebo*	*Lamifiban 1 μg/min ± Heparin*	*Lamifiban 5 μg/min ± Heparin*	*Heparin*	*Tirofiban 0.15 μg/kg-min*	*Heparin*	*Heparin + Tirofiban 0.10 μg/kg/min*	*Tirofiban 0.15 μg/kg-min*	*Placebo*	*Eptifibatide 1.3 μg/kg-min*	*Eptifibatide 20 μg/kg-min*
N	758	755	769	1616	1616	797	773	345	4739	1487	4722
30-day outcome											
Death, %	2.9	3.0	3.6	3.6	2.3	4.5	3.6	(6.1)	3.7	(3.4)	3.5
Non-fatal MI, %	10.6	9.4	10.9	4.3	4.1	9.2	6.6	(9.0)	13.5	(12.0)	12.6
Death or MI, %											
Overall	11.7	10.6	12.0	7.1	5.8	11.9	8.7	(13.6)	15.7	(13.4)	14.2
Relative reduction, %		9	−6		18		27				10
PTCA patients				9.1	7.2	10.2	5.9		16.8		11.8
Non-PTCA patients				6.2	3.6	7.8	3.6		15.7		14.6
6-month outcome											6.4
Death, %	6.6	5.2	6.8			7.0	6.9	(7.2)	16.2		14.7
Non-fatal MI, %	14.3	10.8	12.9			10.5	8.3	(10.1)	15.7		17.8
Death or MI, %	17.9	13.7	16.4			15.3	12.3	(15.9)	19.0		8
Relative reduction, %		23	8				20				3.0
Major bleeding*, %	3.0	3.0	6.0	0.4	0.4	0.8	1.4		1.3		0.1
Intracranial hemorrhage, %	0	0	0.1	0.1	0.1	0	0		0.1		4.4
RBC transfusion†, %	4.4	4.4	8.7	1.4	2.4	2.8	4.0		1.8		0.6
Thrombocytopenia, %	1.1	1.5	1.3	0.1	0.4	0.3	0.5		0.4		

MI = myocardial infarction; PTCA = percutaneous transluminal coronary angioplasty; RBC = red blood cells; Thrombocytopenia defined as platelet count $< 50\,000\,mm^3$.

Numbers in parentheses () are from discontinued treatment arms and are not contemporaneous; these are listed only for completeness, not direct comparisons.

* Major bleeding as defined by intracranial hemorrhage or decrease in hemoglobin $> 5\,g/dL$ not associated with coronary artery bypass grafting (CABG).

† Transfusions reported are not associated with CABG, except for PARAGON. See text for the full names of trials.

Anti-coagulation

Unfractionated heparin

Heparin is the essential co-factor for anti-thrombin III, which is the principal endogenous inhibitor to thrombin. Compared with placebo, several clinical trials have shown the beneficial effects of heparin for the treatment of unstable angina (Table 9). However, when compared with aspirin, evidence for the ability of heparin in combination with aspirin to prevent additional MI or death is less compelling. A meta-analysis of 6 trials with >1300 patients who were randomly assigned to heparin (with the activated partial thromboplastin time titrated to 1.5–2.0 times control) with aspirin (75–650 mg per day) showed that there was only a favorable trend (relative risk = 0.67) during the in-hospital period ($P = 0.06$) and which was lost at follow-up weeks or months later (relative risk = 0.82).

Low-molecular-weight heparin (LMWH)

Heparin has been used widely in the treatment of ACS. Due to its better bioavailability, a more consistent anti-coagulation is achieved with LMWH. In addition, LMWH can be administered subcutaneously, usually without the need for monitoring. Furthermore, they are less likely to activate platelets and have lower incidence of heparin-induced thrombocytopenia. These factors potentially favor the use of LMWH over unfractionated heparin for the treatment of ACS.

Currently, several LMWH are commercially available. Their differences in molecular weight and affinity ratio to coagulation factor Xa to thrombin account for the dissimilarities in biological properties and probable efficacy. Nonetheless, the efficacy of LMWH in treating patients after ACS has been clearly shown in several large-scale clinical trials.

Direct thrombin inhibitors

Direct thrombin inhibitors are another group of agents that were evaluated for the treatment of ACS. *Disrudin* and *bivalrudin* failed to show any benefit in the management of ACS (GUSTO II and TIMI-9). The earlier results showed an excessive bleeding risk. Subsequently, when the dose was reduced there was no difference in the occurrence of the primary

Table 9 Heparin in Unstable Angina

Study (year)	Patient Number	Agents	Regimen	Endpoint Assessment	
				Definition	Frequency (%)
Telford (1988)	214	Heparin	4 × 5000 IU *IV*	Myocardial infarction	3
		Placebo			15
Théroux (1998)	479	Heparin	*IV* (aPTT 1.5 to 5 × control)	Refractory angina,	9.3
		Heparin + ASA	*IV* + 2 × 325 mg	MI,	11.5
		ASA	2 × 325 mg	and death	16.5
		Placebo			26.3
RISC (1990)	796	Heparin	4 × 5000 IU *IV* with bolus	MI,	5.6
		Heparin + ASA	*IV* with bolus + 75 mg	or death	1.4
		ASA	75 mg	at 5 days	3.7
		Placebo			6.0
Théroux (1993)	484	Heparin	*IV* (aPTT 1.5 to 2.5 × control)	Myocardial infarction	0.8
		ASA	2 × 325 mg		3.7
Cohen (1994)	214	Heparin + ASA + warfarin ASA	*IV* (aPTT 2× control) 162.5 mg	Recurrent ischemia, MI or death	10.5 27
Holdright (1994)	285	Heparin + ASA	*IV* (aPTT 1.5 to 2.5 × control) + 150 mg	MI or	27.3
		ASA	150 mg	death	30.5
Serneri (1995)	108	Heparin	*IV* or *SC* (aPTT 1.5 to 2 fold) + ASA	Recurrent ischemia	reduced
		ASA	325 mg		

IV = intravenous; *SC* = subcutaneous; MI = myocardial infarction; aPTT = activated partial thromboplastin time; ASA = aspirin; RISC = Risk of Myocardial Infarction and Death during Treatment with Low-dose Aspirin and Intravenous Heparin in Men with Unstable Coronary Artery Disease.

endpoint between those treated with *hirudin* and *heparin*. Currently, the efficacy of hirudin is being assessed at a moderate dose for the treatment of ACS in the OASIS (Organization to Assess Strategies for Ischemic Syndromes) study. This regimen is based on the favorable results obtained from a pilot study.

Another limiting factor is the reactivation of the thrombotic process following discontinuation of the anti-thrombin agents. The so-called "rebound" phenomenon accounts for the loss of efficacy. Current strategies are directed to provide a longer period of vessel passivation. Until more information is available, the use of direct thrombin inhibitors in ACS is still largely investigative.

Medical Therapy vs. Percutaneous Coronary Intervention in Acute Coronary Syndromes

In ACS, both thrombus and plaque determine the final patency of the lumen. Unlike medical therapy, PTCA expands the lumen and improves CBF, and should theoretically improve the outcome of patients with ACS.

Early studies (TIMI IIIB and VANQWISH (Veteran Affairs non-Q-wave Infarction Strategies in Hospital) did not show an advantage of an early invasive strategy, whereby coronary angiography was performed within 24–48 hours followed by revascularization if the coronary anatomy was suitable, compared with a conservative strategy, whereby coronary angiography was only performed when there was recurrent ischemia or positive stress testing. However, more recent studies (FRISC II and TIMI 18/TACTICS) clearly demonstrated the benefits of early revascularization strategy, particularly among those with high-risk characteristics. All patients were treated with LMWH or a combination of heparin and GP IIb/IIIa blockade.

Conclusions

From the currently available evidence, patients with ACS should be treated medically with anti-ischemic and anti-thrombotic therapy. Anti-ischemic therapy should be tailored to the individual patient, titrating to the optimal dosages tolerated. Unless contraindicated, aspirin should be administered to all patients. The choice between LMWH and unfractionated heparin as the anti-coagulant depends on the availability of these agents,

cost, experience and preference of the attending physician. When a GP IIb/IIIa inhibitor is administered, heparin should be used as well. Subsequent management of low- and intermediate-risk patients may be guided by functional testing. In high-risk patients, such as those with prolonged chest pain, compromised hemodynamics, recurrent ischemia, dynamic ECG ST-segment shift, transient LV dysfunction and elevated biochemical markers of ischemia, with previous MI, CABG or diabetes mellitus, invasive strategy is the preferred treatment modality. Mechanical support with intraaortic balloon counterpulsation is useful as an adjunct to invasive therapy. When early revascularization is planned, a GP IIb/IIIa inhibitor should be administered prior to and after PTCA.

MANAGEMENT OF ACUTE MYOCARDIAL INFARCTION

Pathophysiology of Acute Myocardial Infarction

Generally, when the duration of interruption of coronary flow is >20 minutes, irreversible myocardial damage starts to occur. By 2–3 hours, a large proportion of the myocardium supplied by the occluded or infarct-related coronary artery would suffer irreversible damage. At the end of 6 hours, most myocardial damage will be irreversible if collateral circulation is absent. Clinically, evidence of evolving MI is characterized by serial ECG changes of ST-segment elevation for transmural infarction and ST-segment depression or T-wave inversion for NSTEMI. Cardiac enzyme markers of myocardial necrosis include the very early rise of myoglobin (1 hour), troponin I and T (1–3 hours) and myocardial MB-fraction of CK (CK–MB) elevation (after 6 hours). Currently, the World Health Organization diagnostic criteria of AMI include the presence of 2 of these 3 major diagnostic criteria of chest pain, ECG ST-segment elevation and CK–MB enzyme rise. However, almost a quarter of patients with MI are unrecognized because of the absence of symptoms or presence of atypical symptoms, particularly among those with diabetes mellitus, the elderly and female gender.

In Singapore, the mortality rate from MI has declined steadily from 70 per 100 000 population in 1990 to 48 per 100 000 population in 1997. The incidence of AMI has also decreased steadily. This trend was more evident among patients $\geqslant 65$ years old or a 15% relative reduction in

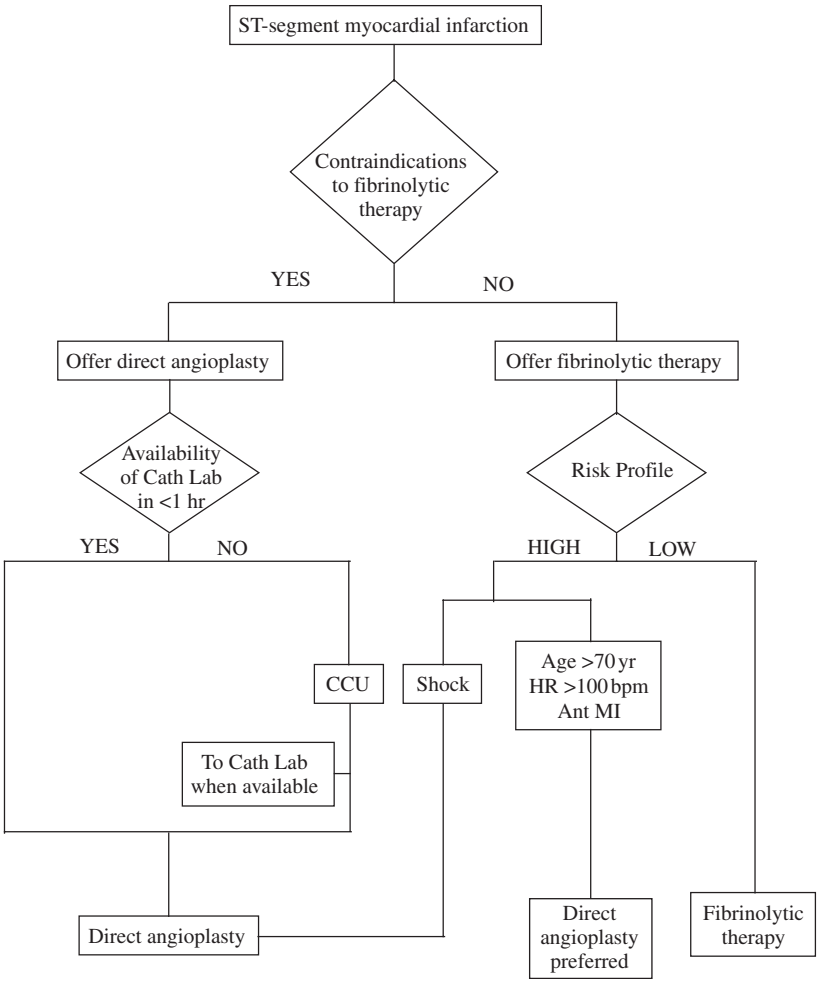


Fig. 6 Treatment strategy for patients with ST-segment myocardial infarction.

Cath Lab = cardiovascular catheterisation laboratory; CCU = coronary care unit; yr = year; HR = heart rate; bpm = beats per minute; Ant MI = anteroseptal myocardial infarction.

mortality. The exact reason for this favorable course is unclear. Improvement in medical therapies as well as the success of primary and secondary prevention strategies in the major risk factors modification may have played a part.

Electrocardiographic Changes of AMI (Table 10)

In the era of fibrinolysis, ECG changes are important in determining the type, site and age of the infarction. Arrhythmic complications associated with MI, such as atrial fibrillation (AF) and ventricular tachycardia (VT), can be confirmed by the ECG.

Biochemical Markers of AMI

Sequential elevation of the myocardial fraction of (CK–MB), aspartate transaminase and lactate dehydrogenase are evidence of recent myocardial necrosis. Several isoforms of these enzymes have been measured to estimate the size of MI. Of these, determination of the mass of the CK–MB is the most widely accepted mode to quantify the infarct size. More recently, raised cardiac components of troponin, particularly the T and I subunits, which are more sensitive and specific, have been used to determine myocardial damage.

Non-invasive Cardiac Imaging

Echocardiography is extremely useful in the evaluation of patients with AMI. M-mode and 2-dimensional echocardiography provides rapid assessment of the area of damage and residual regional and global LV function. Doppler echocardiography provides additional information, such as acute mitral regurgitation and ventricular septal or free wall rupture. Other non-invasive cardiac imaging modalities are less often used in the setting of AMI.

MANAGEMENT OF ACUTE MYOCARDIAL INFARCTION

Management and prognosis of AMI depends largely on the size of the infarction. The Killip Classification is designed to guide therapy and prognosis of AMI (Table 11). Pulmonary artery catheterization used to estimate cardiac output and pulmonary wedge pressure provides an accurate guide to fluid management in AMI, especially in hypotensive patients. Mechanical support with intraaortic balloon pumping is often necessary in Killip class III and IV patients in cardiogenic shock. More

Table 10 Characteristics of Various Biochemical Markers of Myocardial Necrosis

	Myoglobin	Total CK	CK-MB (mass)	MB2/MB1	cTnT	cTnI	LDH
Molecular weight (kDa)	17.8	85	85	NA	33	23.5	135
Cardiac-specific	No	No	++	++	+++	+++	No
Affected by renal function?	Yes	No	Yes	No	Yes	Yes	No
Initial detection (hours)	1–3	4–8	3–4	3–4	4–6	4–6	8–12
Duration of elevation	18–24 hrs	12–24 hrs	24–36 hrs	unknown	10–14 days	7–10 days	10 days
Rapid laboratory assay	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bedside assay	Yes	Yes	Yes	No	Yes	Yes	No

CK = creatine kinase; cTn = cardiac troponin; LDH = lactate dehydrogenase.

Table 11 Killip Classification of Acute Myocardial Infarction

Class	Proportion of Patients (%)	Systolic BP (mmHg)	Third Heart Sound	Urine Output (ml/h)	Chest Rales	Chest X-ray*	Wedge Pressure (mmHg)	CI (L/min/m^2)	Hospital Mortality (%)
I	30–40	>100	-	>30	-	-	<15	>2.5	8
II	30–50	>100	$\pm$	>30	$\pm$	+	<15	>2.5	30
III[†]	5–10	>90	+	>30	++	++	>30	1.8–2.5	44
IV[‡]	10	<90	+	<30	+++	+++	>30	<1.8	>80

*infiltrates
[†]acute pulmonary edema
[‡]cardiogenic shock
BP = blood pressure; − = absent; ± = may be absent or present; + = mild; ++ = moderate; +++ = severe.

aggressive therapy such as direct angioplasty and CABG surgery will improve the otherwise extremely poor outcome in patients with cardiogenic shock.

Early, complete and sustained restoration of coronary flow in the infarct-related vessel is the key to survival and myocardial salvage after AMI. Current pharmacological or catheter-based reperfusion therapies are aimed at improving myocardial salvage and survival in AMI. Patients eligible for reperfusion therapy are offered direct angioplasty, if there are any contraindications to fibrinolytic therapy. The cardiovascular laboratory and the intervention team are made available around the clock. If the patient consents and the laboratory is available within <1 hour, direct angioplasty with coronary stenting is the preferred treatment strategy for recent onset (<6 hours) AMI.

Pharmacological Reperfusion Therapy

In the early 1980s, demonstration of an occlusive thrombus as the underlying mechanism of AMI and the development of potent fibrinolytic agents were crucial in ushering in a new era in the management of patients with AMI. Pharmacologic reperfusion by fibrinolysis has been shown conclusively to improve survival in patients with AMI. This is now standard therapy for patients presenting with AMI onset of up to 12 hours' duration (Fig. 7). Early restoration of patency of the infarct-related vessel salvages myocardium in jeopardy of necrosis and preserves LV function. Late restoration of coronary flow, up to 24 hours, improves LV remodeling and decreases the occurrence of fatal arrhythmias.

Myocardial salvage is highly dependent on the ischemic period. Survival benefit and myocardial salvage is the greatest if reperfusion occurs within the first hour, the so-called "golden hour." Benefits from fibrinolysis diminish sharply after 3–6 hours and are almost non-existent after 12 hours unless collateral coronary circulation is present (Fig. 7).

There are several fibrinolytic agents currently in use. Streptokinase (SK) and anisoylated plasminogen streptokinase activator complex (APSAC or anistreplase) are the first generation fibrinolytic agents to be used. Recombinant tissue-plasminogen activator (TPA or alteplase) is currently the standard fibrinolytic agent against which newer fibrinolytic agents are compared. A new generation of fibrinolytic agents, commonly derived from deletion of some molecular fragments and modification of

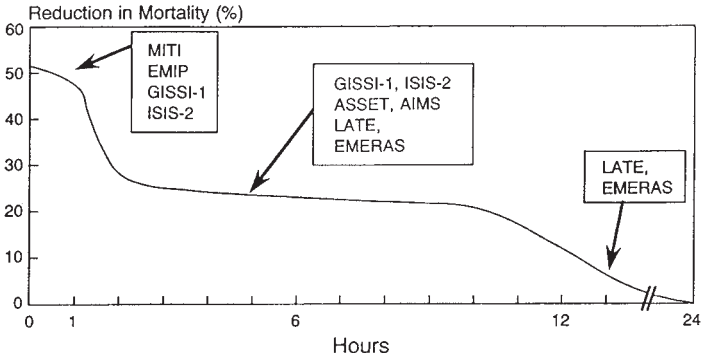


Fig. 7 Mortality reduction (percent) derived from fibrinolytic therapy as a function of time between onset of symptom and initiation of fibrinolytic agent.

Data have been extrapolated. A greater than 50% reduction in mortality is noted during the "golden first hour," after which the mortality benefit declines to a plateau of approximately 25% reduction until 12 hours after symptom onset. Beyond 12 hours, significant benefit has not been demonstrated with fibrinolytic administration. AIMS = Anistreplase Intervention Mortality Study; ASSET = Anglo-Scandinavian Study of Early Thrombolysis; EMERAS = Estudio Multicentrico Estreptoquinasa Republicas de America Del Sur; EMIP = European Myocardial Infarction Project; GISSI-1 = Gruppo Italiano per lo Studio della Sopravvienza nell'Infarto Miocardico-1; LATE = Late Assessment of Thrombolytic Efficacy; MITI = Myocardial Infarction and Triage Intervention Project (reprinted with permission from the *American Heart Association*).

the wild-type TPA, such as reteplase (rPA), lanoteplase (nPA) and tenecteplase (TNK-TPA), saruplase (scu-PA), and staphylokinase have been successfully developed and clinically tested. Compared to TPA, they are more resistant to PAI-1 and therefore, in theory, more potent. Because of their slow hepatic degradation and longer half-life, they can be given as bolus injections instead of infusion. Several trials have been conducted to compare the efficacy of various fibrinolytic agents and have shown them to be comparable to TPA.

A major concern of fibrinolysis is bleeding complications and specific contraindications have been recommended (Table 12). Fortunately, most of these are mild. Intracerebral hemorrhage is the most serious complication but occurred rarely in 0.4% with SK, 0.6% with APSAC and 0.7% with TPA. The main factors associated with intracranial bleeding are advanced age, systemic hypertension, low body weight, history of cerebrovascular accident and use of TPA.

Table 12 Contraindications to Fibrinolytic Therapy

Absolute	Relative
Increased bleeding risk	*Increased bleeding risk*
Active internal bleeding	History of bleeding diathesis
Recent* surgery or trauma as a potential bleeding source	Concurrent use of anti-coagulants surgery or trauma > 2 weeks
Recent* puncture of a non-compressible vessel	Active peptic ulcer disease
	Hemorrhagic retinopathy
Increased risk of intracranial event	*Increased risk of intracranial event*
History of hemorrhagic stroke	Past history of non-hemorrhagic stroke
History of intracranial neoplasm	Severe hypertension > 180/100 mmHg
Recent history† of non-hemorrhagic stroke	
Severe hypertension > 200/120 mmHg	
Prolonged‡ or traumatic CPR	Previous treatment¶ with streptokinase§
Aortic dissection	
Pregnancy	
Allergy to streptokinase§	

*< 2 weeks; †< 3 months; ‡> 10 minutes; ¶48 hours to 2 years.
§use non-streptokinase-related agents or direct angioplasty.

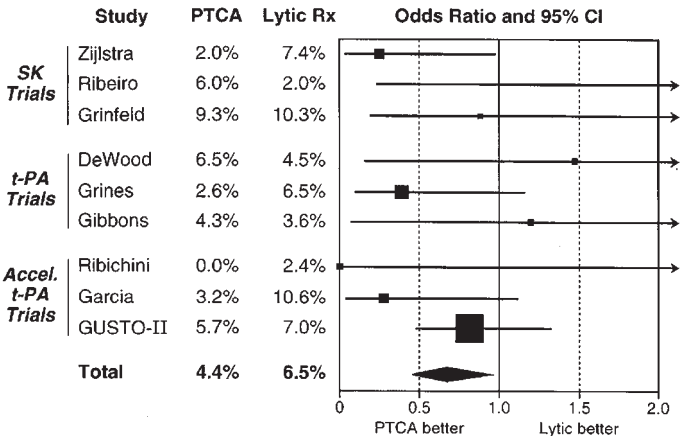


Fig. 8 Mortality at the end of study period in randomized clinical trials comparing direct angioplasty with fibrinolytic therapy.

The rates for each study are shown and are grouped by fibrinolytic drug regimen. The odds ratio and 95% confidence interval (CI) are plotted. SK = streptokinase; t-PA = tissue-type plasminogen activator (reprinted with permission from the *American Medical Association*).

The combination of a GP IIb/IIIa inhibitor (abciximab) with reduced-dose fibrinolytic agent (TPA or rPA) has been shown to improve patency of infarct-related vessel. However, the angiographic improvement did not translate into survival benefit in large-scale studies (GUSTO V and ASSENT 3).

Mechanical Reperfusion Strategy

Compared to fibrinolysis, the major advantages of mechanical lysis using direct angioplasty in AMI are the increased likelihood of restoring normal coronary flow, lower bleeding complications and knowledge of the anatomic coronary disease for better definitive treatment. Furthermore, we have found that direct angioplasty did not activate thrombin compared with fibrinolysis. These advantages may translate into better survival. Many recent randomized trials showed that coronary stenting was superior to balloon PTCA in the invasive treatment of AMI. By providing a larger lumen, coronary stenting improves the number of patients with TIMI 3 flow, decreases the rate of reocclusion and late restenosis. We have also shown that patients with diabetes mellitus derived particular benefit from mechanical reperfusion compared with fibrinolysis. With better outcomes compared with fibrinolysis, direct angioplasty is offered to patients with ST-segment elevation MI who are eligible for reperfusion therapy on a 24-hour basis at the National Heart Centre.

Other Adjunctive Pharmacological Therapy

Anti-platelet Therapy

Aspirin, unless contraindicated, is used routinely for the lifelong treatment of AMI. Ticlopidine and clopidogrel are alternatives as discussed earlier.

Anti-coagulation

Following fibrinolysis with TPA, there is increased thrombin generation and anti-coagulation may prevent reocclusion and subsequent reinfarction. Furthermore, aspirin is not effective in preventing thrombin-induced

platelet aggregation. Indeed, when heparin was used, infarct-related vessel patency was better maintained among patients treated with TPA. Generally, AMI patients are maintained on LMWH or unfractionated heparin anti-coagulation for 48–72 hours. Unless there are other indications, such as LV thrombus or deep vein thrombosis, heparin can be discontinued.

Nitrates (Table 2)

The ability of nitroglycerin to reduce preload, afterload and vasospasm enhances its therapeutic advantage in the treatment of AMI, particularly among patients with congestive heart failure. However, nitroglycerin has not been shown to improve survival.

Beta-adrenergic Blockers (Table 3)

In MI patients, beta-blockers are effective in improving survival, and reducing recurrent ischemia and infarction, infarct size, ventricular arrhythmias, AF and non-fatal cardiac arrest. Heart block, bronchospasm, hypotension and acute LV failure are relative contraindications of beta-blockade in AMI.

Angiotensin-converting enzyme (ACE) inhibitors (Table 13)

Several studies have clearly shown that the efficacy of ACE inhibitors in improving survival following AMI. This benefit is more evident among those with mild or moderate LV dysfunction (Killip class $\geq$II). The favorable effects provided by several of these agents were strikingly similar from a pooled analysis of $>$100 000 patients from 8 randomized trials involving $>$1500 patients. Of note, early administration of ACE inhibitors in AMI confers greater protection. Renal impairment and hypotension are relative contraindications.

Calcium channel blockers (Table 4)

Calcium channel blockers do not play a significant role in the management of AMI. In fact, short-acting dihydropyridines have been shown to have a deleterious effect on the clinical outcome of AMI.

Table 13 Use of Angiotensin converting Enzyme Inhibitors following Myocardial Infarction

Trial	Number	Agent	Inclusion Criteria	Initiation of Therapy	Follow-up	Relative Reduction (%)	*P*-value
SAVE	2231	Captopril	EF ⩽ 40%	3–16 days	3.5 years	19	0.019
AIRE	2006	Ramipril	CCF	3–10 days	1 year	27	0.002
TRACE	1749	Trandolapril	WMI ⩽ 1.2, EF < 35%	1–5 days	2–4 years	18	0.001
SMILE	1556	Zofenopril	Ant MI without fibrinolysis	< 24 hrs	1 year	24	0.198
CCS	13 634	Captopril	All	< 36 hrs	5 weeks	6	0.3
CONSENSUS-II	6090	Enalapril	All	< 24 hrs	6 months	-	-
ISIS-4	58 050	Captopril	All	< 24 hrs	5 weeks	7.0	0.02
GISSI-3	19 394	Lisinopril	All	< 24 hrs	6 months	11	0.03

EF = ejection fraction; CCF = congestive cardiac failure; WMI = wall motion index; Ant MI = anteroseptal myocardial infarction. SAVE = Survival and Ventricular Enlargement Trial; AIRE = Acute Infarction Ramipril Efficacy Study; TRACE = Trandolapril Cardiac Evaluation Study; SMILE = Survival of Myocardial Infarction: Long-term Evaluation; CCS = Chinese Cardiac Study; CONSENSUS-II = Cooperative New Scandinavian Enalapril Survival Study; ISIS-4 = International Study on Infarct Survival-4; GISSI-3 = Gruppo Italiano per lo Studio della Sopravvienza nell'Infarto Miocardico-3.

Other Management Strategies

Alleviation of pain and anxiety

Patients with MI can have severe pain and extreme anxiety. Prompt effective pain and anxiety relief is crucial in the management of these patients. Unless contraindicated, narcotics should be administered.

Rhythm disorders

Cardiac rhythm disturbances are common following MI. Peri-infarct arrhythmias usually occur in the first few days. The onset of VT or fibrillation (VF) is the most serious and may be related to inadequate oxygenation, electrolyte abnormalities and recurrent ischemia or infarction. Episodes with hemodynamic consequences are effectively treated by cardioversion. Less serious episodes of rapid AF or VT can be treated using pharmacological agents such as amiodarone. Patients who suffer from sustained monomorphic VT after 48–72 hours of MI may require electrophysiologic studies.

Heart blocks are generally associated with inferior wall MI. Temporary pacing may be necessary for patients with severe bradycardia and who are hemodynamically compromised. Although peri-infarct heart blocks of secondary degree usually recover spontaneously, some patients with complete heart block who do not recover after temporary pacing may require a permanent pacemaker.

Recurrent ischemia

Patients with recurrent ischemia are of particularly high risk and require urgent attention. Patients with NSTEMI or those treated with fibrinolysis are at a greater risk of this complication.

Acute left ventricular failure

Pump failure is a common complication following MI. The clinical spectrum ranges from congestive heart failure to cardiogenic shock. Although the extent of myocardial damage sustained following coronary occlusion is a major determinant, the baseline myocardial function is crucial. Unlike

patients with chronic heart failure, some of the myocardial injury may be potentially reversible (myocardial stunning) and the resultant LV dysfunction may be transient. However, there may be mechanical disruptions, such as rupture of papillary muscles or ventricular septum, resulting in acute LV failure or cardiogenic shock. These problems usually require surgical intervention. During the acute phase of heart failure, diuretics may be used in conjunction with afterload reducing vasodilators and inotropes. In selected patients, mechanical support with intraaortic balloon counterpulsation may be useful for the relief of symptoms of angina and to reduce the risks of revascularization procedures. Other forms of mechanical LV assisted devices can be used as a "bridge" to revascularization or heart transplantation.

Other complications include pericariditis, post-MI syndrome, deep vein thrombosis, pulmonary embolism, LV thrombus and systemic embolism, LV aneurysm and pseudoaneurysm formation.

Risk factor modification

A very important aspect of the management of patients with AMI is modification of coronary risk factors which will impact on subsequent progression of the disease. Of these, the most important risk modifications following AMI are smoking cessation, lipid lowering and blood pressure control. Stopping cigarette smoking will not only reduce the recurrence risk of IHD, but also save an enormous amount of money. Unfortunately, smokers often suffer from physical and psychological dependence which make smoking cessation difficult.

Cardiac rehabilitation (Table 14)

A comprehensive cardiac rehabilitation program is essentially for any patients recovering from AMI. In fact, all cardiac patients who had undergone PTCA or CABG surgery should routinely be referred to a full post-event cardiac rehabilitation program. It has been shown that cardiac rehabilitation is the single most cost-effective therapy to the society for patients with heart disease. The need for lifestyle changes and coronary risk factor modification will improve the general well-being, if not survival, of the patient.

Table 14 Secondary Prevention Treatment Goals

1) Smoking cessation	5) LDL-C $<$ 100 mg/dL (2.56 mmol/L)
2) BMI $<$ 25 kg/m^2	6) HDL-C $>$ 35 mg/dL (0.9 mmol/L)
Fat intake $<$ 60 g/day	7) BP $<$ 140/90 mmHg (DM: 130/80 mmHg)
Saturated fat intake $<$ 15 g/day	8) DM: HbA1c $<$ 7%
Cholesterol intake $<$ 200 mg/day	9) Aspirin, if no contraindications
3) Regular exercise	10) Beta-blockers, if no contraindications
4) $\leqslant$ 30 ml alcohol: hypt or $\uparrow$TG	11) ACE inhibitor: LVEF $<$ 40%
No alcohol: CHF	12) HRT: Postmenopausal women

BMI = body mass index; hypt = hypertensive; $\uparrow$TG = elevated serum triglycerides; LDL-C = low-density-lipoprotein cholesterol; HDL-C = high-density-lipoprotein cholesterol; BP = blood pressure; DM = diabetes mellitus; HbA1c = glycosylated hemoglobin; ACE = angiotensin-converting enzyme; HRT = hormone replacement therapy.

Table 15 Effect of Risk Factor Modification and Vasoprotective Drugs in Secondary Prevention

Intervention	Decrease in Cardiovascular Mortality (%)
Smoking cessation	45
Cholesterol lowering	40
Estrogen replacement therapy	40
Beta-blockers	25
Aspirin	25
Blood pressure reduction	20
ACE inhibitors	20
Exercise	20

ACE = angiotensin-converting enzyme.

Risk Stratification for Survivors of Myocardial Infarction (Table 15)

Survivors of MI are at a greater risk of subsequent cardiac events. The principal objective in the management of these patients is to identify those who may benefit from various interventions to reduce this risk. Several non-invasive tests, with and without imaging techniques, have been used to stratify a patient's risk profile. Overall, these investigations determine the exercise tolerance of the patients and the presence of ischemic myocardium. Coronary angiography is recommended for those with reversible ischemia to define the coronary anatomy and suitability for revascularization procedures.

Patients with significant arrhythmias, such as VT or those resuscitated from VF, may require electrophysiological studies to locate the focus of ectopy which may be treated by ablative therapy. Various pharmacological therapy of life-threatening arrhythmias is on the whole disappointing and intracardiac defibrillator (ICD) is the preferred treatment for patients resuscitated from VT/VF arrest to prevent future sudden death.

CONCLUSION

Through the past decade, there has been tremendous improvement in the management of patients with IHD. Greater understanding of the pathophysiology of IHD combined with the advent of new pharmacologic agents and mechanical reperfusion technologies has significantly improved the outcome of patients with both stable and ACS. Angiogenesis and cardiomyocyte transformation therapy will be further developed in the next millennium. More precise therapy directed at specific sites of the human genome will halt or modify the genesis of atherosclerosis.

ACKNOWLEDGMENTS

The authors would like to thank the Director, Epidemiology and Disease Control, Ministry of Health, Singapore, for providing the data on ischemic heart disease in Singapore. We appreciate the superb secretarial assistance of Mr. Azharie bin Ismaon.

4

Recent Advances in Diagnosis and Management of Cardiac Arrhythmias

Teo Wee Siong and Ruth Kam

INTRODUCTION

Cardiac arrhythmias are abnormalities in the rhythm of the heart and are extremely common. Arrhythmias may result from abnormalities of impulse initiation (automaticity), conduction (slow conduction, block, reentry), or a combination. The diagnosis is at times difficult because of its transient nature in some patients. The prognosis is benign in some, but in others may be associated with sudden cardiac death and hence the proper diagnosis and treatment is vital.

TYPES OF CARDIAC ARRHYTHMIAS

Cardiac arrhythmias may be due to isolated atrial or ventricular ectopics, or present as a bradyarrhythmia or tachyarrhythmia. Bradyarrhythmia is defined as an arrhythmia with a heart rate of less than 60 beats per minute and tachyarrhythmia is defined as an arrhythmia with a heart rate exceeding 100 beats per minute and may be supraventricular or ventricular.

Atrial or Ventricular Ectopics

Isolated atrial and ventricular ectopics are extremely common and may occur in up to 50% of normal people.[1] The diagnosis is easily made by the ECG and is often asymptomatic in the majority of the patients.

Bradyarrhythmias

The commonest bradyarrhythmia is sinus bradycardia. This in a healthy person is most likely due to increased vagal tone. In the older patients, it may reflect underlying sick sinus syndrome. Sinus bradycardia is commonly secondary to drugs, especially beta-blockers and the calcium antagonists, verapamil or diltiazem. Rarely, bradycardia may be due to associated medical conditions such as hypothyroidism and increased intracranial pressure.

Persistent symptomatic bradycardia is most commonly due to complete heart block. The second most common cause is the sick sinus syndrome. This often presents as a tachybradyarrhythmia with episodes of atrial tachycardia (AT) or atrial fibrillation (AF) followed by pauses, which are symptomatic.

Tachyarrhythmias

The most common tachyarrhythmia is AF. In the Framingham study, the incidence of AF is estimated to be 2–4% of the general population above 65 years old and 10% in patients above 80 years of age.[2] In the younger patients, tachyarrhythmias may be due to a paroxysmal supraventricular tachycardia (previously called paroxysmal atrial tachycardia or PAT). The commonest cause of supraventricular tachycardia (SVT) is an atrioventricular nodal reentrant tachycardia (AVNRT) as seen in Fig. 1, followed by atrioventricular reentrant tachycardia (AVRT) due to an accessory pathway and AT. Less common causes of supraventricular tachyarrhythmias include atrial flutter, which is defined as an AT with a heart rate of 240–350 beats per minute lacking an isoelectric baseline between atrial deflections.[3] Recent evidences have confirmed that atrial flutter is a macro-reentrant atrial tachycardia in the right atrium with a critical isthmus at the tricuspid-IVC region. In a review of 983 patients with paroxysmal supraventricular tachycardia who had undergone electrophysiological studies at the Singapore General Hospital from 1982 to 1997, 601 (61.1%)

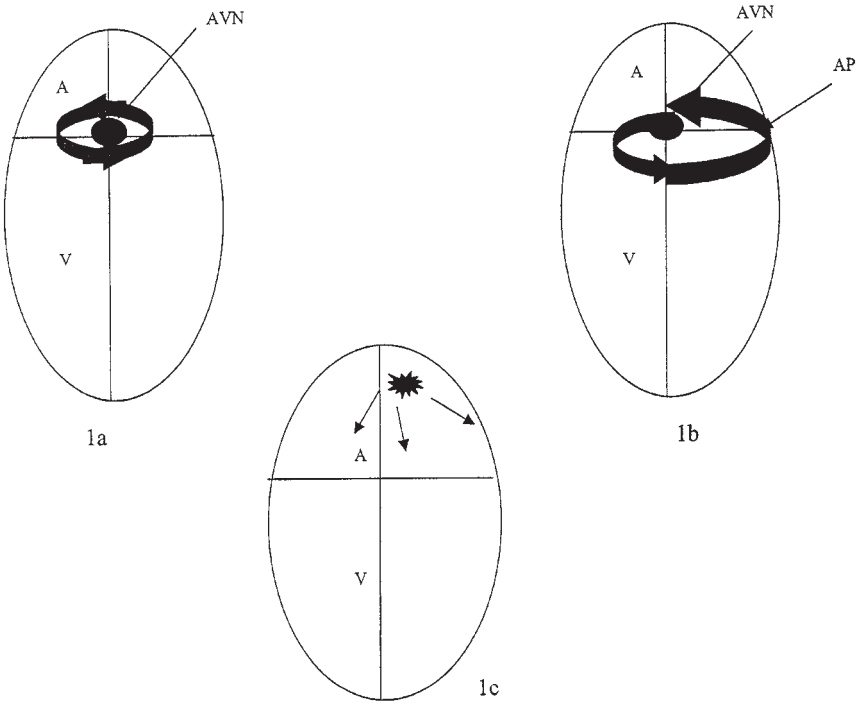


Fig. 1 Common types of supraventricular tachycardia:
1a, AVNRT; 1b, AVRT; 1c, atrial tachycardia; A = atrium, V = ventricle, AVN = AV node, AP = accessory pathway.

had AVNRT; 292 (29.7%) had AVRT due to an accessory pathway; and 87 (8.9%) had AT.

Ventricular tachyarrhythmias are most commonly ventricular tachycardia (VT) and less commonly ventricular fibrillation (VF). VT occurs most commonly in patients with underlying structural heart disease, especially post-MI, but may also occur in patients with normal hearts (idiopathic VT). The idiopathic VT most commonly originates from the right ventricular outflow tract, resulting in a classic left bundle branch block, i.e. inferior axis morphology (Fig. 2). In Asians, however, the idiopathic left VT is relatively common and classically has a right bundle branch block, superior axis morphology. It is often misdiagnosed as SVT with aberrancy because it often has a relatively narrow QRS duration of about 100–120 ms. Other ventricular tachyarrhythmias include torsade de pointes (associated with prolonged QT syndrome) and polymorphic VT (with normal QT interval and often associated with myocardial ischemia). The long QT

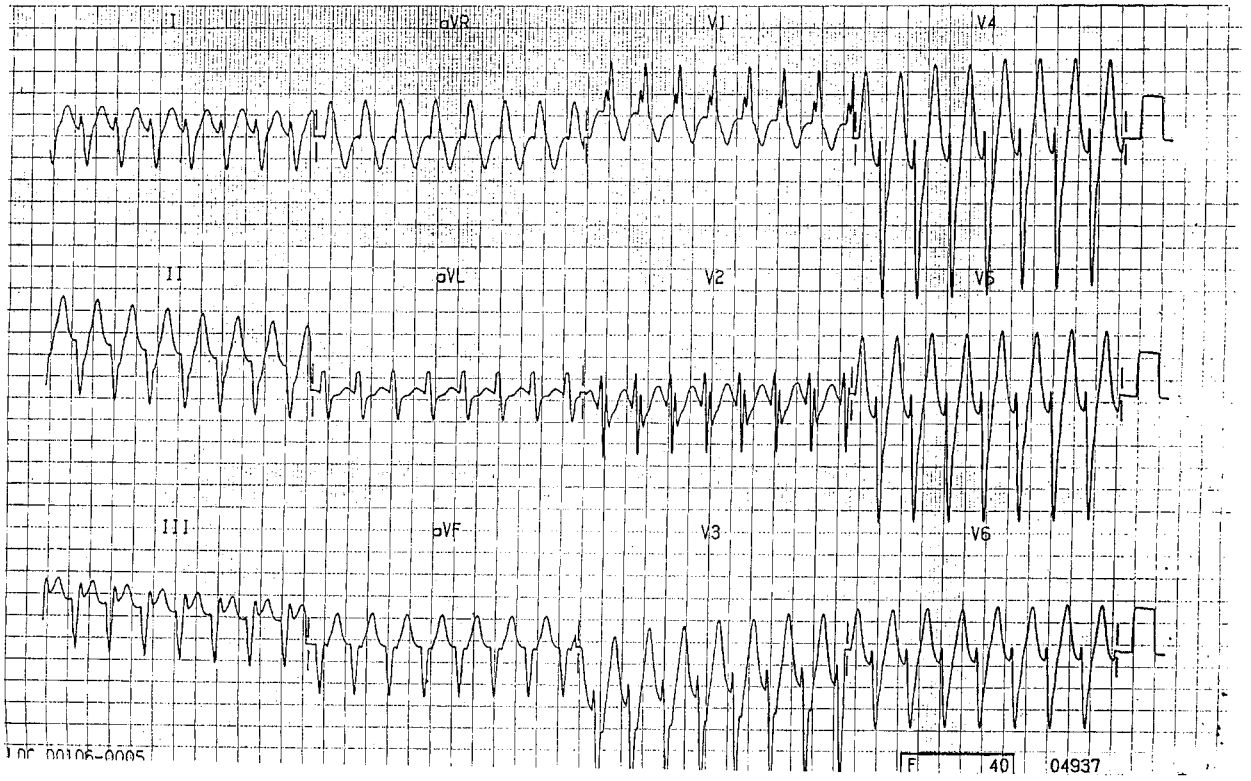


Fig. 2 12-lead ECG of idiopathic left ventricular tachycardia with right bundle branch block and superior axis.

syndrome is most commonly associated with anti-arrhythmic drugs, especially class I anti-arrhythmic drugs such as quinidine and procainamide, but it may also be associated with class III drugs such as sotalol and amiodarone.[4] The episodes of torsades are often pause-dependent. The congenital long QT syndrome has a high risk for sudden cardiac death but is fortunately rare.

BENIGN VS. MALIGNANT CARDIAC ARRHYTHMIAS

Arrhythmias occurring in the absence of underlying heart disease are usually benign. This is especially so for the isolated atrial and ventricular ectopics. However, when sustained SVT or VT occurs, complications may occur. These patients may progress to tachycardia cardiomyopathy and hence be at risk for cardiac failure and sudden death.[5] Rarely, the patient with SVT may be at risk for sudden death. In a series by Wang *et al.*, in patients who were evaluated for sudden cardiac death, SVT was the cause of aborted sudden death in approximately 5%.[6] Patients with the Wolff-Parkinson-White syndrome have been reported to be at risk for sudden cardiac death.[7] Patients with VT usually have a more adverse prognosis with increased risk for sudden death. High-risk groups include those patients with previous myocardial infarction (MI), dilated or hypertrophic cardiomyopathy and patients with a family history of sudden death. Specific syndromes associated with sudden cardiac death include the long QT syndromes and the Brugada syndrome, which has a right bundle branch block with ST-segment elevation in the right precordial leads.[8]

APPROACH TO PATIENTS WITH CARDIAC ARRHYTHMIAS

The role of history taking cannot be overemphasized. Classically, the symptom associated with cardiac arrhythmias is palpitations, which, by definition, is described as an uncomfortable awareness of a beating heart. A history of an occasional thump with a missed beat or a brief fluttering is classical of an isolated atrial or ventricular ectopic. A paroxysmal palpitation that begins suddenly that is described as a rapid racing, often associated with mild symptoms of chest discomfort, feeling weak or sweaty, with or without giddiness, lasting for minutes or occasionally hours is

very suggestive of a paroxysmal SVT, although rarely it may be due to a VT.[9] The history is especially important because of the transient nature of symptoms and signs and hence difficulty in documenting it. A detailed medical history is not only vital for diagnosis but is also helpful for prognosis. Patients with a history of thyrotoxicosis or underlying mitral valve disease who develop palpitations are more likely to have AF. Patients developing palpitations for the first time after a MI are more likely to have VT.[10] However, palpitations are not always present in patients with SVT or VT; the mechanical performance of the heart (left ventricular [LV] function) must be sufficiently good to produce them. Only 6% of 97 patients with sustained, non-syncopal VT caused by an old MI complained of palpitations during the arrhythmia.[10] Other presentations of patients with cardiac arrhythmias may be syncope or near-syncope, breathlessness and chest discomfort. Occasionally, heart failure may be the initial presentation, especially in someone who has incessant tachycardia. The family history is very important, because a family history of sudden death is vital for diagnosis and management. Arrhythmias such as the long QT syndrome are often congenital.

The clinical examination is usually not useful, although it may help sometimes, such as, for example, an obvious thyrotoxic patient with a goiter will help in the diagnosis of the cause of the patient's palpitations. The presence of murmurs suggesting valvular heart disease will suggest AF. Clinical evidence of structural heart disease such as murmurs associated with hypertrophic cardiomyopathy is also useful.

INVESTIGATIONS FOR CARDIAC ARRHYTHMIAS

In view of the transient nature of symptoms and signs, further investigations are very important. This includes the use of invasive and non-invasive tests. The ECG, ambulatory monitoring, esophageal recording, exercise testing, and signal-averaging techniques are the currently available non-invasive tests. Intracardiac electrophysiologic studies and endocardial catheter mapping are invasive techniques.

ECG

The 12-lead ECG is the most important investigation especially during an episode of palpitations. Occasionally, the resting ECG is also useful. The presence of the long QT syndrome (Fig. 3) or preexcitation (Fig. 4) helps

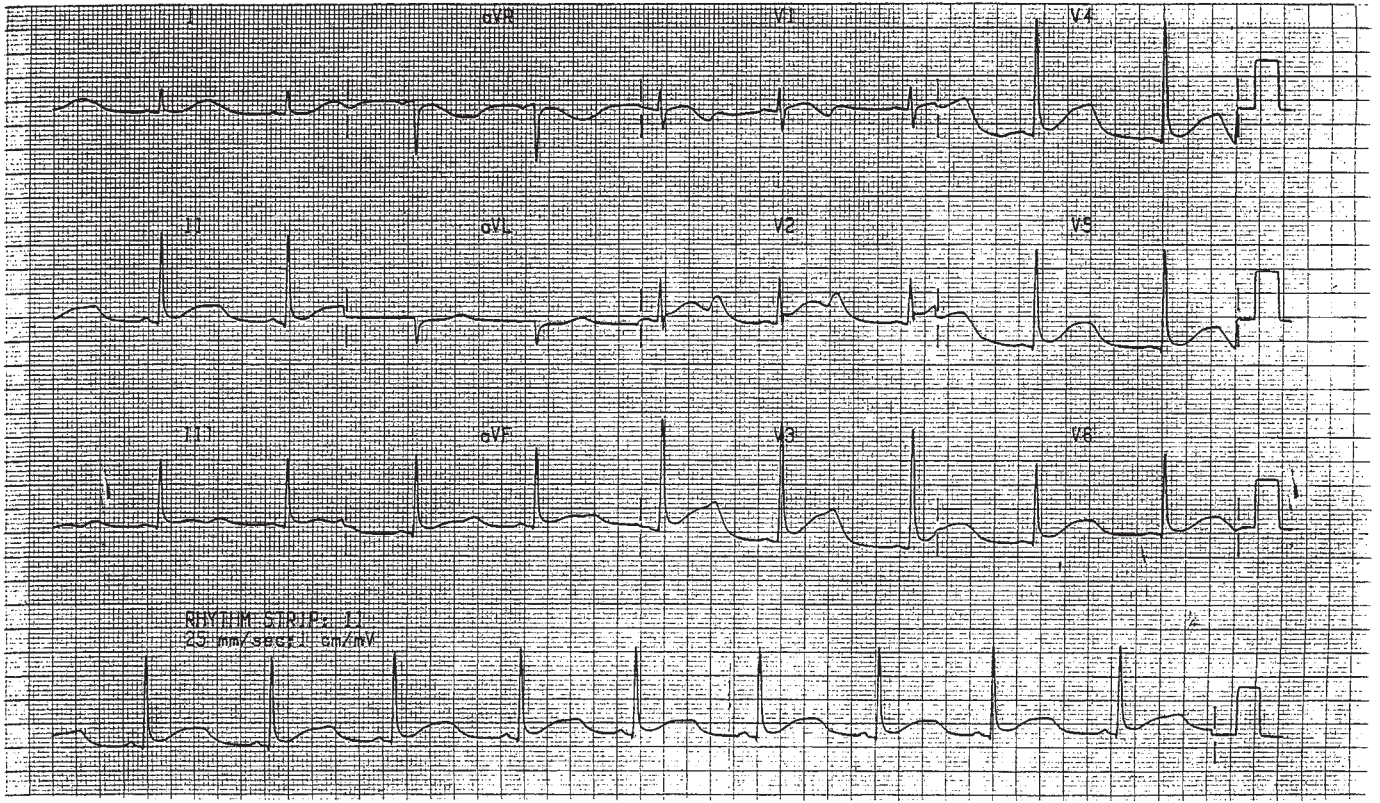


Fig. 3 12-lead ECG showing the long QT syndrome.

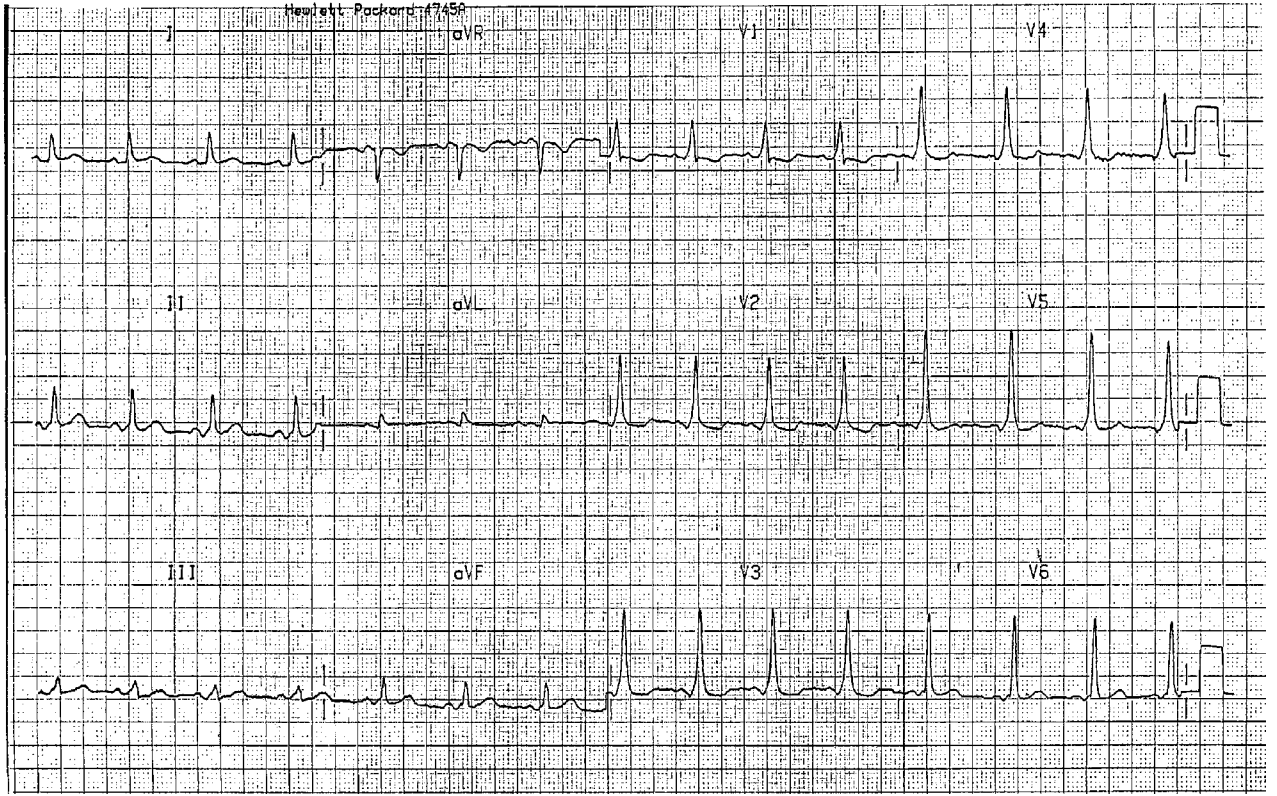


Fig. 4 ECG showing the Wolff-Parkinson-White syndrome, with a short PR interval, widened QRS and delta wave.

to diagnose the underlying arrhythmia. The presence of Q-waves on the ECG suggests previous MI as the substrate for the arrhythmia. The Brugada syndrome (Fig. 5) can be easily diagnosed using the resting ECG, as it shows right bundle branch block with ST-segment elevation in the precordial leads.[8]

Ambulatory Holter ECG and Long-term ECG Recording with Event Recorders or Loop Recorders

The ambulatory 24-hour or 48-hour Holter ECG is useful when arrhythmias occur frequently, but is not documented on the resting ECG. Prolonged monitoring helps to document the arrhythmias associated with the palpitations. It is useful for documenting the sick sinus syndrome (Fig. 6), showing episodes of tachycardia and sinus pauses. Infrequent palpitations may, however, be more easily documented by the use of long-term loop recorders or event recorders which record the ECG when activated.

Signal-Averaged ECG

The signal-averaged ECG records late potentials, which are low amplitude, fractionated activities occurring after the QRS complex of the surface electrogram. They are related to slow and heterogeneous conduction within damaged cardiac tissue after an acute or a chronic ischemic insult to the heart, predisposing patients to develop ventricular arrhythmias. The main utility of signal-averaged ECG is its negative predictive value of $>95\%$ but it has a low positive predictive value ($<25\%$) for ventricular arrhythmias and cardiac mortality.[11]

Exercise Testing

Exercise testing is useful in patients who have complaints of palpitations during exercise. It is also useful to exclude ischemia as an associated abnormality in patients with associated underlying heart disease.

Electrophysiological Studies

The electrophysiological study provides the "gold standard" for the investigation of cardiac arrhythmias and can be used for the investigation of

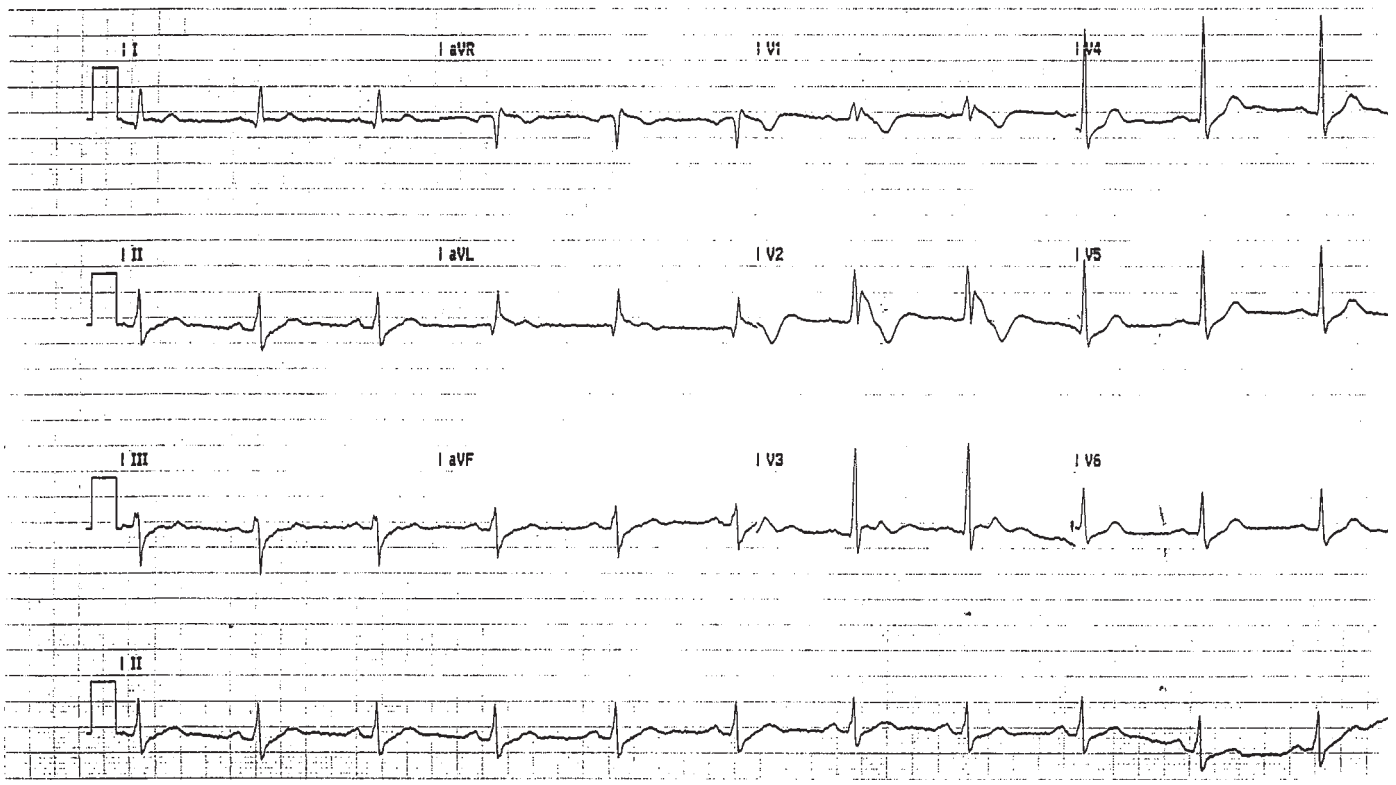


Fig. 5 ECG showing the Brugada syndrome with RBBB and ST elevation in the right precordial leads.

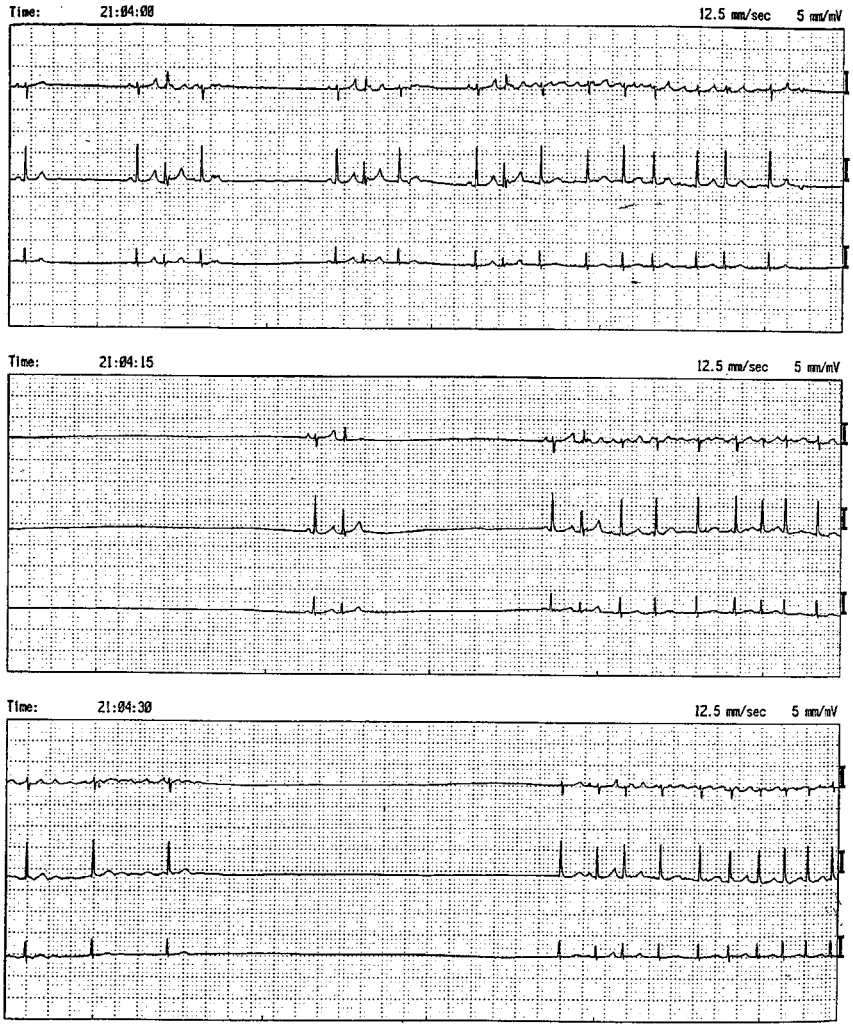


Fig. 6 Holter ECG showing sick sinus syndrome with bradytachyarrhythmias.

bradyarrhythmias and tachyarrhythmias. Briefly, an electrophysiological study involves insertion of electrode catheters into the heart via the femoral vein and the subclavian or internal jugular vein. These electrode catheters allow precise study of the electrical activation inside the heart. The intracardiac electrogram and surface ECG provide detailed information of the

anatomical and electrical activation of the heart. The electrophysiological study is especially important for evaluation of:

1) narrow complex tachycardia;
2) wide complex tachycardia;
3) Wolff–Parkinson–White syndrome;
4) resuscitated sudden death;
5) syncope of unknown origin in patients with underlying structural heart disease or ECG abnormalities, e.g. bifascicular block;
6) recurrent palpitations with a history suggestive of a paroxysmal SVT; and
7) suspected sick sinus syndrome.

Insertable Loop Recorders

These are small devices manufactured by pacemaker companies, which are implanted just underneath the skin over the left pectoral region. They are able to provide continuous ECG monitoring of the heart for up to 14 months. They are especially useful for patients with syncope of unknown origin.

DIFFERENTIAL DIAGNOSIS

The diagnosis of atrial and ventricular ectopics is fairly easy on the ECG, but the narrow complex tachycardia may have multiple causes and may include the following:

1) sinus tachycardia;
2) AVNRT;
3) AVRT involving an accessory pathway;
4) focal AT;
5) macro-reentrant AT;
6) atrial flutter;
7) AF;
8) multi-focal AT; or
9) narrow complex tachycardia.

Wide complex tachycardia are most often due to VT but may also be due to preexcited tachycardia or SVT with aberrancy. The diagnosis

may be suggested by the 12-lead ECG, but occasionally an electrophysiological study is necessary to obtain the correct diagnosis.

TREATMENT OF CARDIAC ARRHYTHMIAS

The definitive treatment of cardiac arrhythmias depends on the precise diagnosis of the arrhythmia. The treatment options include the use of anti-arrhythmic drugs, cardiac pacing (anti-bradycardia, anti-tachycardia), implantable cardioverter defibrillator (ICD), catheter ablation and arrhythmia surgery.

Empiric Treatment with Anti-arrhythmic Drugs

Empiric drug therapy may be used in patients with relatively benign arrhythmias. Anti-arrhythmic drugs, however, have significant risk for proarrhythmia especially in patients with impaired LV function and underlying structural heart disease. The Cardiac Arrhythmia Suppression Trial (CAST) showed that patients taking encainide or flecainide had a 2.5 times rise in mortality in patients treated with anti-arrhythmic drugs when compared to placebo.[12] Similarly, the recent large EMIAT[13] and CAMIAT[14] studies, using amiodarone in post-MI patients, again failed to show significant improvement in total mortality in patients treated with empiric amiodarone. Hence the risk-benefit of anti-arrhythmic drug use must be carefully evaluated before using any drug.

Acute Therapy of Arrhythmias

In the hemodynamically unstable patient, irrespective of SVT or VT, synchronized cardioversion is the therapy of choice. In the stable patient, an attempt should be made to diagnose the arrhythmia so as to decide on the most appropriate therapy. In this respect, a 12-lead ECG is vital. Simple vagal maneuvers should not be forgotten as they can be effective in up to 25% of patients with a paroxysmal SVT.[15] The primary goal of acute therapy for SVT is rate control and termination, if possible. Rate control can be achieved using drugs such as verapamil, diltiazem and beta-blockers. In the patient with heart failure or poor ventricular function, however, the choice is limited to digoxin and amiodarone. Intravenous verapamil and adenosine are very effective for conversion

of paroxysmal SVT to sinus rhythm but are useless and may even be harmful in patients with preexcited AF or VT. In patients with VT, amiodarone and lignocaine are the drugs of choice as recommended by the Advance Cardiac Life Support Committee in Singapore.[16]

Bradyarrhythmias, which are symptomatic, require temporary pacing, which can initially be transcutaneous but because of the discomfort associated with it, usually requires temporary transvenous endocardial pacing. Occasionally, intravenous atropine or even adrenaline may be used initially. Isoprenaline is no longer recommended routinely.[15]

LONG-TERM MANAGEMENT OF CARDIAC ARRHYTHMIAS

Chronic therapy of cardiac arrhythmias has been revolutionized by the advent of non-pharmacological therapies and the realization that drug therapy, while useful acutely, is not curative and long-term drug therapy has significant complications.

Atrial and Ventricular Ectopics

Patients with isolated atrial or ventricular ectopics generally do not require any drug treatment, unless they are very symptomatic. Such patients may be treated with beta-blockers but occasionally class Ic drugs such as propafenone or flecainide or class III drugs such as sotalol may be used, provided there is no evidence of underlying heart disease. In general, most patients should be reassured of the benign prognosis of the ectopics compared with the possible side-effects of anti-arrhythmic drug treatment.

Bradyarrhythmias

For the patient with symptomatic bradyarrhythmia due to complete heart block and the sick sinus syndrome, drugs are no longer appropriate and permanent pacing has become the treatment of choice since it was first introduced in 1958. Our own experience at the Singapore General Hospital shows that of the 1009 pacemakers implanted during the period from 1979 to 1998, complete heart block remains the most common indication for pacing, followed closely by the sick sinus syndrome. Complications are rare and are $<1\%$. Dual chamber pacemakers are also

increasingly used in recent years and comprise almost 50% of all the pacemakers implanted. In general, pacemakers are indicated for patients with:

1) symptomatic AV block, usually third degree AV block, Mobitz II AV block and occasionally, Mobitz type I AV block and marked first degree AV block; or
2) sick sinus syndrome.

Newer indications for pacing include pacing for vasovagal syncope, prevention of arrhythmias, dilated cardiomyopathy and hypertrophic cardiomyopathy.

Atrial Fibrillation

Recurrent AF, which are paroxysmal, may be treated with anti-arrhythmic drugs to try to maintain sinus rhythm. Younger patients with paroxysmal AF should be considered for electrophysiological study because recent evidence suggests that it may be due to a rapidly-firing focus in the pulmonary vein or secondary to other SVT.[17] In the older patients, anti-arrhythmic drugs remain the treatment of choice. The presence or absence of an underlying heart disease influences the choice of drugs. Class Ic drugs may be used when there is no underlying heart disease, but amiodarone is probably the only drug useful when there is significant heart disease, especially heart failure or ischemic heart disease (IHD). When AF is persistent, attempts at conversion to sinus rhythm may be considered with anti-arrhythmic drugs, failing which synchronized external cardioversion should be performed after adequate anti-coagulation for about 3–4 weeks. Occasionally in the hospitalized patients, when rapid conversion is desired, transesophageal echocardiography to exclude the presence of thrombus and heparin anticoagulation may be used before early synchronized cardioversion.[18] Patients with permanent AF remain in AF and further attempts at conversion to sinus rhythm are usually futile. Younger patients with no structural heart disease and a normal left atrial size may be given aspirin. Older patients (especially if >60 years old) with persistent and permanent AF and underlying structural heart disease, must be anti-coagulated (to an INR of at least 1.8–3.0) as the risk of thromboembolism is increased. The Atrial Fibrillation Investigators pooled data from 5 trials in a meta-analysis. The overall risk of stroke was 4.5% per year, of which 2% per year were severe or fatal. The use of

warfarin reduced the overall stroke rate to 1.4% per year, of which 0.6% per year was severe or fatal. The use of warfarin, however, was associated with an increased rate of major hemorrhage from 1% per annum (in controls) to 1.3% per annum (active treatment with warfarin), indicating the need to carefully select patients and closely monitor to prevent over-anticoagulation.[19] When the AF remains rapid, even with drugs such as digoxin, beta-blockers, verapamil or diltiazem, catheter ablation of the AV node or AV node modification to slow down AV conduction, followed by a permanent pacemaker, may be considered.[20–21]

Paroxysmal Supraventricular Tachycardia

Catheter ablation has revolutionized the treatment of SVT because it has become possible to map and ablate the arrhythmogenic focus or critical part of the reentrant circuit without the need for open-heart surgery. The technique is an extension of the electrophysiological study and involves the insertion of a special 4 mm tip catheter, which is able to ablate the arrhythmogenic focus by using radiofrequency energy, which heats the tip of the catheter to a temperature of 50–70°C. The precise mapping enables the arrhythmic focus to be ablated within seconds upon delivery of the radiofrequency energy (Fig. 7). Ablation is now routinely offered to all patients with paroxysmal SVT due to AVNRT, AVRT and AT. Atrial flutter and idiopathic VT are also very successfully cured. Radiofrequency catheter ablation is also increasingly being used for the treatment of AF triggered by a focus in the pulmonary veins and in ischemic VT. It offers curative therapy, thus obviating the need for lifelong therapy and eliminates the risk of sudden death in some patients. Our own experience since 1991 of about 2000 cases to date indicates a success rate of about 98%, with a recurrence rate of less than 5% and complications of less than 1%.[22] New mapping systems that enable three-dimensional mapping of the heart helps us now to map and ablate even more complex arrhythmias successfully.

Malignant Ventricular Tachyarrhythmias

For the patient with malignant ventricular arrhythmias, however, the therapy of choice is now the ICD. Nothing has been more successful or effective in terminating VT or VF (Fig. 8) and preventing sudden cardiac death. The Antiarrhythmics versus Implantable Defibrillators (AVID)

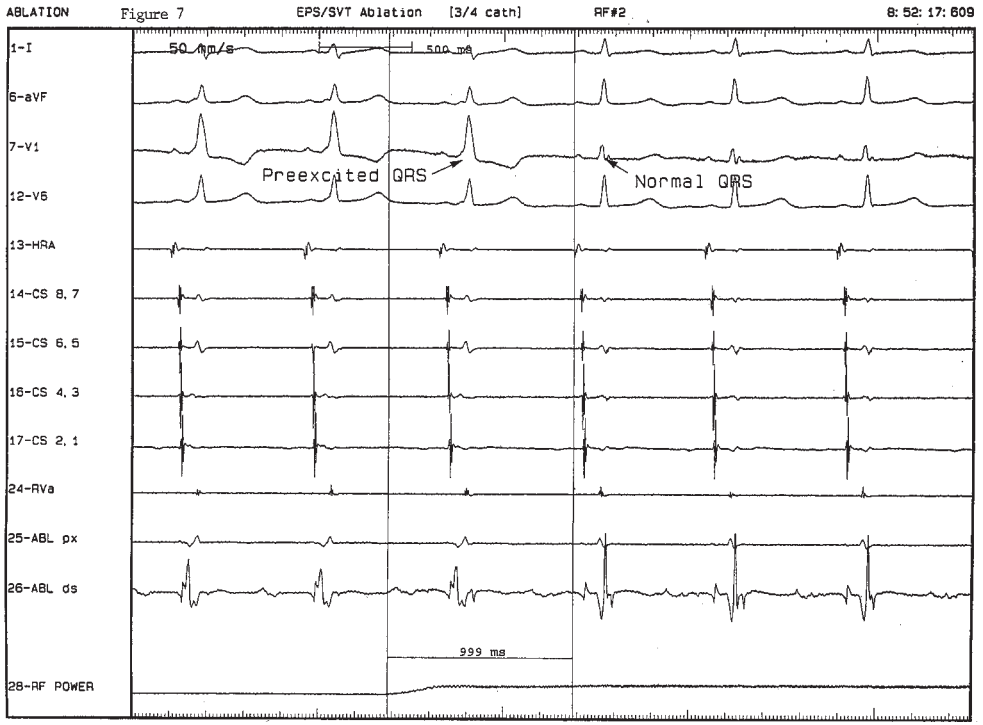


Fig. 7 Radiofrequency catheter ablation of an accessory pathway showing successful ablation of the accessory pathway within 0.9 seconds.
(I,aVF,V1 and V6 are surface ECGs, HRA = high right atrium, CS = coronary sinus, RVa = right ventricular apex, ABL = ablation, px = proximal, ds = distal, RF = radiofrequency).

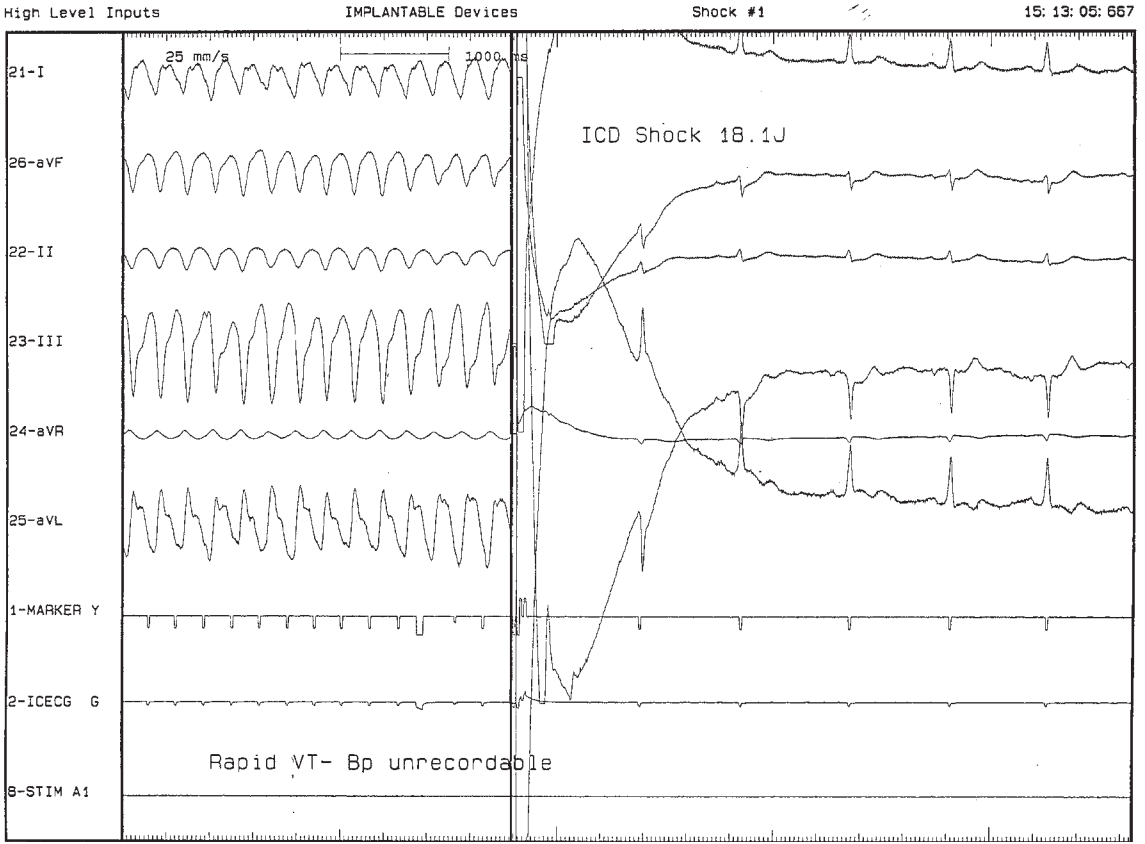


Fig. 8 Termination of ventricular fibrillation by the implantable cardioverter defibrillator.

trial[23] and the Multicenter Automatic Defibrillator Implantation Trial (MADIT)[24] have reported a statistically significant decrease in all-cause mortality in patients randomized to ICD therapy compared to drug treatment with amiodarone or sotalol. AVID studied patients with prior MI who suffered VF, sustained VT with syncope, or sustained VT with significant hemodynamic compromise. Patients were randomized to receive either ICD implantation or conventional anti-arrhythmic drug therapy, primarily amiodarone. Those patients who received an ICD had a 39% reduction in total mortality at 1 year, a 27% reduction at 2 years, and a 31% reduction at 3 years compared to patients in the conventional therapy group ($P < 0.02$). MADIT showed that over the course of 5 years, patients in NYHA functional class I, II, or III with prior MI; a LVEF $\leqslant$ 0.35; a documented episode of asymptomatic unsustained VT; and inducible, non-suppressible ventricular tachyarrhythmia on electrophysiologic study, patients assigned to receive an ICD had a 54% improvement in survival compared with conventional medical therapy. The recently published MADIT II study suggests that in patients with a prior MI and LV dysfunction (LVEF $\leqslant$ 30%), prophylactic implantation of a defibrillator improves survival and should be considered as a recommended therapy.[25] Drug therapy, especially beta-blockers, sotalol and amiodarone, however, still have a role as an adjunctive therapy in these patients in order to reduce the occurrence of arrhythmia and hence frequency of ICD shocks. Catheter ablation may be another alternative if the patient continues to have recurrent shocks with the ICD after anti-arrhythmic drugs. Anti-arrhythmic surgery such as endocardial resection has a limited role because of its significant operative mortality of at least 10–25%.

CONCLUSION

In conclusion, the diagnosis and treatment of cardiac arrhythmias has undergone significant changes in the new millennium. Non-invasive and invasive techniques such as electrophysiological studies have allowed more definitive diagnosis of arrhythmias and risk stratification. Anti-arrhythmic drugs, while still having an important role in the acute therapy of arrhythmias, have become less important in chronic treatment. Non-pharmacological treatment such as pacemakers and catheter ablation are increasingly important as they are curative and the ICDs

have virtually eliminated the risk of sudden death due to ventricular tachycardia or fibrillation.

REFERENCES

1. Brodsky M, Wu D, Denes P, Kanakis C, Rosen KM, Arrhythmias documented by 24 hr ECG in 50 male medical students without apparent heart disease, *Am J Cardiol* **39**:390–395, 1977.
2. Kannel WB, Abbott RD, Savage DD, McNamara PM, Epidemiologic features of atrial fibrillation. The Framingham Study, *N Engl J Med* **306**:1018–1022, 1982.
3. Cosio FG, Arribas F, Lopez–Gil M, Gonzalez HD, Radiofrequency ablation of atrial flutter, *J Cardiovasc Electrophysiol* **7**:60–70, 1996.
4. Tong KL, Lau YS, Teo WS, A case series of drug induced long QT syndrome and Torsade de pointes, *Singapore Med J* **43**:566–570, 2001.
5. Packer DL, Bardy GH, Worley SJ, Smith MS, Cobb FR, Coleman RE, *et al*., Tachycardia-induced cardiomyopathy: a reversible form of left ventricular dysfunction, *Am J Cardiol* **57**:563–570, 1986.
6. Wang Y, Scheinman MM, Chien WW, Cohen TJ, Lesh MD, Griffin JC, Patients with supraventricular tachycardia presenting with aborted sudden death: incidence, mechanism and long term follow up, *J Am Coll Cardiol* **18**:1711–1719, 1991.
7. Teo WS, Klein GJ, Yee R, Sudden death in the Wolff–Parkinson–White syndrome, *Singapore Med J* **33**:247–251, 1992.
8. Teo WS, Kam R, Tan RS, Maglana M, Lim YL, The Brugada syndrome in a Chinese population, *Int J Cardiol* **65**:281–286, 1998.
9. Brugada P, Gursoy S, Brugda J, Andries E, Investigation of palpitations, *Lancet* **341**:1254–1258, 1993.
10. Akhtar M, Shenasa M, Jazayeri M, Caceres J, Tchou PJ, Wide QRS complex tachycardia. Reappraisal of a common clinical problem, *Ann Intern Med* **109**:905–912, 1988.
11. Simson MB, Noninvasive identification of patients at high risk for sudden cardiac death. Signal-averaged electrocardiography, *Circulation* **85**(Suppl-I):I145–I151, 1992.
12. The Cardiac Arrhythmia Suppression Trial Investigators: Preliminary Report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction, *N Engl J Med* **321**:406–411, 1989.
13. Julian DG, Camm AJ, Frangin, Janse MJ, Munoz A, Schwartz PJ, Simon P for the European Myocardial Infarct Amiodarone Trial Investigators, Randomised trial of effect of amiodarone on mortality in patients with left-ventricular dysfunction after recent myocardial infarction: EMIAT. European Myocardial Infarct Amiodarone Trial Investigators, *Lancet* **349**:667–674, 1997.
14. Cairns JA, Connolly SJ, Roberts R, Gent M, Randomised trial of outcome after myocardial infarction in patients with frequent or repetitive ventricular premature

depolarisations: CAMIAT. Canadian Amiodarone Myocardial Infarction Arrhythmia Trial Investigators, *Lancet* **349**:675–682, 1997.

15. Lim SH, Anantharaman V, Teo WS, Goh PP, Tan ATH, Comparison of treatment of supraventricular tachycardia by Valsalva Maneuver and Carotid sinus massage, *Ann Emerg Med* **31**:1–6, 1998.

16. Teo WS, Advanced Cardiac Life Support, *Sing Med J* **Suppl**(1):10–20, 2001.

17. Haissaguerre M, Marcus FI, Fischer B, Clementy J, Radiofrequency catheter ablation in unusual mechanisms of atrial fibrillation: report of three cases, *J Cardiovasc Electrophysiol* **5**:743–751, 1994.

18. Manning WJ, Silverman DI, Gordon SPF, Krumholz HM, Douglas PS, Cardioversion from atrial fibrillation without prolonged anticoagulation with use of transesophageal echocardiography to exclude the presence of atrial thrombi, *N Engl J Med* **328**:750–755, 1993.

19. Atrial Fibrillation Investigators, Risk factors for stroke and efficacy of anti-thrombotic therapy in atrial fibrillation: analysis of pooled data from 5 randomized controlled trials. *Arch Intern Med* **154**:1449–1457, 1994.

20. Feld GK, Fleck RP, Fujimura O, Prothro DL, Bahnson TD, Ibarra M, Control of rapid ventricular response by radiofrequency catheter modification of the atrioventricular node in patients with medically refractory atrial fibrillation, *Circulation* **90**:2299–2307, 1994.

21. Williamson BD, Man KC, Daoud E, Niebauer M, Strickberger SA, Morady F, Radiofrequency catheter modification of atrioventricular conduction to control the ventricular rate during atrial fibrillation, *N Engl J Med* **331**:910–917, 1994.

22. Teo WS, Kam RML, Lim YL, Koh TH, Curative therapy of cardiac tachyarrhythmias with catheter ablation — a review of the experience with the first 1000 patients, *Sing Med J* **40**:284–290, 1999.

23. A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. The Antiarrhythmics versus Implantable Defibrillators (AVID) Investigators, *N Engl J Med* **337**:1576–1583, 1997.

24. Moss AJ, Hall WJ, Cannom DS, Daubert JP, Higgins SL, Klein H, *et al*., Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. Multicenter Automatic Defibrillator Implantation Trial Investigators, *N Engl J Med* **335**:1933–1940, 1996.

25. Moss AJ, Zareba W, Hall J, Klein H, Wilber D, Cannom D, *et al*., Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction, *N Engl J Med* **346**:877–873, 2002.

This page intentionally left blank

5

Rheumatic Mitral Stenosis

Lau Kean Wah

INTRODUCTION

Mitral stenosis (MS) is the prototype of rheumatic heart disease. It is the most common condition related to rheumatic carditis. It continues to be a prevalent and significant cause of morbidity and mortality in developing countries, thereby posing a major problem, in terms of health costs, in these countries. Approximately 25% of all patients with rheumatic heart disease have pure MS and another 40% have mixed MS and regurgitation. Mitral stenosis is rarely caused by other disease processes, such as the carcinoid syndrome, systemic lupus erythematosus, rheumatoid arthritis, or is congenital in origin. The association of MS with atrial septal defect is termed Lutembacher's syndrome.

PATHOLOGY AND PATHOPHYSIOLOGY

Previous rheumatic carditis results in fibrotic thickening and calcification of the mitral leaflets, fusion of commissures, thickening, fusion and retraction of the subvalvular apparatus, as well as progressive narrowing

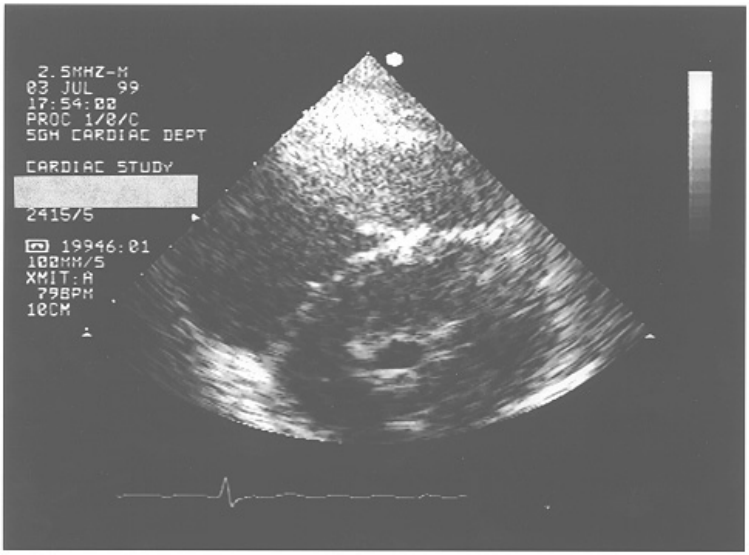


Fig. 1 An echocardiographic view demonstrating the "fish mouth" appearance of a stenotic mitral valve.

of the mitral valve. The stenotic mitral valve typically appears like a "fish mouth" in its shape (Fig. 1).

Hemodynamic abnormalities usually occur about 10 to 20 years after the initial insult, when the valve area is reduced to about half its normal size from 5 cm^2 to around 2–2.5 cm^2. Further reduction to less than 2 cm^2 results in the development of mild to moderate functional impairment, in particular, exertional breathlessness. When the valve area is less than 1 cm^2, severe symptoms are often present. Although obstruction usually occurs at the valve level, it may also take place at the subvalvular level due to chordal fusion and shortening.[2,3]

CLINICAL PRESENTATION

The main symptom of MS is dyspnea, which results from the transudation of fluid into the lungs from increased pulmonary venous pressure and, consequently, reduced lung compliance. In severe cases, the patient develops pulmonary edema, either gradually or suddenly. Other clinical presentations include palpitation from paroxysmal atrial fibrillation (which develops into chronic atrial fibrillation); haemoptysis from

pulmonary edema, rupture of thin-walled, dilated bronchial veins from increased left atrial pressure, pulmonary infarction or chronic bronchitis; thromboembolic phenomenon from a dislodged left atrial thrombus; and infective endocarditis.

Mitral facies, not commonly seen these days and characterized by pinkish-purple cheek discoloration, is usually seen in patients with severe MS with a low cardiac output and systemic vasoconstriction. Other physical signs are summarized in Table 1. A loud S1 and an opening snap (heard just after S2) usually indicate the presence of a pliable stenotic mitral valve. Conversely, their absence may denote calcific MS or associated mitral regurgitation. The *sine qua non* sign in MS is a diastolic murmur best heard with the bell of the stethoscope applied lightly to the apex, with the patient placed in the left lateral position. With increasing severity of stenosis, the diastolic murmur becomes longer and may occupy the entire diastolic period. The opening snap, if present, is also very close to the S2. A decrescendo diastolic (Graham Steele) murmur of pulmonary regurgitation may be heard along the left sternal edge, in patients with secondary pulmonary hypertension, together with a left parasternal heave. When pulmonary hypertension is severe, there may be evidence of right ventricular failure: S3, tricuspid regurgitation murmur, raised JVP, an engorged, palpable liver, and peripheral edema.

The electrocardiogram (ECG), chest X-Ray and echocardiographic findings are summarized in Table 2. Left atrial enlargement, characterized by a widened (more than 0.12 seconds), and often bifid, P-wave in lead II, as well as a biphasic P-wave with a negative terminal deflection in V1, is the principal ECG feature. It is observed in the majority of patients with

Table 1 Physical Findings in Mitral Stenosis

Palpation
• Tapping apex beat
• Right ventricular heave in the presence of pulmonary hypertension
• Apical (Grade $\geqslant 4$) diastolic thrill
Auscultation
• Loud S1
• Opening snap
• Apical diastolic murmur: mid-diastolic, presystolic or pandiastolic
• Loud P2, pulmonary regurgitation (Graham Steele) murmur $\pm$ tricuspid regurgitation in the presence of secondary pulmonary hypertension
• Occasionally concomitant mitral regurgitation murmur

Table 2 Investigatory Findings in Mitral Stenosis

Electrocardiogram
• Left atrial enlargement (P mitrale)
• Atrial fibrillation
• Right ventricular hypertrophy/dilatation: rsR pattern in V1, right axis deviation
Chest X-Ray
• Straightening of the left cardiac border, double contour of the left atrial shadow and widened bronchial carina denoting left atrial enlargement
• Upper lobe diversion, pulmonary edema and Kerley B lines
• Mitral valve calcification
• Pulmonary artery and right ventricular enlargement
Echocardiography
• Thickened ± calcified mitral leaflets with decreased mobility
• Fused ± calcified mitral commissures
• Fused, thickened chordae tendineae
• Narrowed mitral valve orifice (fish mouth appearance)
• Determination of mitral valve area

MS and sinus rhythm. Atrial fibrillation, however, is present in 50–60% of patients and its incidence increases with the chronicity of MS. As with ECG, the chest X-Ray commonly shows left atrial enlargement, although the size of the left atrium correlates more with the chronicity of atrial fibrillation rather than the severity of MS. Occasionally, mitral valve calcification, if gross, may be evident on the chest X-Ray. Upper lobe diversion, pulmonary edema and interstitial edema (Kerley B lines) may be present and usually indicate significant MS. Cardiac catheterization is not indicated in all patients with MS. It is commonly used to exclude coexisting coronary artery disease and to clarify the severity of MS in unechogenic patients. In skilled hands, right and left heart catheterization can be performed expeditiously and with minimal risk to the patient.

NATURAL HISTORY

The natural history of MS is such that there is usually little or no evidence of the disease for the first 10 years after initial infliction with rheumatic carditis. However, in the ensuing one to two decades, the disease progresses and becomes clinically apparent. The course of the disease may be more aggressive in less developed countries, where severe rheumatic MS is not uncommon when the patients are still in their teenage years.

The prognosis of patients with MS basically depends of the severity of the functional class which, in turn, reflects the extent of valve involvement.[1] With conservative management, patients with New York Heart Association (NYHA) Class III can only expect a 60% and 40% five- and 10-year survival, respectively. The long-term outcome is even worse for patients in NYHA Class IV. Patients with MS frequently succumb to cardiac failure, thromboembolism and infective endocarditis. The advent of surgery and, more recently, percutaneous balloon valvuloplasty, however, have improved this grim outcome in patients with severe MS.[2,3]

TREATMENT STRATEGIES

Medical Treatment

All patients with MS should be advised and given prophylaxis for bacterial endocarditis. Asymptomatic or mildly symptomatic patients with MS require little intervention and do well over long periods of time. Medical treatment for symptomatic MS consists mainly of oral diuretics, salt restriction, control of ventricular rate (with oral beta-blockers or digoxin) and long-term, therapeutic oral anticoagulation for patients with atrial fibrillation. Acute pulmonary oedema should be aggressively treated with intravenous diuretics, the precipitating factors corrected and the heart rate slowed down if there is atrial fibrillation with fast ventricular response.

Mechanical Treatment

Basically, there are two therapeutic strategies available for patients with MS who require mechanical intervention: surgery (closed/open mitral commissurotomy or mitral valve replacement) and percutaneous balloon mitral valvuloplasty (BMV). In the past, for a certain subgroup of patients with MS, surgical commissurotomy (open or closed) was the mainstay treatment modality for relieving stenosis. It was indicated for patients with symptomatic severe MS (mitral valve area of $\leqslant 1.0\,cm^2$). However, the introduction of percutaneous BMV in 1984[4] and the very encouraging early and long-term results afforded by this technique have opened up an exciting percutaneous, non-surgical therapeutic avenue for patients with this disorder. It has also broadened the indications for intervention to include several settings not appropriate for surgery (discussed below).

BMV Compared with Surgical Commissurotomy

There is no doubt that surgical commissurotomy, either open or closed, is effective in enlarging the mitral orifice and relieving symptoms. It may also improve survival in a certain subgroup of patients. Although many centers and surgeons in "more developed" countries currently prefer the open to the closed surgical approach, because the former potentially provides a more direct and "controlled", and apparently safer, surgery, in reality there is no convincing data to support the superiority of the open over the closed technique, particularly for pliable, mobile valves without extensive subvalvular disease. On the contrary, numerous reports over the past three decades have attested to the safety and short- and long-term efficacy of closed commissurotomy in experienced centers.

BMV is a relatively new development, but it splits the fused commissures in a way that resembles closed surgical commissurotomy.[5] It is now well-established that BMV is a feasible, safe and effective non-surgical dilatational technique in selected patients. In skilled hands, the success rate of BMV is very high and the risk of major complication is extremely low, averaging less than 1% for cardiac perforation/tamponade and stroke, and a 3–8% incidence of procedure-related, severe mitral regurgitation (Table 3).[6–14]

The short-term and late results are excellent.[9,15–19] With successful dilatation, it is common to experience a 100% increase in the mitral valve area (Table 3) accompanied by a significant and dramatic fall in the transmitral gradient (Fig. 2), left atrial and pulmonary artery pressures. There is also an increase in the cardiac index because of increased flow across the widened valve. Due to the excellent early and late results (Table 4), as well as low rate of complications provided by this less aggressive technique, BMV has now been extended to patients with moderate MS (for MVA of 1–1.5 cm^2).[14]

At least six randomized trials which compared the two treatment strategies in more than 470 patients with favorable valve morphology (non-calcified, pliable valve with minimal subvalvular disease) and no or mild mitral regurgitation have amply illustrated that BMV is as efficacious as, if not better than, surgical commissurotomy, particularly closed commissurotomy, in acutely relieving the obstructed valve and in maintaining the favorable outcome (Table 5).[20–25] Another important finding in all these trials was that complications, in terms of mortality, stroke

Table 3 Acute Outcome of Percutaneous Transvenous Mitral Balloon Valvuloplasty Using the Inoue (IN) and Double-balloon (DB) Technique

References	Number of Patients	Tech	Success (%)	MVA (cm^2)		Complications (%)				
				Pre	Post	Death	Tpn	MR	ASD	Emb
Hung[6]	219	IN	99	1.0	2.0	0.5	0	6	15	1.4
Bassand[7]	232	IN (71)	100	1.1	1.9	0	0	5	1	0
		DB (161)	98	1.0	2.0	0	3	4	3	9
NHLBI[8]	591	DB	97	1.0	2.0	3	4	14	10	4
Lau[9]	105	IN	100	0.8	1.7	0	0	0	NA	0
Iung[10]	1,514	IN/DB	98.6	1.0	1.9	0.4	0.3	3.4	NA	0.3
Hung/Lau[11]	799	IN	99.5	1.0	1.8	0	0.1	4.4	8.3	1.4

ASD: significant atrial septal defect defined by oximetry
Emb: systemic embolism
MR: significant angiographic mitral regurgitation defined as an increase in severity of $\geqslant 2+$ MR or $\geqslant 3+$ final MR
MVA: mitral valve area
NA: no data available
NHLBI: National Heart, Lung and Blood Institute Registry
Pre: before BMV
Post: after BMV
Success: procedure completed without cardiac tamponade or death
Tech: technique
Tpn: tamponade

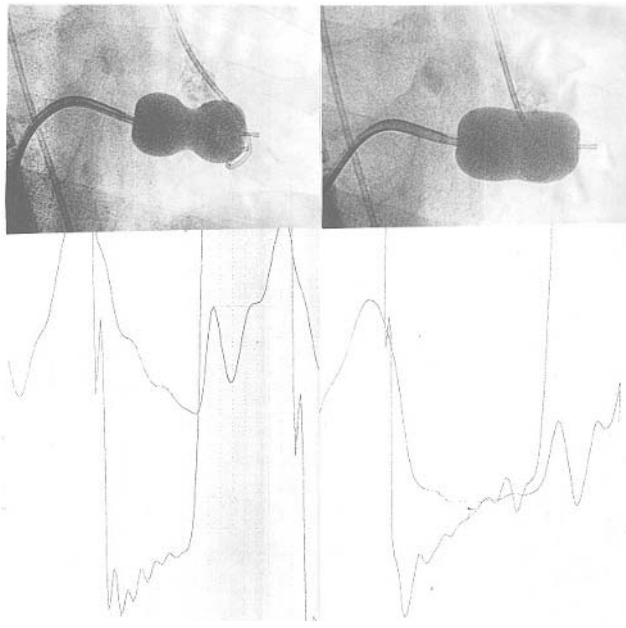


Fig. 2 Successful percutaneous balloon mitral valvuloplasty (BMV) with the Inoue balloon catheter. Note the obliteration of the waist of the Inoue balloon associated with splitting of the fused commissures despite the heavily calcified mitral valve and dramatic reduction in the transmitral valve gradient.

Table 4 Follow-up Restenosis Rate and Clinical Status After Balloon Mitral Valvuloplasty

References	Number of Patients	F-U (Months)	Restenosis Rate (%)	Restenosis Criterion	Favorable Clinical Status*
Lau[9]	102	20	10	A	97
Pan[17]	337	38	16	B	84
Lau[15]	68	44	15	A	91
Chen[16]	85	60	7	B	93

*Favorable clinical status, improved/unchanged (NYHA Class I or II) symptomatic status on follow-up.
F-U: follow-up period
A: loss of $>50\%$ of the initial gain in valve area
B: clinical restenosis

Table 5 Randomized Trials of Balloon Mitral Valvuloplasty (BMV) versus Surgical Commissurotomy

References	Patients/ Tech	Mitral Valve Area (cm^2)			F-U (Months)	Restenosis Rate (%)	Acute Complications (%)		
		Pre	Post	F-U			Death	Stroke	MR
Turi[20]	20/BMV	0.8	1.6	1.6	8	Equal	0	0	5
	20/CSC	0.9	1.6	1.8		At3.5y	0	0	5
Patel[21]	23/BMV	0.8	2.1	—	—	—	0	0	4
	22/CSC	0.7	1.3*	—	—	—	0	0	5
Arora[22]	100/BMV	0.8	2.4	2.0	24	5	2	0	—
	100/CSC	0.8	2.2	1.9		4	2	0	—
Bueno[23]	20/BMV	1.3	2.1	2.0	3	—	0	0	0
	20/CSC	1.3	2.6	2.3		—	0	0	0
Reyes[24]	30/BMV	0.9	2.1	2.4	36	10	0	0	6.6
	30/CSC	0.9	2.0	1.8*		13	0	0	3.3
Farhat[25]	30/BMV	0.9	2.2	1.8	48	7	0	0	3
	30/CSC	0.9	1.6*	1.3*		37*	0	0	0
	30/OSC	0.9	2.2	1.8		7	0	0	0

*Statistically significant for comparison between groups.
CSC: closed surgical commissurotomy
FU: follow-up period
OSC: open surgical commissurotomy
MR: significant mitral regurgitation
Tech: technique

and severe mitral regurgitation, were identical in the treatment groups. This equivalence in results are not that unexpected when one considers the close similarity in the underlying mechanism of valve enlargement effected by both techniques, namely, that of splitting the fused mitral commissures.

In contrast to surgery, BMV has the extra salutary benefits of a much shorter hospital stay, leaves no scar and is, thus, cosmetically more appealing. Furthermore, BMV is less traumatic psychologically and is cheaper than open surgical commissurotomy (and, probably, closed surgical commissurotomy in more "developed" countries). Also, it may be repeated with minimal additional risk in patients with mitral restenosis following previous BMV or surgical commissurotomy.[26] Hence it is unlike repeat surgery, which carries a significantly higher morbidity and mortality risk compared to the initial surgery. In patients who had previous surgical commissurotomy (closed or open), the valve area and hemodynamic improvements obtained with BMV are similar to those without past surgery. In high surgical risk patients, particularly those above 65 years of age, or those with poor left ventricular function, severe pulmonary hypertension or other major comorbid medical conditions, BMV offers a viable, palliative option. Thus, it is a better and more attractive choice.[1,4,27]

SUMMARY

Patients with mildly symptomatic mild MS are best treated medically. In contrast, those with moderate to severe MS, particularly when associated with favorable valve anatomic features and significant functional class impairment, should be considered for BMV, which predictably yields excellent results in such anatomic settings. On the other hand, patients with heavily calcified, immobile valves with extensive subvalvular abnormalities and fused fibrocalcified commissures are best considered for surgery — which is often that of mitral valve replacement rather than commissurotomy — because the early and late results of BMV are generally less favorable in this situation (Table 6). However, BMV may still serve as a short-term palliative measure in patients who refuse surgery or in whom modest relief of the stenosis may improve the risk profile for non-cardiac surgery. Similarly, in patients with end stage MS or those with major comorbid conditions, in whom the risk of surgery is prohibitive, BMV is a practical alternative. In the middle of the disease spectrum, namely,

Table 6 Recommended Treatment Strategies for Mitral Stenosis

Patient Subset	Recommended Treatment
1) Asymptomatic or mildly symptomatic mild MS	Conservative/medical
2) Pliable, non-calcified moderate to severe mitral stenosis with none or mild subvalvular disease, and:	
• Absent or mild mitral regurgitation	BMV
• Moderate mitral regurgitation	Trial of BMV
• Left atrial cavity thrombus	OSC*
3) Non-pliable, grossly calcified mitral stenosis with significant subvalvular disease and:	
• With or without moderate mitral regurgitation	
• Special clinical settings	MVR
— High surgical risk	BMV
— Urgent non-cardiac surgery required	
— Bridge procedure to mitral surgery	
— Patient refusal for surgery	
— Shortened life span from comorbidities	

*With thrombectomy.
MVR: mitral valve replacement
OSC: open surgical commissurotomy
BMV: balloon mitral valvuloplasty

patients with moderately adverse valve anatomy, however, should perhaps be given a trial of BMV initially as the procedure in this subgroup of patients not infrequently produces acceptable results. Should BMV fail to yield durable results in these patients, valve replacement is still feasible.

The presence of a left atrial thrombus, either attached to the interatrial septum or within the cavity, and significant mitral regurgitation ($>$Grade 2+ regurgitation according to Sellers' criteria) are absolute contraindications to BMV because of the heightened risk of embolism and resultant severe mitral regurgitation, respectively. Mitral regurgitation of $\geqslant$ 2+ severity is, however, not considered a contraindication to BMV by some investigators. Patients with MS and concomitant thrombus located within, and not beyond, the confines of the left atrial appendage are not considered contraindications to BMV by skilled interventionalists and have had BMV (usually with the Inoue balloon technique) performed successfully on them without encountering any embolism.

REFERENCES

1. Braunwald E, Valvular Heart Disease, in: Braunwald E (ed.), *Heart Disease: A Textbook of Cardiovascular Medicine*, 3rd ed., pp. 1023–1092, WB Saunders Company, Philadelphia, 1988.

2. Olesen KH, The natural history of 271 patients with mitral stenosis under medical treatment, *Br Heart J*, **24**:349–357, 1962.

3. Kulick DL, Kawanishi DT, Reid CL, Rahimtoola SH, Catheter balloon commissurotomy in adults. Part II: mitral and other stenoses, in: O'Rouke RA (ed.), *Curr Probl Cardiol*, pp. 403–470, Mosby-Year Bk. Inc., St. Louis, 1990.

4. Inoue K, Owaki T, Nakamura T, Kitamura F, Miyamoto N, Clinical application of transvenous mitral commissurotomy by a new balloon catheter, *J Thorac Cardiovasc Surg* **87**: 394–402, 1984.

5. McKay RG, Lock JE, Safian RD, Come PC, Diver DJ, *et al.,* Balloon dilation of mitral stenosis in adult patients: post-mortem and percutaneous mitral valvuloplasty studies, *J Am Coll Cardiol* **9:**723–731, 1987.

6. Hung JS, Chern MS, Wu JJ, Fu M, Yeh KH, Wu YC, *et al.,* Short- and long-term results of catheter balloon percutaneous transvenous mitral commissurotomy, *Am J Cardiol* **67**: 854–862, 1991.

7. Bassand JP, Schiele F, Bernard Y, Anguenot T, Payet M, Ba SA, *et al.*, The double-balloon and Inoue techniques in percutaneous mitral valvuloplasty: comparative results in a series of 232 cases, *Am Coll Cardiol* **18**:982–989, 1991.

8. The National Heart, Lung and Blood Institute Balloon Valvuloplasty Registry participants: multi-centre experience with balloon mitral commissurotomy, NHLBI Balloon Valvulo-plasty Registry report on immediate and 30-day follow-up results, *Circulation* **85**:448–461, 1992.

9. Lau KW, Gao W, Ding ZP, Hung JS, Immediate and long-term results of percutaneous Inoue-balloon mitral commissurotomy using a simple height-derived balloon sizing method for the stepwise dilatation technique, *Mayo Clin Proc* **71**:556–563, 1996.

10. Iung B, Cormier B, Ducimetiere P, Porte JM, Nallet O, Michel PL, *et al.*, Immediate results of percutaneous mitral commissurotomy. *Circulation* **94**:2124–2130, 1996.

11. Hung JS, Lau KW, Lo PH, Chern MS, Wu JJ, Complications of Inoue-balloon mitral commissurotomy: impact of operator experience and evolving technique, *Am Heart J* **38**:114–121, 1999.

12. Lau KW, Ding ZP, Hung JS, Percutaneous balloon mitral commissurotomy: an update, *J Invas Cardiol* **6**:145–153, 1994.

13. Lau KW, Hung JS, A simple balloon-sizing method in Inoue-balloon percutaneous transvenous mitral commissurotomy, *Cathet Cardiovasc Diagn* **33**:120–129, 1994.

14. Lau KW, Hung JS, Ding ZP, Johan A, Controversies in balloon mitral valvuloplasty: the when (timing for intervention), what (choice of valve) and how (selection of technique), *Cathet Cardiovasc Diagn* **35**:91–100, 1995.

15. Lau KW, Ding ZP, Quek S, Kwok V, Hung JS, Long-term (36 to 63 months) clinical and echocardiographic follow-up after Inoue balloon mitral commissurotomy, *Cathet Cardiovasc Diagn* **42**:33–38, 1998.

16. Chen CR, Cheng TO, Chen JY, Long-term results of percutaneous mitral valvuloplasty with the Inoue balloon catheter, *Am J Cardiol* **70**:1445–1448, 1992.

17. Pan M, Medina A, de Lezo JS, Hernandez E, Pan M, Ortega JR, *et al.*, Factors determining late success after mitral balloon valvulotomy, *Am J Cardiol* **71**:1181–1185, 1993.

18. Iung B, Cormier B, Ducimetiere P, Porte JM, Nallet O, Michel PL, *et al.*, Functional results five years after successful percutaneous mitral commissurotomy in a series of 528 patients and analysis of predictive factors, *J Am Coll Cardiol* **27**:407–414, 1996.

19. Dean LS, Mickel M, Bonan R, Holmes DR, O'Neill WW, Palacios IF, *et al.*, Four-year follow-up of patients undergoing percutaneous balloon mitral commissurotomy, *J Am Coll Cardiol* **28**:1452–1457, 1996.

20. Turi ZG, Reyes VP, Raju AR, Kumar DN, Rajagopal P, *et al.*, Percutaneous balloon versus surgical closed commissurotomy for mitral stenosis: a prospective, randomised trial, *Circulation* **83**:1179–1185, 1991.

21. Patel JJ, Shama D, Mitha AS, Blyth D, Hassen F, Le Roux BT, *et al.*, Balloon valvuloplasty versus closed commissurotomy for pliable mitral stenosis: a prospective haemodynamic study, *J Am Coll Cardiol* **18**:1318–1322, 1991.

22. Arora R, Nair M, Kalra GS, Nigam M, Khalilullah M, *et al.*, Immediate and long-term results of balloon and surgical closed mitral valvotomy: a randomised, comparative study, *Am Heart J* **125**:1091–1094, 1993.

23. Bueno R, Andrade P, Nercolini D, Percutaneous balloon mitral valvuloplasty versus open mitral valve commissurotomy: a randomised clinical trial (abstr), *J Am Coll Cardiol* **21**:429A, 1993.

24. Reyes VP, Raju BS, Wynne J, Stephenson LW, Raju R, Fromm BS, *et al.*, Percutaneous balloon valvuloplasty compared with open surgical commissurotomy for mitral stenosis, *N Engl J Med* **331**:961–967, 1994.

25. Farhat MB, Ayari M, Maatouk F, Betbout F, Gamra H, Jarra M, *et al.*, Percutaneous balloon versus surgical closed and open mitral commissurotomy: seven-year follow-up results of a randomised trial, *Circulation* **97**:245–250, 1998.

26. Lau KW, Ding ZP, Gao W, Percutaneous balloon mitral valvuloplasty in patients with mitral restenosis after previous surgical commissurotomy: a matched comparative study, *Eur Heart J* **17**:1367–1372, 1996.

27. Lau KW, Ding ZP, Hung JS, Percutaneous transvenous mitral commissurotomy versus surgical commissurotomy in the treatment of mitral stenosis, *Clin Cardiol* **20**:99–106, 1997.

This page intentionally left blank

6

Echocardiography

Ding Zee Pin

Echocardiography was first introduced in the mid-1950s with M-mode echocardiography. In 1954, Dr. Edler of Sweden published the movement of a cardiac structure: that of the mitral valve. M-mode echocardiography, however, could only provide an ice-peak view of the heart and hence it had limited applications in the evaluation of the heart. In the mid-1970s, two-dimensional echocardiography was introduced and for the first time tomographic images of the heart could be appreciated. Although the development of Doppler echocardiography began in the 1950s, like M-mode echocardiography, it only became clinically utilized in the 1970s. Doppler echocardiography allows velocity to be measured accurately. By using the modified Bernoulli equation, the pressure gradients across chambers and across valves could be reliably estimated. This has been validated in several studies comparing pressure gradients estimated by the Doppler method with that obtained from cardiac catheterization. The cardiac output, as well as shunt ratios, could also be estimated using Doppler echocardiography. In the mid-1980s, the flow of blood was depicted in color. The color-flow imaging further helped to enhance the ability of echocardiography to evaluate flow within the

heart, particularly valvular regurgitation and shunts. These modalities of echocardiography — two-dimensional and M-mode echocardiographies, Doppler and color-flow imaging — enabled echocardiography to evaluate a wide spectrum of heart diseases. It could evaluate not only cardiac morphology and function, but it also provided insight into the hemodynamics of the heart. These include valvular heart disease, ischemic heart disease, various cardiomyopathies, pericardial diseases and pathologies of the proximal pulmonary artery and thoracic aorta.

VALVULAR HEART DISEASE

Echocardiography is an excellent imaging tool in the evaluation of valvular heart disease. Two-dimensional and M-mode echocardiography assess the morphology of the valves. Hence the etiology of the valves can be determined while Doppler echocardiography evaluates the hemodynamics of the heart, such as pressure gradients. In addition, chamber sizes and function can also be easily assessed. Patients with valvular heart disease evaluated by echocardiography can be sent for surgery without further confirmation by cardiac catheterization. In view of the accuracy of echocardiographic findings, cardiac catheterization is only required if there is discrepancy between clinical evaluation of the severity of the valvular lesion and the echocardiographic findings.

Mitral stenosis (MS) is most commonly caused by rheumatic fever. There are, however, other less common causes of MS including congenital etiology, which is almost exclusively seen in infants and young children. Rarely does MS occur in patients with the carcinoid syndrome (although valvular lesions are usually limited to the right side of the heart), Systemic Lupus Erythematosus, Rheumatoid Arthritis and the Mucopolysaccharidoses of the Hunter and Hurdler Syndromes. MS due to rheumatic fever, gives rise to the characteristic features on two-dimensional echocardiography. The thickening of the valve is more marked at the tip of the leaflets. The tip of the anterior leaflet is tethered down, thus giving rise to the characteristic doming of the anterior leaflet (Fig. 1a). The posterior leaflet is usually immobile. In the parasternal short axis, fusion of the commissures is obvious. This gives rise to the "fish mouth" appearance of rheumatic MS (Fig. 1b). From the parasternal short axis view, the mitral valve area can be accurately estimated by the planimetry of the mitral valve orifice and this is the most accurate

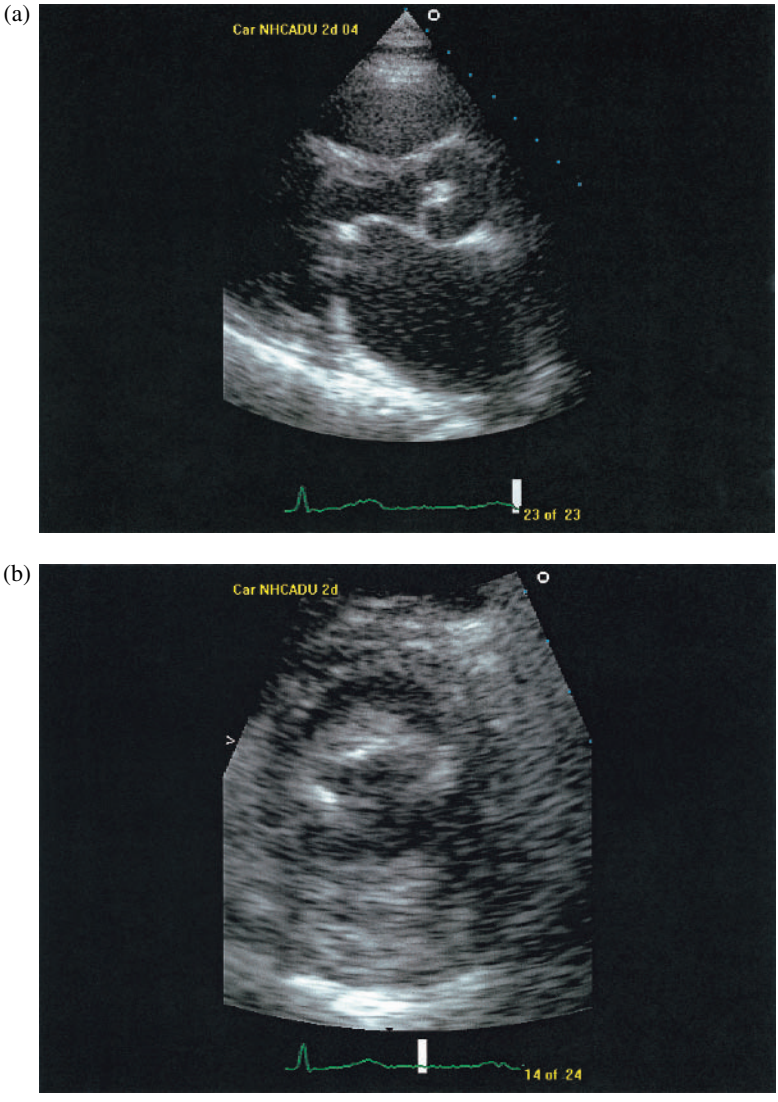


Fig. 1 a) Parasternal long axis view of a patient with mitral stenosis showing characteristically doming of the anterior leaflet and thickening of the leaflets, particularly at the tips. b) Parasternal short axis view of severe mitral stenosis showing "fish mouth" appearance of the mitral orifice due to fusion of the commissures.

method for estimating the mitral valve area in the presence of good images. Doppler echocardiography enables the gradient across the mitral valve to be accurately estimated. The mitral valve area could also be

estimated using the pressure half-time method and PISA method.[1] Color Doppler is useful in viewing the direction of the mitral stenotic jet. It guides the direction of the Doppler beam to ensure accurate estimation of pressure grades and also enables evaluation of concomitant mitral regurgitation (MR). In contra-distinction to rheumatic mitral valves, in MS due to Systemic Lupus Erythematosus, the thickening of the mitral valve is usually uniform throughout. In MS that is due to Hunter and Hurdler's syndrome, there is no fusion of the commissures. This helps to distinguish it from rheumatic MS.

Again, in MR, two-dimensional echocardiography and, to a certain extent, M-mode echocardiography, can help to establish the etiology of MR. The morphology of the rheumatic mitral valve has been described earlier. Prolapse of the mitral valve on M-mode echocardiography is depicted as 1–2 mm posterior displacement of one or both leaflets, or a hollow systolic hammocking of > 3 mm. In two-dimensional echocardiography, Mitral Valve Prolapse is seen as systolic displacement of one or two mitral leaflets in the parasternal view (Fig. 2A). The saddle-shaped mitral annulus precludes a diagnosis of mitral valve prolapse in the apical four-chamber view. Besides a diagnosis of mitral valve prolapse, features of mitral valve prolapse on two-dimensional echocardiography also help to risk-stratify the patients. A number of studies have indicated that mitral valve thickness of more than 4 mm (as measured at the midportion of the mitral leaflets), leaflet redundancy, presence of MR and an increased left ventricular or left atrial size — particularly in men above 45 years of age — are associated with a worse prognosis. When the mitral leaflets are normal in a patient with MR, one must carefully look out for abnormalities in other parts of the mitral valve. The mitral apparatus consists not only of the mitral leaflets, but also the mitral annulus, subvalvular apparatus and papillary muscles. An abnormality in these components can also result in MR. Echocardiography could also assess the size of the mitral annulus, detect the presence of ruptured cordiae — which gives rise to a flail mitral valve and papillary muscle dysfunction. Ischemic mitral regurgitation is recognized as an important clinical entity. Patients with moderate ischemic MR is associated with a poor prognosis. On echocardiogram, the mitral valve is morphologically normal but the geometry of the leaflet are abnormal due to the displaced papillary muscles from LV remodelling leaflets. The leaflet tips are apically displaced, giving rise to tenting appearance of this valve. There is a bending

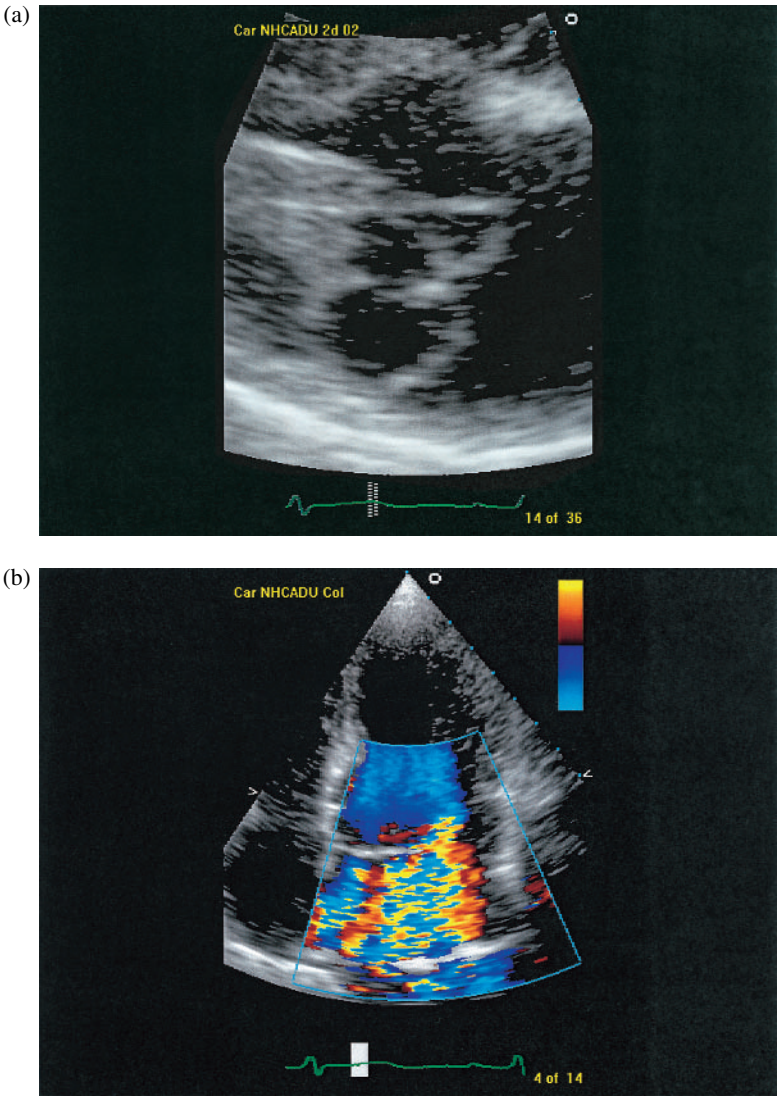


Fig. 2 a) Zoom parasternal long axis view of the mitral valve showing bileaflet prolapse. b) Apical 4-chamber view showing severe mitral regurgitation.

mid-anterior leaflet due to the tethering effect of the secondary chords, from displaced papillary muscle.[8]

The evaluation of the severity of MR has been shown to be comparable to that of cardiac catheterization, which is considered the "gold

standard". The evaluation of the severity of MR can be both semi-quantitative and quantitative. The conventional method of evaluating MR is based on the extent of the mitral regurgitant jet,[3–4] the presence of mitral regurgitant flow in pulmonary veins, and the density of the continuous wave of Doppler signal of MR (as the intensity of the signal is proportional to the number of red blood cells (RBC) and, hence, the degree of MR). A recent method of evaluating MR uses quantitative methods, which calculate the effective regurgitant orifice, most commonly using the PISA method, and regurgitant fraction.

In aortic stenosis, two-dimensional echocardiography and M-mode echocardiography allow the morphology of the aortic valve to be assessed. Aortic stenosis that is due to rheumatic fever is caused by fusion of the commissure of the valve. This can be seen on two-dimensional echocardiography. Often, it also occurs concomitantly with mitral involvement. In the elderly, the most common cause of aortic stenosis is calcific degenerative aortic stenosis. The presence of calcification without commissural fusion indicates a degenerative cause. The severity of aortic stenosis by two-dimensional, M-mode is semi-quantitative. However, on transesophageal echocardiography — which has a clearer view of the aortic valve — the planimetry of the aortic valve can accurately assess the severity of the aortic stenosis. Generally, M-mode and two-dimensional echocardiography allow, at best, a semi-quantitative evaluation of the degree of aortic stenosis. The severity of aortic stenosis is based on Doppler evaluation and this has been shown by several studies that the pressure gradient across the valve, as well as the aortic valve area calculated using the continuity equation, is comparable to that obtained by cardiac catheterization.[5–7] It must be mentioned that the peak instantaneous pressure calculated using the Doppler method is different from the peak to peak pressure gradient that is often assessed in the cardiac catheterization laboratory. The peak instantaneous pressure is often not measured in the cardiac catheterization laboratory. This is because it is time-consuming and tedious, involving two catheters: one in the left ventricle and the other in the aorta in order to measure the peak instantaneous pressure gradient.

In aortic regurgitation, two-dimensional echocardiography and M-mode echocardiography evaluate the morphology of the aortic valve, as well as the size of the annulus and root of the aorta to establish the etiology of the underlying cause for aortic regurgitation. The etiology of

aortic regurgitation is not only limited to aortic cusps, but can also be due to an abnormality of the aortic annulus and the root of the aorta. The severity of aortic regurgitation is based on Doppler and color-flow imaging. These include the width of the color-flow jet just below the aortic valve, the pressure half-time of the aortic regurgitation Doppler signal and the presence of pan-diastolic flow reversal in the descending thoracic aorta and the abdominal aorta. Mitral inflow abnormality may also help to evaluate the severity of aortic regurgitation. In both tricuspid and pulmonary valve disease, the two-dimensional echocardiogram assesses the morphology and, hence, etiology of the valvular abnormality, whereas Doppler and color-flow imaging assess the severity of the valvular lesion.

ISCHEMIC HEART DISEASE

In ischemic heart disease, no abnormalities may be detected on the resting two-dimensional echocardiogram. In patients with recent or past myocardial infarction, regional wall motion abnormalities comprising both akinetic and/or hypokinetic segments may be present. In patients with severe coronary artery disease without infarction, regional wall motion abnormalities may also be seen. Most importantly, echocardiography evaluates the left ventricular systolic function, which provides significant prognostic information. Recent data also suggests that the evaluation of diastolic function may further risk-stratify patients with poor left ventricular function. Generally, a resting echocardiogram does not help in the evaluation of ischemic heart diseases, unless there is past myocardial infarction, cardiomegaly on the chest X-ray or, upon a physical examination, the presence of murmurs and patients with congestive cardiac failure.

CARDIOMYOPATHY

Echocardiography is excellent in the evaluation of patients with suspected cardiomyopathy: Hypertrophic cardiomyopathy, restrictive cardiomyopathy and dilated cardiomyopathy. In hypertrophic cardiomyopathy, the thickness of the ventricles and the site of the hypertrophy could be reliably assessed. Typically, hypertrophy is not concentric but asymmetric. The presence of systolic anterior motion of the mitral valve (SAM) on two-dimensional echocardiography (Fig. 3A) and M-mode echocardiography indicates the presence of left ventricular outflow tract

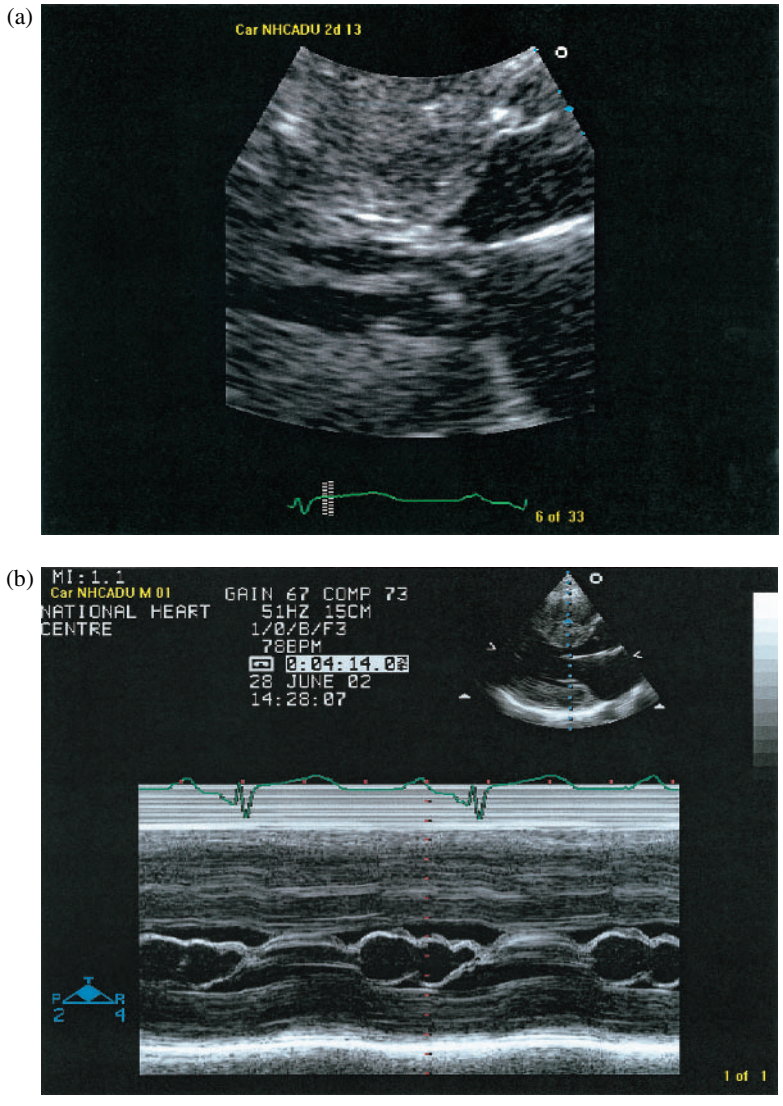


Fig. 3 a) Zoom parasternal long axis view of the mitral valve showing systolic anterior motion (SAM) of the mitral valve. b) M-mode of the left ventricle at the level of the mitral valve demonstrating systolic anterior motion (SAM) of the mitral valve.

(LVOT) obstruction and MR. The systolic as well as diastolic function could also be evaluated. The presence of intracardiac obstruction, including LVOT obstruction, can be accurately estimated using Doppler

echocardiography. This could be serially evaluated following drug therapy. The severity of MR and the morphology of the mitral valve can also be evaluated.

In Restrictive Cardiomyopathy, the ventricles are traditionally of normal size and function. However, the atria are often dilated. The hallmark of restrictive cardiomyopathy is restrictive filling of the left ventricle (LV), which can be evaluated using Doppler echocardiography.

Dilated cardiomyopathy is characterized by a dilated LV with impaired contractility. The right ventricle may be similarly affected. Patients with amyloid heart disease have characteristic sparkling appearance in the myocardium. In addition, Doppler evaluation of the mitral inflow provides prognostic information. Patients with abnormal relaxation patterns are associated with a better prognosis than those with a restrictive filling.

PERICARDIAL DISEASES

One of the most easily diagnosed conditions on echocardiography is pericardial effusion. It is seen as an area of echo-free space surrounding the heart. The presence of pericardial tamponade is suggested by the presence of a significant variation in the mitral inflow. In particular, the E-velocity is lowest at the first beat after onset of inspiration. Constrictive pericarditis, although an uncommon cause of congestive cardiac failure, should be suspected in all patients with congestive cardiac failure and normal LV function as this is an easily treatable condition. Constrictive pericarditis is suggested by the presence of a thickened myocardium. The hemodynamics of the pericardial constriction are also seen on the echocardiography. These include ventricular interdependence on M-mode and two-dimensional imaging, as well as significant mitral inflow variation, particularly with the lowest E-velocity on the first beat after the onset of inspiration, as in cardiac tamponade.

CONCLUSION

Echocardiography remains an important imaging tool in the evaluation of a wide spectrum of cardiovascular diseases. Besides its accuracy, it is non-invasive and hence enables the serial evaluation of patients without

risk of irradiation. The current echocardiography machine is compact and can be brought to the patient's bedside, which is an advantage over other imaging techniques.

REFERENCES

1. Hatle L, Brubakk A, Tromsdal A, Angelsen B, Non-invasive assessment of pressure drop in mitral stenosis by Doppler ultrasound, *Br Heart J* **40**:131–140, 1978.

2. Nakatani S, Masuyama T, Kodama K, Kitabatake A, Fujii K, Kamada T, Value and limitations of Doppler echocardiography in the quantification of stenotic mitral valve area: comparison of the pressure half-time and the continuity equation methods, *Circulation* **77**:78–85, 1988.

3. Helmecke F, Nanda NC, Hsiung MC, *et al*., Colour Doppler assessment of mitral regurgitation with orthogonal planes, *Circulation* **75**:175–183, 1987.

4. Spain MG, Smith MD, Grayburn PA, Harlamert EA, DeMaria AN, Quantitative assessment of mitral regurgitation by Doppler color flow imaging: angiographic and hemodynamic correlations, *J Am Coll Cardiol* **13**:585–590, 1989.

5. Currie PJ, Seward JB, Reeder GS, *et al*., Continuous-wave Doppler echocardiographic assessment of severity of calcific aortic stenosis: simultaneous Doppler-catheter correlative study in 100 adult patients, *Circulation* **71**:1162–1169, 1985.

6. Skjaerpe T, Hegrenaes L, Hatle L, Noninvasive estimation of valve area in patients with aortic stenosis by Doppler ultrasound and two-dimensional echocardiography, *Circulation* **72**:810–818, 1985.

7. Oh JK, Taliercio CP, Holmes DR Jr, *et al*., Prediction of the severity of aortic stenosis by Doppler aortic valve area determination: prospective Doppler-catheterization correlation in 100 patients, *J Am Coll Cardiol* **11**:1227–1234, 1988.

8. Kwan J, Shiota T, Agler DA, *et al*., Geometric differences of the mitral apparatus between ischaemic and dilated cardiomyopathy with significant mitral regurgitation. Real-time three-dimensional echocardiography study, *Circulation* **107**:1135–1140, 2003.

7

Heart Failure

Bernard WK Kwok and Amy SH Ng

DEFINITION

Heart failure is a common condition, being the top non-elective cardiac cause of hospitalization. It is the end-stage of a number of different cardiac diseases. However, there is no universally accepted definition of heart failure. It is, instead, increasingly recognized as a clinical syndrome. Cardinal manifestations are fluid retention, dyspnea, and reduced exercise tolerance. There are also associated characteristic patterns of hemodynamic, renal, neural and hormonal responses.

FRAMINGHAM CRITERIA FOR HEART FAILURE

Over the years, the Framingham criteria remain one of the most clinically relevant.[1] Its mainly clinical criteria, without the need for specialized investigations, makes it easy to use. In addition, the set of criteria has been validated prognostically.

Table 1 Framingham Criteria for Heart Failure

Major Criteria	Minor Criteria
paroxysmal nocturnal dyspnea	dyspnea on exertion
neck vein distension	nocturnal cough
rales	ankle edema
cardiomegaly	hepatomegaly
pulmonary edema	pleural effusion
third heart sound	tachycardia
increased venous pressure >16 cm water	reduced vital capacity by one-third
hepatojugular reflux	
Major or Minor Criteria	
Weight loss ≥4.5 kg in 5 days in response to treatment	

*For a definite diagnosis of heart failure, 2 major or 1 major and 2 minor criteria have to be present concurrently. Adapted from *N Engl J Med* 1971;**285**:1441–1446.

Table 2 New York Heart Association Classification for Heart Failure

Class I	no limitation during ordinary physical activities
Class II	slight limitations, e.g. dyspnea on walking
Class III	marked limitation, symptoms easily provoked
Class IV	breathlessness at rest

NEW YORK HEART ASSOCIATION (NYHA) FUNCTIONAL CLASS

The severity of symptoms in heart failure patients is usually classified according to their functional status. The NYHA functional classification, although subjective, is useful clinically as well as in trials. A modified version is listed in Table 2. This system assigns patients to 1 of 4 functional classes, based on symptoms provoked by different degrees of exertion.

ETIOLOGY OF HEART FAILURE

In Singapore, the common causes of heart failure are coronary artery disease, hypertension, cardiomyopathies and valvular heart disease.

Less common causes are myocarditis, thyrocardiac disease, etc. Locally, the mean age of the hospitalized heart failure patient is 69 years. However, female patients tend to be older (72 years) than male patients (66 years). Gender distribution is almost equal. The epidemiology of heart failure in the local outpatient population is less clear.

RISK FACTORS FOR HEART FAILURE

Epidemiological data from the Framingham study have identified the following modifiable predisposing risk factors for heart failure:[2,3]

Hypertension

Hypertension is the commonest and most important risk factor for the development of heart failure. The increased risk of heart failure in hypertensive subjects was 2-fold in males and 3-fold in females when compared with normotensive subjects. Survival following the diagnosis of heart failure in hypertensive subjects was poor. Fortunately, risk reduction following blood pressure control is possible.

Left ventricular hypertrophy (LVH), a common sequela of hypertension, is itself an independent cardiovascular risk factor. Electrocardiographic LVH has been shown to be a greater predisposing factor to heart failure than X-ray cardiac enlargement.

Diabetes Mellitus

Diabetes predisposes to heart failure in part due to its association with accelerated coronary atherosclerosis. It appears to have a direct deleterious effect on the myocardium, and is an independent risk factor for the development of heart failure. Diabetes therefore predisposes to heart failure whether or not there is interim overt coronary artery disease.[4] The risk is greater in women than in men.

Hypercholesterolemia

Hypercholesterolemia is well-known as a risk factor for coronary artery disease. Multivariate analyses have shown it to be an independent risk factor for heart failure as well.

Obesity

Extreme obesity is recognized to be a risk factor for heart failure. It was recently shown that lesser degrees of obesity also pose a risk.[5] Obese patients have double the risk as compared to patients with normal body-mass index (BMI). For each increment of 1 in BMI, the risk of heart failure increases by 5% in males and 7% in females.

Smoking

Cigarette smoking increases the risk of coronary atherosclerosis in both sexes. It increases the risk of heart failure especially in younger men and older females.

PREVENTION OF HEART FAILURE

Heart failure is the endpoint of the cardiovascular continuum. Interrupting this continuum by treatment and control of the above risk factors reduces the risk of developing heart failure. Heart failure is, therefore, largely a preventable condition. The poor long-term outlook in heart failure patients despite advances in medical therapy testifies to prevention as the best management strategy.

INVESTIGATIONS

The goals of investigations are to confirm the diagnosis and etiology of heart failure, determine the severity and prognosis of patients, and monitor the progression of the condition.[6,7]

General investigations essential in the initial work-up of the heart failure patient include full blood counts, renal panel and thyroid function tests. Frequent assessment of renal function may be necessary in the management of heart failure, especially with the present pharmacological combinations. Cardiac enzymes (creatine kinase, creatine kinase-MB fraction) and troponins may help indicate the presence of coronary artery disease as the etiology of heart failure. Specific cardiac tests should include electrocardiograph (ECG) and chest radiograph. These may be repeated, if necessary, to monitor progression of disease. Of particular interest is the recent use of B-type natriuretic peptides in the diagnosis of heart failure.

Used in conjunction with other clinical information, rapid measurement of B-type natriuretic peptide is useful in establishing or excluding the diagnosis of heart failure in patients with acute dyspnea.[8]

The left ventricular ejection fraction (LVEF) has important prognostic value. More specialized cardiac evaluations for left ventricular function include echocardiography, multigated acquisition blood pool nuclear scanning (MUGA), and magnetic resonance imaging (MRI). Echocardiography, being non-invasive and widely available, is presently the most important investigative tool in the assessment of heart failure. In addition to assessment of LVEF, it may also provide clues to the etiology of the heart failure.

Coronary angiography may be necessary in the evaluation of the heart failure patient due to coronary artery disease. It can help identify the patients who will benefit from revascularization procedures to improve cardiac function.

SYSTOLIC VS. DIASTOLIC HEART FAILURE

It is easy to comprehend systolic dysfunction with impaired LVEF as the underlying reason for heart failure. However, it is increasingly recognized that diastolic dysfunction is almost as common a cause of heart failure. Its prevalence is estimated to be as high as 50%, especially in the elderly. Classic causes of diastolic heart failure are hypertension and hypertrophic cardiomyopathy.

While the definition of diastolic heart failure is still unresolved, expert consensus is that diastolic heart failure is present when there are classical symptoms of heart failure, with evidence of pulmonary congestion, in the presence of normal left ventricular systolic function (LVEF $\geqslant$ 50%).[9]

There are currently no treatment guidelines for the management of diastolic heart failure. Based on pathophysiology, it is postulated that beta-blockers and heart rate-lowering calcium channel blockers may be beneficial in such patients. Diuretics should be used judiciously. The following sections dealing with treatment of heart failure refer to treatment for systolic heart failure.

TREATMENT OF HEART FAILURE

The therapeutic aims in heart failure are to improve quality of life, longevity and to prevent progression of the syndrome. Lifestyle changes

are essential in the successful management of heart failure. Abstinence from smoking and alcohol consumption should be total. The adoption of a healthy diet with low salt (sodium) and low fat content should be continually reinforced. Physical activities need to be tailored to patients' condition.

MANAGEMENT OF ACUTE HEART FAILURE

All patients should undergo evaluation to determine the etiology of the heart failure. Where appropriate, the underlying causes should be reversed or treated to improve cardiac function. Secondarily, precipitating causes should be identified and corrected. Common precipitating causes are listed in Table 3.

Specific therapy in the management of acute heart failure include bed rest, dietary sodium restriction, supplemental oxygen, and diuretics to promote negative sodium and fluid balance. Nitrates help by reducing the cardiac preload and in cases where there is underlying ischemia. Intravenous morphine in severe cases help by increasing the venous capacitance and allaying patients' anxiety.

More severe cases of heart failure with cardiogenic shock may require invasive therapy with a pulmonary artery catheter, an intraaortic balloon pump and intravenous inotropic agents (dopamine, dobutamine, milrinone).

MANAGEMENT OF CHRONIC HEART FAILURE

Angiotensin-converting enzyme (ACE) inhibitors/ Angiotensin receptor blockers (ARB)

ACE inhibitors remain the cornerstone of heart failure therapy. Large clinical trials have firmly established the benefits of ACE inhibition in heart failure patients — in terms of mortality and morbidity. ACE

Table 3 Common Precipitating Causes of Heart Failure

- Ischemia
- Arrhythmia
- Infection
- Anemia
- Drugs, e.g. NSAIDs, steroids
- Poor compliance (salt/fluid/medications)

Table 4 Examples of Some ACE Inhibitor Trials in Heart Failure

Trial	ACE Inhibitor	Starting Dose	Target Dose
SOLVD	Captopril	6.25–12.5 mg tds	25–50 mg tds
CONSENSUS	Enalapril	1.25–2.5 mg bd	10 mg bd
AIRE	Ramipril	1.25–2.5 mg bd	5 mg bd
ATLAS	Lisinopril	2.5–5 mg om	20–35 mg om

inhibitors are now indicated for patients with left ventricular dysfunction, whether symptomatic or not. A mortality reduction of 20–30% is achieved in symptomatic patients, whereas reduced morbidity of up to 30% is seen in asymptomatic patients with left ventricular dysfunction.

In using ACE inhibitors, the dose should be up-titrated to that used in clinical trials, or to the highest tolerable dose (Table 4). Hyperkalemia and postural hypotension remain the major limitations to their use.

A new class of drugs, ARBs, provide ACE inhibition at a different level of the renin-angiotensin pathway. Its efficacy in treating heart failure is believed to be similar to ACE inhibitors'. They are alternatives for patients who develop ACE inhibitor-induced cough. For patients with contraindications to ACE inhibition, the combination of hydralazine and isosorbide dinitrate has been shown to improve mortality, although to a less favorable degree than ACE inhibitors.

Beta-blockers

While the use of beta-blockers seems counter-intuitive to the treatment of heart failure, recent trials have shown the benefits of several agents in this class of drugs. Meta-analyses have shown the degree of benefit to be equivalent, if not greater, than that of ACE inhibitors. Mortality reduction of 30% can be expected with its use. It is important to remember that beta-blockers should only be started in stable patients who are euvolemic. The initial dose should be low and up-titrated not less than at fortnightly intervals, to that of trial doses or patients' maximal tolerable dose. The maxim "start low, go slow" has to be followed closely.

Spironolactone

In patients with advanced heart failure (NYHA class III-IV), low-dose spironolactone improves mortality. When used at the dose of 12.5 mg or

Table 5 Examples of Beta-blockers with Mortality Benefits

Beta-blocker	Starting Dose	Target Dose
Carvedilol	3.125 mg bd	25 mg bd
Bisoprolol	1.25 om	10 mg om
Metoprolol	12.5 mg om	200 mg om

25 mg daily, spironolactone has minimal diuretic properties. The mechanism of benefit is postulated to be due to neurohormonal modulation. Reduction in myocardial interstitial fibrosis is also believed to be a mechanism of its benefit. When used in combination with ACE inhibitors/ARBs, frequent monitoring of serum for hyperkalemia is essential.

Digoxin

Digoxin has long been used to treat heart failure. However, its efficacy has only been recently demonstrated. Although digoxin has no significant effect on mortality, its use is associated with improved exercise capacity and reduced hospitalizations in heart failure patients.

Diuretics

The use of diuretics to provide symptomatic relief in heart failure patients has long been entrenched in our medical practice. Although no clinical trial evidence exists to show the mortality benefits of diuretics in patients with stable chronic heart failure, diuretics remain essential in the management of acute heart failure. Diuretics are potent agents to reduce salt and fluid retention. Diuretics, with salt restriction, are the best therapy for the edematous heart failure patient. The fact that diuretics can acutely improve patients' symptoms means it is unlikely we will ever see a placebo-controlled trial of its use in heart failure. Frusemide, bumetanide and hydrochlorothiazide will remain mainstays in the treatment armamentarium.

Inotropic Drugs

Oral positive inotropic agents increase mortality and should not be used to treat heart failure. Intravenous inotropic agents (e.g. dobutamine, milrinone) can improve symptoms in patients with advanced heart failure. However, they should not be used routinely, as there is some concern that

Table 6 Mortality Benefits in the Medical Therapy of Heart Failure

	ACE I/ARB	Beta-blockers	Spironolactone
NYHA I	+	+/(?)	?
NYHA II	+	+	?
NYHA III	+	+	+
NYHA IV	+	+	+

they may improve symptoms at the risk of increased mortality. They remain agents of last ditch efforts to keep alive patients with severe heart failure.

ARRHYTHMIA IN HEART FAILURE

Heart failure and arrhythmia commonly occur together and aggravate each other. Reversible factors that predispose to arrhythmia should be treated or controlled for the optimal treatment of heart failure. Class I anti-arrhythmic drugs should not be used in heart failure patients. Atrial fibrillation in heart failure requires rate control and full anti-coagulation. Patients with symptomatic ventricular arrhythmia should be considered for an implantable cardiac defibrillator (ICD). Amiodarone and beta-blockers are the alternatives in those who do not qualify for ICDs.

ADVANCED THERAPIES IN HEART FAILURE

Cardiac Resynchronization Therapy

Also known as biventricular pacing, the implantation of this pacemaker with insertion of 3 leads into the right atrium, right ventricle, and left ventricle (via the distal branch of the coronary vein) attempts to restore the electrical synchrony in some heart failure patients. Suitable patients for this modality are those with a widened QRS morphology on ECGs. Meta-analysis of clinical trials suggests modest mortality benefits, in addition to symptomatic improvement.[10]

Ventricular Reconstructive Surgery

Originally developed as potential alternatives to heart transplantation, reconstruction or volume reduction surgery aims to restore the remodeled globular heart to its original conical shape. By removing or excluding parts of the dysfunctional myocardium, it is hoped that the left ventricular

chamber will return to its normal geometry. According to the Law of Laplace, reduction in the internal radius results in decreased wall tension, thereby leading to a more efficient pump.

Ventricular Assist Devices

Present day assist devices are what their name implies — merely assist devices. Implanted in *parallel* to the patients' heart, they augment the heart's pumping action. They function as "bridge-to-transplant," enabling advanced heart failure patients to live functionally while awaiting donor hearts for transplantation. Their availability has effectively allowed heart failure patients to live longer while on the transplant waiting list. The permanent total artificial heart is still technologically in its infancy.

Heart Transplantation

Heart transplantation remains the definitive treatment for end-stage heart failure. Transplant recipients, when rehabilitated, can expect a return to full functional state. One-year and 5-year survival rates of 80% and 65% respectively are no doubt better than the dismal outlook of the end-stage heart failure patient, where a 1-year survival of less than 50% is expected.

END-OF-LIFE ISSUES

Despite many innovations and advances in the therapy of heart failure, the prognosis of such patients remains poor. Five-year mortality of patients diagnosed with heart failure remains high at 40–50%. Sudden death is common, occurring in half of all heart failure cases, with progressive pump failure the mode of death in the other half. Education of patients and family regarding the expected course of the illness, terminal treatment options and plans for "living wills" are important. Discussion of end-of-life issues can help patients and family better accept their physical limitations and the inevitable.

REFERENCES

1. McKee PA, Castelli WP, McNamara PM, Kannel WB, The natural history of congestive heart failure: the Framingham study, *N Engl J Med* **285**:1441–1446, 1971.

2. Kannel WB, Belanger AJ, Epidemiology of heart failure, *Am Heart J* **121**:951–957, 1991.

3. Ho KK, Pinsky JL, Kannel WB, Levy D, The epidemiology of heart failure: the Framingham Study, *J Am Coll Cardiol* **22**:6A–13A, 1993.

4. Kannel WB, McGee DL, Diabetes and cardiovascular diseases: the Framingham study, *JAMA* **241**:2035–2038, 1979.

5. Kenchaiah S, Evans JC, Levy D, *et al.*, Obesity and the risk of heart failure, *N Engl J Med* **347**:305–313, 2002.

6. LeJemtel TH, Sonnenblick EH, Frishman WH, Diagnosis and management of heart failure, in Fuster V (ed.), *Hurst's The Heart*, 10th Ed, McGraw–Hill Companies, Inc., pp. 687–724, 2001.

7. Hunt SA, Baker DW, Chin MH, *et al.*, ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to revise the 1995 Guidelines for the Evaluation and Management of Heart Failure), *J Am Coll Cardiol* **38**:2101–2113, 2001.

8. Maisel AS, Krishnaswamy P, Nowak RM, *et al.*, Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure, *N Engl J Med* **347**: 161–167, 2002.

9. Zile MR, Brutsaert DL, New concepts in diastolic dysfunction and diastolic heart failure: Part I: diagnosis, prognosis, and measurements of diastolic function, *Circulation* **105**:1387–1393, 2002.

10. Bradley DJ, Bradley EA, Baughman KL, *et al.*, Cardiac resynchronization and death from progressive heart failure: a meta-analysis of randomized controlled trials, *JAMA* **289**:730–740, 2003.

This page intentionally left blank

Dermatology

This page intentionally left blank

8

Approach to Skin Disorders

T. Thirumoorthy and John Chiam

The approach to a patient with a skin disorder follows the standard medical procedures of history taking, clinical examination and use of laboratory investigations.

The traditional full medical history is not a necessary requirement before clinical examination. A pertinent screening history is followed by a full dermatological examination. A directed history after the examination for more detailed information then follows.

A patient with a skin disorder may present in any one of the following scenarios:

1) A single lesion the patient had noted and would want simple answers as what it is, is it malignant; what needs to be done with it. Here, a good clinical examination would provide more information than a detailed history.
2) Many patients with dermatologic problems like acne or localized psoriasis are generally healthy; so a directed history and local examination would alone suffice.

3) However, there are some patients who present with a more complicated problem like a drug eruption or vasculitis and there is need for a detailed medical history and a full medical examination.

A screening medical history for a patient with a skin disorder would include:

1) Duration of skin lesions, i.e. when it started.
2) Site of onset of lesion, i.e. where it first started.
3) Pattern of spread, i.e. where did the lesion spread to?
4) Course of dermatoses, i.e. was it progressive, persistent; transient or fluctuating?
5) The character of lesions — this in our environment gets the poorest yield as most patients may not be equipped with sufficient language skills to give a description.
6) Associated symptoms of itch, pain and burning. The character of the itch, severity, effect on sleep and periodicity is useful information.
7) Aggravating and relieving factors, i.e. what made it worse and what had made it better?
8) What treatment was given — self-treatment in skin disorders is common and this can significantly modify the primary problem.

The screening medical history is followed by the examination of the patient. After the examination a directed history on the following is indicated.

1) Past history of skin disorders — allergies atopy.
2) Past medical history — admissions; surgeries.
3) Present medical history and medication consumed.
4) Family history of skin disorders and other medical illnesses.
5) Drug history — prescribed medication, OTC products, herbs and health supplements, history of drug allergies.

PHYSICAL EXAMINATION OF THE SKIN

Two basic requirements are essential for a good dermatologic examination, namely — good lighting and adequate exposure. A hand-held magnifying light with a patient undressed and covered by a cloth sheet is the ideal way of examination. The patient's privacy and modesty must be ensured at all times. All female patients should have an appropriate chaperone.

The examination of the skin involves inspection, palpation and where indicated, measurement. A systematic approach would start with the examination of the hands and nails, then the upper limbs, face, scalp, ears, oral mucosa, the front of the trunk, lower limbs, feet and toe nails, and the external genitalia. The patient is made to turn prone and the examination starts from the scalp to neck, trunk, natal cleft and perianal area, legs, feet and soles. This should cover all skin, nails, hair and mucosa.

MORPHOLOGY, CONFIGURATION AND DISTRIBUTION

While examining the skin one has to observe the morphology of the lesions, their configuration and distribution.

Morphology

In the same manner as one would describe an object as being a square or circle, morphology in dermatology helps us define accurately the skin lesions.

The morphology of the lesions are divided into Primary lesions and Secondary lesions.

Primary lesions are skin lesions that occur as a result of the skin disorder without modification by environmental factors, e.g. macule, papule, etc.

Secondary lesions are a result of modification or alteration from external factors such as excoriation, lichenification, etc. The history is often useful to differentiate the two.

BASIC DERMATOLOGIC LESIONS

Macule: A circumscribed area of skin different in color or texture from its surrounding tissue. A maculopapular rash consists of macules and papules.

Patch: A large macule, more than 2 cm in diameter.

Papule: A small solid elevation of the skin, less than 0.5 cm diameter.

Epidermal papule: A papule composed of localized thickening of the epidermis or of the stratum corneum.

Dermal papule: A papule composed of a localized, solid thickening of the dermis produced by augmentation of normal structures, deposition of metabolic products, concentrations of cells, or other pathologic changes.

Inflammatory papule: A papule produced by inflammation of the dermis, epidermis, or both.

Papulosquanmous lesion: A papule that develops a reactive or degenerative epithelial component resulting in desquamation, which is the loss of epithelial cells or of stratum corneum (scaling).

Nodule: A solid mass of the skin which can be observed as an elevation or can be palpated. Usually 0.5 cm or more in diameter.

Plaque: An elevated area of skin 2 cm or more in diameter. Plaques may have the same subdivisions as mentioned above under papules.

Tumor: An enlargement of the tissues by normal or pathologic material or cells that form a mass. It may be inflammatory or benign or a malignant new growth of cells or tissue.

Papilloma: A nipple-like mass projecting from the surface of the skin.

Wheal: An elevated compressible, evanescent area produced by dermal edema. It is often surrounded by a red, axon-mediated flare.

Vesicle: A circumscribed elevation of the skin less than 0.5 cm in diameter and containing liquid.

Bulla: A circumscribed elevation of the skin over 0.5 cm in diameter containing a liquid. The distinction between vesicle and bulla is arbitrary and depends only on size.

Pustule: A visible accumulation of pus in the skin.

Petechia: A punctate hemorrhagic spot 1–2 mm in diameter.

Purpura: Discoloration of the skin or mucosa due to extravasation of blood.

Ecchymosis: A macule red or purple colored hemorrhage in skin or mucous membrane more than 2 mm in diameter.

Erythema: The redness of the skin produced by vascular congestion or perfusion.

Telangiectasia: A visible vascular lesion formed by dilation of small cutaneous blood vessels.

Scale: A flat plate or flake of stratum corneum. Squama (Latin — a scale or plate-like structure).

Crust: An outer layer from the drying of exudates, secretion, or hemorrhage. Scab.

Excoriation: Any loss of substance of skin produced by scratching.

Erosion: A loss of the epidermis which heals without scarring — partial loss of epidermis.

Ulcers: A skin ulcer is a defect or loss of dermis and epidermis produced by sloughing of necrotic tissue.

Fissures: Any linear gap or slit in the skin surface.

Scar: The fibrous tissue replacing normal tissues destroyed by injury or disease.

Atrophic scar: Papyraceous, or cigarette paper scar — describes thin wrinkled scars.

Hypertrophic scar: An elevated scar with excessive growth of fibrous tissue.

Sclerosis: An induration or hardening of the skin. It is often due to fibrosis.

Stria: A streak or band of linear, atrophic, pink, purple or white lesions of the skin due to changes in connective tissue.

Lichenification: A chronic thickening of the epidermis with exaggeration of its normal markings, often as a result of scratching or rubbing.

Keratoderma: Hyperplasia of the stratum corneum.

Alopecia: Absence of hair from normally hairy areas of skin.

CONFIGURATION (OR ARRANGEMENT)

The configuration or arrangement of lesions is the shape of each lesion and the pattern in which neighboring lesions are arranged in relation to each other.

Four distinctive patterns of arrangement of lesions are: linear, dermatomal, annular (or arcuate) and grouped.

Linear Lesions

1) linear contact dermatitis from plants
2) linear lichen planus as determined by koebnerization
3) linear of developmental origin as in linear epidermal nevus
4) linear as determined by blood vessels — thrombophlebilis

Dermatomal Lesions

Dermatomal or segmental lesions are best exemplified by herpes zoster and segmental vitiligo.

Annular Lesions

Annular describes the lesion appearing in a ringed form. Serpiginous is used when the borders of a ringed lesion is wavy or snake-like. The classical annular lesion is the "ring worm" of dermatophyte infection.

Many other lesions like erythema multiforme, leprosy and cutaneous T cell lymphoma all appear in an annular fashion.

Grouped Lesions

Lesions with greater density or clustered in a localized area. Herpes simplex virus infections are described as grouped vesicles or erosions on an erythematous, edematous base. Insect bites are typically grouped.

DISTRIBUTION OF LESIONS

The distribution of the lesions in a skin disorder gives significant information for the diagnosis of the skin disorder. They can be described as:

1) isolated vs. localized vs. regional vs. generalized;
2) flexural vs. extensor, e.g. atopic eczema vs. psoriasis;
3) bilateral, symmetrical vs. unilateral, asymmetrical;
4) truncal vs. peripheral (acral);
5) exposed areas — photosensitivity.

A young adult with erythematous scaly plaques (morphology) over the scalp and distributed bilaterally and symmetically over the elbows, knees and lower back (distribution) draws one to make the diagnosis of psoriasis.

An elderly person with painful grouped (configuration) vesicles and hemorrhagic bulla (morphology) along the right thoracic 4 dermatome (distribution) is atypical description of herpes zoster.

Investigations

Where the diagnosis is not clear or where confirmatory tests are essential, diagnostic tests are requested.

Common dermatologic investigations include:

1) microscopy — KOH for fungal elements
   — Gram stain for bacteria

— Tzanck's test for multinucleated giant cells of herpes infection
— for identification of scabies, mites and eggs
— for identification of head/pubic lice and nits.

2) cultures are taken from skin lesions, genital secretions, hair and nails for virus, bacteria, and fungus;
3) skin biopsy for histology has high diagnostic value;
4) skin biopsy and blood for immunoflorescence for autoimmune skin disorders like SLE, pemphigus vulgaris;
5) patch test for contact allergy. Prick test for contact urticaria; and
6) woods light examination helps in pigmentry disorders;

Other investigations include serological, hematologic and biochemical tests for diagnosis of specific skin conditions.

CONCLUSION

The approach to a patient with a skin disorder for diagnosis and management must take cognizance that the skin is not only influenced by external factors (microbes, allergens and medicament) but also internal systemic diseases (renal, hepatic, thyroid) and their treatment as well as the psyche of the patient. Mastering the recognition of the dermatologic lesions, their distribution and configuration is the basic requisite to an accurate diagnosis. When the diagnosis is not clear, revisiting the history and clinical examination to review the evaluation of the disease process is prudent.

Diagnosis in dermatology, like in all other specialities, is both an art and a science which needs discipline, perseverance and attention to details.

REFERENCE

Champion RH, Burton JL, in Champion RH *et al*. (eds.), *Diagnosis of Skin Disease in Textbook of Dermatology Vol. 1,* Blackwell Sciences Ltd., pp. 123–137, 1998.

This page intentionally left blank

9

Urticaria and Eczema

Pang Shiu Ming

URTICARIA

Urticaria, also referred to as hives, is a common skin disorder usually associated with exposure to allergenic or physical stimuli. Urticaria may affect both sexes and occur at any age. Up to 20% of the population will experience at least one episode. Urticaria is classified into 2 types: acute or chronic. Most of the cases are acute — lasting a few hours to a few weeks (maximum 6 weeks). Urticaria lasting longer than 6 weeks is defined as chronic urticaria. There is usually an underlying cause in acute urticaria. However, it is difficult to identify an underlying cause in chronic urticaria and no etiological agent can be identified in the majority of chronic urticaria. The course of chronic urticaria is often unpredictable and it may last for months or sometimes years.

PATHOGENESIS

Histamine is the most important mediator of urticaria. Degranulation of mast cells with release of histamine is central to the development of urticaria. When injected into the skin, histamine produces the triple

response — local erythema (vasodilatation), the flare beyond the border of local erythema, and a wheal (leakage of vascular fluid).

There are 2 main types of histamine receptors, H1 and H2 receptors. When H1 receptors are stimulated by histamine, the axon reflex, and vasodilatation and pruritis occur. Activation of H1 receptors causes smooth muscle contraction in the gastrointestinal and respiratory tracts, as well as pruritis and sneezing. These receptors are blocked by H1 antagonists (H1 anti-histamines such as chlorpheniramine). Activation of H2 receptors causes increase in gastric secretion. H2 antagonists such as ranitidine and famodtidine are used primarily for the treatment of gastric hyperacidity. H1 anti-histamines are the mainstay in the treatment of urticaria. The use of H2 anti-histamines in the treatment of urticaria is not well established. A combination of H1 and H2 anti-histamines is no more effective than H1 anti-histamine, except for the treatment of dermographism.

Clinical Aspects

Urticaria is characterized by the appearance of erythematous papules or wheals. Lesions vary in size from 2–4 mm in diameter and from round papules of cholinergic urticaria to a single lesion of many centimeters. Any part of the skin may be affected. The individual lesion may be round or irregular, often with central clearing; or in the form of a linear streak in dermographism. Individual lesions change in size and shape by peripheral extension or regression during the hours that they exist. The edge may become polycyclic if the individual lesions coalesce. They are often intensely pruritic. Lesions usually subside within 2–4 hours (maximum 24 hours), but new ones may appear elsewhere. The evolution of urticaria is a dynamic process. Individual lesions that last more than 24 hours are suggestive of urticarial vasculitis.

Acute Urticaria

The most common causes of acute urticaria are food, drugs and recent viral infection.

History

The history is directed to the exact time of onset of the urticaria and the relation to any food, recent viral infection or medication.

Physical examination

When the patient presents with acute urticaria, the urticaria is usually evident on examination of the skin. Stroke the patient with a wooden tongue depressor on the flexure of the forearm to rule out dermographism.

Investigation

The cause of the acute urticaria is usually evident. Laboratory tests are usually not indicated.

Management

Stop the suspected foods or medication. Since aspirin (salicylates) releases histamine, advise the patient to avoid aspirin. Prescribe non-sedating anti-histamine in the daytime and sedating anti-histamine in the evening. A short course of oral prednisolone is prescribed in difficult-to-control cases.

Chronic Urticaria

Chronic urticaria is a heterogeneous group of disorders. About 30% of this group is physical urticaria. 25–30% is autoimmune urticaria with histamine-releasing autoantibodies. 10–15% is due to urticarial vasculitis, aspirin induced and infection related. The remaining 30% is truly idiopathic. The etiology of chronic urticaria is often unclear even after extensive investigation. Unrecognized physical urticarias may be the underlying cause of the chronic urticaria or may coexist with chronic urticaria. In any patient who presents with chronic urticaria, the presence of physical urticaria should be ruled out by history, physical examination and appropriate tests.

History

History is important in making the diagnosis of urticaria because at the time that a patient presents there may not be evidence of urticaria on physical examination. The diagnosis is made by taking a careful history of the typical characteristics of wheals that come and go. The duration, frequency and the time of occurrence of the wheals are noted. Urticaria tends to appear or become more severe towards the end of the day.

A detailed drug history, including herbal tea or tonic, should be obtained. Any relation of the urticaria to food should be enquired.

Physical examination

On physical examination of the skin, there may or may not be any evidence of urticaria. When the patient is not having an attack of urticaria, examination of the skin may show that it is entirely normal. If urticaria is present, the morphology of the individual lesions and the distribution of the lesions are noted. Stroke the patient on the forearm to rule out dermographism.

A general physical examination is conducted to exclude any associated medical condition that may be the underlying cause of the urticaria.

Investigation

A list of the etiological causes of urticaria is summarized in Table 1. Investigations should be guided according to the findings of the history and physical examination.

A complete blood count and ESR can be used as a baseline screening investigation. Blood eosinophilia should prompt stool examination for parasite infestation. A raised ESR may be non-specific but may be raised in urticarial vasculitis. There is a significant association between autoimmune

Table 1 Etiological Causes of Urticaria

1) *Drugs*: Penicillin and other antibiotics, aspirin, analgesics, codeine (Note: aspirin may exacerbate existing urticaria as well as precipitate an attack).
2) *Foods*: Fish, shellfish, nuts, eggs, chocolate, tomatoes, strawberries, beer (Note: food is frequently a causative factor in acute urticaria but rarely in chronic urticaria).
3) *Food additives*: Salicylates, dyes, preservatives (benzoates).
4) *Infections*: Focal and systemic; bacterial, viral and fungal.
5) *Parasitic infestations*: Protozoa or helminthes.
6) *Inhalants*: Pollen, spices, animal danders and volatile chemicals.
7) *Internal diseases*: Serum sickness, autoimmune thyroid diseases, systemic lupus erythematosus, rhematoid arthritis, lymphomas.
8) *Physical urticarias*: Dermographism, cholinergic urticaria, pressure urticaria, cold urticaria, solar urticaria.
9) *Skin diseases*: Mastocytosis, pre-bullous stage of pemphigoid.
10) *Hormones*: Pregnancy, premenstraul flare-ups.

urticaria and autoimmune thyroid disease. Thyroid function test and thyroid autoantibodies are appropriate in chronic urticaria not responding to anti-histamines. In the occasional patient where food allergy is strongly suspected, dietary exclusion and double-blinded, placebo-controlled oral challenge may reveal certain food as a cause of chronic urticaria. Skin biopsy is essential to confirm the diagnosis of urticarial vasculitits.

Management

The patient should be informed that the course of the disease is unpredictable and should be reassured that oral anti-histamines will decrease discomfort and long-term administration does not decrease efficacy. The aim of treatment is to eliminate itch and not the rash. The medication should be taken on a regular basis and not when the rash is severe. Non-sedating anti-histamine is given in the daytime and sedating anti-histamine in the evening. Doxepin is a tricyclic anti-depressant with very potent H1 anti-histamine action and can be added to the regimen for patients for whom the disease is not adequately controlled with the above regimen.

Physical Urticarias

Physical urticarias are caused by physical stimuli. They account for 30% of chronic urticaria. Care must be taken during physical urticaria challenge testing as the severely affected may develop systemic symptoms. Resuscitative measures should be available during testing.

Dermographism (factitious urticaria)

Dermographism is the most common physical urticaria. It is produced by rubbing or stroking the skin. The patient complains of linear itchy wheals from scratching or friction from clothing. The onset is usually sudden without any precipitating cause. The condition can last from a few weeks to years. The diagnosis is made by drawing a tongue depressor firmly across the flexure of the forearm. A red line occurs followed by the appearance of a wheal. If the response is negative, the patient is asked to check the site a few hours later for delayed dermographism.

Cholinergic urticaria

Cholinergic urticaria typically occurs during or shortly after exercise and consists of round papular wheals 2–4 mm in diameter. The individual lesions may coalesce to become typical hives. The diagnosis is established by running up a flight of stairs several times. Apart from the skin rash, there may be systemic symptoms which include angioedema, hypotension and wheezing. However, the incidence of systemic symptoms is very low.

Cold urticaria

Cold urticaria occurs when there is a drop in the air temperature or exposure to cold water. The hands, ears and skin over bony prominence such as the skin over the wrist and knees are typically affected because they are more vulnerable to exposure to cold. The diagnosis is made by holding an ice cube against the skin for 5 minutes. It is then removed and the site is examined 5 minutes later to allow time for the chilled skin to warm. Development of a wheal at the site of contact confirms the diagnosis.

Solar urticaria

Solar urticaria is induced by exposure to the sun and disappears within 1 hour. Solar urticaria is distinguished from polymorphic light eruption in that the latter appears hours after exposure and persists for several days.

Pressure urticaria

Pressure urticaria is characterized by the appearance of deep, itchy or painful swelling of the skin occurring several hours after a pressure stimulus. The palms, soles, buttocks and waist are commonly affected from holding objects, walking or prolonged standing, sitting on hard surfaces and wearing tight garments. The diagnosis is made by suspending a weight of 4 kg over the shoulder or thigh for 10 minutes. A visible and palpable swelling appearing several hours later is indicative of positive reaction. The urticaria lasts 8–24 hours.

Angioedema

Angioedema is a hive-like swelling caused by transitory subcutaneous edema affecting the skin and mucous membranes. Urticaria and

angioedema commonly occur together. These subcutaneous swelling may cause burning or pain rather than pruritis. The eyelids, lips, palms, soles and genitalia are commonly affected. Involvement of the respiratory tract and gastrointestinal tract causes dyspnea, dysphagia, abdominal pain, vomiting and diarrhea. Attacks of angioedema, frequently affecting the throat as well as the skin usually follow trauma, especially dental extraction.

Drugs Used in the Treatment of Urticaria

Oral anti-histamines

Oral H1 anti-histamines are the mainstay of treatment for urticaria. A combination of H1 and H2 anti-histamines is no more effective than H1 anti-histamine alone, except for the treatment of dermographism. There are 2 main groups of H1 anti-histamines: the sedating (e.g. hydroxyzine) and the non-sedating (e.g. desloratidine). Usually, a non-sedating anti-histamine is prescribed in the daytime with a sedating one in the evening.

Doxepin

Doxepin is a tricyclic anti-depressant with potent H1 anti-histamine action and is used in urticaria resistant to treatment with H1 anti-histamines.

Adrenaline

Severe urticaria or angioedema requires adrenaline which is life-saving in the life-threatening condition of anaphylaxis associated with severe urticaria and angioedema.

Oral corticosteroids

Oral corticosteroids are reserved for severe acute urticaria and short-term use in refractory chronic urticaria.

Immunosuppressive drugs

Immunosuppressive therapy with cyclosporine is becoming an option in some patients with chronic autoimmune urticaria when the disease is disabling, severe and recalcitrant.

ECZEMA

Eczema is the most common inflammatory disease of the skin. A wide range of internal and external factors acting singly or in combination may induce the condition. However, the 2 terms, eczema and dermatitis are used interchangeably.

Stages of Eczema

There are 3 stages of eczema: acute, subacute and chronic. They represent the different stages in the evolution of an inflammatory process of the skin. The eczema in each stage may subside spontaneously or with treatment and may not progress to another stage. The patient may present in any stage or in an acute on chronic stage.

Acute eczema

Acute eczema is caused by contact with specific allergens such as poison ivy or it can occur in other inflammatory processes such as atopic dermatitis. The characteristic features are erythema and edema of the skin associated with tiny, clear, serum-filled vesicles. It is an extremely itchy condition. The vesicles may become confluent to form blisters. As the pressure of the serum-like fluid increases, it leaks to the surface and exudes like serum. The roofs of the small vesicles are often rubbed and scratched because of intense itch. Small crusts of dried serous fluid then cover the lesional skin.

Subacute eczema

In subacute eczema, the erythema and edema are less severe than in acute eczema and the vesicles disappear. This is accompanied by the appearance of scales and crusts and the surface is dry. The borders of the lesions are indistinct. The itch is usually less intense or absent. The symptom is more of a drying and burning sensation.

Chronic eczema

When scratching is not controlled, subacute eczema can be modified and converted to chronic eczema. As a result of long-continued scratching, the

skin becomes thick and leathery. The normal markings of the skin become exaggerated to produce a mosaic composed of flat-topped, shiny quadrilateral facets. This change is called lichenification. The sites most commonly involved are the antecubital fossa, popliteal fossa, ankles, anogenital region and the nape of the neck. Once this itch-scratch-lichenification cycle is established, the eczema often persists.

Classification of Eczema

Eczema is divided into 2 main groups — exogenous eczema and endogenous eczema. Exogenous eczema is caused predominantly by external factors such as chemical or other injurious agent that come into contact with the skin surface. All the other cases are called endogenous eczema. Although some of the terms used for endogenous eczema appear to relate to etiology, cases of endogenous eczema are usually named according to their pattern and distribution over the body surface. Sometimes features of 2 or more types of eczema may appear in the same patient.

Atopic dermatitis

Atopic dermatitis is the most common type of eczema and is an intensely itchy skin disease. The term atopy denotes an inherited tendency to develop bronchial asthma, allergic rhinitis and atopic dermatitis. Atopic dermatitis is a pattern of skin disease associated with atopy. Family and personal history are often positive for 1 or more of these features. It runs a chronic relapsing course which usually begins in early infancy or childhood, although it may occur for the first time at any age. Atopic dermatitis is a clinical diagnosis. Tables 2 and 3 show the guidelines in the diagnosis of atopic dermatitis in infants[1] and the UK Working Party's diagnostic criteria for atopic dermatitis. The distribution of the lesions in atopic dermatitis varies with different age groups. Three phases can be recognized — infantile, childhood and adult phase.

Infantile phase

Atopic dermatitis usually begins between 2–6 months. This is the time the baby develops the coordinated scratching response. The face and the extensor surface of the 4 limbs are affected. The diaper area is often

Table 2 Guidelines for the Diagnosis of Atopic Dermatitis in Infants

Major Features (must have at least 3)

- Pruritus — a required factor
- Typical morphology with facial and extensor involvement
- Chronic, or chronically relapsing dermatitis
- Family history of atopy (asthma, allergic rhinoconjunctivitis, or atopic dermatitis)

Table 3 The UK Working Party's Diagnostic Criteria for Atopic Dermatitis

Major Features (must have at least 3)

- History of involvement of the skin creases such as folds on elbows, behind the knees, fronts of ankles or around the neck (including cheeks in children under 10)
- A personal history of asthma or hay fever (or history of atopic disease in a first-degree relative in children under 4)
- A history of general dry skin in the last year
- Visible flexural eczema (or eczema involving the cheeks/forehead and outer limbs in children under 4)
- Onset under the age of 2 (not used if child is under 4)

spared. The disease is punctuated by remission and exacerbation. Although the disease clears up in about 50% of infants by the age of 18 months, the disease can progress to the childhood phase.

Childhood phase

From 18 months onwards the disease is usually located on the flexure of the elbows and knees, sides of the neck, wrists and ankles. Scratching causes lichenification of the lesional skin and extension of the rash beyond the classical sites of the childhood phase. Although continuous scratching of the itching rash causes distress in children, there is a general tendency towards spontaneous improvement throughout childhood, with most patients going into remission by the age of 30.

Adult phase

Most patients in the adult phase have experienced the disease since infancy or childhood, although there may be an occasional patient without a past history of dermatitis. The characteristics of the disease are essentially similar to the childhood phase with or without lichenification.

However, localization in the nipples in young women and in the anogenital area are more common in the adult atopic dermatitis. Hand dermatitis in adults may be an expression of the atopic diathesis.

Seborrheic (Seborrhoeic) dermatitis

Seborrheic dermatitis is a common inflammatory disease with a characteristic pattern for different age groups. The disease is located over the sebum-rich area of the scalp, face and trunk. In addition to sebum, it is linked to the yeast-like organism *Pityrosporum ovale*, immunological abnormalities and activation of the complement. Changes in humidity, seasonal variation and emotional distress precipitate or aggravate the condition. Seborrheic dermatitis tends to be more severe in bed-ridden and neurological patients such as Parkinson disease and AIDS patients. There are 3 patterns in the different age groups.

Infants (cradle cap)

Infants commonly develop greasy adherent scales on the vertex of the scalp. Minor degree of this cradle cap clears up with frequent shampooing with products that contain salicylic acid. Scales may accumulate and become thick and adherent and may be accompanied by inflammation. Sometimes in the infant the seborrheic dermatitis may be more extensive, involving the folds of the body. In contrast to atopic dermatitis, the infant with seborrheic eczema seems not to be distressed by itch.

Children (tinea amiantacea and blepharitis)

One or more patches of thick, adherent scales appear on the scalp. These whitish thick plates of scales adhere firmly to the scalp and hair and may be associated with inflammation and temporary hair loss. The condition may persist for months and be resistant to treatment. Occasionally, similar white scales may adhere to the eyelashes and lid margins with variable amounts of erythema characteristic of seborrheic blepharitis.

Adults (classic seborrheic dermatitis)

The spectrum of adult seborrheic dermatitis varies from dandruff (pityriasis capitits), which is considered a mild form of seborrheic dermatitis, to

the classic seborrheic dermatitis. In the dandruff form, there is fine, dry white scalp scaling associated with an itchy scalp. In the classic adult seborrheic dermatitis, the scales appear on an inflamed base. The distribution of scaling and inflammation become more diffused and appear on the seborrheic areas: Scalp and scalp margins, eyebrows, nasolabial folds, external ear canals, posterior auricular folds and pre-sternal area on the chest. The axillae, inframammary folds, groins and umbilicus are affected less frequently. The itch is either absent or relatively mild. The disease runs a chronic course with remission and exacerbation.

Asteatotic (xerotic) eczema

Asteatotic eczema occurs after excessive drying (often caused or aggravated by excessive cleansing) and low environmental humidity. It is more common but not limited to the elderly. Patients with atopic diathesis are more likely to develop this type of eczema. The lower legs are commonly affected. The asteatotic skin is dry and scaly. With further drying, red plaques appear with thin and long horizontal fissures. A cracked porcelain pattern of fissuring appears when short vertical fissures connect with the horizontal fissures. The distribution of the rash can be more widespread, involving the arms and trunk especially in the elderly.

Nummular (discoid) eczema

The cause of nummular eczema is unknown but commonly seen in persons with an atopic diathesis. This type of eczema occurs primarily in the middle-aged and elderly. The typical lesion is a well-circumscribed, coin-shaped red plaque of closely set, thin-walled vesicles on an erythematous base. The individual lesions are 1–5 cm in diameter. The lesions progress towards a less vesicular and more scaly stage. This corresponds to the stage of subacute eczema. There may be central clearing with peripheral extension giving rise to a ring-shaped or annular lesion. The back of the hand, extensor surface of the limbs, flanks and hips are the sites commonly affected. The lesions itch and scratching becomes habitual.

Lichen simplex chronicus

Lichen simplex chronicus is also known as neurodermatitis circumscripta. The lesion is a circumscribed, lichenified pruritic patch produced

by habitual scratching. This corresponds to the stage of chronic eczema. The initial stimulus of pruritus is rarely identified. Sometimes the lesion is initiated by a bite or injury. Psychological factors such as stress and anxiety seem to play a part also. Some patients have atopic diathesis. It is more common in adults but may be seen in children. The sites commonly affected are the lower leg and ankle, nape of the neck and the scalp, and the anogenital region.

Prurigo nodularis

Prurigo nodularis is a disease with multiple itchy nodules of unknown cause that may be regarded as an atypical form of lichen simplex chronicus. These nodules measuring 1–2 cm in diameter are usually found on the arms and legs. The number of lesions ranges from a few to many lesions. The nodules are brown or red, hard and dome-shaped. The surface is smooth at first. Itching, which is often intense, causes the surface to become excoriated, fissured or verrucous. The itching is characteristically paroxysmal and unbearably severe and is only relieved by scratching to the point of bleeding and scarring.

Stasis (varicose) eczema

Stasis eczema occurs in the lower legs in some patients with venous insufficiency. The reason why eczema occurs in these patients is unknown. There may be genetic or environmental factors. The affected areas are the lower legs, especially around the medial side, with varying degree of spread to the upper legs and feet. Hyperpigmentation from hemosiderin is often prominent. There are 2 notable features of stasis eczema: the tendency to spread and susceptibility to contact dermatitis.

Hand eczema

Inflammation of the skin on the hands is a common problem that causes discomfort and embarrassment. Occupational hand eczema is a major problem in occupational diseases. The diagnosis and management of hand eczema is a challenge. There is interplay of irritants, allergens and atopic diathesis in the cause of hand eczema. The patient may present in the different stages of the disease: acute, subacute, chronic and acute on chronic, which may affect the management.

Irritant contact dermatitis (housewives' hand dermatitis)

Dryness and chapping are early features. Painful fissures and cracks occur over the joints and the fingertips. The dorsum of the hands is red and swollen while the palmar surface is dry and cracked. Itch occurs when acute eczema superimposes with further irritation. Scratching and excoriation leads to secondary infection. Individuals at risk include young mothers (washing milk bottles and changing diapers); occupations that require repeated washing and drying of the hands (surgeons, dentists, bartenders, hairdressers and dishwashers); occupations that require coming into contact with chemicals; and individuals with atopic diathesis.

Atopic hand dermatitis

Hand dermatitis may be the manifestation of atopic dermatitis in the adulthood. Hand eczema is more common in people with atopic dermatitis. They develop hand eczema independent of contact with irritants. The features of the hand eczema are similar to irritant hand eczema.

Allergic contact dermatitis

Allergic contact dermatitis of the hands is not as common as irritant contact dermatitis. The appearance is not much different from irritant contact dermatitis. Therefore allergic contact dermatitis should always be considered in the differential diagnosis and management of hand eczema. Patch test is used in the diagnosis of allergic contact dermatitis. Nickel (door knobs, handle of kitchen utensils, scissors, industrial equipment); potassium dichromate (leather articles, industrial machines, industrial oils); and rubber (gloves, hoses, cables) are some of the common allergens identified from patch tests in patients with hand eczema.

Pompholyx

Pompholyx is a type of eczema on the palms and soles which is modified by its special site (thick stratum corneum) and in which edema fluid accumulates to form vesicles or bullae. The etiology is unknown. The itch is usually intense. The vesicles dry up in 3–4 weeks' time and are replaced by scales. Chronic eczema may follow and new pompholyx may appear

as acute exacerbation of the chronic hand eczema. Secondary infection is a common complication especially on the feet.

Complications of Eczema

Spread of eczema

Eczema can spread from a single initial site to the other parts of the body by direct extension or indirect extension. Sometimes the spread is symmetrical, which is common in stasis eczema. The spread may be generalized and the whole integument becomes eczematous. Erythroderma or exfoliative dermatitis refers to a universal and confluent eczema.

Secondary infection

Secondary bacterial infection may complicate eczema. This is especially common in atopic eczema. The common bacterium is *Staphylococcus aureus* causing golden yellow crusting. Herpes simplex infection usually occurs on the face and presents as grouped vesicles and pustules that later erode and become crusted Kaposi's varicelliform eruption.

Pigmentation change

There may be hypo- and hyper-pigmentation because scratching damages the melanocytes. Hypopigmentation is usually permanent while hyperpigmentation may improve with time.

Treatment of Eczema

General measures

Identification and removal from contact with a primary or secondary allergen or irritant is the most important measure. When eczema is acute, severe or extensive it is necessary for the patient to rest at home or be admitted to hospital for treatment. The normal barrier function of the skin in eczema is impaired and excessive washing is harmful to the skin. Soap should be avoided. Soap substitutes such as soap-free cleanser or bath oil should be prescribed and emollient/moisturizer should be applied liberally. Temperature change and sweating aggravate eczema and should be avoided.

Topical treatment

In acute eczema, wet compress/soaked with astringent and lotion or cream form of topical medicament are used. In chronic eczema, cream and ointment form of topical medicament are used.

Wet compress/soaking

The common agents used for compress, soak and wet dressing are potassium permanganate, normal saline and Burrow's solution. These measures help to dry up the exudating lesions in acute eczema.

Topical corticosteroid

Topical corticosteroids are the mainstay in the treatment of eczema. They are available in lotion, cream and ointment form and in 3 categories of low, moderate and strong potency. Strong potency steroid such as clobestol proprionate is used for lichenified eczema. Low potency steroid such as hydrocortisone can be used on the face. Moderate potent topical corticosteroid of varying strength may be used on the body. Topical corticosteroids are safe to use under medical supervision. Patients should be educated on the proper use of them and dissuaded from "steroid phobia."

Intradermal corticosteroid

Prurigo nodularis and occasionally lichen simplex chronicus require intralesional injections with triamcinolone acetonide.

Oral antibiotics/antiviral agents

Oral antibiotics such as cloxacillin may be prescribed to clear *Staphylococcus aureus* infection. Oral acyclovir may be required for herpes simplex and Kaposi's varicelliform eruption.

Oral anti-histamines

Sedating anti-histamines such as hydroxyzine control itch and induce sleep. Doxepin is a tricyclic anti-depressant with strong anti-histamine action. Non-sedating anti-histamine such as desloratidine may be used in the daytime.

Systemic corticosteroid

Short-term systemic corticosteroid is used in acute exacerbation and severe cases of eczema. These include acute vesicular eczema with swelling, acute contact dermatitis, rapidly spreading discoid eczema and generalized exfoliative dermatitis.

Phototherapy

Phototherapy with narrowband UVB is effective for mild to moderate atopic dermatitis but the drawback is the need to commute to the clinic or hospital for treatment. Phototherapy usually works better in the chronic dry stage rather than the red inflamed stage.

Oral immune suppressant

Cyclosporine is an effective short-term treatment for refractory atopic dermatitis. However, the disease may relapse despite continuous treatment or may recur after treatment is ceased. Cyclosporine is also useful in erythrodermic eczema and steroid-dependent eczema patients.

Other immune suppresants useful in severe atopic dermatitis include azathioprine and mycophenlate mofetil.

Topical immune suppressant

Topical tacrolimus and pimecrolimus have been shown to have similar anti-inflammatory potency to betamethasone valerate and clobestol propionate, respectively. They are superior to topical steroids because they do not cause skin atrophy. They are promising in the treatment of atopic eczema.

REFERENCE

1. Hanifin JM, Rajka G, Diagnostic features of atopic dermatitis, *Acta Derm Venereol (Stockh)* **92**(suppl):44–47, 1980.

FURTHER READING

Eichenfield LF, Lucky AW, Boguniewicz M, *et al*., Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents, *J Am Acad Dermatol* **46**:495–504, 2002.

Ellis C, Luger T, International consensus conference on atopic dermatitis II (ICCAD II): clinical update and current treatment strategies. *Br J Dermatol* **148** (Supp: 63):3–10, 2003.

Elston DM, Ahmed DD, Watsky KL, *et al*., Hand dermatitis, *J Am Acad Dermatol* **47**: 291–299, 2002.

Feingold S, Huang C, Kristal L, *et al*., Eczemas, *Curr Probl Dermatol* **2**:41–90, 1998.

Grattan CEH, Sabroe RA, Greaves MW, Chronic urticaria, *J Am Acad Dermatol* **46**: 645–657, 2002.

Habif TP, *Clinical Dermatology*, 4th ed. 2004.

Kobza Black A, Lawlor F, Greaves MW, Consensus meeting on the definition of physical urticarias and urticarial vasculitis, *Clin Exp Dermatol* **21**:424–426, 1996.

Sabroe RA, Seed PT, Francis DM, *et al*., Chronic idiopathic urticaria: comparison of the clinical features of patients with and without anti-Fc RI or anti-IgE autoantibodies, *J AM Acad Dermatol* **40**:443–450, 1999.

10

Cutaneous Drug Eruptions

T. Thirumoorthy

INTRODUCTION

An adverse drug reaction (ADR) has been defined as any response to a drug which is noxious, unintended and undesired and which occurs at doses used in humans for prophylaxis, diagnosis or therapy.

Cutaneous drug eruption (CDE) is a noxious, unintended and undesired dermatological condition that results from a systemically administered (oral, intramuscular, intravenous) drug. Topically administered medicaments on the skin resulting in skin inflammation is classified as contact dermatitis and is often excluded in discussions of CDE.

A drug allergy is an immunologically mediated reaction, characterized by specificity, transferability (by antibodies or lymphocytes) and recurs on re-exposure. CDEs are common. Of the 561 ADR reports submitted to the Health Services Authority Singapore in the year 2001, 184 (32%) were macular papular/exanthema eruption; 142 (25%) angioedema; 116 (21%) were urticaria; and 12 (0.02%) were classified as Steven–Johnson syndrome. Of the serious CDEs were Steven–Johnson (12), toxic epidermal necrolysis (6) and allopurinal hypersensitivity (5).

In the Boston Collaborative Study, 2% of hospitalized patients developed CDE.

Often it is difficult for the attending physician to not only determine if the skin eruption is drug-induced but also to identify the offending drug. This is because of the following situations:

1) Any drug may cause CDE. However, some drugs like amoxycillin, ampicillin, cotrimoxazole and cephalosporins have a higher overall cutaneous reaction rate (the number of reactions per exposed patient).
2) The cutaneous eruption (morphological pattern) of CDE is not specific to a single drug. Many drugs have been reported to produce a similar CDE morphologically.
3) The same drug may elicit different cutaneous reactions in different patients. Penicillin can cause the maculopapular-exanthem reactions in addition to urticaria angioedema and Steven–Johnson syndrome.
4) Previous uneventful exposure is not a guarantee against CDE.
5) No single or reliable diagnostic test is available to confirm CDE.
6) CDE may abate and clear while the patient is on the drug and may not be reproducible on challenge.

In this chapter, an attempt is made to provide a framework for a practical approach to diagnosis and management of CDE. Only the most common cutaneous reaction patterns are described. More than 30 common reaction patterns have been described.

CLINICAL REACTION PATTERNS IN CUTANEOUS DRUG ERUPTION

Maculopapular–Exanthem (Morbilliform) Eruption

Maculopapular–exanthems represent the most common type of CDE. The dermatoses are often generalized, symmetrical, erythematous confluent macules and papules, often sparing the face, palms and soles. On the lower limbs purpuric lesions may be noted. The CDE may be accompanied by itch, fever, facial edema, malaise and joint aches.

The eruption may progress to generalized erythrodema or exfoliative dermatitis with continued drug administration. The eruptions usually appear within 1 to 2 weeks after initiation of drug therapy.

The drugs most commonly associated with exanthems are amoxycillin, ampicillin, carbarmazepine, cotrimoxazole, gold, nalidixic acid, NSAIDs, anti-convulsants and allopurinol.

A viral exanthem appears identical to a drug-induced exanthem. Pruritus and eosinophilia may be used as differentiating factors. A skin biopsy will often show a non-specific histology and may be of value for ruling out other dermatoses.

Urticaria Angioedema

Urticaria angioedema is another common and easily recognizable CDE. Urticarial reactions appear as circumscribed (well-defined) raised, edematous, erythematous pruritic wheals. Urticaria is usually generalized and symmetrical. The wheals disappear in a few hours leaving no marks or scaling.

In the case of angioedema, the edema and inflammation is in the deep dermis, subcutaneous or submucosal areas. Angioedema usually affects easily distensible tissues like the eyelids, lips, genitalia, ears and mucus membranes of the tongue, mouth and larynx. Angioedema can persist for 2–5 days.

Urticaria angioedema may appear within minutes of drug administration as in anaphylaxis or several days of drug ingestion and even as long as 10–20 days as in serum sickness syndrome.

Most common causes of drug-induced urticaria include penicillin, NSAIDs, aspirin, radio contrast media, ACE inhibitors, and sulfonamides.

Drug-induced urticaria must be differentiated from food allergy urticaria, acute urticaria related to infections, chronic physical urticaria, hereditary angioedema and chronic idiopathic urticaria by careful history taking. In patients in whom the urticarial lesions last for more than 24 hours with staining of skin on fading, urticaria vasculitis should be considered. A skin biopsy in these instances is deemed useful.

Fixed Drug Eruptions

A fixed drug eruption (FDE) is a hypersensitive reaction characterized by one or more well-demarcated erythematous dusky plaques and sometimes bullous in nature. The drug eruption tends to recur in the same site on re-exposure to the offending drug.

The lesions may vary in size, rarely pruritic but may cause pain or a burning sensation. The lesions may crust, desquamate and on the genitals often erosive. The lesions heal leaving an annular hyperpigmented patch. The common sites are the extremities — hands, feet, face and mucocutaneous junctions like lips and genitals. The lesion develops within 30 minutes to 12 hours following the drug administration.

The drugs commonly implicated in FDE are tetracyclines, ampicillin, aspirin, NSAIDs, sulfonamides, phenolphthalein and chlormezanone.

FDE is believed to be the only CDE that is provoked solely by drugs or chemicals. The diagnosis is made by careful drug history.

Erythema-multiforme (EM) — Steven–Johnson Syndrome (SJS)

Erythema-multiforme (EM) is a rapidly onset dermatoses of well-demarcated bluish-red macular, papular lesions with central vesicles, bullae or purpura (target or iris lesion) predominantly on the limbs. The skin lesion may be accompanied by fever, malaise, and muscle and joint pains. Occasionally, the mucus membranes may be involved. The lesions dry and resolve in the course of the next 2 weeks.

The Steven–Johnson syndrome (EM major) is a severe variety of EM with sudden onset of fever, myalgia, malaise, headache, arthralgia, with ocular, oral and genital erosive lesions. Bullae and erosions are noted on the skin. Painful erosive stomatitis and erosive lesions on the genitals are seen.

Patients are often acutely ill with high fever and the condition may take 6 weeks to heal.

The drugs commonly associated with EM–SJS are allopurinol, anticonvulsant, oral contraceptives, NSAIDs and sulfonamides. Only 10% of EM, and up to 50% of SJS, are reported to be drug-related. Infections like mycoplasma, herpes simplex and streptococcal and chlamydia have been associated with EM–SJS.

Toxic Epidermal Necrolysis (TEN)

This is a form of serious, acute erosive bullous eruption of the skin and mucus membranes. The dermatoses may start as a tender morbilliform

eruption, rapidly progressing to blistering and widespread epidermal denudation. The clinical picture is of an extensive second-degree burn. High fever, fatigue, vomiting and diarrhea are prodromal symptoms. In SJS, skin detachment is usually below 10% of body surface area (bsa), whereas in TEN, it is greater than 10% bsa leading to many dermatologists to consider TEN as a severe form of SJS.

Drugs that are commonly implicated with TEN are allopurinol, ampicillin, amoxicillin, anti-convulsant, NSAIDs, and sulfonamides. The reaction rates of these drugs for TEN are low.

TEN is a medical emergency and carries a high mortality (up to 30%). The patients should be admitted to an intensive care burns unit for management.

Drug Hypersensitivity Syndrome

The dermatologic manifestation starts as pruritus with erythematous, edematous patches spreading to involve the entire skin. The skin may appear swollen and oozing and a few days later, scaly desquamation. Periorbital and facial edema, myalgia, and arthralgia are noted. Pruritus can be very severe. Fever and pharyngitis lymphadenopathy usually preceed the rash. Hepatitis nephritis, myositis pneumonitis and rarely encephalitis, meningitis, pancreatitis and epididymitis are the accompanying manifestations. Eosinophilia and atypical lymphocytosis are noted.

The drug-induced hypersensitivity syndrome usually appears several days to weeks (2–6 weeks) after initiation of the therapy. In patients already sensitized (previous reaction), it may occur within a few days of re-exposure. This condition is sometimes referred to by the acronym DRESS (drug rash with eosinophilia and systemic symptoms).

The drugs commonly implicated in the drug hypersensitivity syndrome are allopurinol, anti-convulsant, sulfonamides, aminopenicillins, chlormezanone, anti-TB drugs (isoniazid, streptomycin), gold, and captapril.

Other forms of generalized exfoliative dermatitis like psoriasis and eczema should be excluded. A skin biopsy would be useful. The lymphadenopathy and fever raises the issues of infectious monoculeosis and lymphomas.

Photosensitivity Dermatoses

Photosensitivity dermatoses appear as erythematous, edematous, eczematous reactions sometimes with vesicles and bullae over the sun-exposed areas like the face, dorsum of the hands and 'V' of the neck.

Although phototoxic drug reactions are more common, it is not possible on clinical grounds to differentiate between phototoxicity and photoallergy. Phototoxicity resembles an exaggerated sunburn, occurring within 5 to 12 hours of sun exposure, whereas a photoallergic reaction, representing an immunological basis, requires a latent period of 24–48 hours for sensitization to occur.

Drugs commonly associated with photosensitivity are sulfonylureas, thiazides, sulfonamides, NSAIDs, nalidixic acid, amiodarone and phenothiazines.

Differential diagnosis of photosensitive dermatoses include primary photodermatitis, polymorphic light eruption, chronic actinic dermatitis, autoimmune conditions like SLE and dermatomyositis and porphyria cutanea tarda.

MANAGEMENT OF CUTANEOUS DRUGS ERUPTIONS

Drug History

A detailed drug history, listing all the medications the patient has been taking at the time of onset of the eruption and 4 weeks before, is essential. A diary of dates, drugs taken and clinical symptoms should be established in a chronological order.

The patient should be questioned on all longstanding medical illnesses (e.g. diabetes, hypertension) and medications taken together with any change of treatment in the preceding 4 weeks.

A history of recent illness, visits to doctors and medications prescribed and taken should be recorded. Over the counter medications, previously prescribed medication for an earlier illness, medications in the medicine cupboard and medications given by friends and relatives are potential sources of CDE. Intermittent medication taken for common symptoms like cough mixtures, anti-constipation tablets, pain-killers, sedatives and muscle relaxants need to be looked into as well.

Finally, vitamins, herbs, tonics, health supplements and traditional medications can cause CDE.

Clinical Diagnostic Criteria for Cutaneous Drug Eruption

A diagnosis of CDE is made based on:

1) establishing a temporal relationship between the drug exposure and onset of eruption — e.g. urticaria can be immediate, exanthem in 2 weeks and drug hypersensitivity in 4 weeks;
2) whether there is an alternative explanation for the dermatoses — e.g. is this coincidental pityriasis rosea or xerotic discoid eczema or could this be just HSV-related EM or chronic physical urticaria;
3) reaction rates of the drugs exposed;
4) clinical improvement on withdrawal of the drug;
5) whether similar reactions have been recorded with previous exposure to the drug or one of its same class.

From the above, it is possible to narrow down the range of suspects but often not possible to identify the offending drug with certainty. One could conclude by linking the cutaneous eruption to the suspected offending drug as definite; probable; possible; and unlikely.

Definite: Where the morphology and temporal relationship is consistent with a CDE; there is no other likely cause for the dermatoses; there is improvement on withdrawal; there was a previous similar episode on previous exposure.

Probable: Where the morphology and temporal relationship is consistent; there is no other likely explanation for the dermatoses; there is improvement on withdrawal.

Possible: Where the morphology and temporal relationship is consistent; there are other possible explanations for the dermatoses; the improvement on withdrawal is not clear cut.

Unlikely: Where the morphology and temporal relationship is not clear cut to drug ingestion; there are other possible explanations.

Diagnostic Test

Skin Prick Test (SPT) may be helpful for diagnosis of IgE-dependent drug reaction (see Table 1). False-positive and false-negative reactions may occur with these skin tests.

Table 1 Skin Prick Test for Diagnosing IgE-dependent allergy

Antibiotics
Penicillin
Cephalosporins
General Anesthesia drugs
Thiopentone
Muscle relaxants
Others
Streptokinose
Insulin
Latex

Provocation or Challenge Test

Although considered as the "gold standard," challenge tests are seldom required. They must be performed only after informed consent, under strict medical supervision with resuscitative equipment available. Low-dose challenges for FDE in the presence of multiple commonly used drugs can be justified.

In Vitro Test

Radioallergosorbent (RAST) for drug specific immunoglobulin E class antibody are available for several drugs, including penicillin, insulin and ACTH.

Patch Test

Patch testing for CDE has not been standardized. Advocates have used it for systemic contact type dermatitis for medicaments, in photosensitivity and FDE.

Skin Biopsy

Skin histology, though not diagnostic, lends support to exclude other causes of the dermatoses.

Treatment of CDE

All but essential medications should be withdrawn. The suspected drug should be stopped and if essential, an alternative non-cross-reacting drug should be substituted.

Minor eruptions can be managed by withdrawal of suspected drug, topical steroids, emollients and anti-histamines.

Drug hypersensitivity syndromes, vasculitis and extensive morbilliform eruptions respond to systemic corticosteroids. Erythrodemic exfoliative dermatitis in the elderly may need measures to treat infection and cardiac failure and must be maintained in an optimal environmental temperature.

TEN is a potentially fatal condition needing intensive nursing care, appropriate dressing, opthalmic care, maintenance of fluids and nutrition.

Diagnostic Outcome and Patient Counselling

Inform the patient of the diagnosis-CDE. A medic alert form should be raised for the patient to acquire a medic alert card. Case notes and patient appointment cards should be labeled and registered both physically and electronically. Notification to the Health Sciences Authority is essential for public health reasons.

Once the diagnosis of CDE is made, the patient should be counselled on it significance. In this situation, a practical guideline on counselling patients is offered as follows:

Definite	Advise no further exposure to drug
	Avoid cross-reacting drugs of same family
	Avoid re-challenge
	Issue medic alert card
Probable	No further exposure. Use alternatives
	Avoid drugs of same family
	Issue medic alert card
	Re-challenge if absolutely necessary for non-serious reaction
	Consider desensitization if necessary to administer drug
Possible	No further exposure if avoidable
	Administer if necessary but stop medication at earliest sign of dermatoses

	Re-challenge only, in non-serious reactions
	Record in patients' notes and appointment card
Unlikely	Take drug as indicated
	Stop as soon as dermatoses appears

CONCLUSION

Cutaneous drug eruptions are common clinical problems. Full history and examination are essential to arrive at a diagnosis. Laboratory and skin tests are not able to identify with certainty the cause of the drug eruptions.

Clinical judgement must be exercised at all levels. Once a diagnosis of CDE is made, counsel the patient on its significance. Mark case notes clearly and offer medic alert cards to patients. Physicians should double-check with patients about drug allergies before prescribing any drugs.

REFERENCES

1. Leung, Greaves, *Allergic Skin Disease*, Marcel Dekker Inc., pp. 307–336, 2000.
2. Breathnach SM, Adverse cutaneous reactions to drugs, *Clin Med JR CPL* **2**:15–19, 2002.
3. Verloet D, Darham S, Adverse reaction to drugs, *BMJ* **316**:1511–1514, 1998.
4. Litt JZ, *Drug Eruption Reference Manual*, Parthenon Publishing, 2001.

Emergency Medicine

This page intentionally left blank

11

The Approach to the Collapsed Patient

Lim Swee Han and V. Anantharaman

INTRODUCTION

Cardiovascular collapse is a dreaded adverse event in a variety of medical emergencies. If not attended to immediately — and with the appropriate procedures — death is almost certain. However, the immediate recognition of cardiovascular collapse and the institution of immediate treatment measures are likely to give the patient a reasonable chance of survival. The array of interventions that are brought to bear immediately on the patient who collapses have, over the years, come to be referred to as "emergency life support". The success of such emergency life support is dependent on the ability of these interventions to be instituted promptly and appropriately. Success is linked to the efficiency with which the chain of survival is implemented (Fig. 1).

The above concept can be applied to a variety of common emergencies, such as:

i) cardiac arrest from an acute myocardial infarction (AMI);
ii) drowning;

Fig. 1 The chain of survival.

iii) electrocution; or
iv) any condition causing unconsciousness.

Amongst these examples, AMI provides the prime example for description of the various aspects of life support and the approach to the collapsed patient and will be so used in this chapter. Variations pertaining to other medical conditions will be highlighted as and when relevant.

EARLY ACCESS

This has a few components:

1) Recognizing that a person is suffering from symptoms of an acute heart attack, bronchial asthma, electrocution or, possibly, drowning. The earlier such a recognition is made, the more rapid other aspects of the chain would be initiated.
2) Calling for help, either from passers-by, the community's emergency ambulance services, or staff in the hospital. It is, therefore, important that there be correct and accurate public knowledge of public emergency numbers. This is 995 in Singapore, 118 in some of the countries in South-East Asia and 911 in the US. Again, the call to the emergency services must be initiated soon after onset of symptoms of the emergency, rather than waiting for collapse to occur.

3) Our emergency services must be allowed to respond as quickly as possible. Having a good emergency despatch service, giving way to ambulances on the roads and ensuring that ambulance crew and hospital staff reach the patient's side with the least amount of delay are all factors that will increase the chances of survival in the event of a collapse and, therefore, are all vital components of this very first link in the chain of survival.

EARLY CARDIO-PULMONARY RESUSCITATION (CPR)

It has been already well documented that in those communities with a high rate of bystander CPR for out-of-hospital cardiac arrest, the rate of survival to discharge from hospital is higher. Thus, King County, Washington, USA, with a bystander CPR rate of 50% has a survival rate of out-of-hospital cardiac arrest of more than 20%, while Los Angeles, with a bystander CPR rate of about 38% has a survival rate of about 5.0%.[1] In contrast, Singapore with a bystander CPR rate of 37% has a survivorship of only 4.1%.[2,3] The earlier CPR is initiated, the sooner mechanically aided blood circulation resumes in a cardiac arrest patient. Though the very best of CPR may only result in a production of 25–30% of normal cardiac output, such an output is usually just enough to keep vital organs, such as the brain, heart and kidneys alive until the arrival of defibrillation to restore the normal heart rhythm and rate. Brain injury begins if blood supply is denied to that organ for at least 4 minutes and is irreversible after a continued absence of flow for up to 10 minutes. Survivals from cardiac arrest decrease at the rate of about 10% for every minute that good CPR is withheld from a victim. A good outcome, therefore, demands that CPR be initiated within 4 minutes of cardiac arrest.

The technique of CPR that is widely practised today in this part of the world closely follows the guidelines promulgated by the American Heart Association,[4] and the Singapore National Resuscitation Council.[5] They are as follows:

1) *Assess responsiveness*. Check if the victim is unresponsive. Tap his shoulders and ask loudly, "Are you OK?"
2) *Call for the ambulance*. Dial 995 (if you are in Singapore) or other appropriate number. Inform the despatch officer that you have an unconscious victim. Give the location of the patient accurately.

Let the despatch officer ask all the necessary questions. Put down the phone only after the despatch officer has hung up.

3) *Position the victim*. If the victim is not lying supine, turn him onto his back. The victim should be placed face up on a firm, flat surface.
4) *Open the airway*. Tilt the victim's head backward and lift his chin, leaving his mouth slightly open. If there are secretions in his mouth, clear them with a finger wrapped around a piece of cloth, or with mechanical oropharyngeal suction.
5) *Check for breathing*. Look at the chest for its rise and fall. Listen for the sound of air escaping from the victim's mouth and nose. Feel for air from the victim's mouth and nose brushing past your cheeks. The presence of at least one of these three features indicates that the victim is breathing. If the victim is breathing, maintain the open airway using either the method described earlier or other methods, such as the jaw thrust, modified jaw thrust or even by placing the victim in the recovery (three-quarters prone) position.
6) *Artificial ventilation*. If the victim is not breathing, artificial ventilation must be provided, either by giving mouth-to-mouth ventilation, mouth-to-mask ventilation or bag mask ventilation. For these techniques to be effective, patency of the upper airway is crucial. When mouth-to-mouth ventilation is given, maintain the head tilt and pinch the nose closed with the fingers using the same hand. Maintain the chin lift with the other hand. Open your mouth wide, take a deep breath, place your mouth over the victim's mouth to make a tight seal and blow into the victim's mouth until you see his chest rise. Then release his nose to let him exhale and repeat the procedure for a second breath.
7) *Check for circulation*. Locate the victim's Adam's apple. Slide your fingers downward to the groove on the anterior border of the sternomastoid and feel gently for the carotid pulse for up to 10 seconds. At the same time, check for absence of other signs of circulation (normal breathing, coughing or movements). If one is not confident that there is circulation, start chest compression.
8) *Locate the hand position for chest compressions*. Run your middle finger along the lower rib margin to the xiphoid notch in the centre of the chest. Place your index finger next to it and the heel of your other hand along the sternum next to the index finger.
9) *Start chest compressions*. Place the heel of your first hand on the back of your second hand and interlace the fingers of both hands. Extend and

lock your elbow. Position your locked arm at 90º above the victim's chest. Use your entire body weight to compress the patient's chest by about 4–5 cm.

10) *Carry out 15 such compressions followed by 2 breaths*. When compressing the chest, count loudly as follows:

    One and two and three and four and five and,
    One and two and three and four and 10 and,
    One and two and three and four and 15.

    Make sure you release pressure on the chest at the end of each compression. Do 4 cycles of 15 compressions and 2 ventilations per cycle before checking the pulse again. If the pulse is again not present, continue the cycles of 15 compressions and 2 ventilations and check the pulse once every few minutes until the ambulance arrives, or at least till the heart's electrical rhythm is known.

For a two-man CPR, if the patient is not intubated, CPR should also be performed with a chest compression to ventilation ratio of 15:2, for both one or two rescuers. If the patient is intubated, chest compression should be done uninterrupted. Ventilations are asynchronous at a rate of 12 per minute.

EARLY DEFIBRILLATION

In pre-hospital cardiac arrest, the 4 most common recorded presenting cardiac rhythms have been ventricular fibrillation (VF), ventricular tachycardia (VT), asystole and pulseless electrical activity (PEA). The frequency of VF as the presenting cardiac rhythm has been reported to be 23.4% in Chicago[6] and 46.6% in Edinburgh.[7] However, a study of 157 cases of sudden cardiac death that occurred during ambulatory Holter monitoring recorded a combined incidence of 83.5% of VF and VT, with the rest having a bradyarrhythmia as the initial rhythm.[8] Time from collapsing to defibrillation is the single greatest determinent of survival for sudden cardiac death. Previously, defibrillators were operated manually. The decision to deliver electrical therapy required an in-depth understanding of ECG wave forms and the ability to read and understand the significance of a variety of rhythm disturbances. Such therapy was therefore delivered by medical practitioners trained in these skills.

Over the years, the development of automated (AED) and semi-automated electrical defibrillators (SAED) has resulted in the removal of the need to interpret ECG rhythms. These AEDs/SAEDs are able to identify rhythms that require defibrillation and can automatically charge the defibrillator to pre-programmed levels. Moreover, they only require the decision of the operator to deliver the current across the chest. AED defibrillation has been the greatest contributor to survivals from pre-hospital cardiac arrest in the last 10 years.

The onset of VF or pulseless VT during an AMI is usually characterized by a seizure lasting a few seconds followed by complete cessation of all vital body functions. If such an event is witnessed by a trained observer, a single precordial thump delivered almost immediately over the centre, or lower half, of the sternum may occasionally convert the patient to a sinus rhythm. The definitive treatment would still be maintenance of an open airway, calling for help, artificial ventilation, CPR and early defibrillation. The first direct current shock to be applied across the chest would be at 200 J, and if this did not convert the rhythm, the defibrillator should be immediately recharged at 200 J or increased to 300 J. A second shock is then delivered. Again, if this does not result in a change in the rhythm, the energy level is increased to the maximum of 360 J and the shock is applied again.

The energy levels described are those employed by monophasic truncated exponential wave electrical defibrillation which has been the conventional waveform used in defibrillation. It is gradually being appreciated that electrical defibrillation using biphasic truncated exponential waveforms hold the promise for use of significantly lower energy levels. Biphasic waveform defibrillation with energy levels at 150 J is now accepted as having at least equivalent efficiency for termination of VF compared with conventional 200 J energy monophasic waveform shock. The efficacy of such low energy defibrillation over high energy monophasic or higher energy escalating biphasic defibrillation is, as yet, not verified.

EARLY ADVANCED CARDIAC LIFE SUPPORT

The 3 links described earlier are nowadays often grouped together and referred to as Basic Cardiac Life Support (BCLS). While the provision of prompt BCLS is crucial to the survival of every patient who sustains a cardiac arrest, the early addition of further measures to stabilise those

so resuscitated has ensured a further decrease in the mortality from cardiac arrest. These further measures, referred to as Advanced Cardiac Life Support (ACLS), comprise of the following:

1) The use of various airway adjuncts and invasive measures to maintain an open airway.
2) The institution of positive pressure ventilation with 100% oxygen and portable devices to provide such ventilatory support in an ambulatory environment.
3) Measures to control and alter various cardiac rhythm abnormalities and cardiovascular hemodynamics through the use of a various maneuvers, drugs and devices, such as pacemakers.
4) The identification of the causes of collapse and various adverse events during cardiac resuscitations and the institution of definitive treatments for them.

Today, ACLS has adopted an algorithmic approach to problem solving in resuscitations.[9,10] The 3 most common algorithms for ACLS are:

1) VF/ Pulseless VT;
2) Asystole; and
3) PEA.

Ventricular Fibrillation/Pulseless Ventricular Tachycardia

Earlier, the need to deliver electrical defibrillation promptly across the chest in instances of VF/pulseless VT has been emphasized. Failure of these 3 stacked shocks to convert the VF/pulseless VT rhythm would result in the steps as described in Fig. 2. Persistent VF or recurrent VT is a common problem in cardiac resuscitations. Its optimal management requires good leadership of the resuscitation team, and an appropriate level of expertise in advanced airway management skills, including emergency endotracheal intubation, positive pressure ventilation with 100% oxygen at a rate of 10–16 per minute, continuing effective cardiac compressions and the use of a combination of vasoactive agents and electrical defibrillation.

The primary drug recommended in the event of persistent VF/pulseless VT is adrenaline. There is some evidence that vasopressin given intravenously at a dose of 40 mg may have at least an equivalent effect as intravenous adrenaline.[11, 12]

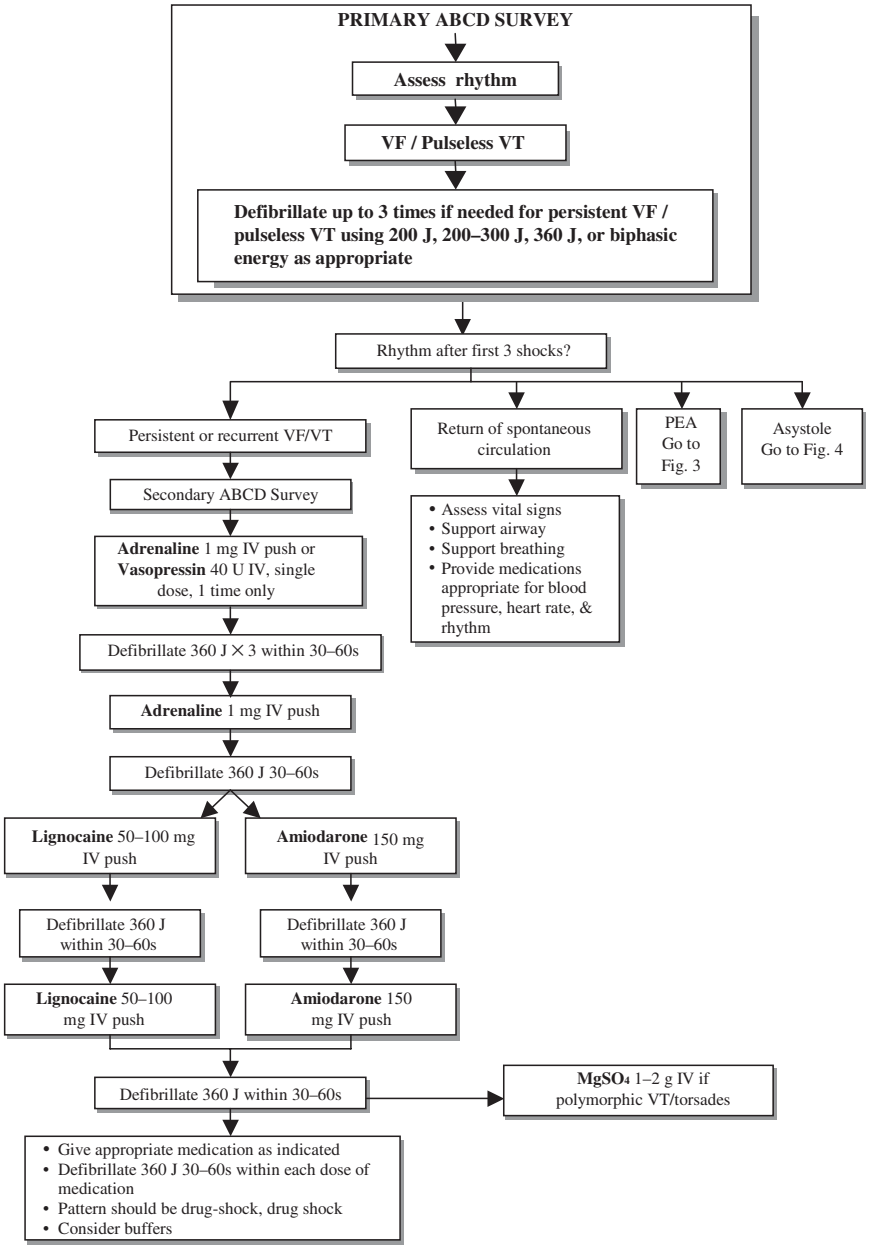


Fig. 2 Ventricular fibrillation/pulseless ventricular tachycardia algorithm.

Other intravenous agents that may be used in the event of failure of return of spontaneous circulation after the use of adrenaline or vasopressin followed by defibrillation are lignocaine, amiodarone and magnesium sulphate. The recommended dosages of these drugs for the management of persistent VF/pulseless VT are as follows:

- IV adrenaline 1 mg push every 3–5 minutes;
- IV vasopressin 40 mg;
- IV lignocaine 1.0–1.5 mg/kg body weight and consider repeating 3–5 minutes later up to a maximum of 3 mg/kg body weight;
- IV amiodarone 300 mg push or in divided doses of 150 mg × 2 at 3–5 minute intervals;[13] and
- IV magnesium sulphate 1–2 g.

After each of these drugs is given, CPR should be continued for at least 30–60 seconds before defibrillation at 360 J or equivalent biphasic.

Asystole

Asystole usually supervenes when there is failure to actively intervene during the phase of VF, pulseless VT or bradyarrhythmia. Characterized by a straight line, it requires continuous active CPR and the use of drugs (Fig. 3). Even then, the mortality of asystole is close to 100% and only extremely well-conducted resuscitations done in the most efficient manner would have any chance of a successful outcome.

Pulseless Electrical Activity (PEA)

This refers to a situation where potentially viable electrical activity is noted on the electrocardiographic monitor in the absence of a palpable carotid pulse. There is a wide variety range of rhythms that can be associated with pulseless electrical activity (PEA). These range from a slow idio-ventricular rhythm to one that resembles a sinus rhythm with near normal or slightly tachycardic rhythm. PEA is said to be associated with the presence of extremely severe forms of at least one or more of conditions (5 H's and 5 T's), the treatment of which is also given in Table 1.

The management of a patient with PEA is the same as for asystole. This is in addition to the rapid identification of as many of the contributing causes as possible and initiating the appropriate emergency treatment.

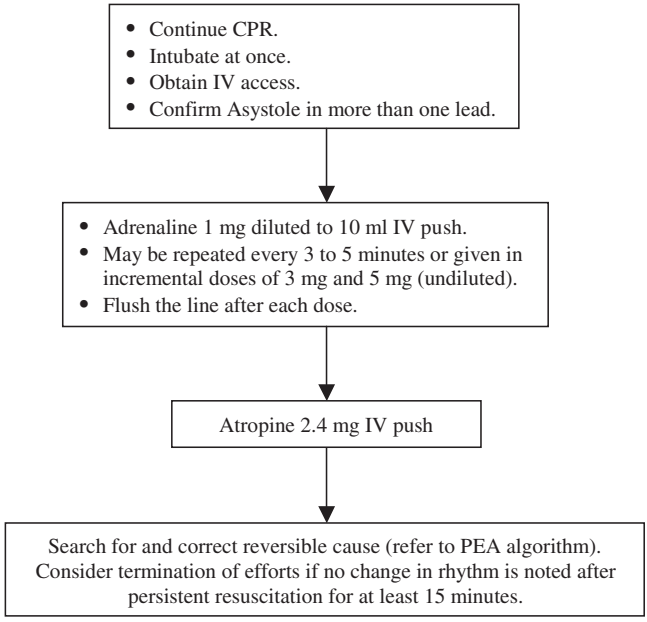


Fig. 3 Asystole treatment algorithm.

Table 1 Causes and Treatment of Pulseless Electrical Activity

Causes of PEA	Treatment
Hypovolemia	Intravenous fluids
Hypoxia	Intact airway system and ventilation with 100% oxygen
Hypothermia	Gradual rewarming
Hyperkalemia	IV $CaCl_2$, IV $NaHCO_3$ insulin/ glucose, dialysis
Hydrogen ion (acidosis)	Good alveolar ventilation and judicious use of bicarbonate
Thrombosis, coronary. AMI with free wall rupture	Emergency open heart surgery
Tablets. Drug overdoses, such as with tricyclic anti-depressants, digitalis, beta-blockers and calcium channel blockers	Treatment with antidotes, if available
Tension pneumothorax	Needle decompression
Thrombosis, pulmonary embolism	CPR/removal of clot
Tamponade, cardiac	Pericardiocentesis

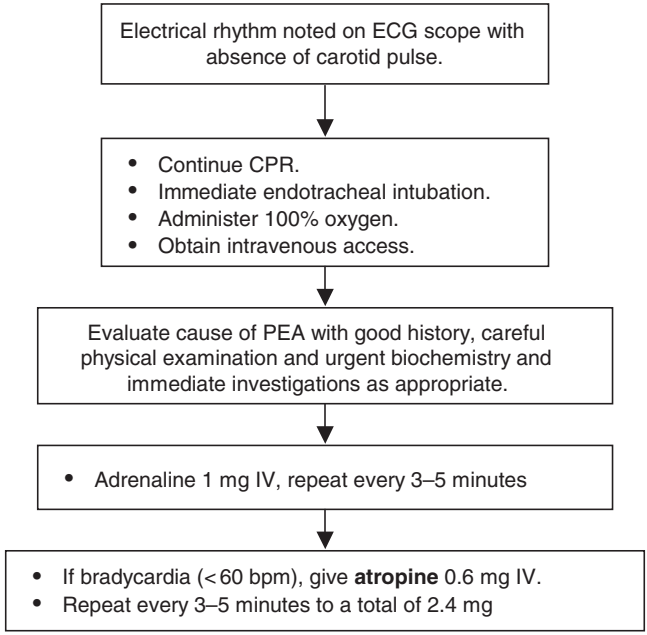


Fig. 4 Pulseless electrical activity algorithm.

The doses of adrenaline and atropine, the need for a definitive airway, 100% oxygen and good CPR are all basic pre-requisites. The treatment algorithm for PEA is shown in Fig. 4.

ORGANIZATION OF MEDICAL RESUSCITATIONS

Resuscitations should be pre-planned events. They are carried out in response to patients collapsing. It is thus not surprising that most resuscitations are chaotic, with plenty of shouting and confusion. However, resuscitations can be an organized affair, with resuscitation teams conducting the event in an orderly and efficient manner to maximise the survivability of the patient. Cardiac resuscitation may be divided into the following phases:

Phase 1: Preparatory Phase

This refers to the phase of training of each member of the ward in the roles of the resuscitation team. At every shift of work, assignments of staff

to specific roles in the team would help to ensure that each member would understand their role if a resuscitation were to be carried out. During this phase, staff would also check the resuscitation trolley and equipment to ensure that everything is available and functioning. One of the most frustrating aspects of a resuscitation is to discover that certain items of equipment are missing or not functioning. The staff will predraw and dilute in 2 syringes of 10 mls each containing 1 : 10,000 adrenaline and 1 syringe containing 2.4 mg of atropine.

Phase 2: Resuscitation Phase

The suggested organization of members of the resuscitation team is as indicated in Fig 5. Each member of the team plays a specific role. The individual roles played by the individual team members during the different phases of resuscitation are explained in Table 2. However, if the situation warrants, the team leader could modify the roles to satisfy the special needs of the patient being resuscitated. Once resuscitation begins, it should flow in an orderly manner. Doctor 1, who is located at the head of the patient, is in charge of airway and breathing. He is also the team leader. He will review briefly the patient's history so that team is well aware of the patient's background. If necessary, he can seek to obtain some initial information about the events preceding the patient's collapse from the other members of the staff or the paramedic team (for resuscitations being carried out at the emergency department). Doctor 1 will also evaluate the state of the airway and assess breathing and circulatory status of the patient. Nurse 1 is the resuscitation nurse-in-charge. Besides assisting Doctor 1 in airway management, she is also responsible for ensuring that the ECG monitoring leads are placed correctly on the patient. Doctor 2 is put in charge of intravenous access and Nurse 2 is standing on a step stool on the left side of the patient, ready to apply external chest compressions when instructed to do so. If the cardiac monitor shows VF or VT, the patient will have to be defibrillated according to the VF protocol (Fig. 2). This will be ordered by Doctor 1 and carried out by Doctor 2.

On the other hand, if the ECG monitor shows PEA or asystole, Doctor 1 will intubate the patient — if this has not already been done — and instruct Doctor 2 to administer intravenous adrenaline and atropine (if necessary). If someone is needed to obtain more information on the patient from the relatives or explain the proceedings to them, this may be

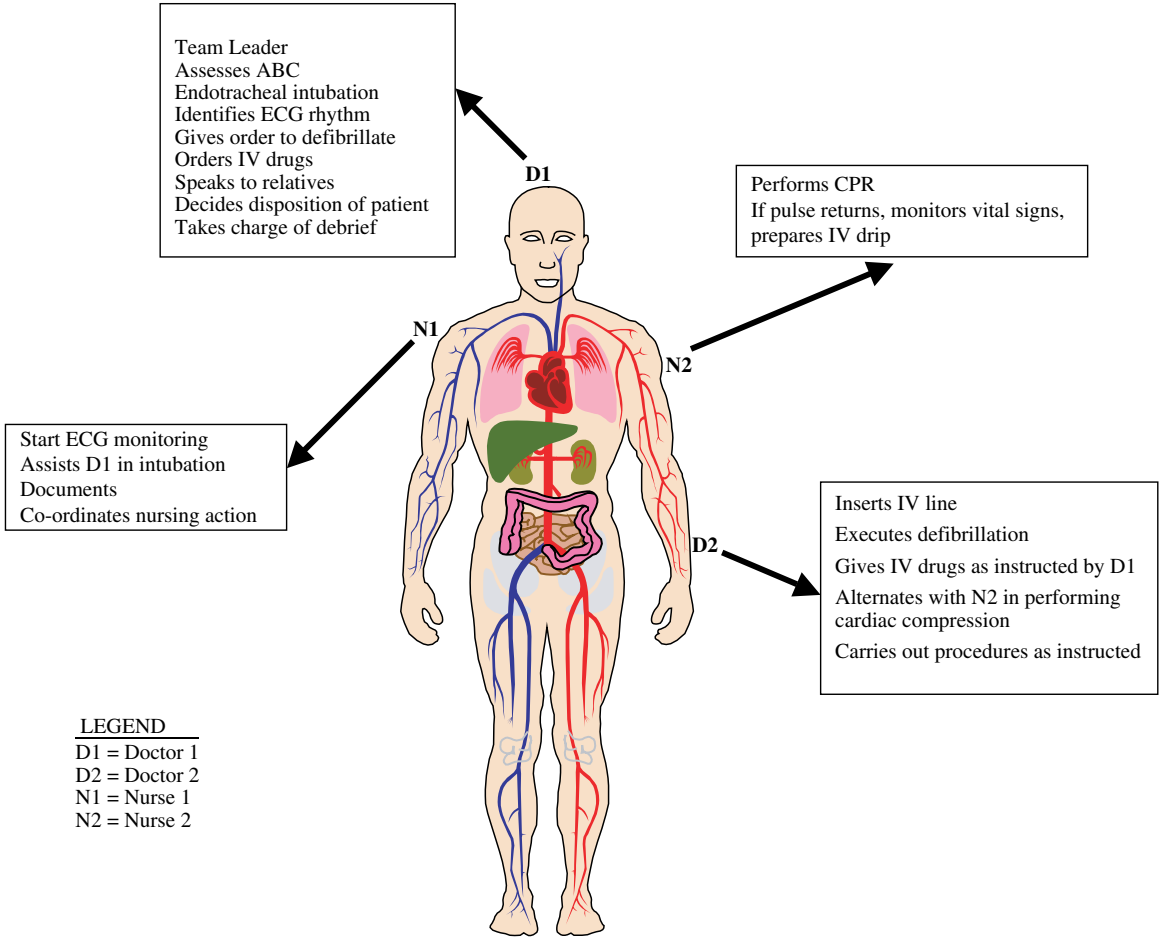


Fig. 5 Basic organization of members of resuscitation team.

Table 2 Organization of Events in a Resuscitation

- Only D1 is allowed to give instructions.
- Only N1 will co-ordinate the instructions from D1.

D1 (Leader)	D2	N1	N2
Resuscitation Phase			
Assess ABC. Ventilate as required. Identify ECG rhythm and give orders accordingly.	Insert IV line — anti-cubital fossa.	Remove patient's clothing. Start ECG monitoring. Assist D1 in intubation and secure ETT after intubation.	Prepare drugs as required. Perform cardiac compressions.
(i) VF			
Ventilate patient. Give order to defibrillate.	Charge defibrillator. Apply defibrillator pads. Execute defibrillation. Give drugs as required.	Record. Prepare drugs as required.	Continue cardiac compressions.
(ii) ASYSTOLE/PEA			
Ventilate patient.	Give: IV Adrenaline 1:10000 10 ml every 5 minutes. IV Atropine 2.4 mg	Document resuscitation sequence. Assist in any procedure ordered.	Continue cardiac compressions.
Insert CV line (optional). Order IV drugs/ shock as indicated.	Take short history from relatives. Alternate with N2 in performing cardiac compressions. Prepare drugs and give IV drugs/ shock as required.		Check vital signs only if pulse return. Prepare IV drip and cardiac pacing as required. Prepare drugs as required.

(Continued)

Table 2 Continued

D1 (Leader)	D2	N1	N2
Post-resuscitation			
Speak to relatives; break the good/ bad news. Inform DIL.	Fill out hospital case record.	Obtain ICU bed.	IV amiodarone/ lignocaine/ dopamine drip. 12-lead ECG.
Decide disposition of patient. Speak to doctor-in-charge of ICU, if necessary	Catheterization, ABG } as indicated	Insert nasogastric tube.	Monitor vital signs of patient every 5 minutes (including SaO_2)
Take charge of debrief.		Get ready documents for transfer to ICU.	

carried out by either of the doctors, a third doctor or a third nurse, if available. This needs to be handled in a tactful and sensitive manner and preferably in a quiet room rather than along a busy corridor.

The resuscitation is recorded by Nurse 1. It will proceed until the patient regains a spontaneous circulation or is pronounced dead.

Phase 3: The Post-Resuscitation Phase

When resuscitation is completed, Doctor 1 should proceed to break the bad news in the event that it failed or to explain the plan for further management when it is successful. For the latter, post-resuscitation investigations, initial treatment, and further monitoring will need to be organised and spelt out clearly.

Phase 4: Debrief

The debrief with the resuscitation team is conducted by the team leader as soon as the event is completed.

CONCLUSION

The provision of emergency life support in the event of collapse is vital as it increases the patient's chances of survival. If the various facets are not provided in an optimal fashion, the end result is usually disappointing. Conducting a resuscitation with dedication, perseverance and meticulous attention to every aspect of the patient is likely to result in a successful outcome. Today, resuscitation has become a core skill that every medical practitioner must master. The personal satisfaction derived from every successful resuscitation is seldom matched by any other event in medical practice.

REFERENCES

1. Eisenberg MS, Howard BT, Cummins RO, Reynolds–Haertle R, Hearne TR, *et al.*, Cardiac arrest and resuscitation: a tale of 29 cities. *Ann Emerg Med* **19**:179–186, 1990.
2. Anantharaman V, Early response to heart attacks, in *Heartline*, Singapore National Heart Association, Singapore, 1–8, 1999.
3. Lim SH, Teo WS, Anantharaman V, Chan YH *et al.*, Results of the first five years of out-of-hospital automatic extend defibrillation project: Singapore in the Ulstein Style, *Ann Emerg Med* **42**(4):548, 2003.
4. The American Heart Association in collaboration with the International Liason Committee on Resuscitation (ILCOR). Guidelines 2000 for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care; An international consensus on science, *Circulation* **102**(I):I1–I11, 2000.
5. Lim SH, Basic Cardiac Life Support (BCLS), *Singapore Med J Supplement* **1**:2–3, 2001.
6. Becker LB, Ostrander MP, Barret J, Kondos GT, *et al.*, Outcome of CPR in a large metropolitan area — where are the survivors? *Ann Emerg Med* **20**(4):355–361, 1991.
7. Guly UM, Mitchell BG, Cook R, Steedman DJ, Robertson CE, *et al.*, Paramedics and technicians are equally successful at managing cardiac arrest outside hospital, *BMJ* **310**:1091–1094, 1995.
8. Luna AB, Coumel P, Ledereq JF, Ambulatory sudden cardiac death: mechanisms of production of fatal arrhythmia on the basis of data from 157 cases, *Am Heart J* **117**(1):151–159, 1989.
9. Cummins RO (ed.), *Advanced Cardiac Life Support*, USA: American Heart Association, Chapter 1, pp. 1–27, 1997.
10. Teo WS, Advance Cardiac Life Support (ACLS), *Singapore Med J Supplement* **1**: 10–20, 2001.

11. Lindner KH, Dirks B, Strohmenger HU, *et al*., Randomised comparison of epinephrine and vasopressin in patients with out-of-hospital ventricular fibrillation, *Lancet* **349**:535–537, 1997.

12. Wenzel V, Krismer AC, Arntz HR *et al*., A comparison of vasopressin and epinephrine for out-of-hospital cardiopulmonary resuscitation, *N Eng J Med* **350**:105–113, 2004.

13. Kudenchuk PJ, Cobb LA, Copass MK, *et al*., Amiodarone for resuscitation after out-of-hospital cardiac arrest due to ventricular fibrillation, *N Engl J Med* **341**(12):871–878, 1999.

This page intentionally left blank

12

The Management of Poisoning

R. Ponampalam and V. Anantharaman

INTRODUCTION

Poisons have been around ever since the existence of humankind. History has numerous examples of the use of poisons for purposes of hunting to political assassinations. In modern day medicine, the first attempt to define a poison was made by Paracelsus[1] in 1564.

The aim of this chapter on poisoning is to give an overview of poisoning and its management.

OVERVIEW OF TOXICOLOGY

Toxicology is the study of the effects on health of various chemicals, both naturally occurring and man-made. The scope is wide and encompasses aspects of environmental, industrial and natural toxins (from plants and animal sources).

The management of toxicological problems are usually complex involving inputs from specialists from various disciplines. The prompt and early management of cases of poisoning, in terms of limiting the amount of

damage done by the poison, cannot be overemphasized. Hence, early diagnosis and interventions that begin with the initial contact by the patient with the physician is extremely important to the outcome. The experience and views of a clinical toxicologist are invaluable in some instances.

Paramedical staff, such as pharmacists, may also be involved as poisons information providers. Their assistance is crucial in sieving through the massive amount of poison and drug information resources.

Social workers may be called upon to help patients with underlying social problems. They will be helpful in the holistic management of the patient.

Environmental scientists and industrial hygienists are usually involved in situations where preventive interventions are essential to limit further damage of environmental contamination with hazardous chemicals.

PRINCIPLES OF POISON MANAGEMENT

The following principles of poison management apply generally to all toxic exposures:

1) Ensuring that the patient will not pose a health hazard to others from the spread of secondary contamination.
2) Taking care of the airway, breathing and circulation (ABC) as well as supportive care form the basis for managing toxic exposures.
3) Prompt and thorough assessment with appropriate interventions is essential to the final outcome.
4) Searching for causes other than poisoning in toxic exposures, e.g. hypoglycemia, in an unconscious patient who has overdosed.

The management of poisoning[2–4] like that of other medical conditions, can be handled in a practical fashion. Seven stages (A–G) have been identified in the management of the poisoned patient.

A. Resuscitation and Stabilization
B. Clinical Evaluation and Definitive Diagnosis (Toxidrome Recognition)
C. Decontamination
D. Enhanced Elimination of Absorbed Poisons
E. Antidotes
F. Supportive Care
G. Disposition of the Poisoned Patient

A. Resuscitation and Stabilization

As in any other medical condition, the primary aim of the physician is to resuscitate and stabilize the patient. This involves paying attention to the level of consciousness, maintenance of an open airway, adequate ventilation, oxygenation and adequacy of circulation. This may involve the use of fluids, chronotropic, and inotropic agents. These have to be expedited to ensure a good outcome. A critical difference in resuscitations[5] involving toxic exposure compared to conventional conditions is the risk to rescuers and healthcare providers from secondary exposure to toxicants from the patient. Hence, it is important to ensure appropriate personal protective measures at all times. Contaminated clothing and toxic vomitus should be handled cautiously and disposed of appropriately. The possibility of off-gassing of toxic gases from the stomach during intubations should be borne in mind and appropriate preventive measures taken.

B. Clinical Evaluation and Definitive Diagnosis (Toxidrome Recognition)

Initial resuscitation and stabilization are carried out almost simultaneously with efforts to obtain a good history, perform a thorough physical examination and carry out the relevant investigations to arrive at a definitive diagnosis.

A thorough history should be obtained, but in many cases this may not be possible as the patient may be unconscious, uncommunicative, deliberately misleading or unintentionally inaccurate. A high index of suspicion to a possible toxic etiology should always be borne in mind in such situations. Circumstantial evidence suggesting a possible toxic exposure may be gathered indirectly by interviewing family, friends, paramedics, police and observers.

The **history** in a case of poisoning has the following components:

1) *Nature of poison/toxin.* This may also involve identification of the poison and includes whether someone has brought either the packaging or some samples of the poison ingested, if available. It is also important to determine if there were any co-ingestants and the formulations in which it was consumed, e.g. sustained release or enteric coated drugs, which will be expected to have altered toxicokinetics as well as a predisposition for formation of concretions in the stomach

with prolonged effects. It is important to remember that in some situations the diluents or additives in the formulations are themselves the poison and not the active ingredient.

2) *Amount of poison/toxin.* This is important to determine mode of decontamination and need for antidote. Further assessment of this may require laboratory tests to determine blood or other tissue levels of the poison.
3) *Route of exposure to poison/toxin.* Although the oral route of exposure is probably the most common, it is important to appreciate the different routes of exposure in any particular poisoning event. These include exposure via the inhalational (respiratory), cutaneous (dermal), parenteral (intravenous, intramuscular, subcutaneous, intraperitoneal, intrathecal, bites and stings from venomous animals, etc) and transplacental routes. The route of exposure to a toxin determines its toxicokinetics in the exposed individual as well as helps determine the type of decontamination procedure to be used. For example, for cutaneous exposure to corrosives, the best decontamination is achieved by copious irrigation of the exposed site with large quantities of water. On the other hand, for exposures via the inhalational route, the best decontamination will be rapid evacuation of the patient from the toxic environment and institution of high flow oxygen. There are also important implications to the rescuers and healthcare providers with respect to the route of exposure.
4) *Time of ingestion or length of contact.* Most substances that are taken orally do not stay in the stomach for more than 2 hours. After 4 hours, the proportion of ingested drug in the stomach is relatively small. An exception is when the amount of ingested poison is extremely large; the poison is salicylate, barbiturates, tricyclic antidepressants; or if it is ingested with a lot of fatty foods. This has implications for the method of gastrointestinal decontamination that may be applied.
5) *Emergency first-aid treatment.* If this has already been applied, one should proceed with the next stage of management as soon as possible. In some communities, ipecac syrup may have been administered as home first-aid for induction of emesis and therefore gut decontamination. It is important to know if this has been carried out as it will affect the initial management of the patient.
6) *The patient's psychological profile.* This forms a very important aspect of the overall management of the poisoned patient. This portion of the history may be difficult to obtain as the patient may be unconscious, unresponsive, uncooperative, or confused.

7) *Symptoms experienced by the patient.* These will shed some light as to the nature of the poison, especially if it is not readily divulged by the patient or other persons.
8) *Presence of comorbid illnesses.* Finally, knowledge of the presence of comorbid diseases such as liver or renal disease, as well as long-term medication use will assist in predicting potential toxicity. The physician will need to take into account altered pharmacokinetics and drug interactions, and then direct specific therapies.

Since the history may not always be forthcoming from the patient, one must seek additional information from other sources, such as the paramedics who brought the patient to hospital, office colleagues, relatives or anyone who can provide any information on the patient or events surrounding the incident.

Often, however, a history may not be available. Therefore, a careful and thorough physical examination[6] will yield valuable clues. The list below gives a common list of physical findings commonly associated with particular poisons. The order of the presentation reflects the order in which the physical examination should be carried out. The signs given in this table should be taken with the findings of systematic examination, which in cases of poisoning must also include the following:

1) vital signs, such as heart rate, respiratory rate, temperature, and blood pressure;
2) neurological state, conscious level evaluation and pupil size and reactivity assessment;
3) dermatological clues including skin color (cyanosis suspicious of methaemoglobinemia, or bright pink coloration suggestive of poisoning by carbon monoxide, cyanide or hydrogen sulfide), needle track marks from intravenous drug use (opiates), blisters due to pressures from prolonged lying in one position (carbon monoxide, barbiturate poisoning), sweatiness (sympathomimetics, cholinergic agents), hot and flushed (anti-cholinergic agents); and
4) smell of toxins on the breath such as alcohol, ketones (diabetic/alcoholic ketoacidosis), oil of Wintergreen (salicylates), bitter almonds (cyanide), garlic (organophosphates, arsenic, selenium, thallium), rotten eggs (hydrogen sulfide, disulfiram), solvent or petroleum (hydrocarbons such as glue, thinner).

TOXIDROMES

Toxidromes refer to a constellation of physical signs and symptoms in a poisoned patient that helps point the treating physician to suspect a particular class of poisons as the etiologic agent. This is a useful tool in aiding the doctor manage a patient poisoned with an unknown agent.

Toxidrome recognition[7] involves the detection of alterations in mental state, vital signs, pupil size, and other signs described previously.

C. Decontamination

The definitive management of poisoning begins from the first contact with the poisoned victim. Following initial assessment of ABC and institution of critical interventions to preserve the ABC, drug elimination becomes the primary concern of the physician. Decontamination refers to the removal of poisons from the portals of entry into the body. The portals of entry correspond to the route of exposure to the poisons. In general, these include oral, inhalational, cutaneous and parenteral exposures. It is important to know the route of exposure because attempts to decontaminate are determined by them. Three areas of decontamination are discussed:

- eye and skin decontamination;
- respiratory decontamination; and
- gastrointestinal decontamination.

Eye and Skin Decontamination

Protection of rescuers/staff

Decontamination of easily accessible areas of the body such as the eyes and the skin should be preceded by measures taken to protect staff of decontamination teams. Such protective gear should include as a bare minimum, disposable hospital gowns, individual chemical resistant suits, plastic goggles, latex gloves, and surgical masks.

Decontamination procedure[8]

1) Move the victim away from a contaminated area so that there will be no further contact with the poison or hazardous chemical.

2) Flush exposed eyes and other body surfaces with copious plain water for at least 2–5 minutes. Eye irrigation should continue for at least a further 10–15 minutes either with plain water or with Normal Saline. Precautions will need to be taken to avoid secondary contamination of the unaffected eye by proper positioning of the unaffected eye and the patient during eye irrigation.
3) In the event of skin contamination, the patient should be stripped of contaminated clothing and copiously irrigated for at least 15 minutes with water and soap solution. If the offending chemical agent is thought to have a blistering effect on the skin, decontamination should be with Fuller's Earth to remove as much of the contaminant as possible, followed by flushing with copious water for 15 minutes. Failure to ensure adequate skin decontamination will result in continuous absorption of residual contaminants through the skin.
4) To ensure completeness of the procedure, even the nails should be scrubbed with a scrub brush or plastic nail cleaner.
5) Following showering, the victim should be dried and re-clothed afresh.

Respiratory Decontamination

The most important aspect of respiratory decontamination is the removal of the victim from the contaminated environment and the initiation of high oxygen therapy to remove residual contamination in the alveoli. Rescuers undertaking such an operation should be properly equipped with protective gear and breathing apparatus before entry into such a toxic environment in order to avoid becoming casualties themselves. Broncho-alveolar lavage may be available and indicated for certain toxins such as hydrocarbons.

Gastrointestinal Decontamination

This has been an area of some of the most interesting changes in toxicology over the last few years. The majority of poisonings encountered in clinical practice occur via the gastrointestinal tract. An understanding of gastric emptying mechanisms would play a part in choosing the technique of gastric decontamination to be employed or in deciding whether enhancing elimination of absorbed poison has to take priority over gastric emptying procedures.

Several factors affect gastric emptying of poisons. Most substances are absorbed in the intestines. Gastric absorption is usually minimal. The factors that are known to affect gastric emptying are:

1) Food in the stomach:
   - fat-rich foods slow gastric emptying;
   - emptying of contents in an otherwise empty stomach is usually complete in 1–2 hours; and
   - a full stomach may take up to 6 hours for near-complete emptying.
2) Amount of poison ingested: The larger the amount ingested, the longer the gastric emptying.
3) Type of poison: Certain poisons such as salicylates, barbiturates and tricyclic anti-depressants slow down gastric emptying. In these instances, significant amount of the poison may be detected in the stomach for up to 6 hours.

Many procedures[6,7,12] have been tried in removing or decreasing absorption of drugs in the gastrointestinal tract, *viz*.

1) **Dilution**
   Water is the best diluent and achieves 2 important results: It helps to reduce the irritation of the gastric mucosa that is induced by many ingested poisons; and by adding bulk to the contents of the stomach it aids the process of emesis, if this is not contraindicated.

   The volume of water recommended is usually 100–200 ml for a child and 200–400 ml for an adult. Water, however, should only be given to conscious victims. It should never be forced.

   Remember that excessive water can distend the stomach wall resulting in premature gastric emptying into the duodenum, making it more difficult for further removal of the poison. In the case of ingestion of solids, e.g. tablets, dilution with water may promote dissolution and enhance absorption of the ingested drug.

   Milk is another diluent readily available in the home. It is indicated most often for ingestion of caustic or irritant substances except phosphorus. Milk may delay the onset of drug-induced emesis.
2) **Emesis** (induced vomiting)
   Traditionally, emesis has been one of the earliest procedures used in gastrointestinal decontamination. There are, however, certain

conditions during which emesis should not be attempted:

a) If the patient is
   - unconscious;
   - does not have a gag reflex;
   - has severe cardiovascular disease or emphysema or severe bleeding tendency; or
   - is under 6 months of age.

b) If the poison is
   - any agent causing rapid decrease in the level of consciousness, seizures, cardiovascular collapse or neuromuscular collapse or neuromuscular paralysis such as cyclic antidepressants, isoniazid, propoxyphene, camphor, beta-blockers and convulsants;
   - has hydrocarbons that are likely to result in severe lung injury if aspirated even in minute amounts; or
   - is a corrosive acid or alkali that will aggravate upper gastrointestinal injury as a result of high pressures created during emesis.

The following are known methods of emesis:

a) *Syrup of ipecac.* This produces vomiting through stimulation of chemoreceptors in the central nervous system. The recommended dosages are:

Age	Dose (ml)
6–12 months	10
1–12 years	15
More than 12 years	30

Though generally safe and well tolerated, its adverse effects include protracted vomiting, diarrhea, lethargy, excessive sweating and fever. The uncontrolled emesis can be extremely distressing both to patients and relatives and further drug elimination procedures are difficult to perform until the bouts of vomiting have ceased, which may not be for at least one hour. Administration of activated charcoal is thus delayed.

b) *Apomorphine*. Though this drug rapidly induces vomiting, it needs to be given by a doctor, and home use is inappropriate. Its adverse

effects such as central nervous depression and hypotension must also be borne in mind.

c) *Mechanically-induced.* The poisoned victim is made to lean forward with extended neck and the back of the tongue or pharynx is stimulated using fingers, spoons, or a tongue depressor. It has, however, been found to be generally less effective than ipecac emesis.

3) **Gastric lavage**

This involves washing out the stomach with solutions and is preferred often because patient compliance is not a feature. It can also be carried out rapidly and the administration of activated charcoal can be facilitated bypassing palatibility issues. However, there are some possible complications from performing gastric lavage. These include aspiration, mechanical injury to the throat, esophagus and stomach, laryngospasm, fluid and electrolyte imbalances and patient discomfort. In addition, the amount of toxin removed by lavage is highly variable and diminishes with time. This has resulted in the current guidelines,[12] which recommended that gastric lavage be performed only if the patient has ingested a potentially life-threatening amount of poison and if the procedure can be done within one hour of the ingestion. The one-hour limit may be extended in certain situations up to 4 hours post-ingestion, in cases involving poisons that have a tendency to remain in the stomach for prolonged periods, e.g. salicylates, anti-cholinergic agents, opiates, etc.

Gastric lavage is contraindicated in the following instances:

- Ingestion of corrosives, e.g. acid/alkali.
- Ingestion of petroleum distillates.
- Presence of convulsions.
- Patients at risk of gut perforation or hemorrhage due to pathology or recent surgery.

The procedure for gastric lavage is as follows:

a) Use the largest-bore tube possible, e.g. a size 32–36 French gauge oro-gastric lavage tube and confirm the position of the tube before beginning lavage.

b) Give the patient a glass of water to drink before passing the tube — if the patient is conscious.

c) Protect the airway. In the comatose patient a cuffed endotracheal intubation is mandatory before passing the tube.

d) Place the patient on his side, usually in the left lateral position, with the head lower than the waist to minimize chances of accidental aspiration.
e) Use either tap water or warmed saline as the lavage fluid.
f) To check correct placement of the tube, place the outer end of the tube in a glass of water. Active bubbling on expiration suggests tracheal placement.
g) Aspirate stomach contents once the tube has entered the stomach.
h) After instillation of 100–300 ml of lavage fluid, manually agitate the stomach and then withdraw the fluid. Repeat until the lavage return is clear.
i) Following completion of the lavage procedure, instil 50 gm of activated charcoal suspension in the stomach prior to removal of the lavage tube.

4) **Activated Charcoal**

Over the last 25 years, activated charcoal has gained an impressive role in the initial management of the poisoned patient. Activated charcoal is an adsorbent of many chemical substances.

There are no absolute contraindications to activated charcoal. However, it is known, by itself, to result in constipation and repeated dosage in patients with ileus may result in vomiting. Though it has no direct adverse effect on the lungs, it is often mixed with oral bacteria and gastric acid, which do cause damage if aspirated. It is therefore, preferably given with a cathartic. Activated charcoal may interfere with endoscopic evaluation in situations where this procedure may be required.

The superior effect of activated charcoal has led to its use as a common gastrointestinal decontaminant. Generally, the following drugs are known to easily adsorb onto activated charcoal:

Acetaminophen
Aliphatic alcohols
Amitryptilline (and other anti-depressants)
Anti-pyrines
Arsenic
Aspirin
Atropine
Chlorpheniramine (and related anti-histamines)
Chlorpromazine (and other phenothiazines)
Dextro-amphetamine
Digoxin
Glutethimide
Imipramide
Iodine
Isoniazid
Meprobamate
Mercuric chloride

Methylsalicylate	Phenylpropanolamine
Morphine	Phenytoin
Nortryptilline	Propoxyphene
Paraquat	Quinidine
Phenobarbitone (and other barbiturates)	Quinine
Penicillin	Salicylates

Activated charcoal has not been shown to actively adsorb the following:

Aromatic alcohols	Hydrocarbons
Boric acid	Heavy metals
DDT (dichloro-diphenyl, trichloroethane)	Malathion
	Methylcarbamate
Ethylene glycol	Cyanides
Iron	Methanol
Lithium	Acids and caustic alkalis

The optimal dose regimen for use of activated charcoal is unknown but data available suggest a dose response relationship that favors larger doses. The ideal dose of activated charcoal to drug ratio is recommended to be 10:1 by weight.

For adults, the first dose of activated charcoal is given in a dose of 30–50 gm in 100–200 ml water either orally or via an oro-gastric tube. Subsequent doses are administered in a dose of 15–25 gm/50 ml water at 2–4 hourly intervals for up to a maximum of 24 hours.

For children, the first dose is 1 gm/kg body weight orally or via gastric lavage tube. Subsequent dosage is at 0.5 gm/kg body weight at 2–4 hourly intervals for up to 24 hours.

Remember that activated charcoal is not absorbed by the gastrointestinal mucosa and is pharmacologically inert.

The use of multiple doses of activated charcoal has been reported to enhance elimination of certain absorbed drugs. The mechanisms for this are thought to be interruption of enterohepatic circulation and a gastrointestinal dialysis-type effect.

Other adsorbents that are gradually coming into clinical practice include bentonite and Fuller's Earth, the prime action of which has been noted in poisoning with paraquat and in lithium poisoning.

5) **Cathartics**

Although cathartics have been used in poison management, there is no proven record of their efficacy in clinical practice. The theoretical advantage of using catharsis as the sole method of gastrointestinal

decontamination has been to increase gastrointestinal transit speed and thus decrease time available for absorption of the poisoning agent. What has been of more recent interest is the ability of cathartics to neutralise the constipating effect of activated charcoal so as to allow more of the activated charcoal to come into contact with and adsorb additional amounts of the poisoning agent.

Cathartics have not been proven to decrease the efficacy of activated charcoal and have therefore come to be included in part of the management regimen following activated charcoal administration.

D. Enhanced Elimination of Absorbed Poisons

Once the poison has gained access into the circulation, it will be distributed to all the body systems where it will start destructive processes at the end organs depending on the specific poison involved. The body tries to cope with the poison by detoxification processes occurring in the liver or enzymatic degradation in the blood. In addition, the normal excretory organs of the body such as the kidneys, lungs, gall bladder, gut and sweat glands of the skin are also employed to enhance elimination of the poisons. It is when these natural systems are overwhelmed or non-functional due to disease processes that the body suffers from permanent injury to the predisposed end organs.

Toxicological management at this stage involves enhancing the natural excretory or detoxification pathways or even promoting the use of an alternative pathway that exists but is rarely employed by the body under normal circumstances. In order to be able to manipulate these to the patient's advantage it is important that the treating physician is well-versed with the pharmacokinetics[9] of drugs in therapeutic doses and its toxicokinetics in overdoses. Three important pharmacokinetic/toxicokinetic principles are relevant here:

Transport across cell membranes

Passive diffusion of drugs across concentration gradients through phospholipid cell membranes requires these drugs to be soluble and in solution. A prime example of this would be organophosphate insecticides and nerve gas agents which, because of their high lipid solubility, are easily absorbed through the skin resulting in poisoning.

Liver enzymes transform lipid-soluble drugs into water-soluble metabolites for excretion by the kidneys

Though these metabolites are usually only mildly active or inactive, occasionally some may be highly active and toxic. The hepatic microsomal mixed functional oxidases can be induced by drugs and other environmental agents. This, however, is time-consuming, and is of no practical benefit during drug overdose situations when speed of action is essential. Some liver enzymes can, however, be inhibited by competition between 2 drugs. Thus, in methanol poisoning, ethanol can be given and competes with methanol for alcohol and aldehyde dehydrogenases so that acetic acid rather than the toxic formic acid can accumulate. Some enzyme activities can be saturated during episodes of poisoning. This can result in depletion of endogenous conjugating substances. Thus, in poisoning with paracetamol, the commonest drug of overdosage in this part of the world, the glucuronidation and sulphation activities are saturated. While the liver mixed function oxidases usually activate only about 6–10% of the paracetamol resulting in a highly toxic metabolite that conjugates with glutathione to form mercapturic acid and cysteine conjugates that are excreted, depletion of the glutathione in overdosage situations allows the toxic metabolite to accumulate. This ends in hepatic necrosis. Increasing glutathione stores to handle the toxic metabolites by addition of N-acetylcysteine provides a rationale for treatment of paracetamol poisoning.

Glomerular filtration of water-soluble drugs and metabolites is dependent on their plasma protein binding

Free drugs get filtered and organic acids (glucuronides, penicillins, uric acid) and bases (choline, neostigmine) are actively secreted into the tubules. Passive reabsorption of fat-soluble drugs occurs in the presence of the appropriate pKa. Therefore, renal elimination will be dependent on the volume and pH of the urine. Manipulation of the volume and pH of the urine can then enhance total body clearance. This principle is used to enhance elimination of drug in salicylate poisoning. Since in overdose situations, liver metabolism and plasma protein binding of salicylates is saturated, and also the active secretion of the drug into the tubules, increasing urine volume will increase the amount actively secreted into

the urine, and alkalinizing the urine will decrease reabsorption of the salicylate by the tubules. Ion-trapping in the renal tubules has been postulated as the mechanism by which forced alkaline diuresis enhances renal clearance of salicylate when that drug is taken in overdosage. All this is provided the patient is able to accommodate the fluid load administered. Similar mechanisms also work in overdosage with amphetamines, alcohol, aniline dyes, long-acting barbiturates, bromides, ethylene glycol, isoniazid, lithium, methanol, penicillin, phencyclidine, quinine, quinidine, strychnine and sulphonamides.

The following are common techniques of enhancing elimination of absorbed drugs:

1) **Forced alkaline diuresis**: This can be achieved by a cycle of 1.5 litres of fluid per 3 hours as follows:
   - 500 ml of Normal Saline
   - 500 ml of 5% Dextrose + 20 ml of 7.45% Potassium Chloride
   - 500 ml of Normal Saline
   - Intravenous Frusemide 20 mg is given at the end of each such cycle. Care should be exercised when employing forced diuresis in elderly patients and in patients with underlying cardiac and renal disease and in those who have consumed poisons which are cardio- or nephro-toxic. Serum pH and electrolytes should be monitored closely and urinary pH should be maintained at a value of 8.0 or above by adjusting the rate of bicarbonate infusion.
2) **Forced acid diuresis**: This is a rarely used procedure and is theoretically useful for poisoning with weak bases such as quinine, phencyclidine, amphetamines, and fenfluramine.
3) **Hemodialysis, hemoperfusion, hemofiltration and peritoneal dialysis**: These other modalities have been utilised as adjuncts to enhance elimination of poisons in severely poisoned patients. Other indications for using these techniques include the failure of the end organ responsible for elimination of the drug and the presence of poisons whose metabolites are equally or more toxic than the drug itself. Knowledge of the pharmacokinetics and toxicokinetics of the poisons is essential in order to predict which poisons will respond to these interventions. Generally, poisons having a small volume of distribution, low protein binding, high water solubility and small size (molecular weight less than 500 daltons) are amenable to this intervention.

Hemodialysis is useful in patients with renal failure when the normal excretory routes are blocked. The added benefit of correcting fluid and metabolic abnormalities makes it an attractive modality. Besides, most large general hospitals usually have access to renal services and physicians who will be able to perform hemodialysis in-house without having to transfer potentially unstable patients. Drugs that may be removable by hemodialysis include sedatives and hypnotics, alcohols, analgesics, metals and others such as theophylline, digoxin and tricyclic anti-depressants.

Hemoperfusion with activated charcoal, or ion-exchange resin is a modification of this intervention combining the adsorbent properties of activated charcoal/resin with dialysis. The method involves setting up vascular access similar to hemodialysis and pumping the blood through a column containing activated charcoal/resin. This again has specific indications for specific poisonings. The use of charcoal hemoperfusion in barbiturate, carbamazepine, salicylate, and theophylline poisonings has been found to be effective in lowering the serum levels. The extraction ratio of toxins is 2–3 times more compared to hemodialysis. The potential complication to be anticipated in this intervention is profound thrombocytopenia, which may require platelet transfusion. Systemic anti-coagulation is also required as in hemodialysis.

Peritoneal dialysis is relatively easier to perform but has poor extraction ratios making it an inefficient process.

Hemofiltration is less invasive with minimal effects on hemodynamics compared to hemodialysis and hemoperfusion and can be performed continuously on the patient making amends for its less effective poison extraction ratio. However, its role in acute management of poisonings is yet to be evaluated.

E. Antidotes

Antidotes[4,6,10,11] are chemicals that generally antagonize the effects of a poison or class of poisons. Although they are nice to have, most poisons do not have an antidote. There are many mechanisms by which antidotes work. Four mechanisms are currently understood.

1) **Chemical antagonism.** These react with the poison to produce a compound that is either non-toxic or of lesser toxicity. An example is the use of sodium thiosulphate in cyanide poisoning. Thiosulphate is a sulfur donor. This enables the conversion of cyanide to the less toxic

thiocyanate. Another example is the use of specific antibodies, such as digoxin-specific antibodies that are administered in the event of digoxin overdosage and act by forming a complex with the digoxin. The digoxin molecule in such a complex is no longer pharmacologically active.

2) **Receptor antibodies.** These compete with the poison for its receptor sites in human tissues resulting in decreased formation of poison–receptor complexes and therefore decreasing the effect of the poison. Examples of such an effect are seen in:
   - naloxone competing with morphine and other narcotics for opiate receptors; and
   - flumazenil which selectively competes with benzodiazepine receptors in the central nervous system.
3) **Chelation.** Chelating agents form water-soluble complexes with heavy metals. Such complexes prevent the heavy metal from exerting its adverse effects and also are excretable through the kidneys thus allowing for more rapid elimination of the absorbed poison. Examples are the use of British Anti-Lewisite (BAL or Dimercaprol), which is the chelating agent for heavy metals such as arsenic, mercury, lead and gold. Penicillamine chelates copper, lead, mercury and arsenic, and desferrioxamine is currently the best available chelating agent for iron overload.
4) **Competitive inhibition.** This refers to the use of agents that compete with the poison for enzymes employed by the poison for its effect on the individual. Examples are:
   - Ethanol, which competes with methanol and ethylene glycol for alcohol dehydrogenase and therefore displaces these poisons from its enzyme substrate, thereby preventing the formation of toxic metabolites;
   - Pralidoxime, one of the antidotes used in organophosphate poisoning, exerts its effect by reactivating phosphorylated cholinesterase enzymes and protecting the enzyme from further inhibition; and
   - Physostigmine, which reversibly inhibits acetylcholinesterase and is used in cyclic anti-depressant overdosage.

Antidotes, just like the poisons they are used to counteract, are chemicals, which may have potential adverse effects. Hence, in using these agents it is wise for the physician to be well aware of the indications and limitations for each antidote.

A list of common antidotes and their recommended uses is given in Table 3.

F. Supportive Care

The initial emergency management of poisoning should be instituted in a timely manner with the initial assessment of the patient. Following this, the patient needs to be assessed to determine the severity, prognosis of the poisoning event and the potential for deterioration recognized.

Supportive measures include general measures and specific measures for specific toxins. During this phase the patient should continue to receive general supportive care which refers to frequent monitoring of vital signs, maintaining fluid and electrolyte balance, providing cardio-respiratory support as indicated and appropriate nursing care to preserve integrity of body systems. Sometimes general supportive care may be all that is required.

G. Disposition of the Poisoned Patient

Generally speaking, all mild poisonings with no recognized potential for deterioration can be discharged following initial assessment and treatment and following a period of observation during which the social and educational aspects should be covered. Where a psychiatric consultation is required, such consultation should optimally be carried out during the initial visit so that further ambulatory management of the patient may be expedited. Protocols should be set in place that would automatically activate these processes.

All severe poisonings will require a period of inpatient monitoring and stabilization before discharge. Admission of such poisonings may be to an intensive care or high-dependency area, with continuous monitoring of vital signs.

Between the two extremes, the moderate poisonings can be assessed, initial treatment provided in the emergency department and the patient then admitted for a few days to tide them over the first one to two days of therapy and monitoring for possible adverse effects of the poison. Discharge of the patient should be with the accompaniment of advice. An example will be the case where a patient suffers a short period of smoke inhalation. The initial symptoms may have resolved by the time he seeks medical treatment. However, the potential for secondary pulmonary edema should be recognized and the patient monitored for a period of

time and perhaps even managed as an in-patient before being discharged with advice to return if symptoms should recur and education to prevent further episodes of exposure.

TOXICOLOGY LABORATORY

In most poisoning events a clinical suspicion of a particular poison is possible from a careful history and a thorough examination. However, a definitive diagnosis and management plan for the poisoned patient may be guided by certain laboratory investigations, and especially by blood levels of the causative agent of poisoning, which will give an objective assessment of the situation. In some situations, there may be a need for repeated sampling of blood for toxins 2–4 hours from the initial sample to help determine the trends in the blood levels which may give a better indication of severity of poisoning, effectiveness of therapeutic interventions carried out and to help prognosticate. It is therefore crucial that one be familiar with investigations pertaining to each poison and the capabilities of the in-house toxicology laboratory.

The following investigations are considered valuable in the management of poisonings:

1) Serum electrolytes and anion gap calculation;
2) Serum glucose;
3) Renal function tests;
4) Arterial blood gases;
5) Hemoglobin co-oximetry (methemoglobin, carboxyhemoglobin levels);
6) Serum osmolality and osmolar gap;
7) Liver function test;
8) Electrocardiogram (ECG);
9) Urinalysis for crystalluria, hemoglobinuria and myoglobinuria;
10) X-rays for radio opaque poisons; and
11) Statistical blood levels of certain poisons/tests that are important for management as shown in the table below.

Requirements for collection of specimens during management of patients with poisoning include:

- gastric aspirate: At least 50 ml (plain);
- blood for drugs screen or cyanide: 10 ml (citrated, oxalated or with heparin);

Table 1 Statistical Blood Levels of Some Poisons/Tests

Poison/Test	*Toxic Concentration
Carbamazepine	> 10 mg/L
Carbon monoxide	> 10% smokers > 5% non-smokers
Cholinesterase, rbc	< 5000 (11 500–21 000)
Cholinesterase, serum	< 2500 (5000–13 500)
Digoxin	> 2 ng/mL
Ethanol	> 200 mg/dL
Ethylene glycol	> 20 mg/dL
Iron	> 350 μg/dL
Isopropanol	> 100 mg/dL
Lead	> 40 μg/dL
Lithium	> 1.5 mmol/L
Methanol	> 20 mg/dL
Methemoglobin	> 20%
Paracetamol	> 140 mg/L at 4 hours post-ingestion
Phenobarbital	> 40 mg/L
Phenytoin	> 20 mg/L
Salicylate	> 300 μg/mL
Theophylline	> 20 mg/L
Valproic acid	> 100 mg/L

*Refers to levels at which active medical interventions are usually necessary.

- blood for carboxyhemoglobin: 5 ml (citrate);
- urine: At least 20 ml (plain); and
- disposable containers and tubes with leak-proof screw caps.

Most communities tend to have arrangements for rapid assays of common poisons for use in emergency situations where the blood level is crucial in determining mode of management.

POISON INFORMATION RESOURCES

The large quantity of poisons including drugs, chemicals and naturally occurring toxins is ever increasing in our modern society. It is difficult for anyone to be knowledgeable about every poison and to keep up to date with changes. Recognizing these limitations, many countries have established Poison Information Centers to act as information resources for the myriad toxins in our environment. In addition, these centers give guidance to physicians on management aspects of poisonings, indirectly

performing a quality control function. The best option would be to be able to discuss poison management issues with a clinical toxicologist who has the back-up Poisons Information Center and is thus in a position to confidently advise in a realistic manner.

REFERENCES

1. Deichmann WD, Henschler D, Holmstedt B, Keil G, What is there that is not poison? A study of the Third Defence by Paracelsus, *Archives of Toxicology* **58**:207–213, 1986.
2. Olson KR, *Poisoning and Drug Overdose*, Appleton & Lange, 1999.
3. Goldfrank LR, *Goldfrank's Toxicologic Emergencies, 6th ed.*, Appleton & Lange, 1998.
4. Ellenhorn M, *Ellenhorn's Medical Toxicology, 2nd ed.*, Elsevier, 1997.
5. Gossel TA, Bricker JD, *Principles of Clinical Toxicology, 2nd ed.*, Raven Press, 1990.
6. Albertson TE, Dawson A, Latorre F, Hoffman RS, Hollander JE, Jaeger A, *et al.*, TOX-ACLS, *Ann Emerg Med* **37**:S78–S90, 2001.
7. Haddad LM, Winches ter JF, *Clinical Management of Poisoning and Drug Overdose, 3rd ed.*, Saunders, 1998.
8. Anantharaman V, *Proceedings of the First Singapore Symposium on Poisoning*, Ministry of Health, 1991.
9. Lee HS, *Pharmacological Principles in Drug Elimination: Proceedings of the First Singapore Symposium on Poisoning*, Ministry of Health, 1991.
10. Chee YC, *Antidotes and Their Application in Poisons Management: Proceedings of the First Singapore Symposium on Poisoning*, Ministry of Health, 1991.
11. Doull J, Klaassen C, *Casarett and Doull's Toxicology: The Basic Science of Poisons, 3rd ed.*, MacMillan, 1990.
12. The AACT/EAPCCT Position Statements: Gut Decontamination, *J Toxicol Clin Toxicol* **35**(7), 1997.

This page intentionally left blank

13

Emergency Airway Management

Evelyn Wong and V. Anantharaman

INTRODUCTION

Problems in airway management are common to many resuscitations. Patients undergoing resuscitation may experience life-threatening upper airway obstruction that require immediate intervention for survival. Medical practitioners need to be adequately proficient in a wide range of airway management skills so that a patient's airway may be secured, allowing adequate oxygenation and maintenance of ventilatory functions. Airway management is usually the first and most important step in basic and advanced patient resuscitation. All other resuscitative procedures to maintain circulation hemodynamics may come to nought if a patient's airway cannot be maintained. There are many causes of airway emergencies and even non-airway related problems may, over a course of time, affect the airway and breathing. A delay in management of the airway and breathing is often a cause of added morbidity to a very ill patient. Good airway management requires firstly, a thorough knowledge of upper airway anatomy and physiology; secondly, practised skills in

basic upper airway management; and thirdly, if these do not adequately serve the cause of the patient, confident use of advanced airway skills including endotracheal intubation and use of appropriate airway adjuncts.

ANATOMY AND PHYSIOLOGY OF THE UPPER RESPIRATORY SYSTEM

The respiratory system is divided into 2 regions:

1) an upper airway comprising the nose, mouth, nasopharynx, oropharynx, hypopharynx, larynx and trachea; and
2) a lower airway consisting of bronchi, bronchioles and alveoli.

Various mechanisms exist to prevent accidental obstruction of the pharynx and stop secretions and other foreign bodies from being aspirated into the lungs via the larynx and trachea. These include the cough and swallowing mechanisms, and the gag reflex.

CAUSES OF AIRWAY EMERGENCIES

In the assessment of the airway, 3 considerations are essential:

- whether the airway (A) is secure or is threatened;
- whether the breathing (B) or oxygenation is adequate; and
- whether there is likely to be a change (C) in A and/or B over a short period of time owing to progression of the disease or to a procedure that will require early airway and ventilatory management (see Table 1).

In an out-of-hospital situation availability of airway devices is limited. Knowledge of basic airway skills would then be essential and often life-saving. Such procedures may be performed by either laypersons or trained medical personnel. Where suctioning is available, this should be resorted to in attempts to clear fluids that contribute to upper airway obstruction.

Table 1 Some Causes of Airway Emergencies

Factors Affecting the Airway (A)	Factors Affecting Breathing/ Oxygenation (B)	Potential Changes Threatening the Airway (C)
Tongue Foreign bodies Secretions, blood Gastric contents Bony fragments Laryngeal edema Epiglottitis Acute laryngotracheobronchitis Quinsy Lugwig angina Tracheal stenosis Trauma to the larynx	Asthma, chronic obstructive pulmonary disease (COPD) Pneumonia and other pulmonary diseases Heart failure Pulmonary embolism Respiratory poisons e.g. cyanide and carbon monoxide poisoning Neuromuscular diseases e.g. Guillain–Barre syndrome, Myasthenia gravis Cervical cord lesions	Gastric lavage in a comatose patient Decompensating patient with progressive decline in conscious level or vital signs

SYMPTOMS AND SIGNS OF UPPER AIRWAY OBSTRUCTION

In a conscious patient, the principal symptom of upper airway obstruction or impending obstruction is inspiratory stridor. This is characterized by a harsh, raspy noise produced by the flow of air through a partially obstructed airway. Stridor occurring during both phases of respiration indicates a lesion in the trachea and stridor occurring only in the expiratory flow of respiration indicates lower airway obstruction. A stridorous person would be likely to be dysphonic and experiencing drooling of saliva.

MANAGEMENT OF THE UPPER AIRWAY

The initial action on contact with a patient with a possible upper airway obstruction is to determine if an open airway exists. In the event of spinal injury this must be done in conjunction with appropriate cervical spine stabilization. In most cases, manual airway control, ventilation and oxygenation should precede attempts to place an endotracheal tube. Such manual efforts allow an initial correction of hypoxic or hypercarbic states that may be present.

Airway management may require that one place his/her hands into a patient's mouth to perform a variety of maneuvers. Protective latex gloves should be worn. In addition, whenever there is a chance that bodily fluids will be splashed, protective eyewear, and possibly protective overalls may need to also be worn.

BASIC AIRWAY MANAGEMENT

The following measures described constitute basic maneuvers to open and maintain the airway:

Heimlich maneuver (Fig. 1)

This maneuver is used in the relief of upper airway obstruction, caused by a foreign body, in a still responsive adult. It involves the employment of

Fig. 1 Heimlich maneuver.

Fig. 2 The recovery position.

abdominal thrusts carried out with the rescuer's arms wrapped around the victim's waist. The rescuer places the thumbside of his fist against the victim's abdomen, in the midline slightly above the umbilicus. This fist is grasped by the other hand and the rescuer then presses his fist into the victim's abdomen with quick inward and upward thrusts repeatedly until the object is expelled from the airway or the victim becomes unresponsive.

For an unresponsive victim, standard Cardio-Pulmonary Resuscitation (CPR) as described in Chapter 11 is carried out.

Recovery Position (Fig. 2)

This position is recommended for patients who are unresponsive but able to breathe and having a pulse. In this position, the tongue falls forward and oral secretions are able to drain passively from the oral cavity outwards. Patients who need to be in this position for a long time would require turning to the opposite side at roughly 30-minute intervals.

Steps in converting an unresponsive patient from a supine to the recovery position:

1) The victim's arm nearest the rescuer should be straightened and tucked palm upwards under the victim's thigh.
2) The other arm is then brought across the victim's chest and the back of his palm placed against the cheek supported by the rescuer's one hand.
3) The rescuer's other hand should then bend the victim's far knee to an acute angle.

4) The victim is then rolled towards the rescuer until he has been turned more than a right angle and his near cheek is resting on the back of the hand that was earlier placed against it.
5) The victim's crossed-over knee is then allowed to rest on the ground bent at about 90 degrees. The victim's other leg is straightened out.
6) The victim's hand that was initially tucked under his thigh is now straightened out behind him with palm facing upwards.

Head-tilt/Chin-lift (Fig. 3)

This will lift the tongue and epiglottis up and out of their obstructing positions. The rescuer's hand is placed on the victim's forehead and the head tilted back by applying firm downward pressure with the rescuer's palm. With the other hand the chin is gripped, the thumb placed on the anterior part of the mandible and the index finger on the inferior portion of the mandible. The jaw is then lifted forwards to open the airway.

In the trauma victim, concern for integrity of the cervical spine dictates that a head-tilt not be carried out and only a gentle chin-lift be performed.

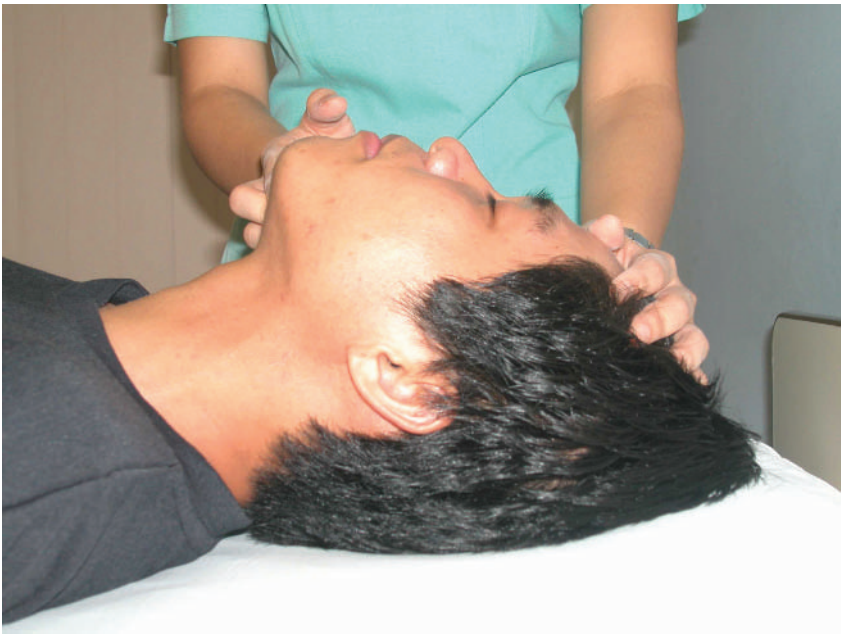


Fig. 3 The head-tilt, chin-lift technique.

Jaw Thrust (Fig. 4)

The rescuer is positioned at the top of the patient's head. The fingertips of each hand are placed on the angles of the patient's lower jaw. The jaw is firmly displaced forward, while gently tilting the patient's head backward.

Retraction of the lower lip with the thumbs to open the mouth will result in what is called the "triple airway maneuver".

Modified Jaw Thrust (Fig. 5)

Sometimes, in trauma, there is concern that excessive forward, backward or sideway movement of the head or the neck may occur with the jaw thrust. A modified jaw thrust maneuver may be performed instead.

The rescuer places both hands parallel to each other on both sides of the patient's neck. The fingers then slide just under the trapezius on the side of the neck so that the neck can be easily supported. Both thumbs are fully abducted till they push both angles of the patient's mandible forward, lifting the tongue from the back of the pharynx.

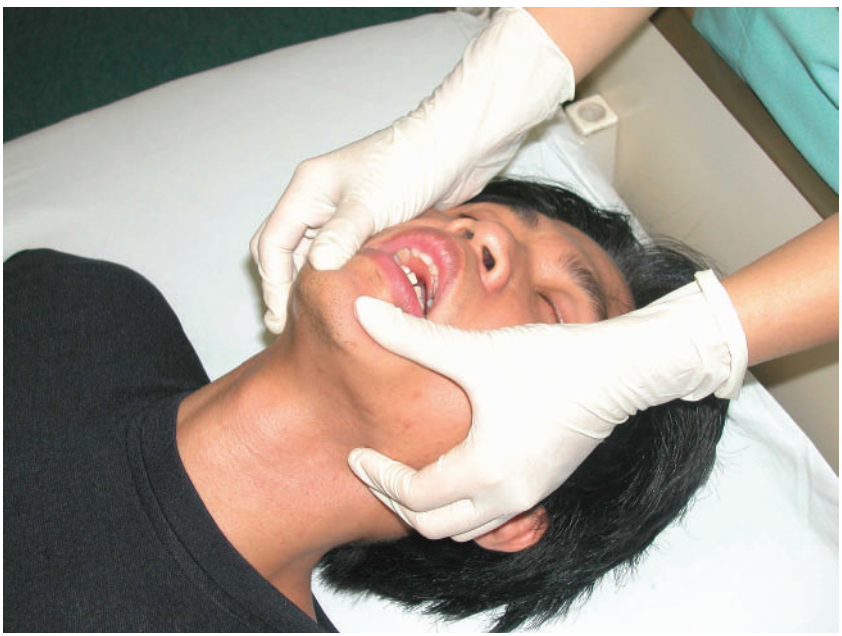


Fig. 4 The jaw thrust technique.

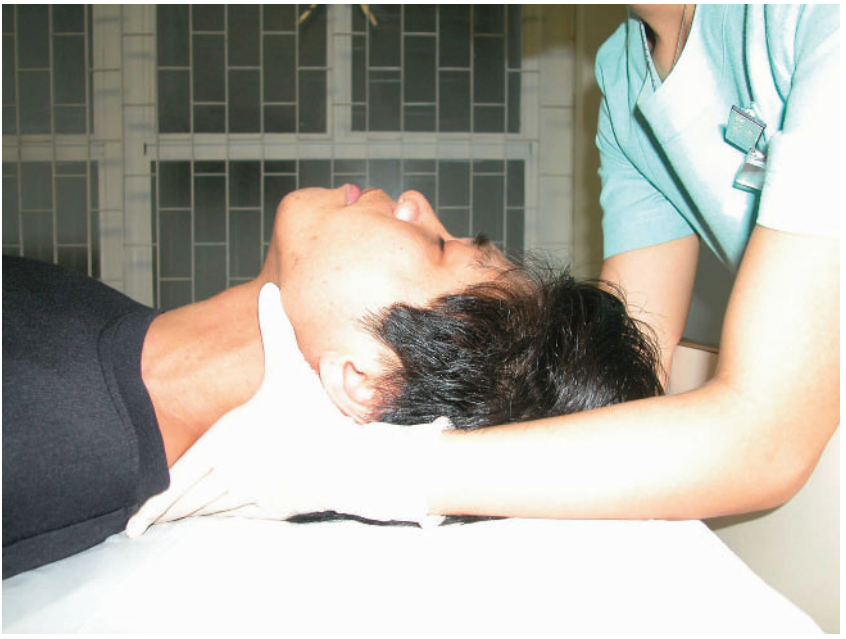


Fig. 5 The modified jaw thrust.

This maneuver minimizes side-to-side and anterior-posterior movement of the neck while maintaining an open airway.

Jaw Lift (Fig. 6)

This technique should be employed carefully as it involves placing fingers into the patient's mouth. The jaw is grasped with a gloved hand with the thumb placed on the lower incisor and the index finger on the lower mandible. The lower jaw is then pulled anteriorly to open the airway.

One or more of the above techniques may be tried before airway adjuncts are employed. Ideally, at least 2 resuscitation team members should be involved, one to perform the maneuver, and the other to provide suction or ventilatory assistance.

OROPHARYNGEAL/NASOPHARYNGEAL AIRWAYS

Both these airways are designed to be interposed between the tongue anteriorly and the back of the pharynx posteriorly.

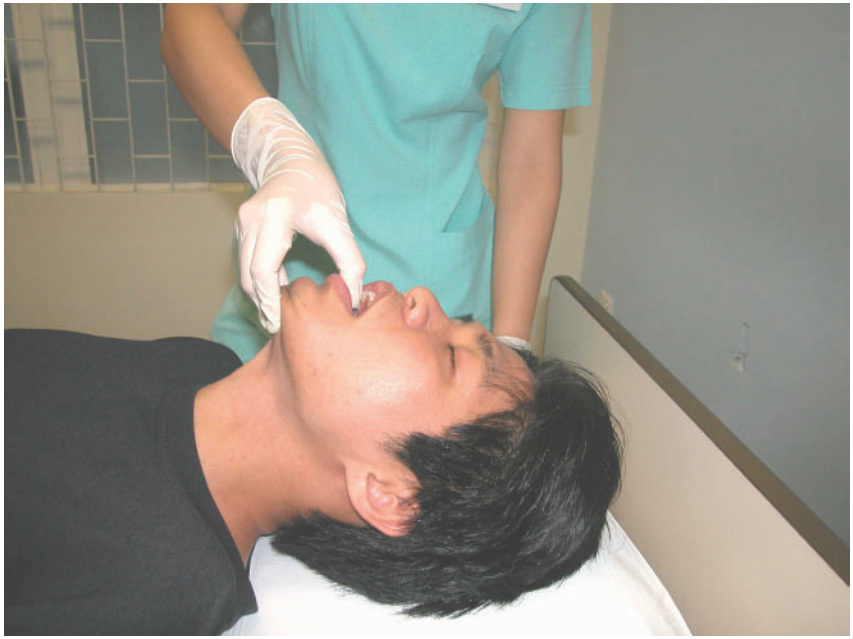


Fig. 6 The jaw lift.

The Oropharyngeal Airway (Fig. 7)

Also referred to as the "Guedel" airway, it is a semicircular device made up of plastic or rubber and designed to fit the curvature of the palate. It is useful when managing an unconscious patient who is breathing spontaneously or who needs to be ventilated by a bag-mask device or other ventilatory apparatus. Suctioning of the pharynx is easy as a large catheter can pass on either side. It may also be used as a block to prevent a patient from biting down on an endotracheal tube (ETT). This airway device does not isolate the trachea and is easily dislodged. It cannot be inserted if the teeth are clenched or if the gag reflex is present. Oropharyngeal airways come in a variety of sizes (#0 for infants and #4 for adults). The required size is measured by determining the distance from the angle of the jaw to the angle of the mouth. The airway may be inserted by pointing the tip of the airway towards the palate and sliding it down along the curvature of the palate as far down as possible, then gently turning the airway around 180 degrees (Fig. 8). Alternatively, it may be inserted directly using a tongue depressor or a laryngoscopic blade. Improper insertion of the

Fig. 7 Oropharyngeal airways.

oropharyngeal airway may cause the tongue to be pushed posteriorly and obstruct the airway.

The Nasopharyngeal Airway (Fig. 9)

This is an uncuffed tube made of soft rubber or plastic ranging from 17–20 cm in length and 20–36 Fr in diameter. The proximal end is funnel-shaped which prevents it from slipping completely into the patient's nose. The distal end is bevel-shaped to facilitate passage. The tube follows the normal curvature of the nasopharynx and extends from the external nostrils to the posterior pharnyx just below the base of the tongue. The nasopharyngeal airway is used in situations where an oropharyngeal airway is not advised, e.g. in the presence of clenched teeth, oral injury or gag reflex. Like the oropharyngeal airway, the nasopharyngeal airway does not isolate the trachea. Complications of its use include epistaxis and pressure necrosis during prolonged usage. Relative contraindications to its use include nasal injury and base-of-skull

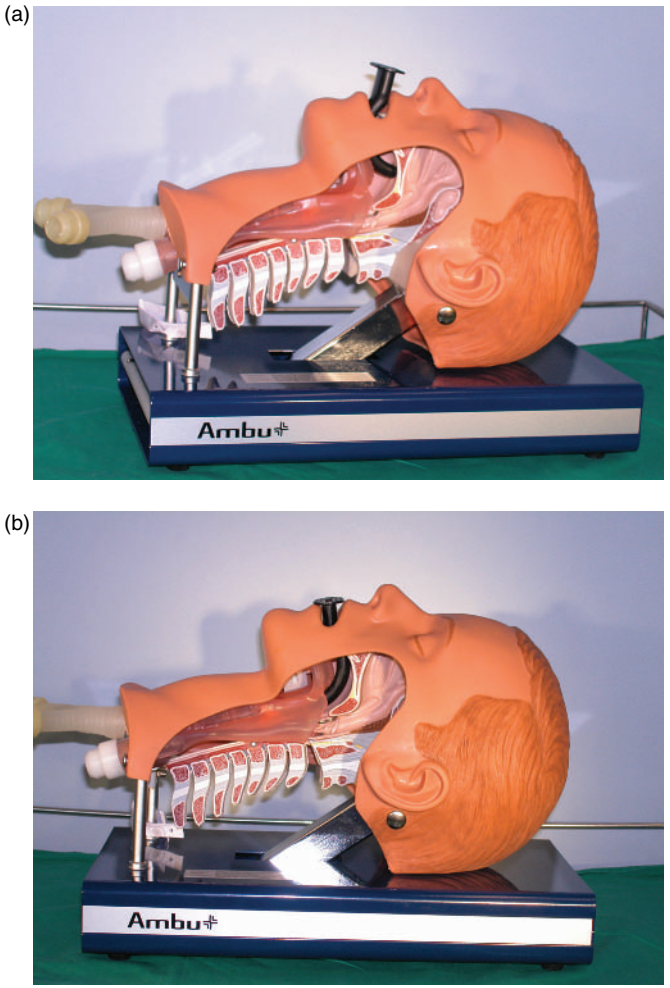


Fig. 8 The oropharyngeal airway insertion technique.

fractures. An adequate sized nasopharyngeal tube is slightly longer than the distance from the patient's nose to his earlobe. Its calibre should be one size less than if an endotracheal tube is to be used. The nasopharyngeal airway is lubricated externally and inserted with the bevel towards the septum of the nose. It should be inserted in the direction similar to the plane of the hard palate.

Once the airway has been secured, a variety of ventilatory techniques using various adjuncts may be employed to deliver oxygen-rich air to the

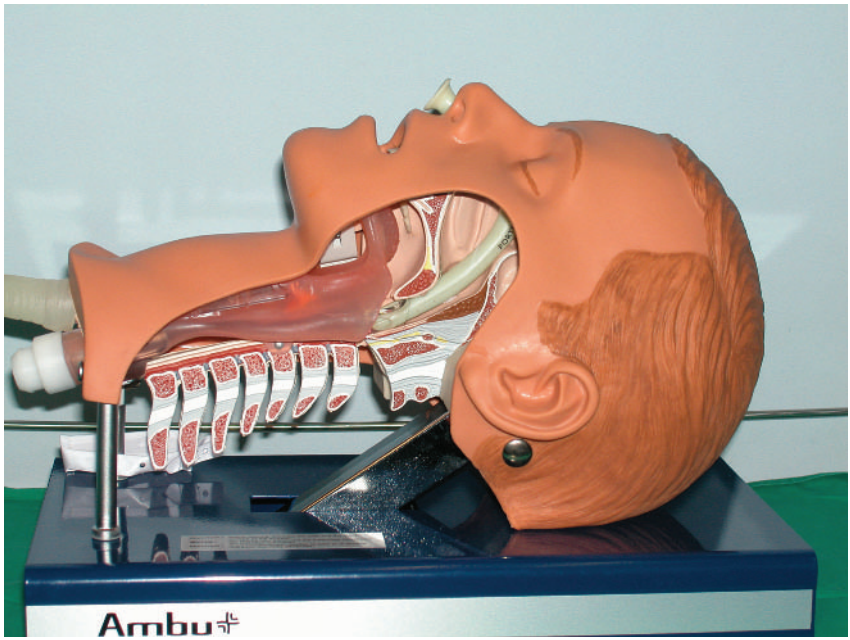


Fig. 9 Nasopharyngeal airway in position.

alveoli. Three basic ventilatory techniques are described. All these are commonly used in the pre-hospital or in-hospital environment.

Oropharyngeal Suctioning

As mentioned earlier, the upper airway may be obstructed by a variety of fluids such as vomitus, blood and secretions. These secretions may be removed by oropharyngeal suctioning. Devices that are available for suctioning may be either portable (hand-, foot-, oxygen- or battery-powered) or stationary units (e.g. vacuum-powered). In an emergency setting, vacuum units should generate at least 300 mmHg vacuum when the distal end is occluded.

The choice of suction catheters should also be appropriate. In the resuscitation situation, large-bore, rigid, suction catheters, such as the Yankauer type which has a large tip and multiple holes at its distal end, are able to remove large amounts of secretions rapidly under direct vision, especially prior to insertion of definitive airway adjunct devices. Following endotracheal intubation, or in attempting to clear naso-pharyngeal secretions, long,

flexible, whistle-tip suction catheters may be utilized. These may be inserted through the nares, into the oropharynx or nasopharynx, through a nasopharyngeal or oropharyngeal airway, or through an ETT, and may even be able to reach portions of the lower respiratory tract.

Airway suctioning temporarily reduces a patient's access to oxygen. Therefore, each attempt at suctioning should not exceed 10 seconds. Patients should be hyperventilated with 100% oxygen prior to and following each suctioning effort. Suctioning should not be applied during the process of inserting a suction catheter. Only when the suction catheter is properly positioned should suction be applied, and subsequently the catheter should be withdrawn.

Occasionally, secretions may be extremely thick and viscous, and may obstruct flow through the suction catheter. This may be minimized by suctioning water through the tubing in between suctioning attempts. This dilutes the secretions and facilitates flow to the suction cannister.

Mouth-to-mouth Ventilation

Mouth-to-mouth or even mouth-to-nose breathing can, over short time periods, provide effective ventilatory support to a patient. No adjuncts are required. The main limitation of this technique is the provision of limited oxygen as expired air contains only 17% oxygen. In addition, the victims often have large amounts of copious secretions and vomitus, and may be suffering from infectious diseases, which makes this technique unattractive to the rescuer. To date, however, there has been no documented instance of infection having been transmitted from one person to another as a result of mouth-to-mouth ventilation. Barrier devices, such as face shields, are available.

Mouth-to-mask Ventilation

This makes use of a pocket mask, which is usually made of clear plastic. It sits on the patient's face and covers the mouth and nose. The mask prevents contact between the rescuer's mouth and the victim's mouth/nose, thus reducing the fear and risk of infection. Such masks may be disposable or reusable. Most have a one-way valve that prevents the patient's expired air, or other oral secretions, from coming into contact with the rescuer. Many also have an oxygen inlet that allows the administration of

supplemental oxygen. A combination of mouth-to-mask ventilation and supplemental oxygen at 10 litres per minute would result in delivery of approximately 50% oxygen to the patient. An oropharyngeal airway may also be used in conjunction with mask-to-mouth breathing.

Bag-valve-mask Ventilation

A bag-valve-mask unit, as shown in Fig. 10 below, comes in various sizes for neonates, pediatric patients and adults.

Without supplemental oxygen this unit will deliver 21% oxygen. With an oxygen source attached via the air inlet and flow at 12 litres/min, up to 60% oxygen may be delivered to the victim. Addition of a reservoir bag will increase the oxygen delivered to 90–95%. The addition of a demand valve device to the air inlet will allow delivery of up to 100% oxygen. The bag-valve-mask unit may be used for direct ventilation or via an ETT.

ADVANCED AIRWAY MANAGEMENT

Advanced airway management involves securing and placement of a definitive airway.

Fig. 10 A bag-valve-mask unit.

Table 2 Methods of Placing a Definitive Airway

1)	Direct orotracheal intubation
2)	Nasotracheal intubation
3)	Orotracheal intubation via an intubating laryngeal mask
4)	Orotracheal intubation via a fiberoptic device
5)	Retrograde intubation
6)	Cricothyrotomy
7)	Tracheostomy

Table 3 Advanced Airway Adjuncts

1)	Laryngeal mask airway
2)	Esophageal-tracheal combitube
3)	Needle cricothyrotomy

If there is difficulty placing a definitive airway, then adjuncts may be used to isolate the airway and provide oxygenation while awaiting the arrival of more expert help.

Direct Orotracheal Intubation (OTI)

Direct orotracheal intubation using laryngoscopy is the most common method of placing a definitive airway. OTI may be achieved without the use of drugs, by rapid sequence intubation, or by sedation and topical anesthesia.

1) In a patient with respiratory arrest, near cardiovascular collapse or without a gag reflex, the ETT is placed immediately without the use of any drugs.
2) Patients who require OTI but who may be still responsive or have a gag should be given sedation (or induction agent) and a muscle relaxant (paralytic agent) to facilitate placement of the endotracheal tube. The emergency intubation of a patient, who is presumed to have a full stomach, using a strong sedating or induction agent and a neuromuscular blocking agent is called rapid sequence intubation (RSI).
3) Patients with potentially difficult airways should not be paralyzed pharmacologically. These include patients where bag-mask ventilation (Table 4) and/or endotracheal intubation (Table 5) are difficult. They should receive sedation or induction and topical anesthesia of the oropharynx to diminish the gag reflex. This can be administered

Table 4 Features Associated with Potentially Difficult Bag-mask Ventilation (mnemonic: BONES)

B = Beard
O = Obesity
N = No teeth
E = Elderly
S = Smashed face (maxillofacial trauma)

Table 5 Features Associated with Potentially Difficult Intubations

Obesity
Small mouth (admits less than 3 fingers)
Receding chin (less than 3 fingers space between the mentum and the hyoid)
High larynx (less than 2 fingers space between the floor of the mouth and the superior aspect of the thyroid cartilage)
Obstruction of the oropharynx
Neck immobility
Inability to visualize the whole uvula and pharynx (Mallampati 3 and 4, reference 4)
Inability to visualize part or whole of the glottis on laryngoscopy (Cormack–Lehane 3 and 4, reference 4)

Table 6 Equipment Required for Endotracheal Intubation

Oxygen source
Yankauer suctioning device
Oral and nasopharyngeal airways
Non-rebreather masks
Bag-valve-mask with oxygen reservoir
Endotracheal tubes (ETT) of different sizes (6–8 for female, 7–9 for male)
Stylets
Water-soluble lubricant
Laryngoscope
Curved MacIntosh blades sizes 3 and 4
Straight Miller blades sizes 2 and 3
McGill's forceps
10 cc syringe
Stethoscope
Endtidal CO_2 detector or bulb/syringe esophageal detection device
Tape
Gauze
Head-stabilizer/small pillow
Rescue airway devices, e.g. bougie, laryngeal mask, combitube

using a lignocaine spray or by nebulization of 5 ml of 4% lignocaine for 2–3 minutes.

Adequate preparation for direct OTI is essential as this reduces the stress in performing this procedure.

Steps to performing direct OTI in a patient in cardiac arrest or without a gag reflex:

1) The patient is placed in a supine position while the intubator stands at the head of the patient. Visualization of the vocal cords is essential. In a non-trauma patient or one without any cervical spine pathology, this is made easier by aligning the oral, pharyngeal and laryngeal axes using a head-stabilizer or a small pillow at the occiput (see Fig. 11), placing the head in a morning-sniff position.
2) The patient is ventilated using a bag-mask device with 100% oxygen while external cardiac massage is performed if indicated.
3) A second assistant prepares the equipment. A stylet is lubricated and introduced into the ETT. The ETT is lubricated at the distal end. The balloon is tested.
4) An assistant performs cricoid pressure to prevent regurgitation of gastric contents risking aspiration.

Fig. 11 Morning-sniff position.

5) A laryngoscope, attached to an adequate sized curved blade, is held with the left hand. It is inserted at the right angle of the mouth down to the base of the tongue just anterior to the epiglottitis (the vallecula), taking care not to traumatize the lips, teeth, tongue and pharynx. The tongue is swept to the left and the jaw is lifted at an angle of 45 degrees in the upward and forward direction. This movement lifts the epiglottis and exposes the glottic opening.
6) Any oral contents are suctioned out under direct vision using the Yankauer suction device or removed using McGill's forceps.
7) The ETT is then passed through the glottic opening. The balloon is inflated with 8–10 cc of air. The endtidal CO_2 ($ETCO_2$) device followed by the bag-valve-oxygen device are then attached to the proximal end of the ETT.
8) The bag is compressed and placement of the tube is confirmed by the following steps: (i) observing air-condensation on the tube; (ii) chest expansion; (iii) auscultation of the chest; (iv) $ETCO_2$ detector or esophageal detection device (ODD); (v) pulse oximetry; and (vi) chest radiography. Auscultation is performed at the right and left apices, right and left bases and epigastrium. The presence of breath sounds mainly in the right chest indicates that the tube's position is in the right main bronchus while breath sounds in the epigastrium indicates that the tube is in the esophagus. The $ETCO_2$ detector confirms that the tube is in place if it is continuously positive for carbon dioxide during expiration. Alternatively an esophageal detection device may be used. Resistance on pulling the plunger indicates that the tube is in the esophagus.
9) The ETT is then secured using tape.
10) The patient is then connected to a mechanical ventilator or bagged with 100% oxygen.

Steps in performing RSI for direct OTI:

1) Preparation for intubation is made.
2) The patient is preoxygenated with 100% oxygen using a non-rebreather mask for 3–5 minutes.
3) An intravenous line is established. Cardiac and vital signs monitors are connected to the patient.
4) About 3 minutes before intubation, pre-treatment medication may be given to the patient if indicated.

Table 7 Initial Ventilator Settings[3]

Ventilate with 100% oxygen
Tidal volume of 10–15 ml/kg
Respiratory rate of 12–15 breaths per minute
Inspiratory time of 2 seconds
In patients with severe COPD, lower respiratory rate to 6–8 per minute to prevent auto-PEEP

5) After 3 minutes of preoxygenation, an induction agent, e.g. etomidate, is given, followed by a neuromuscular blocking agent, e.g. succinylcholine, in rapid sequence.
6) After 45 seconds to 1 minute, the jaw is tested for relaxation and the patient is intubated.
7) Tube placement is confirmed. The tube is secured and the patient is connected to the ventilator.
8) Further long-acting sedation and neuromuscular blockade may be administered to ensure patient comfort and synchronous breathing with the mechanical ventilator.

Blind Nasotracheal Intubation

This form of intubation is rarely practised nowadays and its only indication is for a spontaneously breathing patient who has a difficult airway not due to pathological abnormalities, e.g. strictures or mechanical obstruction, and not a progressively deteriorating airway, e.g. laryngeal edema or epiglottitis. The difficult airway is likely to be static and physiological, e.g. obesity, short neck, receding chin. The patient may be intubated in a sitting or supine position. The procedure should be explained to the patient to ensure co-operation. A nasal decongestant is applied. The pharynx is sprayed with local anesthetic, e.g. lignocaine. The largest ETT that fits the most patent nostril is selected. This is usually one size smaller than the one used for orotracheal intubation. It is passed through the nostril to the back of the throat and just above the glottic opening. The tube is manipulated until the breath sounds are heard the loudest at the proximal end and the tube is inserted during inspiration. When a cough is heard and the patient loses ability to speak, the tube is in the trachea. Tube placement can also be confirmed with an $ETCO_2$ detector. This procedure should not be attempted if intubation has to be performed immediately. It should be abandoned if the patient is combative or if he deteriorates clinically.

Intubating Laryngeal Mask (ILMA)

This device is usually not readily available in the emergency setting. Practice and skills retention are essential. Hence it is usually used by anesthetists in a controlled environment. The ILMA may be used to bag a patient or an ETT may be inserted through it into the trachea. The ILMA may be used for primary intubation in the operating room or as a rescue airway in a difficult intubation.

Fiberoptic Intubation

This device is also not readily available in most emergency settings. Again practice and skills retention is an important issue. This procedure cannot be performed very quickly. The fiberoptic scope is passed into the

Table 8 Drugs Commonly Used in RSI (Readers are Encouraged to Look Up Full Drug Dosages, Indications, Contraindications and Side-effects)

Drug	Indications, Contraindications, Side-effects	Dosage (IV)
Lignocaine	Pre-treatment for raised intracranial pressure, reactive airways	1–1.5 mg/kg
Fentanyl	Pre-treatment for raised intracranial pressure, cardiac disease	2 μg/kg
Atropine	Pre-treatment in children before a second dose of succinylcholine in adults/children; prevents bradycardia	0.02 mg/kg
Etomidate	Induction, most hemodynamically stable	0.3 mg/kg
Midazolam	Induction/sedation. May cause hypotension. Variable response — no fixed dosage; titrate according to effect	0.05–0.2 mg/kg
Ketamine	Induction in asthma, shock, avoid in raised intracranial pressure or ischemic heart disease	1–2 mg/kg
Thiopentone	Induction, especially raised intracranial pressure, a/w hypotension, avoid in asthma, porphyria; histamine-releasing	3–4 mg/kg
Succinylcholine	Short-acting neuromuscular blocker. Avoid in patients with hyperkalemic risk, personal/family history of malignant hyperthermia	2 mg/kg
Rocuronium	Intermediate-acting neuromuscular blocker, for hyperkalemic patients	1 mg/kg
Pancuronium	Long-acting neuromuscular blocker, post-intubation paralysis, a/w tachycardia	0.08 mg/kg
Vecuronium	Intermediate-acting neuromuscular blocker, post-intubation paralysis, cardiac stable	0.1 mg/kg

trachea by direct visualization and the ETT is passed over the scope. This device is also used in cases of difficult intubation.

Retrograde Intubation

This device is used in difficult intubations. Under sterile conditions and local anesthesia, the cricothyroid membrane is punctured and a guide-wire is inserted through and directed superiorly towards the mouth. The guide-wire is then retrieved through the mouth and the ETT is passed over it into the trachea. Once the ETT abuts the cricoid membrane, the wire is retrieved through the proximal end of the ETT while the ETT is continued through the trachea. This procedure is rarely performed in an emergency situation.

Cricothyrotomy

In the case of a failed orotracheal intubation without any rescue devices available or an airway obstruction proximal to the glottis, cricothyrotomy can be performed. Practice using an animal model is recommended. Commercial cricothyrotomy kits are available. The neck is extended, if possible, to stretch the skin and facilitate localization of the cricothyroid membrane. Under sterile conditions and local anesthesia, an approximately 5 cm midline longitudinal skin incision is made over the region of the thyroid cartilage and proximal trachea. The cricothyroid membrane is located between the thyroid and cricoid cartilages using the left index finger with the left middle finger and thumb immobilizing the trachea on either side. A size 11 scalpel is used to incise the membrane transversely. A tracheal hook is used to lift the thyroid cartilage towards the skin. An artery forceps is used to dilate the incision and a cricothyrotomy tube is placed through it. The tube is secured and the patient is then ventilated through it. Tube placement is confirmed by an $ETCO_2$ detector. Alternatives to using a specialized cricothyrotomy tube include size 6 ETTs or tracheostomy tubes. Complications include bleeding, creation of a false passage and damage to surrounding structures.

Tracheostomy

This is usually performed by surgeons in cases of airway obstruction which are distal to the cricothyroid membrane. The complication rate is

high in inexperienced hands and these include delayed tracheal stenosis and bleeding.

ADVANCED AIRWAY ADJUNCTS

Laryngeal Mask Airway (LMA)

The LMA was created by British anesthetist, Dr Archibald Brain. It is a device consisting of an inflatable elliptical silicon mask and a tube. The mask has a laryngeal side where the distal opening of the LMA is, and a pharyngeal side. The inflatable mask fits around the larynx and air or other gases are passed through the tube directly into the larynx and trachea. It is not a definitive airway device as the seal around the larynx is not tight and air may leak out around the mask while gastric and oral contents may leak in. However, it is more effective than the bag-mask device in providing a more direct route for gases to pass through and it decreases the risk of aspiration. The LMA, however, cannot be used in patients who have a gag response. Hence they must be obtunded or deeply sedated or pharmacologically paralyzed.

The LMA can be used in the pre-hospital setting or in the hospital. It can replace the bag-mask device, buying time for a more definitive airway placement. This is particularly useful since training and skills retention for the LMA is less demanding than that for endotracheal intubation. Pre-hospital personnel can use the LMA at an early stage to give effective ventilation to very ill patients as well as minimize the risk of gastric distension that often accompanies bag-mask ventilation. Doctors in the hospital can use the LMA as a temporizing measure while awaiting more expert help in the event that endotracheal intubation is difficult. Complications of the LMA include occasional aspiration of gastric contents and inadequate ventilation in patients with low pulmonary compliance.

Steps to placement of a LMA:

1) The LMA is prepared by deflating it with the mask on a flat surface until the mask is quite flat but not crinkled. Water-soluble lubricant is applied on the pharyngeal side of the mask.
2) The practitioner stands at the head of the patient, holding the LMA like a pencil with the right hand.
3) The patient's mouth is opened with the left hand, holding the tongue and the inside of the jaw with the left thumb.

4) The laryngeal mask is introduced into the mouth with the pharyngeal part pressed against the hard palate and the axis of the tube in the midline.
5) The tube is inserted between the hard palate and tongue as deeply as it can go until resistance is felt.
6) Both hands are released and the balloon of the LMA is inflated with about 30–40 ml of air. A slight upward shift of the tube is to be expected.
7) Tube placement is confirmed by observing chest expansion, auscultation and $ETCO_2$ detection.
8) The tube is then secured with tape.

Esophageal-tracheal Combitube (OTC)

The Combitube is a double lumen-double balloon tube device that is introduced into the esophagus. One balloon occludes the esophagus hence preventing gastric distension and regurgitation while the other occludes the pharynx, preventing air-leak and aspiration of oral contents. The glottic opening is isolated between the 2 balloons. Air is passed through one of the tubes through side holes between the esophageal balloon and pharyngeal balloon. The air then passes through the glottis into the trachea. Tube placement is confirmed by auscultation and $ETCO_2$ detection. In the rare event that the OTC is inserted into the trachea, the patient can be ventilated through the other lumen.

Like the LMA, the Combitube is used as a temporizing measure in pre-hospital airway management and in cases of difficult endotracheal intubation. Eligible patients must not have a responsive gag reflex. It is also contraindicated in patients with esophageal disease. Complications of the Combitube include local trauma and bleeding and ventilation of the wrong lumen.

In Hong Kong, the Combitube is used as a second-line pre-hospital airway device when the LMA is inadequate. Other places where the Combitube has been used as a pre-hospital airway device include Japan, Canada, UK and the USA.

Needle Cricothyrotomy

This life-saving temporizing technique is used in cases of upper airway obstruction, e.g. epiglottitis or laryngeal edema, when OTI cannot be

performed or has failed, and if the practitioner does not know how to perform a cricothyrotomy. A size 14 G needle is inserted through the cricothyroid membrane. This is connected to a high flow wall oxygen source. The oxygen is pulsed at an inspiratory to expiratory ratio of 1:3 seconds. The patient can be oxygenated for between 30–45 minutes. Prior training on a mannequin is strongly advised.

Non-invasive Ventilatory Support (NIVS)

Under special circumstances, instead of mechanical ventilation via endotracheal intubation (invasive ventilation), an acutely dyspneic patient with a rapidly reversible condition may be ventilated using a nasal or a facemask. The typical scenario would be a patient with chronic obstructive pulmonary disease, status asthmaticus or congestive heart disease who is still conscious and co-operative but has inadequate ventilation. Apnea and inability to protect the airway are contraindications to NIVS. The most common mode of NIVS is bilevel positive airway pressure (BL-PAP or Bi-PAP).

Bi-PAP combines inspiratory pressure support ventilation (PSV) and continuous positive airway pressure (CPAP). This is achieved by adjusting the inspiratory positive airway pressure (IPAP) and expiratory positive airway pressure (EPAP) on the Bi-PAP ventilator. The IPAP must be set higher than the EPAP and the difference is the pressure support provided. The effect of EPAP is analogous to the CPAP or positive end expiratory pressure (PEEP).

Patients on NIVS require close vital signs and cardiac monitoring. They should not be given medication that may depress respiration.

Table 9 Recommended Initial Settings for Bi-PAP[4]

IPAP of 8 cm H_2O
EPAP of 3 cm H_2O
4 L oxygen
IPAP should be raised in 2 cm H_2O steps
Keep IPAP: EPAP ratio at 2.5:1
EPAP should be less than 8–10 cm H_2O

REFERENCES

1. Lim SH, Basic Cardiac Life Support, *Singapore Medical Journal* (Supp) **1**:2–3, 2001.
2. Adult Basic Life Support Guidelines 2000 for cardiopulmonary resuscitation and emergency cardiovascular care, International consensus on science, *Circulation* **102**: 22–59, 2000.
3. Adjuncts for oxygenation, ventilation and airway control. Guidelines 2000 for cardiopulmonary resuscitation and emergency cardiovascular care, International consensus on science, *Circulation* **102**:95–104, 2000.
4. Walls RM (ed.), *Manual of Emergency Airway Management*, Philadelphia, Lippincott, Williams and Wilkins, 2000.
5. Benumof JL (ed.), *Airway Management and Clinical Practice*, St Louis, Mosby, 1996.
6. Berry AM, Brimacombe JR, Verghese C, The laryngeal mask in emergency medicine, neonatal resuscitation and intensive care medicine, *International Anesthesiology Clinics* **36**(2):91–109, 1998.
7. Brimacombe JR, Problems with the laryngeal mask airway: prevention and management, *International Anesthesiology Clinics* **36**(2):139–154, 1998.
8. Airway and ventilatory management, *Advanced Trauma Life Support. American College of Surgeons Committee on Trauma: Instructor Course Manual* 61–96, 1997.

This page intentionally left blank

14

Acute Medicine in the Pre-hospital Environment

V. Anantharaman and Fatimah Lateef

Diseases in Their Early Stages

Chronologically, any disease commences at some point, progresses and terminates, either with treatment or without, with a favorable or unfavorable outcome. The beginning is usually not clear, but as the disease progresses along its natural course, symptoms begin to manifest. In the case of diseases that manifest *de novo* acute symptoms may be distressing, e.g. chest pain, abdominal pain, difficulty in breathing, vomiting, diarrhea or disordered function of one or more body parts. Since diseases in early phases may not manifest with all the gamut of symptoms and signs, diagnosis may be a challenge. Patients with early symptoms of disease often brush these off as minor insignificant ailments. It is only if the symptoms recur with increasing frequency or severity, persist, or progressively worsen, that a patient realizes that there may be significant illness requiring care. Only when this trigger stage is reached will the patient and relatives access the Emergency Medical Services (EMS).

All illnesses are time-sensitive, *viz*. the earlier they can be diagnosed, the greater the potential for treatment to resolve the illness. There are

some illnesses, though, that have a tendency to progress rapidly to the effect that if the patient were to delay presentation to an appropriate medical facility, the potential for severe morbidity and death is great. Some such illnesses are:

1) *Acute coronary syndromes, such as myocardial infarction and unstable angina pectoris.* These usually present with chest pain or discomfort. If the duration of the initial symptoms is short, or not accompanied by the classical constellation of precordial pain with shortness of breath, diaphoresis, and radiation of pain down the left arm, or if they present with discomfort in the radiation areas, instead of the precordium, these may be attributed to tiredness, gastritis or anxiety. Failure to seek immediate medical care may result in heart failure, irreversible damage to the myocardium, or ventricular arrhythmia resulting in sudden death. Delayed presentation to the EMS will render patients ineligible for the variety of emergency coronary reperfusion procedures that are available today.
2) *Acute bronchial asthma/acute exacerbation of chronic obstructive airway disease.* Patients usually present with breathlessness, which may be mild at the early stages. Ignoring these early symptoms may be dangerous, and failure to seek medical treatment may mean a presentation to the hospital with a more severe attack, and even perhaps, in extremis. The initial presentation of a mild cough may be mistaken for an upper respiratory infection, and one would have to consider more potentially serious conditions when attending to such patients.
3) *Acute cerebrovascular accidents* generally tend to begin with one or more of a constellation of neurological symptoms consisting of facial weakness, slurring of speech and unilateral muscular weakness. These symptoms are often attributed, by the lay public, to general tiredness, and only persistence of these symptoms for a number of hours would lead patients or relatives to suspect that a more serious pathology was present. It is nowadays believed that education of the public and pre-hospital medical personnel[1] of the constellation of such symptoms would lead them to seek medical assistance earlier and therefore become available to time-sensitive treatments, such as use of intravenous thrombolytic and neuroprotective agents.

4) *Pneumonias and other major infectious diseases*. These tend to present with acute fever, often with symptoms referable to individual body systems, but sometimes with generalized symptoms. Diagnosis may not be possible in the early stages of the illness and the patient, in the absence of clear evidence of major organ involvement, may attribute such fever to mild viral illnesses. It is usually the persistence of symptoms over a period, or the failure of the symptoms to resolve with initial treatment provided by primary health practitioners that drive these patients to seek care at hospital emergency departments.
5) *Acute poisoning* is an emergency. Often, regardless of the circumstances or route of poisoning, symptoms tend to be mild in the initial few hours. This is, however, the crucial period when the toxic agent is gradually being absorbed into the circulation, and simple interventions would be able to have maximum effect in preventing or minimizing the absorption of the poison. The mild symptoms of nausea or occasional vomiting may lull patients into thinking that the effects of the poison on the body are minimal. The lack of severe symptoms in the initial phases of many poisoned victims is the most common reason for delays in seeking emergency medical care in such instances.
6) *Acute diabetic emergencies*, such as diabetic ketoacidosis, hypoglycemia and hyperosmolar non-ketotic diabetes mellitus, seldom present in classical fashion to the hospital because the early symptoms of persistent polyuria and polydypsia, early fatiguability, shortness of breath and acute confusional states are not always recognized early enough as potentially catastrophic. The habit of checking blood sugar during periods of ill-health is not yet so widely practised as to lead many of these patients to seek earlier medical attention. Even then, at an early stage, medical practitioners may not identify the potentially developing condition. The early recognition and management of such conditions pose a challenge to physicians.

Because all these are eminently treatable diseases, and earlier treatment results in stabilization of the patient, faster resolution of symptoms, shorter duration of hospitalization, and a better functional state of the involved body system, it is important that the public be aware of the severity of symptoms and initiate attempts to access the EMS system.

An Environment with Little Medical Support

Facilities for the identification and management of the bulk of medical emergencies are largely lacking in the pre-hospital environment. Diagnostic aids are generally unavailable, except for home glucometers for rapid blood sugar measurements, mini peak flow meters for asthma and urine tests for pregnancy. We are gradually moving toward times when more forms of early diagnostics may become available in the home, such as hand-held electrocardiograph meters, which transmit electrocardiographic waveforms to remote monitoring and alarm centers via telephone cables and provide the patient with possible instant diagnoses. Home telemedicine is still many years from becoming a commonplace reality.[2–4] Family practice clinics and primary health polyclinics are mainly able to offer very basic diagnostic and therapeutic services. Common acute medical ailments may be managed by such clinics. Chronic illnesses, once stable, may be managed by practitioners working in such clinics as well. However, major and more serious problems still require hospitals for diagnosis and definitive management. Since diagnosis is always urgent, it is understandable that many people tend to move toward Emergency Departments of hospitals for the initial diagnostic efforts. This is compounded by the fact that many do not have a regular dedicated family practitioner to service them. There also seems to be a tremendous degree of confidence by the public in the ability of hospitals to come to the bottom of the medical problem and initiate steps that will lead to a patient's recovery. In this kind of environment, it would be understandable for medical institutions to provide a range of early diagnostic facilities for the initial diagnosis of the wide range of medical emergencies that afflict the community at large. Therefore, emergency departments play a vital role in acting as gatekeepers to determine the nature of the problem, sieving off those who can be managed as outpatients, identifying those who require urgent interventions and initiating those interventions to promote immediate improvement that may allow them to continue being managed in the ambulatory environment, and determining those who will definitely need to be managed as inpatients. For such inpatients, it is imperative that emergency treatments be initiated to resuscitate and stabilize them and allow the further expertise of the various clinical services of the hospital to guide them towards full recovery over the subsequent days.

IMPORTANT ASPECTS OF HISTORY IN ACUTE PRE-HOSPITAL MEDICINE

History in the pre-hospital environment is obtained from the following sources:

1) the patient;
2) relatives — a large number of whom may be present in the home;
3) office colleagues — who may not be always available at the hospital to provide a firsthand account of what happened on-site; and/or
4) passers-by, who may have seen events happening but are often neither willing nor available to hospital staff.

Recognition of potentially serious medical problems by patients is frequently delayed owing to the patient's ignorance or denial of the basic symptoms of disease. The usual heart attack victim often waits an average of 3 hours before seeking help.[5] Even bystander recognition of major illness and their intervention to expedite care to the sick in the pre-hospital environment is uncommon, and only those few communities that have been able to show a high degree of involvement, as in use of bystander CPR, have been able to show improvement in morbidity and mortality.[6,7]

In such a situation, it is important that the paramedic documents the information he gathers from various sources. This information should be actively sought by hospital-based medical and nursing staff if patient history is to be as complete as possible. An example would be a man who "fainted while swimming at the beach" and is noted to have lower limb weakness on arrival at the hospital. The clue is out when a paramedic reveals that it had rained at the beach and bystanders reported multiple flashes of lightning. Closer inspection of the patient then reveals singeing of scalp hair and features of lightning injury resulting in rhabdomyolysis and muscle weakness.

The history is often even more complete when the paramedic or the patient's relatives or office colleagues bring with them some of the evidence, e.g. the empty bottle of pills found close to an unconscious patient.

All this emphasizes the fact that in acute situations, there is a need to closely question anyone who can throw some light on the events surrounding or leading to the incident. One should picture the situation at the site

the patient came from and be better able to appreciate the environment in which the emergency evolved. This has implications for assessment and management planning.

Paramedics usually adopt the following ten-point approach in history taking of the chief complaint:[8]

1) *Location*: Identify the site where the problem is experienced. If pain is the problem, does it radiate? If so, where, and under what circumstances?
2) *Quality*: If there is pain, describe what it feels like. Is it sharp, dull, piercing, throbbing or crushing? Is there impaired function, e.g. of walking or breathing? If so, what is the degree of impairment?
3) *Intensity*: How bad is the symptom, say, the pain?
4) *Quantity*: Sometimes the problem may be quantified with questions such as: how many, how often, how long, how much, how big?
5) *Chronology*: This describes the temporal development of symptoms, e.g. time of onset, duration, frequency, or how the symptoms progressed over a period of time.
6) *Setting*: Did the problem arise because of injury? Under what circumstances did a medical symptom develop? Is it related to exertion or environmental changes?
7) *Scenario of the first symptom*: Where and how did the first symptom occur? What was the patient doing at that time?
8) *Aggravating and relieving factors*: What, if anything, makes the symptoms better or worse?
9) *Associated complaints*: Ask for symptoms that may be related to the main symptom(s).
10) *Attempts to modify symptoms*: What has the patient, relatives or colleagues done to modify the symptoms, e.g. taken some medications or modified some form of behavior?

In seeking further history, the paramedic would likely adopt the AMPLE approach,[8] *viz.*:

A = Allergies
M = Medications the patient is taking
P = Previous medical history
L = Last meal time. This is especially relevant in trauma situations.
E = Events surrounding the main problem or complaint

SUPPORT SYSTEMS FOR ACUTE PRE-HOSPITAL CARE

Community Awareness of Emergencies

For the patient every illness is potentially an emergency. One cannot blame the patient or relative for thinking so, when we know that such is possible. It is to the credit of most patients that they see a primary care doctor during the early phases of their illness. In this country, the 2 principal providers of primary medical care are family practitioners (80%) and government polyclinics (20%).[9,10] Yet, every now and then one hears of Emergency Ambulance Services being deluged with large numbers of non-emergency patients. Though this is common in many communities, some Ambulance Services and even Emergency Departments have managed to decrease their non-emergency workload through a process of regular patient education on the correct use of Emergency Services. The EMS serves a wide variety of patients. Since resources are finite, there is, frequently, a need to prioritize patients to receive emergency medical care. Acuity would naturally form the basis for such prioritization. While different communities have varying triage practices (with the number of triage categories of patients varying from 2 to 6) the various public hospitals and the Emergency Ambulance Service in Singapore recognize that, for practical purposes, 4 categories of patients may be delineated. These gradations are referred to as the Patient Acuity Categorization Scales (PACS)[11] and are as follows:

PACS 1: These are patients who are either already in a state of cardiovascular collapse or in imminent danger of collapse and would, therefore, be required to be attended to without a moment's delay. They would be likely to require the maximum allocation of staff and equipment resources for initial management.

PACS 2: These patients are ill and non-ambulant, and in various forms of severe distress. They would appear to be in a stable state on initial cardiovascular examination and are not in danger of imminent collapse. The severity of their symptoms requires very early attention, failing which early deterioration of their medical status is likely. They would be trolley-based.

PACS 3: These patients have acute symptoms, but are ambulant, have mild to moderate symptoms and require acute treatment that will result in resolution of symptoms over time.

PACS 4: These are non-emergency patients. They should not be presenting to Emergency Departments or using the Emergency Ambulance Services in the first place and should more appropriately be managed in a primary healthcare setting such as family practice or public polyclinics. They may have an old injury or condition that has been present for a long time. They do not require immediate treatment. There is no immediate threat to their life or limb.

Early Access to Emergency Medical Services

Once a person is already suffering from a medical emergency, early access to the EMS is crucial to allow early diagnosis and treatment to minimize mortality and morbidity. This is best exemplified in patients with acute myocardial infarction and in those with cardiac arrest due to ventricular fibrillation. Both require rapid access to the EMS, but with some differences in priority. A patient in cardiac arrest would need someone to recognize that an emergency that requires immediate attention has occurred, and for the immediate application of first-aid such as CPR followed by rapid defibrillation. This requires speed and simplicity of action. A patient having an acute myocardial infarction requires a responder to understand the implications of cardiac chest pain, with the subsequent availability of a competent medical team to offer pain relief, accurate diagnosis, reperfusion therapy and early treatment of complications. This requires precision in diagnosis and initiation of complex therapies. An appropriate combination is an important requirement of a responsive EMS. Current delays in accessing the EMS include the following:

1) *Patient decision time*: Symptoms are often interpreted incorrectly owing to mechanisms such as denial, or displacement and rationalization. Responses are influenced by severity of symptoms, the emotional reactions to them and the degree of target organ dysfunction.
2) *Doctor decision time*: For an increasing number of emergencies, interventions that are time-sensitive are required, but at the same time only available in the setting of acute general hospitals. Unless general practitioners are able to initiate definitive treatment, patients may have to be increasingly relied upon to make the initial judgment to go directly to hospital or call for the ambulance to take them directly to the Emergency Department, instead of via referral from the family doctor. This especially applies to the set of common major

emergencies typified by acute coronary syndromes (acute myocardial infarction and unstable angina), acute strokes, severe bronchial asthma and acute epilepsy.

3) *Lack of a universal access number*: The United States has a 911 system; Singapore has a 995 system. Different countries utilize different access codes. In many countries the Emergency Call access number receives a large number of prank calls that many ambulance stations would like to do away with and thus resort to 7-digit numbers, with the eventuality that usage will decrease. This amounts to a less responsive Emergency Ambulance Service. Strict implementation of such a system with mechanisms to control abuse will lead to its reliability.
4) *Road etiquette towards emergency vehicles*: Ambulances, fire engines, etc. should have right of way on our roads and expressways. Prompt access to major emergencies would then be assured. Road users should be educated about the importance of "lights and sirens". Emergency vehicles should be given priority access and right of way and be allowed to bypass red lights with caution if the need arises. Use of equipped motorcycles, as a first-tier response for certain emergencies, appears to be an option for overcoming such delays.
5) In *high-rise environments*, delays in ambulance crew reaching the patient are common events, as have been noted in Singapore, New York, and Chicago.[12,13] The use of express elevators for emergency response personnel, and building up of local community emergency response systems may help to save valuable minutes.

Facilities in Ambulances

Table 1 provides a list of the facilities generally available in all emergency ambulances. These equipment are checked frequently by the ambulance crew. Utilized items are replaced accordingly. In Singapore, the composition of the equipment in the emergency ambulances is a regular subject of review by the Medical Advisory Committee of the Emergency Ambulance Service. This is to ensure they are kept current and updated according to evolving practices.

Emergency Medical Dispatch

The first link in the chain of survival, i.e. early access, involves accessing the community's Emergency Ambulance Service. The officer answering

Table 1 List of Equipment in Emergency Ambulance

S/No	Description	Quantity
1	Main Stretcher	1
2	Collapsible Stretcher	1
3	Orthopedic Stretcher	1
4	Short Back Board	1
5	Extrication Device	1
6	Oxygen Cylinder Complete w/Holder	1
7	Lift-off IV Pole	1
8	Head Restraint	1
9	Extrication Cervical Collar	5
10	First-aid Box	1
11	Trauma Bag	1
12	Maternity Bag	1
13	Fetoscope Aluminum Type	1
14	Stethoscope	1
15	Digital Blood Pressure Set	1
16	Ventilator Resuscitator	1
17	Manual Bag Resuscitator	1
18	Portable Suction Unit	1
19	Entonox Set (Analgesic Gas)	1
20	Delivery Set	2
21	Fracture Strap Set	1
22	Fracture Immobilizer	2
23	Safety Helmet	3
24	Raincoat	3
25	Search Light	2
26	Oxygen Therapy Set	1
27	Pillows + case	2
28	PVC Hand Carrying Stretcher	3
29	Linen Set	1
30	Aluminum Long Backboards	1
31	Sam Splints (Orange Color)	3
32	Luminous Vests	3
33	Medical Torch	2
34	Instant Thermometer	1
35	Digital Thermometer	1
36	Blood Glucose Meter	1
37	Defibrillator	1

the 995 call is referred to as "Emergency Medical Dispatcher" (EMD). A trained EMD has a fairly detailed knowledge base of medical terms and first-aid. When receiving a call for assistance from a member of the public, the dispatcher acts as an interrogator, radio dispatcher, triager, logistics

coordinator, resource provider, and pre-arrival aid instructor.[14] A trained EMD uses one of two dispatch systems, *viz*. a computer-aided dispatch system, or one based on a series of dispatch cards.[15] In deciding on priority of dispatch, the dispatcher has to determine whether the problem is a true time-priority case such as cardiac arrest, respiratory difficulty, unconsciousness, severe trauma, or other emergencies requiring a response within 5 minutes. The EMD has to decide whether a motorcycle unit be sent first to render life-saving initial care and whether the ambulance adopts a "lights-and-sirens" approach. The protocols used by the dispatcher consist of short essential questions, pre-arrival instructions, and dispatch priorities.[16]

Some dispatch centers adopt a multi-level priority determinant system. The category chosen determines the types of response vehicles dispatched. Older dispatch systems utilize a diagnostic list of medical conditions, such as heart attack, asthma attack, stroke, appendicitis, major trauma, etc. to make dispatch decisions. The latter would, however, require the dispatcher to initially make a guess as to the diagnosis. Since the EMDs are not trained in the art of diagnosis, complaint-based or incident-based systems have been adopted.[16]

Key questions that a dispatcher may ask include:

1) Is he alert (able to talk normally)?
2) Is he breathing normally?
3) Does he take insulin? If yes, did he take it today?

Pre-arrival instructions usually follow the key positions, such as:

1) Ensure open airway — tilt the head backward gently, and lift the chin forward.
2) Check for breathing — look at the chest for movements and listen for the sound of breathing.
3) Treat the patient for shock. Elevate his legs on 2 pillows. Loosen his belt.
4) Call back if his condition gets worse.

Sometimes, instructions on performance of CPR may be given over the telephone, with reported success.[17] Therefore, when a dispatcher activates a paramedic crew, he would pass on information pertaining to the

incident, e.g.:

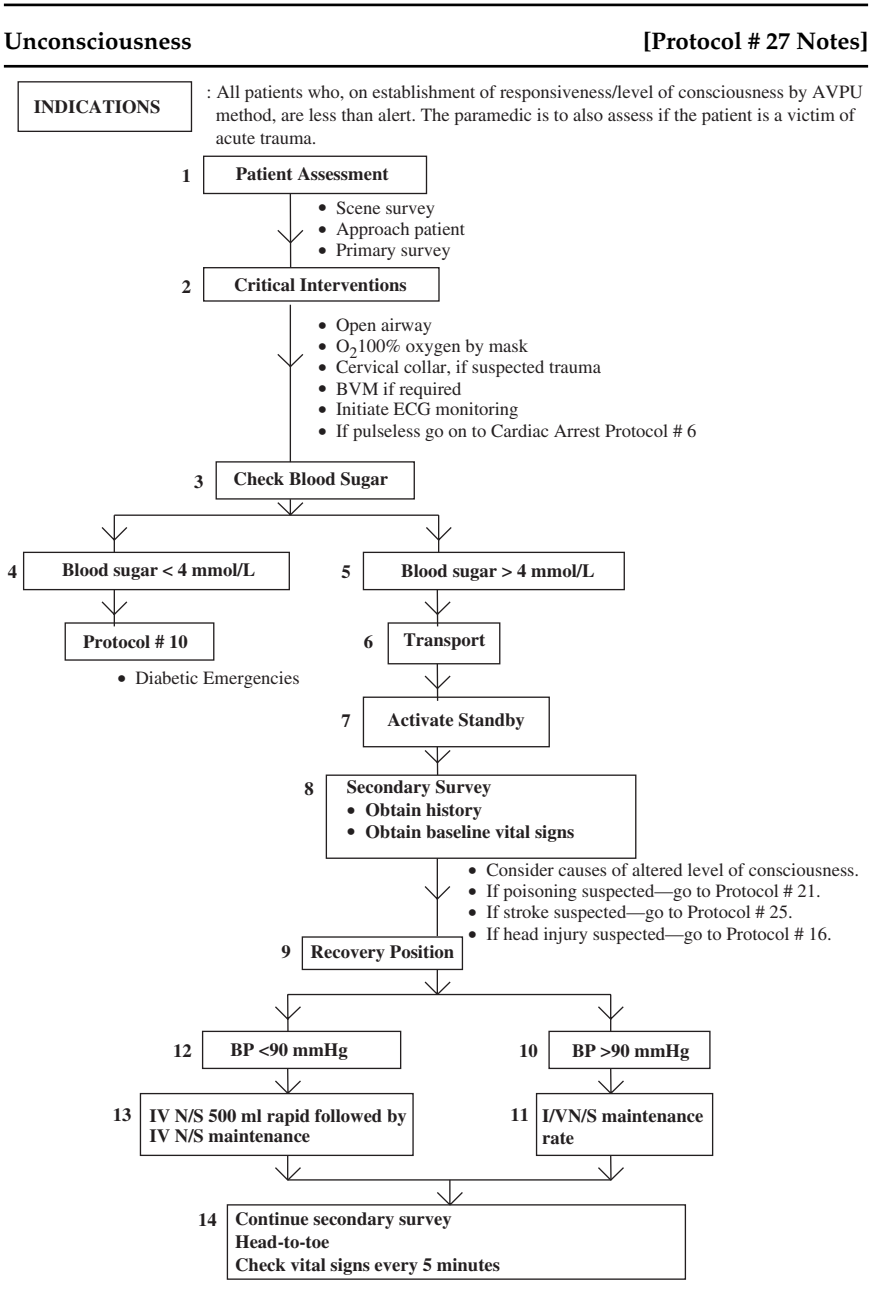
"A 56-year-old Chinese male has fainted by the foot of Block 123 at 456, High Lane. He is not conscious, but apparently breathing. There is a cut on his left forehead."

This information helps the ambulance crew to prepare for their arrival at the site of the incident and anticipate the kinds of problems they may have to handle. At the same time, the dispatcher continues to communicate with the caller and provide further pre-arrival instructions as deemed necessary. In the course of their work, dispatchers have to interact with various types of callers: The anxious, the uncooperative, and sometimes even the hysterical. In handling such callers, the dispatchers have to use their abilities to anticipate the actions of the caller, assist the caller in regaining control of himself and convert the caller into a calm person who can not only provide useful information, but also follow given pre-arrival instructions.

Emergency Ambulance Protocols

In the early days of Emergency Ambulance Services, the ambulance crew had to obtain permission from a physician, who was usually Emergency Department-based, before carrying out any interventions. Variations in the kind of instructions issued by different physicians led to numerous inconsistencies in pre-hospital care practices, since most physicians had little or no concept of the realities of pre-hospital care. The 1970s began to see the introduction of a number of written protocols for very specific and common cases handled by ambulance crews, but these still did not permit much room for paramedics to initiate treatment without direct orders by the physician. With the development of structured paramedic training programs and the introduction of greater supervision in the ambulances, indirect orders that enabled paramedics to perform more interventions based on guidelines drawn up by recognized and authorized physicians and taught to paramedics subsequently became common. Today, medical protocols exist for a large variety of problems that paramedics encounter in their daily work.[18] These allow them to act more promptly in the face of a rapidly-developing emergency.[18,19] Fig. 1 is an example of one such protocol currently in use by the Emergency Ambulance Services of the Singapore Civil Defence Force. A protocol is followed by a set of

Fig. 1 "Unconsciousness" protocol of Emergency Ambulance Service.



(Continued)

Fig. 1 (Continued)

Unconsciousness	**Protocol # 27 Notes**
Notes	**Remarks**
1) Patient Assessment	
• Need to exclude trauma early as cervical spine stabilization needs to be included if trauma is considered a likely cause of the unconsciousness.	Procedure # 9
• Scene survey should actively seek a traumatic cause. Use AVPU method to determine level of responsiveness.	
2) Critical Interventions	
• As per Procedure # 9.	Procedure # 9 — Item 4
• In addition, initiate ECG monitoring.	
• If respirations are compromised, bag-mask ventilation is to be done.	
• All patients to receive 100% oxygen by mask.	Drug Monograph # 5
• In addition to these, further management depends on the state of the pulse.	
3) Check Blood Sugar Stat	
• Use the glucose reflectance meter.	
• Blood from finger prick should be taken.	
• Ensure that drop of blood is adequate.	
4) Blood Sugar <4 mmol/L	
• Patient is presumed hypoglycemic. Go on to Diabetic Emergencies Protocol # 10.	Protocol # 10
5) Blood Sugar >4 mmol/L	
• Many conditions can result in non-hypoglycemic unconsciousness in a patient not in cardiac arrest. Most of them are urgent problems. Therefore all these patients are considered "load and go" patients.	Protocol # 10
6) Transport	
• Urgent transportation is required. Other procedures such as IV infusion, etc. can be carried out en route to hospital.	Procedure # 9
7) Activate Standby	Procedure # 9
8) Secondary Survey	
• The complete head-to-toe examination should aim to look for the following causes of altered level of consciousness:	
➔ head trauma — Protocol # 16	Protocol # 16
➔ poisoning/drug overdosage — Protocol # 21	Protocol # 21
➔ stroke — Protocol # 25	Protocol # 25
➔ alcoholic odor in the breath	

➔ patient in shock — if so, secondary survey should aim to determine the cause of shock, e.g. blood loss, severe hyperglycemia, uremia, liver disease or others.

- Care should, therefore, be taken in performing the secondary survey.
- Place the BP cuff on the upper arm that supports the head.
- The BP cuff should remain on the arm till arrival in hospital.

9) Recovery Position

- If trauma is excluded, the unconscious patient should, after examination, be managed in the recovery position so as to ensure an open airway.
- The recovery position (or three-quarters prone position) involves the following:
  - ➔ turn patient onto one side (either right or left);
  - ➔ push the arm below to the back;
  - ➔ place the back of the palm of the arm above just beneath the side of the head to support it;
  - ➔ flex the knee and hip of the leg above and place it on the ground;
  - ➔ push the leg below backwards; and
  - ➔ the patient will now be in the three-quarters prone position with the airway open.

10) Systolic Blood Pressure >90 mmHg

- This indicates that patient is not likely to be in shock

11) Start IV N/S at Maintenance Rate Procedure # 5

12) Systolic BP <90 mmHg

- Patient is in shock. The secondary survey should actively determine cause of shock. This information also needs to be passed on to the receiving hospital.

13) IV N/S Infusion

- Initiate rapid N/S infusion with IV cannula at antecubital veins. Check blood pressure at end of rapid infusion. Continue with maintenance N/S infusion. Monitor patient every 15 minutes.

14) Continue Secondary Survey

- The head-to-toe examination should be completed.
- Vital signs should be checked every 5 minutes.

---

explanatory notes for the paramedic to refer to when in doubt. In addition to such clinical protocols, procedures and medications are also written out in similar fashion so that paramedics have a simple and rapid reference system in the ambulances. Protocols set a minimal standard of

clinical care for the Ambulance service and a benchmark for audit of the performance of every ambulance run.

Ambulance Documentation and Communications

The Ambulance Case record provides the principal evidence of the information obtained by ambulance crew and the treatment carried out there. Most such records are captured in the written form. Occasionally, when dealing with extremely ill patients, the ambulance crew will inform the receiving hospital of the imminent arrival of such patients. Such information is conveyed by a radio-transmission system using Ultra-high frequency (UHF) radio communications. Occasionally, cellular telephones are utilized. In Lancashire, England, a system of transmitting video pictures of patients being managed by ambulance crew has recently been successfully piloted.[20] However, such systems require direct medical control (i.e. the presence of a doctor at the hospital, either watching a video screen or handling radio calls and advising the paramedics). While the word is not out yet on the usefulness of such video transmission, some prolongation of pre-hospital times is anticipated. Most ambulance systems have moved to various forms of indirect medical control, whereby paramedics are taught a series of protocols to use without direct medical guidance. Subsequent audit will assess the propriety of the medical care provided in the ambulances.

While such records aim to produce complete documentation of what happens during ambulance runs, these are still written records. Audit of ambulance runs using such written records is a tedious process and various attempts at doing such audit have been either random events or carried out because of some form of adverse feedback of a particular run.

A completely electronic ambulance case record has been piloted in Singapore. These records are transmitted electronically to the receiving hospital prior to the patient's arrival. The system (called the Hospital & Emergency Ambulance Link, or HEAL) is also able to transmit vital signs information, Glasgow Coma Scale scores, ECG waveforms and calculated Trauma Scores and Revised Trauma Scores.[21] The receiving hospital is warned at the time of message reception *via* an audio-indicator to indicate the severity of the patient being transported. The advantages that have surfaced with such a system are:

1) For extremely ill patients ("load and go" situations — PACS-1 equivalent) — (see Fig. 2), a documentation rate of 95.5% of critically

Fig. 2 "Load and go" situations of Singapore's Emergency Ambulance Service.

1) There are certain situations that require hospital treatment within minutes if the patient is to have any chance for survival. Load and go situations must be identified during the primary survey. Then only, critical interventions may be performed prior to transport. Non-life-saving procedures (such as splinting and bandaging) must not hold up transport.
2) It is easy to remember load and go situations as situations that compromise the Airway, Breathing and Circulation as well as Level of Consciousness (Mental Status).

A — Airway
- obstruction
- stridor
- facial burns/burns in the airway
- traumatic facial fractures
- allergic or anaphylactic reaction

Altered Mental States — from injury, metabolic or other causes

B — Breathing
- breathlessness with respiratory rate <10/minute
- breathlessness with respiratory rate >30/minute
- severe asthmatic attack
- oxygen saturation (SpO_2) < 90% even with oxygen supplement
- chest trauma with suspicion of
  - ➔ tension pneumothorax
  - ➔ open, sucking chest wound
  - ➔ flail chest
  - ➔ cardiac trauma
  - ➔ massive hemothorax

— Burns
- burns involving > 20% Body Surface Area with/without primary, secondary or tertiary blast injury

C — Circulation
- Shock from whatever cause when the systolic blood pressure is < 90 mmHg (for patients > 10 years of age), and pulse rate > 140/min.
- Major blood loss from suspected
  - ➔ fractured pelvis
  - ➔ bilateral fractured femur
  - ➔ abdominal injury
  - ➔ any combination of the above.
- chest pain — suggestive of unstable angina or acute myocardial infarction

C — CNS
- head injury with deteriorating conscious level
- head injury with GCS < 12
- head injury with seizures, unequal pupils or lateralizing signs
- status epilepticus
- acute stroke within 6 hours of onset

(Continued)

Fig. 2 (Continued)

D — Others (Trauma)
- penetrating injury such as gunshot wound or stab wound to the head, neck, chest, abdomen or pelvis
- 2 or more proximal long bone fractures
- high velocity injuries resulting from major transfer of forces, e.g. falls from more than 2 metres height
- prolonged entrapment/extrication (> 20 minutes)
- ejection from a vehicle
- pedestrian hit a speed of > 30 kph
- death of same-car occupant
- major trauma with coagulopathy

Others (Non-trauma) — any patient who is critically ill.

3) Load and go situations may be detected during either the primary or secondary survey portions of the patient assessment. To be able to identify such situations, it is imperative that the procedure of patient assessment be followed meticulously.
4) If the patient is assessed to be stable and there are no indications for immediate transport, the secondary survey (head-to-toe examination) could be done on the scene. It is important to note that a patient can become unstable, from stable to load and go in seconds. Findings in the secondary survey could change a previous assessment to a load and go patient. Hence, monitoring and reassessment must be continued en route to the hospital.
5) If there is a doubt that a patient has indications for load and go, it is better to err on the side of precaution and transport immediately to the hospital.
6) Critical interventions that may be instituted at site for such patients may be one or more of the following, as appropriate:
   a) Open airway with or without available airway adjuncts.
   b) Cervical spine stabilization, without traction.
   c) Oxygen 10–15 L/min using non-rebreather mask.
   d) Bag-mask ventilation.
   e) Nebulization to be started, if appropriate.
   f) Control of external bleeding and of shock.
   g) ECG monitoring, where appropriate.
   h) Spinal immobilization, if required.
7) A critical load and go patient must never be sent to the hospital without prior alerting of the Emergency Department. Pertinent information should be given such as age, sex, medical history and the present condition of the patient. The receiving hospital should be informed that they are being activated for standby.
8) Monitor vital signs and continually reassess patient status.
9) Non-critical interventions such as intravenous infusions and splinting of fractures may be done during transportation.
10) A detailed head-to-toe examination should be carried out. Look out for any load and go/standby situations as in para 2 above.
11) Vital signs such as BP, pulse rate, respiratory rate as well as GCS of the patient must be reviewed every 5 minutes.

important data (age, sex, chief complaint, expected time of arrival) has been achieved as opposed to an average of only 65% for verbal messages transmitted by UHF radio transmission.

2) For PACS-2 patients, at least 68% have been able to transmit more than half of the complete ambulance record prior to arrival of the patient. Without the electronics, there had previously been no transmission of information of PACS-2 equivalent patients prior to arrival in hospital. This transmission of information has allowed such patients to be registered before arrival and to approximately halve waiting time to doctor.
3) The system has also been able to prompt ambulance paramedics to carry out essential patient management procedures based on the patient's chief complaint resulting in compliance of ambulance paramedics to emergency ambulance protocols increasing from 58% in the non-electronic ambulances to 95% in the HEAL ambulances.
4) The system has for the first time been able to offer a search mechanism to pick out instances when specific ambulance care protocols had not been followed, allowing these specific runs to be subject to special audit. It has also allowed the performance of individual paramedics to be monitored electronically.

Such electronic systems will revolutionize ambulance run documentation and information transmission and open up a new era, which will allow the level of ambulance medical care to be monitored closely and improved in greater leaps.

Future Directions in Pre-hospital Acute Medicine

Many countries of the world are still beset by problems of poverty, rural transportation, and communications. Until these conditions improve, major delays will continue to be expected in the provision of pre-hospital emergency medical care. The delays that result will contribute to the high morbidity and mortality seen in the acute phases of the pathophysiology of many emergency medical conditions. Gradually, technology and organization of pre-hospital care systems in many communities is allowing greater access to emergency medical resources. The advent of information technology is hastening such access and will set new standards in the level of pre-hospital medical care that will become available. For doctors working in a hospital environment, knowledge of pre-hospital care

systems will allow a greater understanding of the type of care that has been provided prior to the patient presenting in an inpatient environment, and in planning outpatient care subsequent to discharge.

REFERENCES

1. Eisenberg MS, Howard BT, Cummins RO, *et al.*, Cardiac Arrest and Resuscitation: a tale of 29 cities, *Ann Emerg Med* **19**:179–186, 1990.
2. Laures MF, Lacleva MJ, Christensen JP, Health in the New Communications Age: Healthcare Telematics for the 21st Century. Washington DC, IOS Press, Ohmsha, pp. 1–28, 1995.
3. Meade B, Barnett P, Emergency care in a remote area using interactive video technology: a study in prehospital telemedicine, *J Telemed Telecare* **8**(2):115–117, 2002.
4. Levine SR, Gorman M, Telestroke — the application of telemedicine for stroke, *Stroke* **30**(2):464–469, 1999.
5. Ho KK, Lateef F, Anantharaman V, *et al*., Acute coronary syndrome: causes of delayed presentation. Annals, *Acad Med Singapore* **31**(3):287–293, 2002.
6. Mosesso V, Be the catalyst: transferring the bystander into a citizen responder, *J Emerg Med Serv* **27**(3) Suppl 2:16–20, 2002.
7. Engdahl J, Borg A, Herlitz J, Factors affecting short and long term prognosis among 1069 patients with out-of-hospital cardiac arrest and pulseless electrical activity (PEA), *Resuscitation* **51**(1):17–25, 2001.
8. Campbell JE (with Alabama Chapter, American College of Emergency Physicians), Chapter 2: Assessment and Initial Management of the Trauma Patient, in *Basic Trauma Life Support*, 4th ed., pp. 23–34.
9. *Annual Report*, Ministry of Health Singapore, **28,** 1999/2000.
10. *Health Facts*, Ministry of Health, Singapore, 2002.
11. Anantharaman V, Singapore Patient Acuity Categorization Scale, Letter to Ministry of Health on Recommendations of A&E Services Committee, March 2001.
12. Lateef F, Anantharaman V, Delays in the EMS response to and evacuation of patients in high-rise buildings in Singapore, *Prehospital Emergency Care* **4**(4):327–332, 2000.
13. Lateef F, Anantharaman V, Bystander CPR in prehospital cardiac arrest patients in Singapore, *Prehospital Emerg Care* **5**(4):387–390, 2001.
14. Clawson J, Martin R, Iloyd B, The EMD as a medical professional, *J Emerg Med Serv* **21**(5):68–71, 75–77, 1996.
15. Neely KW, Norton, Schmidt TA, The strength of specific EMS dispatcher questions for identifying patients with important field findings, *Prehosp Emerg Care* **4**(4): 322–326, 2000.

16. Clark JJ, Culley LL, Eisenberg MS, Accuracy of determining cardiac arrest by EMD, *Ann Emerg Med* **23**(5):1022–1026, 1994.

17. Rea TD, Eisenberg MS, Culley LL, Dispatcher assisted CPR and survival in cardiac arrest, *Circulation* **104**(21):2513–2516, 2001.

18. Brown LH, Collins JK, Gough JE, An evidence-based evaluation of EMS protocols, *Prehosp Emerg Care* **3**(1):31–35, 1999.

19. Janing J, Assessment of a scenario-based approach to facilitating critical thinking among paramedic students, *Prehospital Disaster Med* **12**(3):215–221, 1997.

20. Curry GR, Harrop N, The Lancashire Telemedicine Ambulance, *J Telemed Telecare* **4**(4):231–238, 1998.

21. Anantharaman V, Lim SH, Hospital and Emergency Ambulance Link: using IT to enhance emergency pre-hospital care, *Int J Med Informatics* **61**:147–161, 2001.

This page intentionally left blank

Endocrinology

This page intentionally left blank

15

Management of Type 2 Diabetes Mellitus

Lim Heok Seng

Diabetes mellitus (DM) is a heterogeneous disorder of metabolism due to absolute or relative deficiency of insulin. Although the hallmark of diabetes is hyperglycemia, the disease is often associated with dyslipidemia. It is a common condition that, in Singapore, affects 9% of the population aged 18–69 years[1] causing multi-organ damage and dysfunction, notably of the eyes, kidneys, nerves and cardiovascular system. The risk of these complications increases with duration of the disease. A local study showed that about 20% of patients with diabetes have retinopathy, of which 11% have sight-threatening disease.[2] In Singapore almost 60% of hospital admissions requiring acute dialysis are due to diabetic nephropathy.[3] In the year 2000, diabetic nephropathy made up 47% of all end-stage renal failure entering dialysis.[4] Diabetes is associated with 9.3% of all deaths,[5] the major causes of which are coronary heart disease, strokes, renal failure and infection.

DM is currently classified as Type 1 diabetes, Type 2 diabetes, gestational diabetes and a miscellaneous group.[6] This latter category includes diabetes caused by chronic pancreatitis, drugs such as steroids, hemochromatosis, acromegaly, and certain monogenic defects of insulin

secretion. Type 1 and Type 2 diabetes are primary diabetes, while most conditions in the "miscellaneous" category are secondary diabetes. Gestational diabetes is diabetes that occurs, or is first recognized, during pregnancy, regardless of the probability of its existence prior to pregnancy, or whether it reverts to a non-diabetes state after that.

Type 1 diabetes is due to severe insulin deficiency caused by immune destruction of the pancreatic β-cells. Glutamic acid carboxylase (GAD) antibody and islet cell antibody are often present at diagnosis. The typical patient is below 30 years old, underweight and requires insulin for survival. Type 2 diabetes is more common. It is due to varying degrees of insulin resistance and insulin deficiency and usually affects people above 40 years who are overweight. It can initially be controlled without insulin. In recent years the development of Type 2 diabetes in younger individuals is becoming more common.

A form of diabetes of indeterminate classification is being increasingly recognized. Called latent autoimmune diabetes of adults or LADA,[7] it is characterized by intermediate age of onset and presence of glutamic acid decarboxylase (GAD) antibody and islet cell antibody at diagnosis. Although LADA is somewhat different from either Type 1 or Type 2 diabetes, in its early stage it behaves like classic Type 2 diabetes but requires insulin for control relatively early.

There is also a subset of atypical diabetes who at onset are severely insulin deficient, islet autoantibodies negative and, after a short period of insulin dependence, are subsequently controllable without insulin.[8]

Table 1 Diagnostic Criteria[6]

	Impaired Fasting Glycemia	Impaired Glucose Tolerance	Diabetes
Fasting plasma glucose	6.0–6.9 mmol/L	< 7.0 mmol/L	$\geqslant 7.0$ mmol/L
2 hrs after a 75 g oral glucose tolerance test*		7.8–11.0 mmol/L	$\geqslant 11.1$ mmol/L

*The oral glucose tolerance test is not always necessary for diagnosis. Repeated fasting plasma glucose $\geqslant 7.0$ mmol/L and/or random plasma glucose $\geqslant 11.1$ mmol/L in an asymptomatic individual are enough for diagnosis. For symptomatic individuals only one of these values is required.

Current diagnostic cut-off points for diabetes are based on population studies showing that above fasting plasma glucose of 7.0 mmol/L and 2-hour post-prandial plasma glucose of 11.1 mmol/L the prevalence of diabetes-specific retinopathy rises steeply.[9] The criteria for diabetes and pre-diabetes states are shown in Table 1.

CLINICAL PRESENTATION OF TYPE 2 DIABETES

Although Type 2 diabetes is often asymptomatic, polyuria, polydipsia and lethargy may develop gradually. The majority of patients are overweight, but weight loss may develop in severe cases. Because it is often not diagnosed for many years, complications may be present at first detection. It is, therefore, not uncommon that a late complication, such as retinopathy, myocardial infarct or lower limb ulcers, is the first indication of the condition. Known risk factors are a positive family history, hypertension, overweight, previous gestational diabetes, impaired glucose tolerance (IGT), plasma HDL-cholesterol less than 0.9 mmol/L and plasma triglyceride above 2.8 mmol/L.

PATHOPHYSIOLOGY

Insulin Resistance, Insulin Deficiency, Glucose Toxicity and Lipotoxicity

Type 2 DM is commonly clustered with insulin resistance, central obesity, hypertension, atherosclerosis, low HDL-cholesterol, high triglyceride and small dense LDL-cholesterol in what has variously been described as the Syndrome X, the insulin resistance syndrome, or the atherogenic lipoprotein phenotype (ALP).

The influence of disease-susceptibility genes in the causation of Type 2 diabetes is strong but not well understood. The odds of offspring Type 2 diabetes developing from maternal diabetes is 3.4; paternal diabetes, 3.5; and biparental diabetes, 6.1.[10]

A fundamental early defect in Type 2 diabetes is insulin resistance, which is said to be present when the steady-state blood glucose is higher than is expected for a given level of plasma insulin. In the face of insulin resistance, the pancreatic islet β-cells secrete more insulin, resulting in hyperinsulinemia (Fig. 1). As long as insulin response is able to keep up with prevailing blood glucose, homeostasis is maintained. In time,

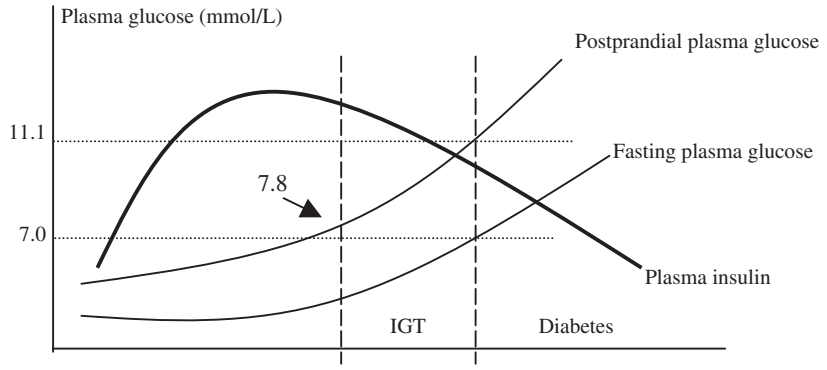


Fig. 1 Progression from normoglycemia to Type 2 diabetes mellitus. IGT = Impaired Glucose Tolerance.

postprandial glucose rises as first-phase, followed by second-phase, insulin release after a meal gradually declines, signalling progressive β-cell failure. Failure of insulin secretion to keep up with insulin resistance thus results in impaired glucose tolerance and then frank diabetes. Eventually, not only is the postprandial plasma glucose high, the fasting plasma glucose also increases as the liver is unable to modulate overnight hepatic glucose production.

Although insulin resistance and insulin deficiency coexist in Type 2 diabetes, the primacy of one over the other has not been completely settled. The balance of current evidence tends to favor a primary role for insulin resistance. This defect is determined by indeterminate genetic susceptibility, age, obesity, physical inactivity, excessive calorie intake and glucose toxicity. Intrauterine malnutrition and growth retardation with low birth weight have been implicated in the development of insulin resistance (as well as insulin deficiency) in later life.

Insulin resistance occurs at the level of the skeletal muscles, the liver and the adipose tissues in descending order of magnitude. The mechanisms of insulin resistance are believed to involve receptor and postreceptor defects. In the muscles, glucose uptake, phosphorylation and glycogen storage are impaired leading to reduced glucose disposal. In the liver, insulin resistance leads to unrestrained glycogenolysis and gluconeogenesis leading to increased hepatic glucose output. In the adipose tissues, insulin normally inhibits lipolysis. Resistance to this action of insulin presents the liver with excessive amounts of free fatty acids (FFAs) which indirectly channel pyruvate into gluconeogenesis. This mechanism

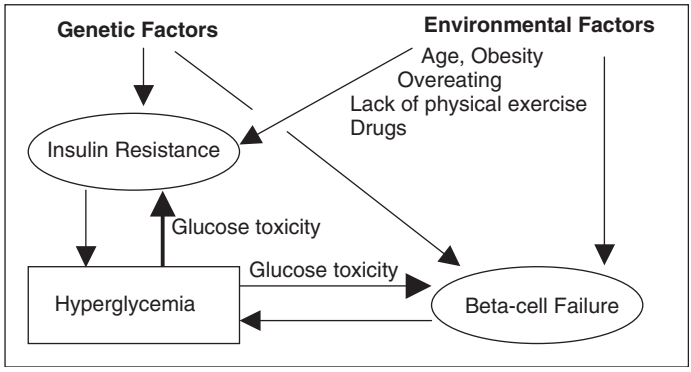


Fig. 2 The relationship between insulin resistance, β-cell dysfunction, glucose toxicity and Type 2 diabetes mellitus.

has been called lipotoxicity. In this regard, visceral adiposity plays a crucial role, as the FFAs derived from visceral fats are conveyed directly to the liver via the portal venous system.

Although insulin deficit probably plays second fiddle to insulin resistance, its role is crucial as it is the "turning point" in the progression from normal glucose tolerance to IGT to diabetes. A genetic basis for insulin deficit is suggested by the observation that many of the co-twins of Type 2 diabetics have subclinical defects of insulin secretion.[11] Environmental insults to the β-cells are not well understood. Intrauterine malnutrition is believed to also cause insulin deficiency in later life. The presence of amyloid polypeptide in the pancreatic islets provides a further clue to the underlying pathology.

As shown in Fig. 2, glucose toxicity participates in a vicious cycle between hyperglycemia on the one hand, and insulin deficiency and insulin resistance on the other. Under prolonged exposure to hyperglycemia, the β-cells respond less effectively to prevailing glucose. It is believed that oxidative stress is involved in this process.[12]

DIABETIC COMPLICATIONS

DM is associated with microangiopathy and macroangiopathy — small and median vessel defects largely responsible for diabetic retinopathy, nephropathy, autonomic and somatic peripheral neuropathy and atherosclerotic vascular disease. The latter comprises the triad of coronary

artery disease, cerebral thrombosis and peripheral arterial disease. The presence of Type 2 diabetes confers the same cardiovascular risk as having had a previous myocardial infarct,[13] i.e. diabetes is a cardiovascular risk equivalent. The risk of cardiovascular disease is 2–3 times higher in diabetics than in non-diabetics. Furthermore, in diabetics, coronary heart disease not only occurs at a younger age but also carries a worse prognosis. Modifiable risk factors for these complications are hyperglycemia, dyslipidemia, insulin resistance, hypertension and smoking. These complications are also interrelated. For example, the presence of albuminuria, a feature of nephropathy, is an independent risk factor for cardiovascular disease.

MANAGEMENT OF TYPE 2 DIABETES

Prevention

Diabetes can fundamentally be prevented in susceptible individuals if insulin resistance is corrected and β-cell reserve preserved. In this regard, weight reduction and regular exercise in obese subjects with IGT have been shown to prevent progression of IGT to diabetes.[14–16] Pharmacological agents, such as metformin,[16] acarbose[17] and orlistat,[18] in conjunction with lifestyle changes, have been shown to be likewise effective (see later section for the modes of action of these drugs).

Goals of Treatment and Global Approach

Treatment strategies should aim at: 1) alleviating symptoms; 2) averting acute sequelae, such as infections and life-threatening metabolic decompensation (diabetic ketoacidosis and hyperosmolar hyperglycemic nonketotic states); and 3) preventing or delaying the onset of microvascular and macrovascular complications to avert further catastrophes — limb amputation, cardiac crippling, strokes, blindness, and renal failure. Although it is fundamental to reverse the 2 most pronounced manifestations of Type 2 diabetes (hyperglycemia and dyslipidemia), it is equally vital to attend to other coexisting risk factors like smoking, hypertension and central obesity. At all levels of treatment, psychosocial dysfunction associated with the disease needs to be attended to (see Conclusion).

Management Strategy

Optimum diabetes care requires a strategy that includes: 1) a thorough initial clinical and biochemical evaluation upon diagnosis; 2) appropriate pharmacological intervention; 3) patient education on diet, self blood glucose monitoring (SBGM) and other aspects of self-care; and 4) periodic monitoring and review of treatment (Table 2).

Targets of Control

Guidelines developed by the Ministry of Health for optimizing glycemia and lipidemia[19,20] are shown in Tables 3 and 4. Apart from metabolic control, weight and blood pressure control is equally important. A normal body mass index (BMI) of 20–23 kg/m^2 should be targeted at, with emphasis on

Table 2 Management of Diabetes: Evaluation and Monitoring Scheme

1) Initial Evaluation and Stabilization
- Complete history
- Physical examination (focus on weight, BMI, pulse, blood pressure, cardiovascular disease, nephropathy, foot examination for neuropathy and peripheral vascular disease, eye examination* for retinopathy)
- Biochemistry: Blood glucose, HbA1c, lipids, creatinine, Urine glucose, ketones, albumin, microalbumin, microalbumin/creatinine ratio
- Start diabetes education
- Nutritional advice
- Self blood glucose monitoring (SBGM)
- Medications, if indicated

2) Subsequent Reviews to Stabilize Control (intervals vary)
- Weight, blood pressure, blood glucose, HbA1c
- Review diet, treatment, SBGM

3) At 3 Months
- Weight, blood pressure
- Blood glucose, HbA1c
- Lipids (if initially abnormal)
- Urine: Albumin, microalbumin, microalbumin/creatinine ratio (if initially abnormal)
- Foot examination (in those at high risk)
- Review treatment
- Continue diabetes education

4) Annually
- Physical examination as in initial visit
- Weight, BMI, blood pressure
- Eye examination
- Foot examination
- Biochemistry: As in initial visit
- Review diet, treatment, SBGM

*Visual acuity, fundoscopy or retinal photography. Screen at diagnosis for Type 2 diabetes (within first 5 years of diagnosis for Type 1 diabetes).

Table 3 Targets of Glucose Control[19]

	Levels of Glucose Control			
	Ideal (Non-Diabetic Levels)	**Optimal (Target Goal for the Majority)**	**Suboptimal (Adequate Goal for Some)**	**Inadequate (Action Required in All Patients)**
Preprandial glucose (mmol/L)	4.0–6.0	6.1–8.0	8.1–10.0	>10.0
2-hr postprandial glucose (mmol/L)	5.0–7.0	7.1–10.0	10.1–13.0	>13
Glycated Hb, HbA1c (in absolute %)	4.5–6.4	6.5–7.0	7.1–8.0	>8.0

Table 4 Targets of Lipid Control for Diabetic Individuals[20]

Lipids	**Levels to Initiate Action**	**Targets to Achieve**
Total cholesterol (mmol/L)	$\geqslant 5.2$ ($\geqslant 6.2$)	$\leqslant 4.0$ ($\leqslant 5.2$)
LDL-cholesterol (mmol/L)	$\geqslant 3.4$ ($\geqslant 4.0$)	$\leqslant 2.5$ ($\leqslant 3.4$)
Triglyceride* (mmol/L)	$\geqslant 2.3$ ($\geqslant 4.5$)	$\leqslant 1.7$ ($\leqslant 2.3$)
HDL-cholesterol (mmol/L)		$\leqslant 1.0$ ($\geqslant 1.0$)

Numbers without brackets are for: (a) diabetics above 45 years; and (b) those below 45 years with risk factors for coronary artery disease.

Numbers in brackets are for diabetics below 45 years without other risk factors for coronary artery disease.

*Lowering triglyceride is recommended, especially if HDL-cholesterol <0.8 mmol/L, but benefits of triglyceride-lowering await ongoing trials.

reduction of central obesity. Hypertension should be treated to a target of less than 130/85 mmHg, as the UK Prospective Diabetes Study (UKPDS) has shown that blood pressure control reduces the risk of microvascular and macrovascular complications.[21]

GLYCEMIC CONTROL

Data from the Diabetes Control and Complications Trial (DCCT)[22] have shown that in Type 1 diabetes, microvascular diseases can be significantly reduced by tight glycemic control. For each 10% reduction of HbA1c (e.g.

9.0% to 8.1% or 7.0% to 6.3%), there is approximately a 40% reduction in sustained retinopathy and 30% reduction in microalbuminuria and clinical neuropathy.[23] More recently, the UKPDS showed that in Type 2 diabetes, there is a significant 25% reduction of microvascular complications with intensive glucose control and a borderline significant 16% reduction of cardiovascular complications.[21] Combined fatal and non-fatal myocardial infarction is reduced by 18% for every percentage point reduction in HbA1c (e.g. 9% to 8%).[24] Furthermore, post-trial analyses 10 years after the DCCT and 5 years after the UKPDS have shown that initial tight glycemic control continued to have a protective effect, suggesting that optimized treatment should begin as early as possible[25] (also reported at the 18th International Diabetes Federation Congress, 24–29 August 2003).

The goal of therapy, therefore, should be a HbA1c of less than 7%. Although HbA1c is a composite of both fasting/premeal and postmeal blood glucose, it correlates better with postmeal glucose.[26] Furthermore, some studies have found a positive correlation between postprandial glycemia and cardiovascular mortality,[27] suggesting that postprandial blood glucose be specifically targeted at.

Achieving Glycemic Control

In diabetes, automatic glucose homeostasis is lost. To achieve glycemic control, there should be a delicate balance between carbohydrate consumption, energy expenditure and the restoration of insulin availability and action. This is achieved through: 1) an appropriate meal plan; 2) a suitable exercise program; and 3) pharmacological intervention. Whether one or more of these modalities are required at the outset will depend on the severity of insulin deficiency and insulin resistance; the severity of the glycemia; the type of diabetes; the patient's weight; and the presence of concomitant illnesses.

Stepped Care Approach vs. Early Combination Therapy

At diagnosis, a stepped approach, progressing from diet-cum-exercise therapy to pharmacological intervention, is the rule (Fig. 3). However, it may be necessary to use pharmacological agents from the outset, depending on the blood glucose. Generally, if the fasting plasma glucose is less than 11 mmol/L and random, less than 14 mmol/L, nutritional therapy alone for 2–3 months could be tried first. However, if the fasting plasma glucose is

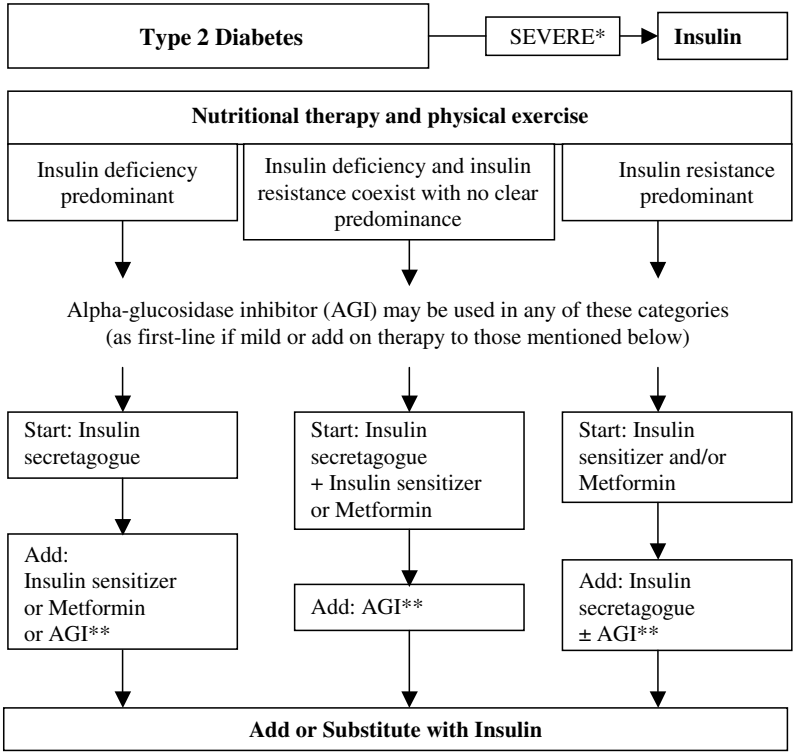


Fig. 3 Algorithm for glycemic control.
* Especially if symptomatic, FBh > 16 mmol/L, RGB > 16 mmol/L and ketonuric.
** Although AGI is mentioned as second-line therapy, it may be used as first-line if condition mild and post prandialhyperglycemia the predominant problem.

11–14 mmol/L and random plasma glucose, 14–16 mmol/L, oral drugs are usually required. Plasma glucose levels above this range usually require the use of insulin, at least temporarily. The prompt correction of hyperglycemia might also restore β-cell function and tissue insulin sensitivity.

The choice of drugs is based on whether the problem is primarily one of insulin resistance, insulin secretion, or both. Oral monotherapy should ordinarily be tried before combination oral therapy and insulin therapy (either in combination with oral therapy or as full monotherapy), in a step-wise manner to achieve glycemic goals.

This scheme, however, is not meant to be rigid. A paradigm shift is growing that since Type 2 diabetes is characterized by both insulin resistance and insulin deficiency, initial therapy could rationally incorporate

2 synergistic oral agents that address these 2 defects to achieve a better glycemic response compared to oral monotherapy.[28]

Therapy adjustments are often necessary in response to changing situations. As the effects of glucose toxicity dissipate, it might be possible to decrease the dose. Furthermore, glycemia may be influenced by a host of other factors that may worsen or compensate for the underlying pathophysiology. Acute stress events (e.g. ischemic heart disease, infection, trauma), steroids and some anti-hypertensive agents, hyperthyroidism and Cushing's syndrome may worsen control. Hypopituitarism, hypothyroidism and hypoadrenalism, on the other hand, are associated with heightened insulin sensitivity and, therefore, may require a scale-down of hypoglycemic therapy. About 40–50% of the insulin reaching the circulation is metabolized by the kidneys as are some oral agents and their active metabolites. Hence, renal impairment necessitates a scale-down of insulin therapy as well as a review of these oral agents.

NUTRITIONAL THERAPY

Physicians and nurses should work closely with professional dietitians to evaluate and advise on the patient's dietary requirements. Nutritional recommendations should attempt primarily to achieve glucose, lipid and blood pressure goals with weight reduction in the overweight as a secondary goal. The focus should be on proper food choices. Daily caloric requirements will depend on age, sex, level of physical activities, current weight, special situations like adolescence and pregnancy, lipid status and renal function.

Fifty percent to 60% of daily calorie needs should come from carbohydrates; 30–35% from fats; and the rest from proteins. Not more than 10% of total fats should be in the form of saturated fats. A relatively high carbohydrate diet may increase plasma triglyceride. Fiber consumption is ideally 20–30 g per day from legumes, roots, leafy green vegetables, fruits and whole meal grains. The use of legume fibers attenuates the hypertriglyceridemia and low HDL-cholesterolemia that result from taking a relatively high carbohydrate/low fat diet.

Replacing saturated fats with polyunsaturated fats can be expected to lower LDL-cholesterol but tends to also lower HDL-cholesterol. Monounsaturated fats, on the other hand, do not produce as much reduction of HDL-cholesterol and should be emphasized. Animal and butter oils, which are relatively high in saturated fats, are best avoided. Olive oil and canola oil have more monounsaturated fats relative to polyunsaturated

and saturated fats and are therefore healthier choices. Most other cooking oils would lie somewhere between these extremes. *Trans*-unsaturated fatty acids are vegetable oils that are hydrogenized to make them more solid. Like saturated fats, they raise plasma LDL-cholesterol and lower plasma HDL-cholesterol. Therefore, intake of *trans*-fatty acids should be limited.

Certain essential fatty acids of the omega-3 class found in fish have been shown to lower trigylceride and elevate LDL-cholesterol. They have no significant effect on glycemia. Two to three servings of fish per week provide dietary n-3 polyunsaturated fat and can be recommended.

With the onset microalbuminuria (24-hour urinary albumin excretion between 30 and 300 mg), protein restriction to 0.8–1.0 g per kg body weight per day may slow the progression of nephropathy. Further restriction to 0.8 g per kg body weight per day is recommended in the presence of overt nephropathy.

PHYSICAL ACTIVITIES

Unless contraindicated, exercise should be a regular part of a diabetic person's routine. The benefits of physical activities are improved glycemic control, reduction of cardiovascular risk factors such as hypertension and dyslipidemia, improvement in insulin sensitivity and better psychological well-being.

Aerobic exercise is preferred and should be sufficiently strenuous over 30–45 minutes to produce 60–70% of maximum allowable heart rate for age, at a frequency of 3–4 times a week. Physical activities should be slowly built up for those above 40 years and those who are just starting out. The American Diabetes Association recommends that such individuals should undergo exercise test prior to starting.

The intensity of exercise can be low (walking, gardening, shopping); moderate (brisk walking, cycling, swimming, water walking, aerobic dancing); or heavy (jogging, vigorous cycling and competitive badminton and tennis). Patients on less than 600 calorie daily diet should only do low intensity exercise. Those on 600–1200 calorie diet should perform low to moderate exercise, while those on more than 1200 calorie could do moderate to heavy exercise. Allow 5–10 minutes of warming up and a similar duration for cooling off at the end of the activity.

Hypertensive patients should choose lower limb activities in preference to upper limb activities and avoid heavy weight lifting and valsalva-like

maneuvers. Walking, cycling, swimming, but not running, are preferred for those with neuropathic or ischemic lower limbs. Proper footwear should be worn to avoid injuries. Those with untreated or recently treated proliferative retinopathy should avoid exercises associated with increased intraabdominal pressure, valsalva maneuvers, rapid head movements and risk of eye trauma. The patient should be cautioned on the risk of hypoglycemia during and after exercise and taught how to avoid and treat it. This may mean a reduction of medication dosage, taking a carbohydrate snack an hour before the scheduled activity, checking blood glucose during the exercise (if prolonged) and after.

A common reason for insufficient exercise or failure to maintain it is the lack of time and opportunity. If structured activities cannot be pursued, the patient should be encouraged to build physical exercise into his daily activities, such as using the stairs more often and taking long detours on foot! Setting realistic goals, engaging an exercise partner and varying the activities could maintain interest and morale.

PHARMACOLOGICAL TREATMENT OF EXCESS WEIGHT

As obesity is common in Type 2 diabetes, there is a place for pharmacological intervention as adjuvants to nutritional therapy. Two are currently available. Sibutramine inhibits both 5-hydroxytryptamine (serotonin) and noradrenaline at the hypothalamus. It promotes satiety, reduces energy intake and maintains metabolism that normally declines with weight reduction. Its efficacy is modest, it is non-addictive and has a good safety profile. The starting dose is 10 mg per day increasing to 15 mg after 4 weeks, if necessary, with blood pressure monitoring. Side-effects are dry mouth, headache and insomnia, raised blood pressure and tachycardia. It is contraindicated when the blood pressure is above 140/90 mm Hg. It should not be used with monoamine oxidase inhibitor and other centrally acting appetite suppressant.

Orlistat or tetrahydrolipostatin, an inhibitor of pancreatic lipase, prevents the hydrolysis of dietary triglyceride to fatty acids and monoglycerols. Besides weight reduction it lowers glycemia, total cholesterol, LDL-cholesterol, LDL/HDL ratio and apolipoprotein B. It is indicated for the obese (BMI $\geqslant$ 30) and moderately overweight (BMI $\geqslant$ 28) with cardiovascular risk factors (hypertension, diabetes, dyslipidemia). The recommended

dose is 120 mg taken just before, with, or up to one hour after each main meal. Side-effects are mainly local and include increased fecal fat but this can be minimized by concomitant reduced fat intake and psyllium. It may reduce the absorption of beta-carotene (by 30%) and vitamin E (by 60%) and supplements of these may be needed. It is contraindicated in chronic malabsorption syndrome and cholestasis.

PHARMACOLOGICAL TREATMENT OF HYPERGLYCEMIA

Based on the pathophysiology of the disease, glycemic control by pharmacological agents can be attempted at several levels:

1) Restricting exogenous carbohydrates from entering the circulation (α-glucosidase inhibitors).
2) Reducing insulin resistance (insulin sensitizers, biguanide).
3) Boosting endogenous insulin availability (insulin secretagogues).
4) Supplying insulin from an external source (insulin injection).

It must be noted that the drugs mentioned do have some actions outside of their primary function. The number of drugs available has proliferated in the last 5 years with the development of new analogs. The therapeutic choice should be based on: a) the disease: The predominant pathophysiology and severity; b) drug characteristics: Its action, efficacy, ease of use, potential for modifying macrovascular risk factors and cost; and c) patient factors: Age, comorbid conditions and affordability.

Insulin Secretagogues — Sulphonylureas (Table 5)

Sulphonylureas have been in use since the 1950s. Although their primary action is the stimulation of β-cells to release insulin in response to prevailing blood glucose, the second and third generation sulphonylureas may improve insulin sensitivity. Predictors of their efficacy include age above 40 years, diabetes less than 5 years, weight between 110–160% ideal and fasting blood glucose below 11 mmol/L. Sulphonylurea failure occurs with time at a rate of 5–10% a year.

The side-effects of sulphonylureas are hypoglycemia and weight gain. Severe hypoglycemia is more common with glibenclamide and chlorpropamide and less with tolbutamide. Because of its long duration of action, hypoglycemia from chlorpropamide is especially troublesome,

Table 5 Insulin Secretagogues — Sulphonylureas

	Dose and Administration	Supplementary Notes
Tolbutamide	250 mg bid to 1 g tid 1/2 hr before meal	First generation SU Relatively more suitable for elderly patients because of short duration of action
Chlorpropamide	125–250 mg once daily, 1/2 hr before meal	First generation SU Longest duration of action of all OHAs, therefore hypoglycemia may be prolonged Partially metabolized (some metabolites active) Most excreted unchanged in urine Caution in elderly and renal impairment May cause hyponatremia due to SIADH
Glibenclamide	2.5 mg once daily to 10 mg bid 1/2 hr before meal	Second generation SU Some metabolites are partially active and excreted via the kidneys Contraindicated in renal impairment
Glipizide	2.5 mg once daily to 15 mg bid 1/2 hr before meal	Second generation SU Inactive metabolites Relative good choice in renal impairment
Gliclazide	40 mg once daily to 160 mg bid 1/2 hr before meal	Second generation SU Metabolized by liver to inactive metabolites Diamicron MR is recently introduced as a long-acting once a day form of gliclazide (dose 30–120 mg od)
Glimepiride	1–6 mg once daily (exceptionally, 8 mg) Shortly before or with meal	Third generation SU, Excretion: 60% renal, 40% liver Less binding to myocardium and vascular smooth muscles (? less cardiotoxic) Reportedly relatively less weight gain

SU = sulphonylurea, OHAs = oral hypoglycemic agents, SIADH = syndrome of inappropriate ADH secretion.

and should best be avoided in older patients. Other side-effects are allergic rash, liver damage and leucopenia.

All the sulphonylureas undergo hepatic metabolism and should be used with caution in the presence of liver dysfunction. The partially active

metabolites of glibenclamide are cleared via the kidneys, while 20% of chlorpropamide are similarly cleared. Therefore, these drugs should not be used in the face of renal impairment. Glipizide and glimepiride are less likely to cause hypoglycemia in this scenario, as their metabolites are either inactive or have minimal hypoglycemic potency. Dose titration, however, is recommended in the presence of severe renal dysfunction.

Insulin Secretagogues — Non-sulphonylurea (Table 6)

Two non-sulphonylurea insulinotrophic agents, whose mechanism of action is similar to that of sulphonylureas, have recently been introduced: 1) Repaglinide, a derivative of meglitinide which is a moiety present in the sulphonylureas; and 2) nateglinide, a derivative of phenylalanine.

They are rapidly absorbed and rapidly eliminated and therefore should be taken immediately before a meal for optimum efficacy. Their primary effect of lowering the postprandial glucose surge has found popularity following studies showing the correlation between postprandial hyperglycemia and cardiovascular mortality.[29]

Like sulphonylureas, their side-effects are weight gain and hypoglycemia. As their metabolism is via the liver, they should be used with caution in the presence of liver dysfunction, whereas renal impairment is not a routine contraindication.

Table 6 Insulin Secretagogues — Non-sulphonylurea

	Dose and Administration	Supplementary Notes
Repaglinide	0.5–4 mg Immediately before meal	Onset of action 5–10 minutes Peak action 1 hour Duration of action 4 hours Number of doses depends on number of meals
Nateglinide	60–180 mg Immediately before meal	Reportedly less late absorptive hypoglycemia Allows for flexibility of dosing Appropriate for elderly who are prone to hypoglycemia Consider using it when meal times irregular Avoid using in obesity

Compared to repaglinide and the sulphonylurea glibenclamide, nateglinide is more selective in its closure of the β-cell K_{ATP} channels (relative to similar actions on vascular channels).[30] Since potassium channel closure in the myocardium and coronary arteries can result in an increased incidence of cardiac events,[31] this difference may confer upon it a higher safety profile.

Metformin (Table 7)

Although relatively new to the USA, metformin has been widely used in Asia, Europe and Canada for a longer period. Its actions are reduction of hepatic gluconeogenesis and increased peripheral glucose uptake. Metformin Retard (Glucophage Retard) is a longer acting version of its parent drug. Metformin can be used as a monotherapy or in combination with other oral agents and insulin. The UKPDS suggests a potential benefit of metformin monotherapy in reducing cardiovascular risk.[21]

As a monotherapy, it does not usually cause hypoglycemia, but will aggravate existing hypoglycemia. Its side-effects include anorexia, abdominal discomfort and diarrhea. These are usually transient and can be minimized by gradual dosing. Rarely, it may cause lactic acidosis, especially in the presence of renal, liver, respiratory and cardiac impairment, conditions

Table 7 Metformin

	Dose and Administration	Supplementary Notes
Metformin	250 mg bid to 850 mg tid (exceptionally 1 g tid) With or after meal	Has anoretic effects, promotes weight loss Side-effects: abdominal discomfort, diarrhea, anorexia May rarely cause lactic acidosis Caution in renal, liver and cardiac impairment Excreted mainly in urine Stop 2–3 days before surgery and radiological procedures that may cause dehydration May cause B12 deficiency Avoid in liver dysfunction and renal impairment Consider using it in the overweight
Glucophage Retard	850 mg once daily to 850 mg bid With or after meal	

in which metformin is contraindicated or, at least, should only be used with extreme caution. It should be stopped 2 days before and after the intravenous administration of contrast media to mininize the risk of lactic acidosis should renal failure follow the radiological procedure.

Insulin Sensitizers (Table 8)

These are a group of drugs called the thiazolidinediones which interact with certain nuclear receptors called peroxisome proliferator activation receptors-gamma (PPARγ) on the adipose tissues (and to a lesser extent, muscles) to reduce insulin resistance. Troglitazone was the first to be approved by the FDA in 1997, but because of severe hepatotoxicity, it was removed from the market first in the UK and later in the USA in March 2000. Others in this category considered safe are rosiglitazone and pioglitazone.

Table 8 Thiazolidinediones

	Dose and Administration	Supplementary Notes
Rosiglitazone	4–8 mg once daily Any time of day	No causally-related liver toxicity in clinical trials As a precaution, contraindicated in liver disease with Child-Pugh B/C score > 6 or if ALT > 2.5 times ULN Monitor ALT 2-monthly in first year, and thereafter periodically Not indicated in NYHA class 3 and 4 cardiac status Avoid in obesity and liver dysfunction Relatively safe in renal impairment
Pioglitazone	15–45 mg once daily Any time of day	No causally-related liver toxicity in clinical trials As a precaution, contraindicated in liver disease and if ALT > 2.5 Monitor ALT 2-monthly in first year, and thereafter periodically Not indicated in NYHA class 3 and 4 cardiac status Avoid in obesity and liver dysfunction Relatively safe in renal impairment

The glucose lowering effect of rosiglitazone and pioglitazone monotherapy may be seen only after 2 weeks, with maximal benefit probably after 2–3 months. Sufficient time must therefore be allowed before dose increment. The thiazolidinediones are particularly effective when used in combination with metformin. They can also be effectively combined with sulphonylureas and insulin.

The observed reduction in the level of hsCRP in association with rosiglitazone and pioglitazone, independent of glucose lowering, suggests that they inhibit vascular inflammation and may be useful in the prevention of atherogenesis.[32] They also have the potential for inhibiting vascular smooth wall proliferation.[33]

Although rosiglitazone and piglitazone are not associated with significant hepatotoxicity, close monitoring of liver function is recommended by the manufacturers (see Table 8). All the thiazolidinediones cause weight gain, but the fat accumulation is peripheral rather than central. In fact, visceral, hepatic, and intramyocellular fats are reduced, and this may reduce the risk for cardiovascular disease.[34] Edema and a drop in hemoglobin and hematocrit as well as weight gain may result from increased plasma volume. A low-dose thiazide diuretic and/or moderate-dose loop diuretic may be used to alleviate the problem. Nonetheless, the thiazolidinediones should be avoided in patients with New York Heart Association (NYHA) class III and IV heart failure and other conditions of fluid overload. Like metformin, used alone, they do not cause hypoglycemia, but may do so when added to an insulin secretagogue or insulin. In this situation, the dose of these drugs may have to be lowered.

α-Glucosidase Inhibitors

These are drugs that bind with polysaccharides and oligosaccharides in the upper small intestines and thus prevent the digestion of these saccharides by intestinal saccharidases and amylases. The retardation of carbohydrate absorption prevents postprandial glycemic surge, allowing the limited available insulin to cope with glucose disposal.

Available α-glucosidase inhibitors are acarbose and miglitol. They are suitable for monotherapy or in combination with the other hypoglycemic drugs. In a study on a multi-ethnic Asian population, acarbose 100 mg 3 times a day as a monotherapy for diet-failure Type 2 diabetics reduced HbA1c by 0.7%, fasting plasma glucose by 0.37 mmol/L; one-hour postprandial plasma glucose by 0.7 mmol/L; and body weight by 1.3 kg.[35]

The place of α-glucosidase inhibitors, like that of repaglinide and nateglinide, has been enhanced in recent years in the light of evidence that postprandial hyperglycemia is a significant independent risk factor for macrovascular complications.[29]

Very little of acarbose and its metabolites are absorbed. Miglitol, on the other hand, is totally or partially absorbed, but its therapeutic action is not contributed by systemic absorption. It is excreted mostly unchanged in the urine. Because undigested carbohydrates are ultimately broken down lower down in the intestinal tract by bacteria, common undesirable effects of α-glucosidase inhibitors are flatulence, abdominal bloating and diarrhea. These can be minimized by gradual dosing. They are contraindicated in patients with inflammatory bowel diseases. The presence of liver disease is not a contraindication to their use. In the event of hypoglycemia due to concomitant use of other hypoglycemic agents, complex carbohydrates may be ineffective in reversing the hypoglycemia, and only glucose should be used.

Combination Oral Hypoglycemic Agents

Combination 2-in-1 pills are available in some countries. Glucovance, for example, contains metformin 500 mg with either 1.25 mg, 2.5 mg or 5 mg of glyburide (a sulphonylurea). In October 2002, the US Food and Drug Administration (FDA) approved Avandamet™ (rosiglitazone maleate and metformin HCl) for the treatment of Type 2 diabetes. It comes in 3 tablet strengths of rosiglitazone/metformin, respectively: 1 mg/500 mg, 2 mg/500 mg and 4 mg/500 mg. The additive effect of the combination of rosiglitazone and metformin has been shown to be significantly better compared to metformin alone. Better compliance can be expected with the use of these combination oral hypoglycemic drugs.

Insulin

The key actions of insulin are suppression of hepatic glycogenolysis and gluconeogenesis, facilitation of glucose uptake by muscles and fats and inhibition of the breakdown of triglyceride into fatty acids. It is indicated in Type 2 diabetes under the following situations: 1) severe glycemia at diagnosis; 2) failed oral therapy; 3) intercurrent illnesses that worsen hyperglycemia; 4) acute hyperglycemic crises; and 5) gestational diabetes inadequately controlled by diet alone.

Insulin therapy for failed oral therapy should be individualized. For some, it is due to severe β-cell failure that obligates permanent insulin dependence for adequate control. For others, it could be a temporary measure to dampen the effects of glucose toxicity and hence restore some measure of insulin secretory capacity and insulin sensitivity that will lead to better response to oral therapy.

When it is used for failed oral agent therapy, the starting regimen could be a simple once-a-day intermediate-acting insulin regimen (e.g. Insulatard, Humulin N, Humulin L) on its own or in combination with oral hypoglycemic agents (see the next page). More complex regimens involve twice-daily intermediate-acting insulin or twice-daily intermediate-acting plus short-acting insulin. The latter regimen could incorporate either regular short-acting insulin (examples: Actrapid, Humulin R) or rapid-action insulin (examples: Insulin Lispro [Humalog], Insulin Aspart [NovoRapid]). Premixed insulins, in various ratios of intermediate-acting and regular short-acting or rapid-action insulin can be used for those keeping a more routine daily schedule of meals and physical activities.

When the regime consists of a rapid-action insulin, meals must be taken within 10 minutes of the injection. This is because it begins to work within 5–15 minutes of injection, twice as fast as regular short-acting insulin, which begins to act 30 minutes after injection. Moreover, it achieves peak activity in about 60–90 minutes and have a duration of action of approximately 4 hours, compared with 2–4 hours and 6–8 hours or longer, respectively for regular short-acting insulin.

The usual initial total daily dose when oral agents are completely replaced by insulin is empirically 0.5–1.0 units per kg body weight, depending on glycemic severity, degree of insulin resistance (usually judged by BMI) and physical inactivity. For twice-daily insulin regimens, the rule of two-thirds/one-third is empirical but convenient. The daily dose is usually split into two-thirds in the morning and one-third in the evening before dinner. Each is split again into two-thirds as intermediate-acting insulin and one-third regular or rapid-action insulin. Dose changes are then made based on blood glucose response. With restoration of metabolic control, the dose may have to be scaled down as glucose toxicity abates and endogenous insulin production improves.

Other than the intermediate-action insulins mentioned above, insulin therapy may incorporate ultralente insulin which is given once a day because of its long duration of action. However, its absorption from the

injection site is erratic. More recently, insulin glargine (Lantus), a long-acting insulin analog that has a flat, peakless profile of activity that lasts for 20–24 hours, has proven to be effective as a once-daily substitute for twice-daily intermediate-acting insulin. Insulin glargine is formed by the addition of 2 arginine residues to the end of the C-terminus of the B chain of the insulin molecule and the replacement of asparagine at the 21st position of the A chain by glycine. Given at bedtime, it reduces the risk of nocturnal hypoglycemia compared with regimens using twice-daily intermediate insulin.[36] SBGM is strongly encouraged when insulin is started.

Under certain circumstances, insulin therapy in Type 2 diabetes requires a different regimen from the above. This often involves intravenous insulin (for acute hyperglycemic crises and surgery), sliding scale subcutaneous short-acting insulin (for acute illnesses), or up to 3–4 multiple daily injections (in pregnancy). In these situations, insulin therapy is short-term unless there is already severe β-cell failure.

Except during pregnancy and hyperglycemic crises, insulin therapy for the Type 2 diabetic can be combined with oral agents (see below).

Combination Oral Hypoglycemic Agent and Insulin Therapy

Signs that the Type 2 diabetic person is severely insulinopemic are weight loss, marked hyperglycemia and ketonuria, although these should not be the sole criteria. Traditionally, when oral agents fail, the treatment is switched completely to insulin. The success of such a strategy is, however, often achieved at the expense of higher insulin dose, hyperinsulinemia, excessive weight gain and more frequent hypoglycemia.

Combining insulin with one or more oral agents will minimize these disadvantages. The use of an insulin secretagogue takes advantage of the remaining endogenous insulin reserve. Using insulinotrophic agents to harness what remains of endogenous insulin, the latter is transported directly to the liver via the portal venous system (thus minimizing peripheral hyperinsulinemia) to control postprandial blood glucose. As for metformin, insulin sensitizers and α-glucocosidase inhibitors, although they do not contribute to the circulating insulin pool, their specific extrapancreatic actions, mentioned before, are complimentary to those of insulin.

In combination therapy, 2 situations arise. Usually, insulin is added to the oral agent(s). Not uncommonly, though, an oral agent is re-added

after oral therapy has been switched to insulin monotherapy to avoid further need for increase of the insulin dose.

In the first situation, an intermediate-acting insulin or insulin glargine can be given at bedtime. Oral agents are continued at either the current maximum dose or at half of it, or the number of oral drugs to be retained can be reduced. Although the starting dose of a bedtime insulin is determined empirically on a trial and error basis, a generally safe starting dose is the numerical equivalent of the fasting blood glucose in mmol/L. Alternatively, a starting dose of 0.1 unit per kg body weight is initiated and this is then increased gradually to 0.3 units per kg body weight. Dose adjustments of 2 units at a time are made every few days based on the response of the fasting plasma glucose. If the regimen is started with a single bedtime intermediate-acting insulin, the need for subsequent addition of a second injection in the morning is determined by day-time blood glucose.

Unless β-cell failure is far advanced, the efficacy of combination therapy is well-established when insulin is judiciously combined with a sulphonylurea, metformin, acarbose, or thiazolidinedione. Equivalent glycemic control can be achieved with a smaller dose of insulin in combination therapy. The risk of hypoglycemia is smaller. With metformin or acarbose in the combination, weight gain is more effectively contained, but this is harder to achieve with a sulphonylurea or thiazolidinedione. Most patients would more readily accept insulin therapy if given once a day in combination with oral therapy. However, the costs of combination therapy may be higher and the continued need to take multiple doses of oral drugs may not be acceptable to some patients. Some patients may actually opt for full conversion to insulin therapy.

CONTROLLING DIABETIC DYSLIPIDEMIA

Dyslipidemia plays a major role in the development of diabetic macrovascular disease, especially coronary artery disease. Classic diabetic dyslipidemia is characterized by a high triglyceride level and low HDL-cholesterol level (which predisposes to an increase in the more atherogenic small, dense LDL particles) and slightly elevated LDL-cholesterol levels. All patients should have their yearly lipid profile checked and if they are on treatment for dyslipidemia, more frequent tests are indicated to optimize the treatment towards therapeutic goals.

Targets of lipid control (see Table 4) are more stringent for diabetics than for non-diabetics. Diabetes should be considered on a par with pre-existing coronary heart disease as a risk factor in determining the lipid level at which to initiate pharmacological intervention. Guidelines on lipid targets of control are mostly based on secondary prevention studies.

The first objective of dyslipidemia therapy is to lower LDL-cholesterol. Once this is achieved, the next priority is to raise HDL-cholesterol and lower triglyceride. The benefits of this three-pronged attack on diabetic dyslipidemia in the prevention and treatment of coronary heart disease has been well established by large-scale clinical trials. A recent study suggests that the target of LDL-cholesterol may have to be lowered below current recommendations.[37]

Lifestyle Changes

Cessation of smoking should be part of the global management strategy. It can reduce LDL-cholesterol and increase HDL-cholesterol.[38] The expertise of a clinical psychologist working within the framework of a smoking cessation program may work when all else fails. Weight loss and aerobic exercise, apart from lowering blood pressure and blood glucose, may lower triglyceride and LDL-cholesterol and raise HDL-cholesterol levels. Dietary modification should consist of appropriate calorie balance and not more than 200 mg dietary cholesterol per day. Although 50–60% of daily calorie needs should come from carbohydrates, relatively higher carbohydrates at the expense of fats tends to increase triglyceride and VLDL, but there may be no alternative in the obese patient. In the normal weight subject, it might be more appropriate to recommend a relatively higher fat intake enriched with monounsaturated fats.

Correction of Glycemia

The management of dyslipidemia in Type 2 diabetes should go hand in hand with correction of hyperglycemia. Improved glycemia lowers triglyceride. This might result in a favorable change in LDL composition. Furthermore, LDL-cholesterol may decrease modestly (up to 10–15%) with the achievement of optimal glycemic control. The effect of glycemic control on HDL-cholesterol is more variable. Some oral hypoglycemic agents, such as metformin, have a more favorable effect on the diabetic

Table 9 Effects of Oral Hypoglycemic Agents on Macrovascular Risk Factors Independent of Glycemia

	Wt	LDL	HDL	TG	Plasma Insulin	Insulin Resistance	Pro-coagulant State
Sulphonylurea	↑	↔	↔	↔	↑		?
Metformin	↓	↓		↓	↓	↓	↓ (PAI-1)
Thiazolidinedione	↑	↑	↑	↓	↓	↓	↓
α-glucosidase inhibitor	↔	↔	↔	↔	↓	↔	↔
Repaglinide	↑				↑		

↑ = elevated; ↓ = decreased; ↔ = no change; ? = uncertain.

lipid profile than others (Table 9). The long-term implications of rosiglitazone's effects on lipids are not known.

Lipid-lowering Drugs

For the majority of patients, lifestyle modifications either cannot be sustained or are not sufficient to achieve recommended lipid targets, and the use of lipid-lowering agents then becomes necessary. Often, this can only be achieved with combination therapy.

Lipid-modifying drugs include nicotinic acid, fibrates, HMG-CoA reductase inhibitors (statins), resins and omega-3 fish oils. (Please refer to Chapter 17, *Management of Hyperlipidemia*.) Suffice to mention here that although nicotinic acid lowers LDL-cholesterol and triglyceride and raises HDL-cholesterol, it may worsen glycemia due to the increased production of FFA which then leads to or aggravates insulin resistance. In this regard, its analog, acipimox, however, does not have this drawback.[39] Newer formulations such as Niaspan may also be more appropriate for diabetics.[40] Resins effectively lower cholesterol but raise plasma triglyceride. Being non-absorbable, it is, however, the safest cholesterol-lowering drug in the presence of renal impairment. Fish oils, a useful adjunct for triglyceride-lowering, tend to promote weight gain. With regards to the statins, which are the first choice for lowering LDL-cholesterol, the Heart Protection Study (HPS) has recently shown that patients with diabetes who were on statin therapy had a reduction in cardiovascular risk regardless of the level of their initial LDL-cholesterol.[37] A new agent, ezetimibe,

retards the intestinal absorption of cholesterol and synergizes effectively with statins to lower LDL-cholesterol.[41,42]

TREATING THE HYPERCOAGULABILITY STATE OF TYPE 2 DIABETES

Studies on anti-platelet therapy for the primary and secondary prevention of vascular events in diabetes have shown benefits, prompting the American Diabetes Association to recommend that enteric-coated aspirin in doses of 81–325 mg/day be given to diabetic persons with a history of myocardial infarction, vascular bypass procedure, stroke or transient ischemic attack, peripheral vascular disease, claudication, and/or angina. It should also be considered for primary prevention in high-risk patients (positive family history of coronary artery disease, smoking, hypertension, obesity, albuminuria, HDL-cholesterol < 55 mg/dL, triglyceride > 200 mg/dL) in those aged above 30 years.[43] There is no benefit of combining aspirin with other anti-platelet agents.

POTENTIAL NEWER TREATMENT MODALITIES

As more is known of the pathophysiology underlying Type 2 diabetes, intense search for more effective or adjunctive therapy has turned up quite a few potential candidates.

One of these is chromium, an essential trace element required for carbohydrate and lipid metabolism. It is found in small quantities in potatoes, grains, chicken, carrots and spinach. It is believed to enhance insulin action and glucose uptake. Several studies have shown improved glycemic control when therapy is supplemented with chromium picolinate.[44,45]

A peptide that is co-secreted with insulin by the β-cells is amylin, which is deficient in Type 2 diabetes. Amylin slows gastric emptying, reduces postprandial glucagon secretion and reduces postprandial glucose surge. A synthetic analog of amylin, pramlintide, when given by subcutaneous injection to Type 2 diabetics on insulin, has been shown to improve glycemic control and cause weight loss.[46]

Newer forms of insulin delivery are also being tested. Notably, intrapulmonary insulin by inhalation has been shown to have faster onset of action than subcutaneous regular insulin.[47] Its short-term use in combination with oral hypoglycemic agents in Type 2 diabetics has been shown to

improve glycemia with no significant adverse effects.[48] However, its long-term safety has yet to be established. Oralin is an oral spray form of insulin that can be absorbed through the buccal mucosa. Used in combination with oral hypoglycemic agents, it has been shown to be effective in controlling postprandial glucose.[49] The convenience of pulmonary and buccal mucosa delivery of insulin is a potential edge over traditional injections.

CONCLUSION: BIOLOGICAL AND NON-BIOLOGICAL APPROACH

This chapter has thus far focused mainly on what the physician could do for the diabetic person through therapeutic intervention of known pathophysiology of the disease. There is now an expansive array of drugs and supplements that can be mixed and matched to meet the needs of the patient. The pharmacotherapeutic choice for Type 2 diabetes is wider than for Type 1 diabetes, as Type 2 diabetes is a disease not only of insulin deficiency but also insulin resistance. The ideal metabolic and non-metabolic endpoints of treatment are now clearly defined and very stringent at that.

A purely biological approach to the dysmetabolism of diabetes has a higher chance of success in the long-term only if the psychosocial issues of the disease are also attended to. Treatment plans may fail for various reasons. Enthusiasm on the part of both the physician and the patient is a *sine qua non*. Although physician effort is essential, it alone is often insufficient. The physician must not encourage a situation in which the patient sees the health professional as the main or, worse, the only answer to his disease. Medical empowerment must be shared with the patient. In a chronic disease like diabetes, the long-term locus of control should not stay solely with the care-provider, but should gradually shift to the patient. This can only be accomplished if the patient is well-informed of his condition. He has to understand the rudiments of diabetes and its complications, and be motivated to give up unhealthy activities, such as sedentary routines, overeating and smoking.

It should be recognized that some patients take a longer time to accept and come to terms with their disease. As long as he is still in the denial or revolt phase, the success of self-care education is limited. Sufficient time is needed for the patient to go through these phases in the education effort.

One of the biggest obstacles facing the diabetes care team is changing the patient's health beliefs that often are deeply rooted in cultural and religious influences. He has to understand and accept that diabetes is a serious condition despite the absence of symptoms in the early stage, and that there is long-term benefit in therapy. However, treatment handed down to the patient must not be perceived as unrealistic and prohibitive. Recurrent hypoglycemia due to a callous, indeed, over-zealous, treatment regimen, and financial burden are common reasons for a negative attitude on the part of the patient.

As many patients may not be able to cope with trying to achieve too many objectives simultaneously, well-meant instructions on too many issues may be counterproductive. Members of the diabetes care team would do well to recognize the different educational levels, circumstances and needs of their patients. For some, choosing the right food comes instinctively, while for others, SBGM is the only way to gauge the right amount and type of foods to eat, and this quickly becomes a source of depression leading to non-adherence to therapy.

In the follow-up of the patient, therapeutic strategies have to be reviewed over and over. Therapy burnt-out and depression are real problems in need of an understanding, yet firm, approach on the part of the care-provider in helping the patient to sustain long-term interest in the treatment plan.

At every level of care, an integrated strategy involving the physician, dietician, diabetes nurse educator, podiatrist, clinical psychologist, social counsellor, other specialists, and the patient himself is necessary. It is only with such a global approach tending meticulously to biological and non-biological issues that we can hope to reduce the mortality and morbidity of one of the modern scourges that is Type 2 DM.

REFERENCES

1. National Health Survey 1998, *Epidemiology and Disease Control*, Department of Ministry of Health, Singapore.
2. Lau HC, Voo YO, Yeo KT, Ling SL, Jap A, Mass screening for diabetic retinopathy — a report on diabetic retinal screening in primary care clinics in Singapore, *Singapore Med J* **36**:510–513, 1995.
3. Goh HK, Lee EJ, Profile of admissions to an acute dialysis care unit, *Singapore Med J* **43**(9):447–451, 2002.

4. Lee G, End-stage renal disease in the Asian-Pacific region, *Semin Nephrol* **23**(1): 107–114, 2003.

5. Cutter JL, Diabetes mortality in Singapore, 1991 — a preliminary study, *Singapore Med J* **39**:311–318, 1998.

6. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, *Diabetes Care* **20**:1183–1197, 1997.

7. Tuomi T, Groop LC, Zimmet PZ, Rowley MJ, Knowles W, Mackay IR, Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease, *Diabetes* **42**:359–362, 1993.

8. Tan HH, Lim HS, Lim SC, Severe diabetes in remission: a Singapore's perspective, *Ann Acad Med Singapore* **30**(6):615–619, 2001.

9. Engelgau MM, Thompson TJ, Herman WH, Boyle JP, Aubert RE, Kenny SJ, Badran A, Sous ES, Ali MA, Comparison of fasting and 2-hour glucose and HbA1c levels for diagnosing diabetes: diagnostic criteria and performance revisited, *Diabetes Care* **20**:785–791, 1997.

10. Meigs JB, Cupples LA, Wilson PW, Parental transmission of type 2 diabetes: the Framingham Offspring Study, *Diabetes* **49**(12):2201–2207, 2000.

11. Vaag A, Henriksen JE, Madsbad S, Holm N, Beck-Nielsen H, Insulin secretion, insulin action and hepatic glucose production in identical twins discordant for non-insulin-dependent diabetes mellitus, *J Clin Invest* **95**:690–698, 1995.

12. Robertson RP, Harmon J, Tran PO, Tanaka Y, Takahashi H, Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection, *Diabetes* **52**(3):581–587, 2003.

13. Haffner SM, Lehto, Ronnemaa T, *et al.,* Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction, *N Engl J Med* **339**:229–234, 1998.

14. Eriksson KF, Lingarde F, Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise. The 6-year Malmo feasibility study, *Diabetologia* **34**:891–898, 1991.

15. Pan XR, Li GW, Hu YH, *et al*., Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study, *Diabetes Care* **20**:537–544, 1997.

16. Knowler WC, Barrett-Connor E, Fowler SE, *et al*., Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin, *N Engl J Med* (United States) **346**(6):393–403, 2002.

17. Chiasson J-L, Josse RG, Gomis R, Hanefeld Markolf, Avraham Karasik, Markku Laakso, from the STOP-NIDDM Trial Research Group, Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial, *Lancet* **359**:2072–2076, 2002.

18. Sjostrom Lars, XENDOS (XENical in the prevention of Diabetes in Obese Subjects, 9th International Congress on Obesity in Sao Paulo, Brazil.), August 26, 2002.

19. Clinical Practice Guidelines for Diabetes Mellitus, Dec 1999, Ministry of Health, Singapore.

20. Clinical Practice Guidelines for Lipid Disorders, 2000, Ministry of Health, Singapore.

21. UK Prospective Diabetes Study Group, Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS33), *Lancet* **352**:837–852, 1998.

22. Diabetic Control and Complications Trial Research Group, The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, *N Engl J Med* **329**:977–986, 1993.

23. The DCCT Research Group, The relationship of glycemic exposure (HbA1c) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial, *Diabetes* **44**:968–983, 1995.

24. American Diabetes Association: Implications of the U.K. Prospective Diabetes Study, *Ann Rev Diabetes* pp. 68–74, 1999.

25. Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group, Effect of intensive therapy on the microvascular complications of Type 1 diabetes mellitus, *JAMA* **287**:2563–2569, 2002.

26. Avignon A, Radauceanu A, Monnier L, Nonfasting plasma glucose is a better marker of diabetic control than fasting plasma glucose in type 2 diabetes, *Diabetes Care* **20**(12):1822–1826, 1997.

27. Hanefeld M, Fischer S, Julius U, Schyulze J, Schwanebeck U, Schmechel H, Ziegelasch HJ, Lindner J (The DIS Group), Risk factors for myocardial infarction and death in newly detected NIDDM: the Diabetes Intervention Study, 11-year follow-up, *Diabetologia* **39**:1577–1583, 1996.

28. Garber AJ, Larsen J, Schneider SH, Piper BA, Henry D, Simultaneous glyburide/metformin therapy is superior to component monotherapy as an initial pharmacological treatment for type 2 diabetes, *Diabetes Obes Metab* **4**(3):201–208, 2002.

29. Donahue RP, Abbott RD, Reed DM, Yano K, Post-challenge glucose concentration and coronary heart disease in men of Japanese ancestry, Honolulu Heart Program, *Diabetes* **36**:689–692, 1987.

30. Hu S, Wang S, Fanelli B, *et al.*, Pancreatic beta-cell K(ATP) channel activity and membrane-binding studies with nateglinide: a comparison with sulfonylureas and repaglinide, *J Pharmacol Exp Ther* **293**:444–452, 2000.

31. Bell DSH, University group diabetes program — "déjà vu all over again"? *Endocr Pract* **4**:64–65, 1998.

32. Satoh N, Ogawa Y, Yamada K, Kuzuya H, Nakao K, Antiatherogenic effect of thiazolidinediones in type 2 diabetic patients, mediated via mechanisms distinct from its antidiabetic effect, Abstract 568–P, in *Program and Abstracts of the 63rd Scientific Sessions of the American Diabetes Association*, June 13–17, 2003, New Orleans, Louisiana.

33. Chan JCN, Chan K-WA, Ho LLT, Fuh MMC, Lee CH, Sheaves R, Panelo AA, Kim D-K, Embong M, An Asian multicentre clinical trial to assess the efficacy and tolerability of acarbose compared with placebo in type 2 diabetic patients previously treated with diet, *Diabetes Care* **21**:1058–1061, 1998.

34. Yki-Jarvinen H, Dressler A, Zieman M, The HOE 901/3002 Study Group, Less nocturnal hypoglycemia and better post-dinner glucose control with bedtime insulin glargine compared with bedtime NPH insulin during insulin combination therapy in type 2 diabetes, *Diabetes Care* **23**:1130–1136, 2000.

35. Heart Protection Study Collaboration Group, MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo-controlled trial, *Lancet* **360**:7–22, 2002.

36. Terres W, Becker P, Rosenberg A, Changes in cardiovascular risk profile during cessation of smoking, *Am J Med* **97**(3):242–249, 1994.

37. Fulcher GR, Catalano C, Walker M, Farrer M, Thow J, Whately-Smith GR, *et al*., A double blind study of the effects of acipimox on serum lipids, blood glucose and insulin in non-obese patients with type 2 diabetes mellitus, *Diab Med* **9**:908–914, 1992.

38. Grundy SM, *et al*., Efficacy, safety, and tolerability of once-daily niacin for the treatment of dyslipidemia associated with type 2 diabetes: results of the assessment of diabetes control and evaluation of the efficacy of niaspan trial, *Arch Intern Med* **162**(14): 1568–1576, 2002.

39. Tonkon M, Simons L, Lee M, Maccubbin D, Gumbiner B, Effects of ezetimibe added to ongoing statin therapy on the lipid profile of hypercholesterolemic patients with metabolic syndrome, Abstract 920–P, in *Program and Abstracts of the 63rd Scientific Sessions of the American Diabetes Association*, June 13–17, 2003; New Orleans, Louisiana.

40. Turley SD, Dietschy JM, The Intestinal Absorption of Biliary and Dietary Cholesterol as a Drug Target for Lowering the Plasma Cholesterol Level, *Prev Cardiol* **6**(1):29–33, 2003.

41. American Diabetes Association Position Statement on Aspirin Therapy in Diabetes, *Diabetes Care* **25**:S78–S79, 2002.

42. Anderson RA, Cheng N, Bryden NA, Polansky MM, Cheng N, Chi J, Feng J, Elevated intakes of supplemental chromium improve glucose and insulin variables in individuals with type 2 diabetes, *Diabetes* **46**(11):1786–1791, 1997.

43. Houweling ST, *et al*., Effects of chromium treatment in patients with poorly controlled, insulin-treated Type 2 diabetes mellitus. A randomised, double-blind, placebo-controlled trial, Abstract 756, in *18th IDF Congress*, August 24–29, 2003, Paris.

44. Ratner RE, Want LL, Fineman MS, Velte MJ, Ruggles JA, Gottlieb A, Weyer C, Kolterman OG, Adjunctive therapy with the amylin analogue pramlintide leads to a combined improvement in glycemic and weight control in insulin-treated subjects with type 2 diabetes, *Diabetes Technol Ther* **4**(1):51–61, 2002.

45. Fishman RS, Guinta D, Chambers F, Quintana R, Shapiro DA, Insulin administration via the AeroDose™ inhaler: comparison to subcutaneous injected insulin, Abstract 30-OR,

in *Abstract Book, 60th Scientific Sessions of the American Diabetes Association*, June 9–13, 2000.

46. Cefalu WT, Skyler JS, Kourides IA, *et al.*, Inhaled human insulin treatment in patients with type 2 diabetes mellitus, *Ann Intern Med* **134**:203–207, 2001.

47. Kidron M, *et al.*, Treatment with Oralin in comparison with regular and rapid insulins in Type 1 diabetic patients evaluated by euglycemic clamp technique, Abstract 806, in *18th IDF Congress*, August 24–29, 2003, Paris.

16

Autoimmune Thyroid Disease

Ho Su Chin and Daphne Khoo

THYROTOXICOSIS

Definitions

Thyrotoxicosis is defined as a state caused by an excessive amount of circulating thyroxine (T4) or triiodotyronine (T3) or combination of both, leading to biochemical and clinical features of increased thyroid hormone action at the tissue level. Therefore, the mere presence of increased T4/T3 levels is not sufficient for the diagnosis of thyrotoxicosis. For instance, in resistance to thyroid hormone, thyroid hormone levels are elevated while clinically, the patients may be euthyroid or even hypothyroid. Hyperthyroidism refers to the overproduction of thyroid hormones from the thyroid gland.

Prevalence

Thyrotoxicosis is a common disorder and it is estimated that 1.28% of the Singapore population is thyrotoxic.[1] There is also a female preponderance for the disease with a female to male ratio of 7:1.

Causes

The causes of thyrotoxicosis are listed in Table 1. Graves' disease (GD) remains by far the commonest cause, accounting for more than 90% of cases locally. In other series, frequent causes of thyrotoxicosis include toxic multi-nodular goiter, autonomously functioning toxic adenomas and the various types of thyroiditis. The prevalence of non-GD forms of thyrotoxicosis varies in different populations.[2] This may be related to the iodine intake. Populations with low iodine intake are at risk of nodular thyroid disease while high iodine intake predisposes to autoimmune thyroid disorders. Our Singapore population is iodine replete. Thus, non-autoimmune toxic multinodular goiter and toxic adenomas appear to be comparatively rare locally. Factitious thyrotoxicosis should also be borne in mind especially when there is a history of drug usage as illustrated by the recent reports of Slim10-induced thyrotoxicosis caused by the presence of thyroid extracts in the pills. One rare but interesting cause of hyperthyroidism is due to germline mutation of the thyrotropin (TSH) receptor resulting in hereditary non-autoimmune hyperthyroidism. To date, 2 families have been reported in Singapore.[3,4]

Table 1 Causes of Thyrotoxicosis

Common
Graves' disease
Toxic multinodular goiter
Solitary toxic nodule
Thyroiditis
subacute
silent
post-partum
Factitious hyperthyroidism
Rare
Exogenous iodide administration (jod-Basedow)
radiographic contrast
iodination program
drugs, e.g. amiodarone
Hereditary non-autoimmune hyperthyroidism
Excess secretion of TSH
pituitary adenoma
pituitary resistance to thyroid hormone
Struma ovarii
Choriocarcinoma and hydatidiform mole
McCune-Albright syndrome

Clinical Features

The clinical features usually develop insidiously and most patients have symptoms of 3–6 months' duration at presentation. Thyroid hormone excess affects most systems in the body and they are listed in Table 2.

Table 2 Symptoms and Signs of Thyrotoxicosis

General
Weight loss with increased appetite
Irritability, hyperactivity, apathy
Heat intolerance, sweating
Muscle weakness and easy fatigability
Polydipsia and polyuria
Neuromuscular
Fine hand tremors
Proximal myopathy
Periodic paralysis
Chorea
Depression, psychosis
Cardiovascular
Palpitations, poor effort tolerance and angina
Widened pulse pressure
Tachycardia, atrial fibrillation, conduction abnormalities
Cardiac failure
Gastrointestinal
Nausea, vomiting
Frequency of stools
Raised liver enzymes
Occasional splenomegaly or hepatomegaly
Reproductive
Loss of libido
Gynecomastia
Oligomenorrhea, amenorrhea
Infertility
Dermatological
Pruritus
Hair loss
Palmar erythema
Moist warm skin
Skeletal
Osteopenia
Osteoporosis

Investigation and Interpretation of Thyroid Function Tests

A number of different thyroid function tests can be used to confirm thyrotoxicosis. Traditionally, total T4 levels and the free thyroxine index (FTI) have been used. In recent years, these have gradually been replaced by free T4 and TSH measurements. The elevation of free T4 and suppression of TSH confirm the presence of thyrotoxicosis. With the use of current, extremely sensitive TSH assays, increasing numbers of patients with 'subclinical thyrotoxicosis' are being detected. This occurs when the thyroid hormone levels are normal but TSH is suppressed. Occasionally, the thyroid gland secretes excess amounts of T3 but normal quantities of T4. These patients are said to have "T3-toxicosis" and the diagnosis is made on the basis of elevated T3 or free T3 levels. Occasional patients with high free T4 or T3 levels may have normal TSH values, a situation known as inappropriate secretion of TSH. Such patients should be screened for TSH-secreting pituitary tumors or they may have an unusual condition termed "resistance to thyroid hormone". During illness, changes in thyroid hormone binding and metabolism may result in abnormalities in thyroid function tests—the "euthyroid sick syndrome". The high free T4 variant of this syndrome may occasionally be confused with thyrotoxicosis, but the T3 and free T3 levels in these patients are low.

One useful investigational method to differentiate between the various causes of thyrotoxicosis is the technitium 99 thyroid scan. Hyperfunctioning of the gland will cause an increase in uptake, whilst a reduced uptake of tracer is seen in conditions associated with thyroid cell damage or suppression of function due to an exogenous source of thyroid hormones (Table 3).

Table 3 Conditions Associated with Increased and Decreased Tracer Uptake of Tc 99 m in a Thyroid Scan

Increased uptake
Homogeneous uptake
Graves' disease; hereditary non-autoimmune hyperthyroidism; excess TSH secretion
Localized/patchy uptake
Toxic adenoma; toxic multinodular goiter
Decreased uptake
Thyroiditis; factitious hyperthyroidism; struma ovarii

Because GD is the major cause of thyrotoxicosis and hyperthyroidism, the remaining section of this chapter will be devoted to the discussion of this disease.

GRAVES' DISEASE

Hyperthyroidism was first described by Caleb Perry (1755–1822). However, it was Robert Graves, an Irish physician, to whom credit is usually attributed. The disease is also known as Basedow's disease on the European continent to recognize the description by Karl A. von Basedow (1799–1854). Graves' disease is a syndrome characterized by a hyperfunctioning goiter, ophthalmopathy, pretibial myxedema, and rarely, thyroid acropachy. These features, which are specific to GD, are present in 90%, 60%, 1–5% and <1% of patients respectively and are caused by autoimmunity *per se*. The thyrotoxicosis that occurs in GD does not differ from that induced by other causes of thyroid hormone excess. The disease frequency peaks in the fourth decade and is rare in childhood. There is a female preponderance of 5:1 and the racial distribution is found to be 82% Chinese, 15% Malay and 3% others in a local cohort of 89 patients.[5] There may be a family history of thyroid disease. It is also associated with other autoimmune disorders such as type 1 diabetes mellitus, pernicious anemia and myasthenia gravis.

Clinical features

Usually, Graves' disease is clinically evident. Many patients have a diffuse, non-tender goiter although a proportion, in particular the elderly patients, have no goiter. The extra-thyroidal features of GD: Ophthalmopathy, pretibial myxedema, thyroid acropachy and thyrotoxic periodic paralysis are virtually pathognomonic of the disease. About 30% of Singaporean patients will have clinically evident ophthalmopathy at the time of presentation. The most common findings in Graves' ophthalmopathy are eyelid retraction followed by exophthalmos and extraocular muscle involvement.[6] In some instances, patients may have orbital features highly suggestive of Graves' ophthalmopathy but are not hyperthyroid. These patients are said to have euthyroid ophthalmopathy or euthyroid GD. Pretibial myxedema is very rare in our population. These lesions are caused by dermal infiltration by glycosaminoglycan and are usually located in the lower limbs above the ankles. They are pink or salmon in

color and are at times pruritic. If confluent, it can cause a swollen, beefy appearance to the limbs. Rarely, there is associated clubbing of the fingers and osteoarthropathy, including periosteal new bone formation. These 2 phenomena constitute thyroid acropachy. Thyrotoxic periodic paralysis is a condition characterized by sudden transient episodes of paralysis and hypokalemia seen primarily in Oriental males. The exact prevalence is unknown. The condition is not uncommon in Chinese and Malay males in Singapore but is extremely rare in Indians.

Pathogenesis

GD is an autoimmune disorder characterized by the presence of TSH receptor antibodies (TRAB) directed against the ectodomain of the receptor protein on the thyrocytes. The binding of these antibodies mimics the effects of TSH and results in increased biosynthesis and secretion of thyroid hormone. They may also stimulate thyrocyte proliferation leading to goiter formation. While the hyperthyroidism of GD is clearly due to these antibodies, the pathogenesis of Graves' ophthalmopathy, pretibial myxedema and periodic paralysis are still poorly understood but are thought to be immune mediated.

Diagnosis

The diagnosis of thyrotoxicosis due to GD is dependent on the detection of elevated TRAB levels and/or presence of other anti-thyroid antibodies. TRAB levels greater than 10 U/L are present in 79% of local GD patients. Besides TRAB, antibodies directed against other thyroid antigens could also be present in GD. Up to 78.3% of patients with GD have antibodies directed to the "microsomal antigen" in the thyroid, known now to be thyroid peroxidase while 53.2% are positive for anti-thyroglobulin antibodies.[7]

Treatment

Three modalities of treatment are commonly used in the management of GD:

1) blocking of hormone synthesis by anti-thyroid drugs;
2) destruction of the thyroid by radioiodine ^{131}I; and
3) partial surgical ablation of the thyroid.

The selection of treatment depends on multiplicity of factors. The patient's age, complications such as ophthalmopathy and thyrocardiac disease, the desire for conception, a history of failure with medical therapy, patient's undue concern over the hazard of irradiation and compliance to strict medical regimen are taken into considerations when choosing the treatment of choice.

Medical Treatment with Anti-thyroid Drugs

When do you use anti-thyroid drugs?

Medical treatment with anti-thyroid drugs offers the advantage that it avoids induced damages to the thyroid (and parathyroids or recurrent nerves) by radiation exposure or surgery. In a recent study patients with thyroids under 40 g weight, with low TRAB levels, and age over 40, were most likely to enter remission (in up to 80%).[8] The difficulties with drug treatment are the requirement of compliance to a medical schedule for prolonged period, frequent clinic visits, adverse drug reactions, and, most importantly, a disappointingly low permanent remission rate. Therapy with anti-thyroid drugs is used as the initial modality in people under age 18–20, in many adults through age 40, and in most pregnant women. Medical therapy remains the most popular form of treatment in Singapore. They are used as the sole modality of treatment or are used to control the thyrotoxicosis until surgery or radioiodine.

Which anti-thyroid drugs?

The anti-thyroid drugs most commonly used are carbimazole and propylthiouracil. Methimazole, the active metabolite of carbimazole, has recently been introduced. Anti-thyroid drugs reduce thyroid hormone levels by inhibiting thyroid peroxidase, and propylthiouracil has the further beneficial action of inhibiting T4 to T3 conversion. They are also thought to have immune-modulatory effects because treatment with anti-thyroid drugs also produces a prompt reduction in the TRAB and anti-thyroid antibody levels as well as a reduction in the activated T cell level.

The choice of carbimazole/methimazole or propylthiouracil is largely one of personal preference and familiarity with the drug. Both classes of drugs are well absorbed from the gastrointestinal tract and peak serum levels are reached 1–2 hours after ingestion. The serum

half-lives of methimazole and propylthiouracil are approximately 5 hours and 1 hour respectively. Despite the relatively short half-lives, carbimazole/methimazole is retained in the thyroid for longer duration, which allows longer duration of action and hence can be prescribed as a single dose. In contrast, propylthiouracil is less effective and in our experience, most patients require administration of at least twice a day to achieve and maintain euthyroidism. Clearance is delayed in patients with renal and hepatic disease, and dosage in these patients should be reduced.

Although it requires more frequent administration, propylthiouracil is preferred in 2 situations: When rapid control of thyrotoxicosis is required and in pregnancy/lactation. Propylthiouracil has additional action of blocking T4 to T3 conversion and thus, theoretically, should control hyperthyroidism more rapidly. It has traditionally been preferred in pregnancy because of lower fetal bioavailability. Methimazole, but not carbimazole, has also been linked to aplasia cutis.[9] Although case reports of other congenital abnormalities have also been associated with usage of carbimazole and methimazole, these cases are extremely rare. In general, however, anti-thyroid drugs in pregnancy appear to be safe and in lactating mothers, propylthiouracil is preferred.

How do you give the anti-thyroid drugs?

The usual starting doses are 30–45 mg a day for carbimazole, 20 mg a day for methimazole and 300–600 mg a day for propylthiouracil. Euthyroidism can be achieved usually within 6–8 weeks and is dependent on goiter size, disease activity, starting dosage as well as patient compliance. Once thyroid hormone levels normalize, the drugs are tailed down to a maintenance dose of carbimazole/methimazole 5–10 mg a day or propylthiouracil 50 mg twice a day with the occasional patient requiring higher doses. These regimens are commonly referred to as decremental or titration regimens. In the "block and replace" regimen, high doses of anti-thyroid drugs were combined with L-thyroxine to maintain euthyroidism. Subsequent studies have shown no benefit of these regimens. The duration of anti-thyroid treatment has been much studied and ranged from 6–24 months, with most centers using an 18-month treatment period.

Prognosis

In the Singapore population, we have found no difference in remission rates when patients were treated with a decremental regimen compared to a block and replace regimen. One year after completing 12 months of anti-thyroid drugs, 50.0% of patients who had been taken off treatment had relapsed, 12.5% had subclinical thyrotoxicosis and 37.5% were euthyroid. Of patients who had previously failed medical therapy and who then received a second course of therapy, 67.7% relapsed at one year, 12.9% had normal thyroid hormone but suppressed TSH levels and 19.4% were euthyroid.

Side-effects

The commonest side-effects are pruritus and maculopapular rash. These may resolve with anti-histamine treatment despite continued therapy. Other less common adverse reactions are listed in Table 4.

Neutropenia and agranulocytosis are the most serious complications. The incidence of agranulocytosis in a large series of patients was 0.4%. Reactions tend to be most frequent in the first few months of therapy but can occur at any time, with small doses of drug, and in patients of all ages. A white cell count must be taken whenever there is any suggestion of a reaction, and especially if the patient reports malaise, fever or a sore throat. In the event of severe neutropenia or agranulocytosis, the patient should be monitored closely and given antibiotics if infection develops. Administration of recombinant human granulocyte colony stimulating factor appears to hasten neutrophile recovery in most patients who start with neutrophile counts $> 0.1 \times 10^9$/L.

Table 4 Adverse Effects of Anti-thyroid Medication

Minor	Major
Pruritus	Agranulocytosis
Maculopapular rash	Cholestatic jaundice
Alopecia	Hepatitis
Abnormal sensation of taste	Lymphadenopathy
Arthralgia	Lupus-like syndrome
	Aplastic anemia
	Thrombocytopenia

Radioiodine (131)I Therapy

Radioiodine has been used to manage hyperthyroidism for almost 50 years. In North America, it is used as a primary therapy. It is safe, easy to administer and effective. Early concerns with regards to increased risks of leukemia or thyroid cancers were unfounded since several large trials have failed to find any statistical association. There is also no risk of infertility and congenital malformation when it is avoided in pregnancy. It reduces thyroid hormone production by inducing follicular cell necrosis in the early phase, leading to vascular and stromal fibrosis in the late phase. The transient thyrotoxicosis that may accompany the necrotic phase is usually asymptomatic but may be life-threatening in elderly patients particularly those with coexisting cardiac disease. In Singapore, it is common practice to treat patients with anti-thyroid drugs prior to administration of (131)I. This decreases the thyroid hormone stores and hence the likelihood of severe post-radiation thyrotoxicosis. These drugs should be omitted 3 days before and after radioiodine therapy to ensure maximal organification. The disadvantage of using anti-thyroid drugs prior to I^{-131} administration is that uptake is decreased. It is very effective in the treatment of hyperthyroidism as evidenced by a study done by S Tavintharan *et al*., which showed a 99.2% success rate. The main concern of radioiodine therapy is the development of hypothyroidism. In the local series, the incidence of hypothyroidism was 47.4% in the first year and 4.8% from the second to fifth year and 1% thereafter.[10]

Surgery

Surgery is indicated in the following categories of patients. They are: 1) patients who do not respond after prolonged anti-thyroid drug therapy, or who suffered major adverse reactions to the drug and for whatever reason are unsuitable for (131)I therapy; 2) patients with huge glands, which frequently do not regress adequately after (131)I therapy; and 3) patients with thyroid nodules that raise a suspicion of carcinoma. The major advantage of surgery is that definitive management is often obtained over an 8- to 12-week period, including pre-operative medical control, and the majority of adult patients are euthyroid after operation. Its well-known disadvantages include high expense, the surgery itself, and the risks of recurrent nerve and parathyroid damage, hypothyroidism, and recurrence. On average only about 50% maintain long-term

euthyroidism after surgery. A small percentage (<10%) remain hyperthyroid or eventually relapse. In contrast, a substantial proportion of patients have subclinical or overt hypothyroidism. The rates of relapse depend on the thyroid remnant size. While total or near-total thyroidectomies are less likely to result in subsequent relapses, the likelihood of hypothyroidism increases with more extensive procedures. Reoperation is rarely carried out for patients who relapse after previous thyroid surgery. However, recent data suggests that the procedure is safe.

HYPOTHYROIDISM

Hypothyroidism is defined as state of decreased thyroid hormone production and secretion. It may arise due to thyroid, pituitary or hypothalamic disorders. Thyroid causes are by far the most common and the term primary hypothyroidism is used to describe this group. These patients have low free thyroid hormone levels, which in turn stimulate a rise in TSH production. In mild cases, free thyroid hormone levels may be normal but TSH is elevated. This is commonly referred to as "subclinical hypothyroidism". As many patients with "subclinical" disease do in fact have symptoms or signs, there has been a gradual move to rename this condition "mild hypothyroidism". Large studies indicate that elevated TSH levels are present in 4.6–9.5% of the American population. The prevalence of hypothyroidism in the Singapore population is unknown. However, a survey of all cases of autoimmune thyroid disease referred to the Department of Endocrinology, Singapore General Hospital in 1997 revealed a total of 1478 cases of hyperthyroidism and 144 cases of primary hypothyroidism. The racial distribution of GD patients was 91.0% Chinese, 6.6% Malay, 1.1% Indian and 1.2% others, while for hypothyroidism the figures were 62.5% Chinese, 5.6% Malay, 29.2% Indian and 2.8% others respectively. If these figures are representative of the population, they would suggest that hypothyroidism is far less common than hyperthyroidism in Singapore, and that Chinese appear to be predisposed to GD while Indians are more likely to develop hypothyroidism.

Causes

Common causes in the local population are shown in Table 1. Worldwide, iodine deficiency is the most frequently implicated cause, but hypothyroidism is not seen in Singapore which is iodine replete. Most cases of

hypothyroidism in Singapore are due to chronic lymphocytic thyroiditis (Hashimoto's) or as a consequence of therapy for GD. Medical, surgical and radioiodine treatment of GD may result in transient or permanent hypothyroidism. Transient hypothyroidism is also seen in subacute or post-partum thyroiditis. Congenital hypothyroidism in Singapore is seen in 1:5000 live births. Less commonly, low thyroid hormone levels arise due to decreased TSH production as a result of pituitary or hypothalamic disease and this is termed central hypothyroidism. Central hypothyroidism in Singapore is seen most often in patients who received irradiation for nasopharyngeal carcinoma some 5–10 years before.

Treatment

The mainstay of therapy remains L-thyroxine. L-thyroxine requirements may vary according to the etiology of the hypothyroidism. In cases where the hypothyroidism is expected to be transient, L-thyroxine therapy may not be necessary unless patients are severely symptomatic. In mild cases of hypothyroidism in which TSH levels are less than 5 mU/L, withholding therapy is not unreasonable particularly if patients are asymptomatic. In such cases, TSH levels can be repeated after 3 months. In general, the doses required by Singaporean patients with hypothyroidism appear to be lower than those required in Caucasians perhaps due to lower body weights. The dose necessary to normalize TSH levels in Singaporeans averages 75–100 μg a day but requirements may range from 50–125 μg. Higher doses appear to be necessary particularly in thyroid cancer, patients with congenital hypothyroidism and the rare cases of hypothyroidism due to TRAB. The usual starting dose is about 50 μg a day although doses of 100 μg may be used in patients with more severe hypothyroidism. In elderly patients or those with cardiac disease, smaller does of 12.5–25 μg a day are used and this is gradually increased by 25 μg every 6 weeks until euthyroidism is achieved. Once the dose has been adequately titrated, patients can be monitored once every 6 months. In general, requirements remain fairly stable. However, on occasion, a previously well-controlled patient may inexplicably develop elevated TSH levels despite compliance. A careful history in such cases often reveals the consumption of new medications. Even estrogen, ferrous sulphate, antacids, dietary or fiber supplements may be the culprit. L-thyroxine requirements typically increase by 25–50% in pregnancy and

failure to treat hypothyroidism in pregnancy may compromise eventual IQ of the offspring.

A percentage of patients continue to complain of symptoms even after TSH levels have normalized. In our experience, many of these cases have normal free T4 and TSH levels but low free T3 values. The value of combining L-thyroxine with T3 has been long debated. In a recent study by Bunevicius and associates,[11] 33 hypothyroid patients on LT4 replacement were enrolled in a randomized, double-blind, crossover study. For 5 weeks they received either their usual LT4 dose or a regimen in which 50 μg of LT4 was replaced by 12.5 μg of T3 and then they were switched over to the alternative treatment for the next 5 weeks. Patients on combination LT4/T3 had less fatigue, less depression and less anger. Of the 33 patients, 20 preferred combination, 11 had no preference and 2 preferred LT4 alone.

However, the pharmacokinetics of oral T3 are not ideal from a therapeutic standpoint. T3 is rapidly absorbed, penetration into the volume of distribution is slow and the circulating half-life is short. T3 values can therefore fluctuate widely through the day. In our population, the use of 10 μg or more of T3 in a single dose when combined with LT4 almost invariably results in TSH suppression. Since TSH suppression is associated with a three-fold increased risk of atrial fibrillation in the elderly, the doses used by Bunevicius *et al.* may not be appropriate for our population. Until further data is available, T3 should not be routinely prescribed for hypothyroidism.

REFERENCES

1. Yeo PPB, Thai AC, Hughes K, Lun KC, Wang KW, Sothy SP, Thyroid diseases in Singapore, *J Med Assoc Thai* **69** (Suppl 3):48, 1986.
2. Williams I, Ankrett VO, Lazarus JH, Volpe R, Aetiology of hyperthyroidism in Canada and Wales, *J Epidemiol Community Health* **37**(3):245–248, 1983.
3. Lee YS, Poh L, Loke KY, An activating mutation of the thyrotropin receptor gene in hereditary non-autoimmune hyperthyroidism, *J Pediatr Endocrinol Metab* **15**(2): 211–215, 2002.
4. Khoo DH, Parma J, Rajasoorya C, Ho SC, Vassart G, A germline mutation of the thyrotropin receptor gene associated with thyrotoxicosis and mitral valve prolapse in a Chinese family, *J Clin Endocrinol Metab* **84**(4):1459–1462, 1999.
5. Tan CE, Wang KW, Chew LS, Rajasoorya C, Manifestations and outcome of treatment of thyrotoxicosis: a clinical study, *Ann Acad Med Singapore* **19**(6):798–801, 1990.

6. Bartley GB, Fatourechi V, Kadrmas EF, Jacobsen SJ, Ilstrup DM, Garrity JA, *et al*., Clinical features of Graves' ophthalmopathy in an incidence cohort, *Am J Ophthalmol* **121**(3):284–290, 1996.

7. Khoo DH, Fok AC, Tan CE, Koh LK, Lim SC, Eng PH, Ho SC, Thyroid stimulating hormone receptor antibody levels in Singaporean patients with autoimmune thyroid disease, *Ann Acad Med Singapore* **26**(4):435–438, 1997.

8. Vitti P, Rago T, Chiovato L, Pallini S, Santini F, Fiore E, Rocchi R, Martino E, Pinchera A, Clinical features of patients with Graves' disease undergoing remission after antithyroid drug treatment, *Thyroid* **7**:369, 1997.

9. Burrow GN, Thyroid function and hyperfunction during gestation, *Endocr Rev* **14**(2):194–202, 1993.

10. Tavintharan S, Sundram FX, Chew LS, Radioiodine (I-131) therapy and the incidence of hypothyroidism, *Ann Acad Med Singapore* **26**(1):128–131, 1997.

11. Bunevicius R, Jakubonien N, Jarkevicius R, Cernicat J, Lasas L, Prange AJ Jr, Thyroxine vs thryroxine plus triiodothyronine in treatment of hypothyroidism after thyroidectomy for Graves' disease, *Endocrine* **18**(2):129–133, 2002.

17

Management of Hyperlipidemia: The Singapore Perspective

Tan Chee Eng

INTRODUCTION

The use of statins (HMG-CoA reductase inhibitors) for the treatment of hypercholesterolemia has become widely accepted, both in primary and secondary prevention of coronary heart disease (CHD). There is a wealth of good clinical trial data and the benefits of treatment in those with high coronary risk have been clearly demonstrated. Large prospective studies such as 4S, CARE, LIPID, WOSCOPS and AFCAPS/TexCAPS have shown unequivocally that statins reduce not only major coronary event rates in primary and secondary prevention, but also all-cause mortality in secondary prevention. In these trials, statins demonstrate a relatively rapid improvement in clinical outcomes beginning approximately 1 or 2 years after treatment. The recently released Heart Protection Study[1] has further shown that benefits of treatment are seen, regardless of gender, age and baseline LDL-cholesterol levels. The recent withdrawal of Cerivastatin (Lipobay®) due to increased incidence of rhabdomyolysis in combination with Gemfibrozil should not detract from the tremendous benefits of statins when used

appropriately. There is no need to stop using all statins because of the withdrawal of one in its class as the benefits and safety of statins are well proven.

About a quarter of Singaporeans in the 1998 National Health Survey (NHS'98) were found to have high cholesterol levels (more than 6.2 mmol/L or 240 mg/dL).[2] Even more alarming was the rising trend compared with the 1992 Survey where 19% had high cholesterol levels.[3] The National Cholesterol Education Program (NCEP)[4] as well as the European Atherosclerosis Society Guidelines[5] has recommended that desirable cholesterol levels should be less than 5.2 mmol/L (200 mg/dL). That being the case, about half of Singaporeans have cholesterol levels above the desirable range.

Singapore has one of the highest rates of CHD in the Asia-Pacific region. A major contribution to this rapid rise in CHD is the cholesterol level in the Singapore population. The NHS'98 showed that the age-adjusted mean cholesterol level in the non-diabetic population was 5.47 ± 0.02 mmol/L. Amongst diabetics, the mean cholesterol level was even higher (5.69 ± 0.05 mmol/L). Furthermore, the NHS'98 also showed that the prevalence of diabetes mellitus (DM) was 9.0%, with more than half previously undiagnosed. This was again higher than any South-East Asian countries and comparable to Hong Kong (9.5% in males and 9.8% in females). There is no doubt that DM is a major contributor to the high CHD rates in Singapore and women with DM have the same CHD rates as in males. The majority of DM patients, especially the type 2 DM, suffer from microvascular and macrovascular complications, regardless of duration of DM. The macrovascular complications do not appear to be reduced by good glycemic control. In another recent study on acute myocardial infarction (AMI) patients admitted to all the major hospitals in Singapore, one-third of patients were either known or undiagnosed diabetics (unpublished data).[6] This is further proof that amongst patients with CHD, DM is a significant contributor to excess CHD risk.

There were also ethnic differences in total cholesterol, with Indian males and Malay females having the highest cholesterol levels. The NHS'98 survey showed that 94.2% of all diabetics and 89.9% of impaired glucose tolerance (IGT) subjects had LDL-cholesterol above the desirable target of 2.6 mmol/L (100 mg/dL).

OBJECTIVES OF TREATMENT

The purpose of lipid-lowering therapy is the prevention of CHD in patients with established CHD, other atherosclerotic disease and high-risk individuals. Primary prevention refers to individuals without known CHD whilst secondary prevention addresses those with established CHD or have had AMI. Management of other risk factors such as hypertension, obesity and DM must be addressed in conjunction with lipid-lowering therapy but they are beyond the scope of this discussion. Lifestyle and dietary changes are an integral part of the overall management of hyperlipidemia. These include cessation of smoking, reduction of weight, increasing of physical activity and reduction of total fat and cholesterol intake. Pharmacotherapy should be considered in those failing to reach target lipid levels after non-pharmacological measures. There is now evidence to suggest benefits of early initiation of pharmacotherapy concurrently with other measures in secondary prevention (4S follow-up).

PRIORITIES FOR TREATMENT OF LIPID ABNORMALITIES

Secondary Prevention

All with established CHD should be treated for dyslipidemia as well as managed for overall global risk reduction.[4,5] Benefits of cholesterol reduction have been clearly demonstrated in various studies (4S, CARE, LIPID).[7–9]

Diabetics/CHD Risk Equivalent

Diabetics with and without CHD should be treated as aggressively as those with established CHD. The ATP III guidelines[4] and the Clinical Practice Guidelines in Lipids (Ministry of Health, Singapore)[10] have both recommended that diabetics be treated as CHD risk equivalents. As a group, diabetics are at high risk of premature atherosclerosis[11] and exhibited benefits of treatment with statins that were comparable to subjects with established CHD.[1] Subjects with peripheral artery disease and cerebrovascular disease should also be treated as CHD risk equivalents.

Table 1 Targets for Good Control

Parameters (mmol/L)	Primary Prevention (2 or More Risk Factors)		Secondary Prevention & CHD Risk Equivalents	
	Initiation	Target	Initiation	Target
Total cholesterol	$\geqslant 6.2$	< 5.2	$\geqslant 5.2$	< 4.1
LDL-cholesterol	$\geqslant 4.1$	< 3.4	$\geqslant 3.4$	< 2.6
Triglyceride	$\geqslant 4.5$	< 2.3	$\geqslant 2.8$	< 2.3
HDL-cholesterol		$\geqslant 1.0$	< 0.9	$\geqslant 1.0$

Primary Prevention in Those with High Coronary Risks

Individuals with 2 or more risk factors are considered to be at high risk of developing premature atherosclerosis and should be treated for dyslipidemia when present.[12] The risk factors are enumerated in Table 2.

Familial Hypercholesterolemia

Men with familial hypercholesterolemia often suffer from premature CHD in their fourth to fifth decade whilst women get it about 10 years later. Such individuals should be treated aggressively with lipid-lowering therapy even in the absence of other risk factors.

When Do We Initiate Drug Treatment?

In treating lipid disorders, therapeutic decisions are based upon the patient's global risk for CHD. The risk conferred by any cholesterol level is highly dependent on the presence of other risk factors as enumerated in Table 2. A distinction is made between primary and secondary prevention in that the threshold for initiation of pharmacotherapy and target level is lower in secondary prevention. There are good placebo-controlled trial evidence supporting the benefits of reducing total cholesterol and LDL-cholesterol and CHD prevention. The benefits of TG reduction are not as convincing as those seen with LDL-cholesterol reduction. Nonetheless, there is sufficient evidence from meta-analysis to show that TG is an independent CAD risk factor and the need to treat hypertriglyceridemia.[13] Hence, our current recommendation of treating high TG, particularly in the presence of HDL-cholesterol < 0.9 mmol/L.

Table 2 Risk Factors

Parameters	Description
Age	Male $\geqslant 45$ and female $\geqslant 55$ or premature menopause without HRT
Family history	Premature CHD in 1st degree relatives, males $\leqslant 55$ and females $\leqslant 65$
Smoking	Current cigarette smoking
Hypertension	BP $\geqslant 140/90$ mmHg or taking anti-hypertensive medications
HDL-cholesterol	< 0.9 mmol/L
Negative risk factor	
HDL-cholesterol	$\geqslant 1.6$ mmol/L

DM has been removed as a risk factor because it is now considered a CHD risk equivalent. Other CHD risk equivalent groups includes cerebrovascular or peripheral artery disease.

LABORATORY INVESTIGATIONS

Fasting blood samples are generally defined as venepuncture after at least 12 hours without food/calorie intake but plain water is permitted. Fasting is not required for total cholesterol or HDL-cholesterol. However, when measurements of triglyceride undertaken, fasting is mandatory. LDL cholesterol in most laboratories are calculated values derived by the Friedewald's formula $\{LDL = Total\ cholesterol - (HDL\text{-}cholesterol + TG/2.2)\}$ and hence fasting TG is often needed to derive LDL-cholesterol. The formula does not hold true when TG is greater than 4.6 mmol/L. Some research laboratories in Singapore are able to do LDL-C by ultracentrifugation methods whilst other reference laboratories employ direct LDL assays that can measure LDL in the presence of high TG (up to 17 mmol/L).

Two consecutive measurements of the lipid panel showing similar derangement should be obtained before initiation of any therapy. This is to avoid laboratory errors, failure to fast prior to blood tests and other confounding factors. Lipid values change after major events such as surgery and AMI and it is recommended that the lipid panel should be deferred until at least 2–3 months post-events. Blood taken within 24 hours of an AMI may match the pre-infarction state.

SECONDARY CAUSES OF HYPERLIPIDEMIA

Diseases such as hypothyroidism, severe liver disease, biliary obstruction nephrotic syndrome and chronic renal failure may result in hyperlipidemia. Addressing the underlying disease state may help resolve the hyperlipidemia. However, when hyperlipidemia persists despite treatment of the underlying cause, there may be indications to start lipid-lowering therapy in the presence of high global risks. Blood tests for secondary causes should be done if suspicions are high.

DIET RECOMMENDATIONS

Most physicians are too busy to speak to patients about dietary changes and often relegate such duties to the dieticians. In some patients, formal dietary counselling is not even done. The previously mentioned post-MI studies[6] also showed that the majority of patients with hyperlipidemia have never received formal dietary counselling as distinct from diabetics, who invariably received dietary counselling. The physician has an important role in endorsing the counselling given by the dieticians, as local patients take dietary counselling lightly unless the advice is reinforced by their physicians.

There is a link between dietary saturated fat, cholesterol and CHD. Diets that are low in saturated fats and high in monounsaturated fats have low CHD rates. A reduction in total calories will reduce plasma TG whilst a reduction in saturated fats and cholesterol will reduce total cholesterol. Dietary changes can reduce total cholesterol by 20–25%. In dietary counselling, it is important to emphasize that lifestyle changes are required rather than simply a diet since that connotes in the patient's mind a short-term change. A lifelong change in eating habits and selection of food is the goal in dietary counseling. It is important to emphasize to the patients the type of food they should be eating rather than the foods they should avoid. I have found that a positive approach invokes a better response than a list of food they should avoid.

EXERCISE

There is a common belief amongst hyperlipidemia patients that exercise would directly lower the total cholesterol level. To date, evidence to

suggest that exercise lowers cholesterol is unconvincing. However, exercise has been shown to raise the good HDL-cholesterol and in so doing, reduce the total cholesterol/HDL ratio. There is also indirect evidence to suggest that exercise results in weight loss, improved glucose tolerance and insulin sensitivity with resultant improved lipid profile.

Aerobic exercises are generally recommended and the patients should target for exercise at least 3 times per week, each lasting for at least half an hour. The preferred maximum heart rate during exercise should be 60–75% of the maximal heart rate. Maximal heart rate is determined by a formula: 200 − age of subject (in years).

PHARMACOTHERAPY: THE CHOICE OF DRUGS HYPERCHOLESTEROLEMIA

Classes of drugs that can be used in hypercholesterolemia are 1) statins; 2) bile acid sequestrants; and 3) nicotinic acid derivatives, in order of preference.

Statins

When the predominant lipid disorder is hypercholesterolemia, whether it is polygenic or familial, the choice of drug would be HMG-CoA reductase inhibitors (statins). Types of statins currently available include Atorvastatin, Fluvastatin, Lovastatin, Pravastatin and Simvastatin (in alphabetical order). There is good evidence of benefit, in both primary and secondary prevention. Statins are effective in lowering total cholesterol and LDL-cholesterol. Recent evidence has shown that statins at higher doses can significantly raise HDL-cholesterol and lower triglyceride levels. The recommended daily dosages for the various statins are as follows: Atorvastatin (10–80 mg), Fluvastatin (20–40 mg), Lovastatin (10–80 mg), Pravastatin (10–40 mg), and Simvastatin (10–80 mg). It is generally accepted that 40 mg Fluvastatin ≅ 20 mg Pravastatin ≅ 20 mg Lovastatin ≅ 10 mg Simvastatin ≅ 5 mg Atorvastatin. Cerivastatin had recently been voluntarily withdrawn from the market by its manufacturers because of increased incidence of rhabdomyolysis. The latest statin to be available in Singapore is Rosuvastatin and is twice as potent as Atorvastatin, i.e. 10 mg of Rosuvastatin being equivalent to 20 mg of Atorvastatin.

Side-effects of Statins

The use of statins can occasionally result in a rise in liver enzymes, particularly the transaminases. Whether this response represents hepatic toxicity is unknown. The hepatic enzymes almost invariably return to normal on cessation of statin therapy. The incidence of abnormal liver function is proportional to dose. Another side effect of statins is myopathy. This can occur in 3 forms. Patients could complain of intermittent muscle weakness and tenderness, which may not be accompanied by elevations in creatine kinase (CK). A second group could have intermittent increases in CK, with or without muscle symptoms. A third group could present with severe myopathy, manifesting in extreme muscle weakness accompanied by extreme elevations of CK, myoglobinuria and even acute tubular necrosis. This severe reaction is more likely to occur with concomitant use of other drugs such as cyclosporine, fibric acid derivatives, nicotinic acid and also in the presence of liver disease. The Heart Protection Study has shown that incidence of transaminitis and elevated CK occurs in less than 0.1% of patients treated with statins.[1]

Indications for Stopping/Switching Statins

Elevations in transaminases (ALT and AST) above 3 times the upper limit of normal ranges requires cessation of statin therapy. Statins can be re-introduced at a lower dose when liver functions have normalized. Otherwise, consider switching between lipophilic (e.g. simvastatin, atorvastatin) and hydrophilic statins (e.g. pravastatin, cerivastatin), albeit at lower dosages. There is currently no evidence to suggest that incidence of transaminitis differs between lipophilic and hydrophilic statins although anecdoctal experience suggest there may be differences. Elevations of CK of more than 10 times the upper limit of normal ranges are also an indication to stop statin therapy. Patients who are troubled by muscle symptoms, even in the absence of raised CK, may benefit from stopping statin therapy, reduction of dosage or switching between lipophilic and hydrophilic statins. Occasionally, minor elevations of CK could be due to heavy exercise just prior to blood sampling and patients could be advised to repeat the tests a day or two after stopping physical exercise.

Monitoring for Side-effects of Statins

Transaminases and CK should be checked from 8 weeks to 3 months after initiation of statin therapy. If they are normal, there is no need to check them again unless symptoms arise or dosages are increased or combination therapy is initiated.

Resins

Bile acid sequestrants (resins) such as cholestyramine and colestipol are also effective in lowering total and LDL-cholesterol. Both are powders that must be dispersed in water or fruit juices before ingestion. 4 g of cholestyramine is equivalent to 5 g of colestipol and these drugs are usually given twice daily but sometimes up to 3 times. Recommended dosages are as follows: Cholestyramine (4 g bd up to 8 g tid), Colestipol (5 g bd up to 10 g tid). It is recommended that they be taken with meals (preferably the largest meals of the day) to avoid dyspepsia.

Side-effects of Resins

The major side-effects of the resins are gastrointestinal such as constipation, bloating, nausea, epigastric fullness, flatulence and occasional diarrhea. As a consequence of these side-effects, tolerance of resins is usually poor and they are no longer recommended as first-line drugs for hypercholesterolemia. The use of resins in diabetics can be problematic as it could aggravate gastrointestinal symptoms in diabetics with autonomic dysfunction. It can also interfere in the absorption of other drugs such as digoxin, warfarin, thyroxine, diuretics and beta-blockers. Concomitant administration of such drugs should be 1 hour before or 4 hours after administration of resins. Paradoxical rise in serum triglyceride has also been documented in treatment with resins.

Nicotinic Acid

Nicotinic acid in pharmacological doses can be used as a cholesterol-lowering agent. The best responses to nicotinic acid therapy are obtained with intakes in the range of 3–6 g/day. An intake of 1.5 g tid is often used.

Side-effects of Nicotinic Acid

Only about 50–60% of patients are able to tolerate nicotinic acid on a long-term basis because of the numerous side-effects. Our experience on the use of nicotinic acid suggests that tolerance amongst local patients may be even lower. The side-effects include gastric irritation, nausea, epigastric discomfort, flushing and itching of the skin. Flushing of the face is particularly common shortly after starting the drug, and may disappear over a period of time as it is related to prostaglandin release. We have also found that patients are better able to tolerate the nicotinic acid if dosages are built up gradually rather than giving them full doses at the initiation of therapy. The flushing may also be reduced by the concomitant use of aspirin. In rare instances, patients may develop acanthosis nigricans. The most common manifestations of liver function abnormality are raised transaminases. Therapy should be withheld in patients with significant and persistent elevations of transaminases. Nicotinic acid also worsens glucose tolerance and is not recommended in diabetic patients. It is also not recommended in patients with gout as it increases uric acid and can precipitate an acute gouty attack.

Combination Therapy in Hypercholesterolemia

In patients failing to achieve set target values despite maximal doses of statins, combination therapy is an option, with the addition of resins or nicotinic acid. Patients with familial hypercholesterolemia often require combination therapy and even then they may not achieve therapeutic target. Recently, a new class of cholesterol lowering agents have been made available. This class of agents are the selective cholesterol absorption inhibitors (Ezetimibe). By itself, ezetimibe only lowers LDL cholesterol by approximately 15–20%. However in combination with any of the statins, it can reduce LDL cholesterol between 45% to 60%.

HYPERTRIGLYCERIDEMIA

Fibrates

Fibrates are the drug of choice in hypertriglyceridemia. In moderate hypertriglyceridemia, the fibrate drugs can effectively lower triglyceride levels significantly. They also have the added benefits of increasing HDL-cholesterol. LDL-cholesterol-lowering effects are in the region of 10–15%.

Fibrates have also been shown to change the quality of LDL particles from the atherogenic, dense form to a less atherogenic, buoyant LDL particle.[14] Available fibrates include Bezafibrate, Ciprofibrate, Clofibrate, Fenofibrate and Gemfibrozil (in alphabetical order). Recommended dosages of fibrates are as follows: Bezafibrate (200–600 mg), Ciprofibrate (100–200 mg), Fenofibrate (160–400 mg), and Gemfibrozil (600–1200 mg).

Side-effects of Fibrates

First generation fibrates like gemfibrozil can lead to increased risk for cholesterol gallstones because of supersaturation of bile with cholesterol. There is no documented increased incidence of gallstones with second generation fibrates like fenofibrate and ciprofibrate. Another possible side-effect is myopathy with complaints of muscle weakness and tenderness, together with elevations of CK. The danger of myopathy is increased in renal failure and caution must be exercised in these groups of patients. The myopathy and CK usually normalizes with cessation of fibrate therapy. Elevations in liver enzymes such as transaminases and gastric irritation have also been documented with fibrate therapy.

Omega-3 Fish Oils

In severe hypertriglyceridemia (>10 mmol/L), there may not be increased CHD risk but there is risk of pancreatitis. Fibrates may not adequately lower the levels and there is a role for the use of marine fish oils rich in omega-3 polyunsaturated fatty acids in severe hypertriglyceridemia. Omega-3 fatty acids can be used in dosages of 3 g, up to a total of 12 g per day. The use of fish oil is not contraindicated in diabetics but glycemic control can worsen if the total calorie intake is not reduced to compensate for the calories resulting from the fish oil.

MIXED HYPERLIPIDEMIA

There are no specific trials on lipid-lowering therapy in mixed hyperlipidemia. However, we have strong evidence on the benefits of reducing LDL-cholesterol with regards to CHD prevention. Trials are currently ongoing to determine the benefits of triglyceride-lowering on CHD prevention (FIELD trial). Until such trial results are made available, we

would recommend that individuals with mixed hyperlipidemia be treated with a statin as the drug of first choice, given the fact that statins reduce CHD events when LDL-cholesterol is reduced. Furthermore, statins have shown significant triglyceride-lowering at higher doses. The magnitude of TG reduction by statins is dependent on the baseline TG levels, i.e. the higher the baseline TG, the greater the reduction by statins. When triglyceride remains elevated despite the use of statin, the use of fibrate can be added on in combination. In clinical practice, mixed hyperlipidemia in diabetics often require a combination of statins and fibrates before achieving target levels.

Combination Therapy in Mixed Hyperlipidemia

A combination of 2 lipid-lowering drugs often controls the hyperlipidemia better than either drug alone. Patients with familial combined hyperlipidemia or DM often fail to achieve target values on monotherapy alone. When lipid levels remain outside set target values despite maximal doses of statins, fibrates can be added. It would be prudent to start at low doses of fibrates and increase the dosage gradually to achieve target values. Monitoring of transaminases (ALT and AST) and CK should be done within 6–8 weeks of initiation of combination therapy. Although there is theoretically greater potential for side-effects compared with monotherapy, local clinical experience (more than 100 patients on combination therapy) suggests that side-effects may not be any more than in monotherapy if used judiciously and with careful monitoring.

Other possible combination includes nicotinic acid but this must be used with caution in diabetics, where glycemic control can be adversely affected.

DURATION OF TREATMENT

Most patients on lipid-lowering therapy require the medications for life, especially in inherited dyslipidemia such as familial hypercholesterolemia, familial combined hyperlipidemia and Type I and V hyperlipidemia. In hypercholesterolemia and hypertriglyceridemia secondary to diet and lifestyle indiscretion, there may be a place for discontinuing pharmacotherapy when target levels are reached. Lipids should be monitored after 3–4 months and therapy restarted if a rising trend is demonstrated.

SPECIAL CONSIDERATIONS

Diabetes

The lipid abnormality in diabetics is usually an elevation in triglyceride and low levels of HDL-cholesterol. LDL-cholesterol in type 2 diabetics can also be increased. There are also qualitative changes in the LDL-cholesterol, with a predominance of dense LDL particles.[15] Such dense LDL particles are associated with increased coronary risks. Diabetics have high risk of developing premature atherosclerosis and Haffner *et al*.[11] demonstrated that diabetics without previous MI have as high a risk as non-diabetic patients with previous MI. Diabetics also have greater case fatality, poorer prognosis and develop more complications after myocardial events. Given the fact that 9.0% of our population have diabetes, and the majority being type 2 DM, our current lipid guidelines recommend that diabetics be treated as aggressively as patients with previous MI or established ischemic heart disease, regardless of age.

Target LDL-cholesterol levels in secondary prevention should be less than 2.6 mmol/L (100 mg/dL) (Table 1) and we recommend that DM subjects should be treated at those levels. The NHS'98 showed that 89.9% of IGT subjects and 95.4% of newly diagnosed DM subjects had LDL-cholesterol above 2.6 mmol/L (unpublished data). Given such data, setting lipid levels in diabetic subjects as in secondary prevention would have a tremendous implication on healthcare cost. However, we believe that these group of diabetics would benefit most from risk reduction, the cost of treatment notwithstanding.

Mixed Hyperlipidemia in Diabetics

Patients with DM often fail to achieve target values on monotherapy alone. When lipid levels remain outside set target values despite maximal doses of statins, fibrates can be added. The combination of statin with fibrate may be the best option, given the relative contraindication of nicotinic acid and resins in diabetics.

Children

Dietary intervention is the primary treatment in children with elevated cholesterol. Drug therapy may be considered in children above 10 years

and older if they are proven to have familial hypercholesterolemia and if dietary intervention fails to bring LDL-cholesterol < 5.2 mmol/L. Resins are generally recommended for hypercholesterolemia in this group of patients. Statins may be used in the setting of a specialized lipid clinic.

Severe chylomicronemia (triglyceride >10 mmol/L) should also be treated by dietary means. The purpose of intervention is to reduce the risk of acute pancreatitis. In patients with triglyceride persistently above 10 mmol/L, use of omega-3 fatty acid may be considered. There is currently no consensus on the desirable triglyceride levels in children but we should aim to keep triglyceride levels below 10 mmol/L.

Women

Postmenopausal women and those with early menopause secondary to surgery and other causes should be considered for lipid treatment as in the male population. Pre-menopausal women with hyperlipidemia and high global risks should also be treated. Others with low risk should be treated by non-pharmacological methods.

Pregnancy

There is no indication to treat hypercholesterolemia or any form of dyslipidemia in the pregnant female. The only exception is severe chylomicronemia (TG >10 mmol/L), where the aim is to reduce the risk of acute pancreatitis. Clinical experience suggests that the chylomicronemia worsens with progression of the pregnancy. However, the only therapy we would recommend would be the use of omega-3 fish oils, after intensive dietary changes have failed to bring the triglyceride levels below 10 mmol/L.

Elderly

Elderly subjects are at high risk of getting coronary events and would derive great benefits from cholesterol reduction. The decision to start pharmacotherapy should be based on the global risk assessment, the life expectancy as well as the quality of life of the patient. Age is not a contraindication to pharmacotherapy if indicated. The Heart Protection

Study[1] has shown that the elderly derived similar benefits from treatment with a statin. We await the results of other trials such as the PROSPER study to demonstrate the benefits of statin in improving cognition and preventing dementia.

Renal Disease

Patients with chronic renal failure and hyperlipidemia can be treated with statins. However, caution must be exercised as treatment can potentially worsen renal function, and hence starting doses of statins should be low and monitored closely. During therapy, both CK and renal functions should be monitored. There is a relative contraindication to fibrates and nicotinic acid in chronic renal failure because of reported cases of myopathy. Fibrates could still be used in mild to moderate renal failure but dosages should be reduced and with stringent monitoring for side-effects, especially myopathy. When creatinine clearance is less than 10 ml/min, fibrates should not be used.

Patients with nephrotic syndrome often have hypercholesterolemia, requiring intervention with statins. Again, the high cholesterol levels must be assessed in the context of global risks of the patient.

Renal transplant patients on immunosuppressive may also develop hyperlipidemia. Treatment decisions are again based on global risk assessment. However, caution must be exercised in patients on cyclosporin because of the risk of myopathy. Transplant patients may periodically suffer from infections, particularly fungal infection, requiring use of anti-fungal agents. Care must be taken to monitor for myopathy even if previously stable because of the interactions between the anti-fungal agents and statins.

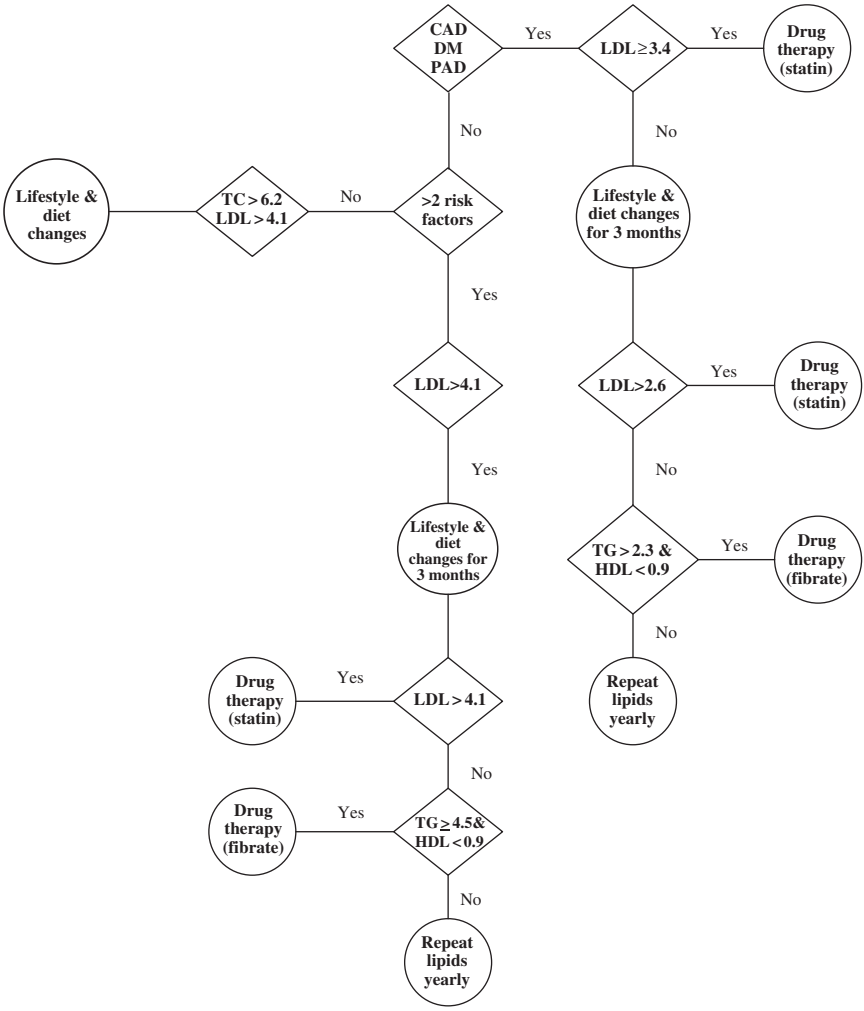
Liver Disease

Chronic active liver disease such as hepatitis B carrier, is common in this part of the world. The presence of chronic liver disease, whether from hepatitis or alcoholic abuse, is not an absolute contraindication to lipid-lowering therapy. We recommend screening the liver functions (especially ALT and AST) on 2 consecutive occasions, in known hepatitis carriers and alcoholics prior to initiation of therapy. If levels of ALT and AST are consistently under 3 times the upper limit of normal ranges,

statins (for hypercholesterolemia) or fibrates (for hypertriglyceridemia) can be initiated, albeit starting at low doses and checking for side-effects after 4–6 weeks.

Fatty liver is another condition that is not uncommon, often associated with hypertriglyceridemia. In such instances, treating the underlying lipid problems may resolve the pathology but such patients often have elevated ALT and AST prior to starting therapy. We recommend an ultrasound of

Fig. 1 Management of hyperlipidemia — an evidence-based approach.



the hepatobiliary system to exclude other underlying causes and tests for hepatitis markers before labeling such as fatty liver. If a concomitant high TG is found, there may be indications to start fibrate therapy even if transaminases are elevated above 3 times the upper limit. Such therapy should preferably be carried out under close supervision by trained lipidologists. Preliminary data have shown potential benefits for the use of thiazolidenediones such as rosiglitazone and pioglitazone in reducing fatty liver and improving the lipid profile.

Resistant Hyperlipidemia

For the category of patients who remain outside target values despite adequate dietary changes and maximal doses of pharmacotherapy, we recommend referral to specialized lipid units for further management.

REFERENCES

1. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial, *Lancet* **360**:7–22, 2002.
2. Cutter J, Tan BY, Chew SK, Levels of cardiovascular disease risk factors in Singapore following a national intervention programme, *Bull World Health Organ* **79**:908–915, 2001.
3. Tan CE, Emmanuel SC, Tan BY, Jacob E, Prevalence of diabetes and ethnic differences in cardiovascular risk factors. The 1992 Singapore National Health Survey, *Diabetes Care* **22**:241–247, 1999.
4. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III), *JAMA* **285**:2486–2497, 2001.
5. Wood D, De Backer G, Faergeman O, Graham I, Mancia G, Pyorala K, Prevention of coronary heart disease in clinical practice: recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention, *Atherosclerosis* **140**:199–270, 1998.
6. The Asian Pacific CHD Risk Factor Collaborative Group, The Asian Pacific collaboration study on CHD risk factor intervention, *Atherosclerosis* **136**(Suppl):S37, 1998.
7. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S), *Lancet* **344**:1383–1389, 1994.
8. Sacks FM, Pfeffer MA, Moye LA, *et al.*, The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators, *N Engl J Med* **335**:1001–1009, 1996.

9. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels, The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group, *N Engl J Med* **339**:1349–1357, 1998.

10. Lipids, in Clinical Practice Guidelines, Ministry of Health, 7/2001.

11. Haffner SM, Lehto S, Ronnemaa T, Pyorala K, Laakso M, Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction, *N Engl J Med* **339**:229–234, 1998.

12. Shepherd J, Cobbe SM, Ford I, *et al.*, Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group, *N Engl J Med* **333**:1301–1307, 1995.

13. Austin MA, Epidemiology of hypertriglyceridemia and cardiovascular disease, *Am J Cardiol* **83**:13F–16F, 1999.

14. Tan CE, Chew LS, Tai ES, *et al.*, Benefits of micronised Fenofibrate in type 2 diabetes mellitus subjects with good glycemic control, *Atherosclerosis* **154**:469–474, 2001.

15. Tan CE, Chew LS, Chio LF, *et al.*, Cardiovascular risk factors and LDL subfraction profile in Type 2 diabetes mellitus subjects with good glycaemic control, *Diabetes Res Clin Pract* **51**:107–114, 2001.

18

Type 1 Diabetes

Peter Eng, Tan Hwee Huan and Goh Su Yen

INTRODUCTION AND EPIDEMIOLOGY

Diabetes mellitus (DM) is now classified according to the etiology rather than descriptions based upon age of onset and insulin requirements.[1] Type 1 DM refers to diabetes that is due to destruction of the pancreatic beta-cells (β-cells) leading to insulin deficiency. Type 1 diabetics are prone to ketoacidosis and require insulin for survival. Type 1 DM was previously also referred to as insulin-dependant diabetes (IDDM) or juvenile-onset diabetes. The vast majority of Type 1 DM is due to an immune-mediated destruction of the β-cells while a small minority has idiopathic destruction of the β-cells. Type 1 DM comprises between 5–10% of all diabetics worldwide with an extremely varied geographic distribution that seems to be racially dependent. At one end of the spectrum are certain countries in Scandinavia, which have a very high incidence of Type 1 DM at $>20/100\,000$ population. In contrast, China and South America have a very low incidence of Type 1 DM of $<1/100\,000$. The incidence of Type 1 DM in Singapore children is relatively low with an annual incidence of $3.8/100\,000$ in 1994, which is similar to that of Hong Kong and

Japan.[2] In Singapore, there seems to be a preponderance of female Type 1 DM compared to males, with a male to female ratio of 1:1.85. This has also been seen in other Asian populations. This contrasts with Western populations where the male to female ratio approximates 1:1. Although Type 1 DM typically presents in childhood or in adolescence, it can present at any age from childhood to the ninth decade. In the long-term, type 1 diabetes is complicated by microvascular complications of nephropathy, neuropathy and retinopathy. There are also macrovascular complications of coronary artery disease and stroke. Optimal glycemic control is important not only for preventing complications but also to allow the patient to lead as normal a life as possible.

PATHOGENESIS

Type 1 DM results from an immune-mediated destruction of the β-cells of the pancreas. The process occurs in background of genetic susceptibility to the disease. Postulated environmental triggering events then lead to development of autoantibodies to the pancreatic β-cells. The end result is loss of β-cell mass and declining insulin secretion. When there is sufficient loss of β-cell mass, symptomatic diabetes occurs.

Identical twins of patients with Type 1 DM have a 50% risk of also developing the disease. There is considerable heterogeneity in the manifestation of the disease and the risk is greater if the first twin develops the disease at a younger age. The risk of developing Type 1 DM is 5% if a sibling or parent has the disease. The genetic risk of developing Type 1 DM is closely related to the genes coding for the human leukocyte antigens (HLA), located on chromosome 6. HLA is expressed on the surface of cells and their functions include immune recognition and antigen presentation. The most important determinants of Type 1 DM seem to be the HLA DR and DQ alleles. While there is a positive association between type 1 diabetes and DR3 and DR4 alleles in the Caucasian population, only DR 3 seems to be operative in the Chinese population, with the B17 allele conferring the strongest risk in Singapore Chinese.[3] Other non-HLA genes on chromosomes 2, 11 and 15 may also be important in conferring genetic risk for Type 1 DM.

The triggering events that lead to the development of pancreatic autoimmunity are not fully known. Putative environmental triggers include viruses (e.g. coxsackie) and dietary factors (bovine milk, nitrates

in water). Congenital rubella is associated with an increased risk of developing diabetes that is greater than 1/5 and has been postulated that the congenital infection damages the developing immune system.

CLINICAL PRESENTATION AND DIAGNOSIS

Type 1 diabetes in children usually presents with classical symptoms of polyuria, polydipsia, weight loss and tiredness. The duration of symptoms is usually short and may be less than 2–3 weeks. Some patients may present very acutely in diabetic ketoacidosis with nausea, vomiting and dehydration (see chapter 20 on endocrine emergencies). On urine testing, type 1 diabetics are ketone-positive. The hallmark of Type 1 DM is autoimmunity against components of the pancreatic β-cell. Antibodies against glutamic acid decarboxylase (GAD antibodies) and islet cell antibodies (ICA) are the antibodies commonly tested. Although GAD and ICA antibodies are positive in 60–90% of Type 1 DM in Caucasians at the time of diagnosis, there is a much lower incidence of GAD and ICA positivity in Asians. In a recent study in Singapore involving 41 children with Type 1 DM, only 41.5% of them tested positive for either or both GAD and ICA.[4] It is postulated that other autoantibodies yet to be identified may be responsible. Alternatively, non-autoimmune processes may be responsible for a significant proportion of Type 1 DM in Singapore.

In adults with Type 1 DM, the clinical presentation can sometimes be much less acute and patients may initially be diagnosed as type 2. These patients are usually 25 years or older, non-obese and are maintained in good metabolic control for several years on oral hypoglycemic agents before insulin dependency. This form of Type 1 DM with slowly progressive β-cell destruction is termed as latent autoimmune diabetes in the adult (LADA). In a recent study profiling LADA patients in Singapore, it was found that they were mostly females, had a mean body mass index of 17.2 kg/m^2, were diagnosed to have type 2 diabetes in the forth decade of life and became insulin-dependant after a mean of 2.5 years.[5]

CLINICAL COURSE AND LONG-TERM COMPLICATIONS

After initiating insulin treatment in newly diagnosed diabetics, many patients may proceed to a "honeymoon" phase, during which excellent

glycemic control can be achieved with very low doses of insulin. This phase may last from several months to over a year. Hyperglycemia in itself impairs β-cell function and normalizing the blood glucose level results in the temporary improvement in β-cell function as seen during the honeymoon phase. After the honeymoon phase, the insulin requirement increases and insulin is needed for the rest of the patients' lives.

Type 1 diabetics are prone to microvascular complications of retinopathy, neuropathy and nephropathy. The development of these complications is related to both the duration of diabetes and the glycemic control. Microvascular complications start to occur about 5 years after the diagnosis of diabetes. With regards to retinopathy, over 90% of Type 1 DM patients will develop some degree of retinopathy 20 years from diagnosis of DM. In contrast, nephropathy will only affect 35–45% of type 1 diabetics. Unlike retinopathy, the prevalence of nephropathy does not rise continually with the duration of diabetes. Genetic factors may predispose a subset of patients to developing nephropathy. The incidence rates of persistent proteinuria peak at 10–15 years after diagnosis and then decline. Overall, 75% of patients with persistent proteinuria will develop end-stage renal failure in 20 years. Most epidemiological studies were done before the publication of trails showing the benefit of tight glycemic control and hence the long-term outcomes for patients with very good glycemic control may be significantly better (see next page for section on intensive insulin therapy).

Type 1 diabetics are also prone to macrovascular complications. The risk of cardiovascular disease, cerebral vascular disease and peripheral vascular disease are estimated to be 2–4 times that of the non-diabetic population. Macrovascular disease, especially cardiovascular disease, accounts for the majority of deaths in type 1 diabetics.

MANAGEMENT

Insulin Requirements

The daily insulin production is between 24–36 units. Type 1 diabetics with no endogenous insulin require between 0.5–1 units/kg/day. This is divided into basal and prandial needs. Basal requirements are between 40–60% of the total daily dose, and is usually given as 2 injections of an intermediate-acting insulin or 1 injection of a long-acting insulin. The prandial requirements are given as pre-meal injections of short-acting

insulin 2–3 times a day. After acute stabilization of the glycemic control, the newly diagnosed type 1 diabetic can be started on a regimen with a total dose of between 0.2–0.4 units/kg/day. This is then adjusted based on glucose monitoring results. Requirements may diminish drastically during the honeymoon phase but it is recommended to keep these patients on a small dose of insulin.

Intensive Insulin Therapy

Hyperglycemia has always been thought to be the pathogenesis of long-term complications but it was not until 1991 that the Stockholm Diabetes Intervention Study (SDIS) first convincingly suggested that intensified conventional insulin treatment might actually retard microvascular complications.[6] Two years later, the Diabetes Control and Complications Trial[7] (DCCT) provided statistically significant evidence of the beneficial effect of intensive therapy on the prevention of long-term microvascular complications. Since the release of its results in 1993, intensive management of type 1 diabetes has gradually become the expected norm. Both continuous subcutaneous insulin infusion (CSII) via an insulin pump and multiple daily insulin injection therapy are effective means of implementing intensive diabetes management with the goal of achieving near-normal levels of blood glucose and improved lifestyle flexibility.

Prior to the trial, conventional therapy for Type 1 DM would have meant a regimen consisting of 1 or 2 daily injections of insulin, including mixed intermediate and regular insulin. Conventional therapy did not usually include daily adjustments in the insulin regimens. In the DCCT, the intensive regimen was designed to achieve blood glucose values as close to the normal range as possible with 3 or more daily insulin injections or treatment with an external insulin pump (CSII). The insulin dosage was adjusted according to the results of self-monitoring of blood glucose performed at least 4 times per day, dietary intake and anticipatory exercise. The goals of intensive therapy included pre-prandial blood glucose between 3.9–6.7 mmol/L, postprandial concentrations less than 10 mmol/L, a weekly 3 am measurement greater than 3.6 mmol/L and hemoglobin A1c (HbA1c) measured monthly within normal range (less than 6.05%). This therapy was conducted by an expert team of diabetologists, nurses, dietitians and behavioral specialists. The time, effort and cost required were considerable.

The DCCT showed that intensive therapy of patients with type 1 diabetes delayed the onset and slowed the progression of clinically important retinopathy, including vision-threatening lesions, nephropathy and neuropathy, by a range of 35% to more than 70%. The main drawbacks of intensive insulin therapy is the increase in incidence of severe hypoglycemia, which in the DCCT was approximately 3 times higher in the intensive therapy group compared to the conventional group.[8] Weight gain was another problem encountered with the intensive therapy, with an increase of 33% in the mean adjusted risk of becoming overweight.

In the DCCT, 42% of those in the intensive regimen group was placed on CSII. Experience with CSII therapy indicates that candidates for CSII must be strongly motivated to improve glucose control and willing to work with their healthcare provider in assuming substantial responsibility for their day-to-day care. CSII therapy should be provided by a skilled professional team familiar with CSII therapy and capable of supporting the patient.[9]

Insulin Analogs

Type 1 diabetics are prone to marked fluctuations in glycemic control partly due to very poor β-cell reserve and the inability of conventional insulin injections to mimic the normal insulin secretory function of the pancreas. Over the last few years, insulin analogs have been developed to provide more physiological insulin action, reduce hypoglycemic episodes and provide greater flexibility. The analogs are modifications of the insulin molecule to either make it faster-acting or to give it a sustained action without a peak. Two rapid-acting analogs currently available are insulin lispro and insulin aspart. These rapid-acting insulins have an onset of between 5–15 minutes and are given just before the meal. Insulin glargine is a long-acting "peakless" insulin analog that is injected once a day. A common combination would be a single long-acting analog to maintain a constant basal insulin level and several injections of rapid-onset insulin just before meals.

Prevention and Management of Diabetic Complications

Nephropathy

In the DCCT, intensive glucose control has been shown to decrease the occurrence of microalbuminuria by 39% and that of clinical albuminuria

by 54%. In the Collaborative Study Group trial, the addition of captopril to Type 1 DM patients with overt proteinuria reduced the endpoints of doubling of serum creatinine end-stage renal failure or death by 50%.[10] In a meta-analysis of Type 1 DM patients with microalbuminuria that are normotensive, angiotensin-converting enzyme (ACE) inhibitors significantly reduced the progression of microalbuminuria to overt proteinuria and increased the chances of regression to normoalbuminuria.[11] Hence tight glycemic control is recommended for majority of Type 1 DM patients, perhaps with the exception of those with advanced renal disease. ACE inhibitors are recommended for all type 1 patients with microalbuminuria or overt proteinuria, even if the patients are normotensive. Hypertension in a type 1 diabetic often develops when there is nephropathy. Control of blood pressure is essential to delay the progression of nephropathy and recommendations are to target for a systolic blood pressure of less than 130 mmHg and a diastolic pressure of less than 80–85 mmHg.

Retinopathy

The DCCT showed that intensive glucose control marked reduced the development or progression of retinopathy in type 1 diabetics. The Early Treatment Diabetic Retinopathy Study (ETDRS) showed the beneficial effects of laser photocoagulation before the stage of high-risk proliferative retinopathy.[12] Treatment of hypertension may also retard the progression of retinopathy. Trials with aldose-reductase inhibitors and aspirin have not shown any benefit.

Neuropathy

Meticulous glycemic control achieved in the DCCT reduced clinical neuropathy by 60%. Unfortunately, no other treatment has been proven effective in retarding or preventing diabetic neuropathy. Tricyclic antidepressants are helpful in treating the nocturnal symptoms of painful neuropathy but do not alter the disease process. Presently, the most benefit is derived from improvement of foot-care services for detection of the neuropathic foot.

Macrovascular Disease

Unfortunately, the benefit of intensive control on reduction of macrovascular complications is unproven, though several studies including the

DCCT show a trend toward benefit of intensive insulin therapy in reducing the extent of macrovascular events. Since diabetes is now considered an ischemic heart disease equivalent, all type 1 diabetics should have aggressive management of concomitant cardiovascular risk factors.

FUTURE PROSPECTS OF PREVENTION AND CURE FOR TYPE 1 DM

Modern diabetes management will involve new insulin analogs that provide more physiologic insulin delivery, as well as inhaled and oral insulin formulations. Closed-loop insulin administration strategies and approaches for replacement of the insulin-producing β-cells (either by pancreatic organ, islet cell or stem cell transplantation) may offer improved treatments, which could delay or prevent diabetes complications. Other treatment strategies under investigation include the use of non-activating humanized monoclonal antibodies against CD3,[13] as well as gene therapy targeted at tolerance induction and modification of antigen presentation.

Ultimately, however, prevention of Type 1 DM in susceptible individuals represents the best chance for reducing the burden of the disease. This will require reliable methods for early diagnosis of predisposition to the disease, using improved genetic and serological screening on a wide scale. Researchers from the Diabetes Prevention Trial-type 1 (DPT-1) reported that low-dose insulin injections do not prevent Type 1 DM in people with impaired insulin secretion who have a high ($\geqslant 50\%$) risk of developing diabetes in 5 years.[14] The same study group is now testing whether oral insulin can prevent Type 1 DM in people at moderate (25–50 %) risk of developing Type 1 DM within 5 years. Results are also pending for the European Nicotinamide Diabetes Intervention Trail (ENDIT) which is assessing the effects of high-dose nicotinamide in the prevention of Type 1 DM.[15]

REFERENCES

1. Alberti KG, Zimmet PZ for the WHO Consultation, Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation, *Diabet Med* **15**:539–553, 1998.

2. Lee WW, Ooi BC, Thai AC, Loke KY, Tan YT, Rajan U, Tan CL, The incidence of IDDM in Singapore children, *Singapore Med J* **39**:359–362, 1998.

3. Lee BW, Chan SH, Tan SH, Wee GB, Yap HK, Wong HB, Tan CL, Tan KW, HLA-system in Chinese children with insulin-dependent diabetes mellitus: a strong association with DR3, *Metabolism* **33**:1102–1105, 1984.

4. Lee YS, Ng WY, Thai AC, Lui KF, Loke KY, Prevalence of ICA and GAD antibodies at initial presentation of type 1 diabetes mellitus in Singapore children, *J Pediatr Endocrinol Metab* **14**:767–772, 2001.

5. Tan HH, Lim SC, Latent autoimmune diabetes in adults (LADA): a case series, *Singapore Med J* **42**:513–516, 2001.

6. Reichard P, Berglund B, Britz A, Cars I, Nilsson BY, Rosenqvist U, Intensified conventional insulin treatment retards the microvascular complications of insulin-dependent diabetes mellitus (IDDM): the Stockholm Diabetes Intervention Study (SDIS) after 5 years, *J Intern Med* **230**:101–108, 1991.

7. DCCT Research Group, The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus, *N Engl J Med* **329**:977–986, 1993.

8. DCCT Research Group, Epidemiology of severe hypoglycemia in the Diabetes Control and Complications Trial, *The American J Med* **90**:450–459, 1991.

9. American Diabetes Association, Clinical Practice Recommendations 2002, Continuous Subcutaneous Insulin Infusion, *Diabetes Care* **25**:S116, 2002.

10. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD, The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy, The Collaborative Study Group, *N Engl J Med* **329**:1456–1462, 1993.

11. The ACE Inhibitors in Diabetic Nephropathy Trialist Group, Should all patients with type 1 diabetes mellitus and microalbuminuria receive angiotensin-converting enzyme inhibitors? A meta-analysis of individual patient data, *Ann Intern Med* **134**:370–379, 2001.

12. Early photocoagulation for diabetic retinopathy, ETDRS report number 9, Early Treatment Diabetic Retinopathy Study Research Group.Ophthalmology, **98**(5 Supp l): 766–785, 1991.

13. Herold KC, Hagopian W, Auger JA, Poumian-Ruiz E, Taylor L, Donaldson D, Gitelman SE, Harlan DM, Xu D, Zivin RA, Bluestone JA, Anti-CD3 monoclonal antibody in new-onset type 1 diabetes mellitus, *N Engl J Med* **346**:1692–1698, 2002.

14. Diabetes Prevention Trial — Type 1 Diabetes study group, Effects of insulin in relatives of patients with type 1 diabetes mellitus, *N Engl J Med* **346**:1685–1691, 2002.

15. Schatz DA, Bingley PJ, Update on major trials for the prevention of type 1 diabetes mellitus: the American Diabetes Prevention Trial (DPT-1) and the European Nicotinamide Diabetes Intervention Trial (ENDIT), *J Pediatr Endocrinol Metab* **14** (Suppl 1): 619–622, 2001.

This page intentionally left blank

19

Disorders of Calcium Metabolism

Leonard Koh

OSTEOPOROSIS

The most serious complication of osteoporosis is fractures. As the incidence of fractures increases with age, and the proportion of the elderly in many countries is increasing rapidly, osteoporosis, and attendant fractures, will constitute an evolving public health problem. In Singapore, 6% of the population in 1990 was above the age of 65 years, but by 2030, this figure is projected to rise to 19%.

In Singapore, hip fracture incidence rates among men and women above the age of 50 years have risen 1.5-fold and 5-fold respectively since the 1960s, a trend consistently documented in many other countries. The age-adjusted rates among Singapore women over the age of 50 years are currently among the highest in Asia, and approaching those of the West. Vertebral fracture rates among Asian populations are already similar to Caucasian populations.

Among patients who sustained osteoporotic hip fractures in Singapore, a mortality of 20% was demonstrated at 2 years. Of the survivors, 20% became semi- or fully dependent, and 42% became less or

non-ambulant. The main social and economic burden appears to be borne by patients' families, as only 8% were cared for by chronic healthcare facilities.

Pathophysiology

Osteoporosis results as a culmination of events affecting the skeleton: Failure to attain genetically programmed peak bone mass, increased loss of bone subsequently, or a combination of both, and also, deficits in other determinants of bone strength such as bone geometry and bone quality (e.g. architecture, turnover and mineralization).

The main determinant of peak bone mass is genetic factors. Twin and family studies have suggested that 40–80% of the variance in peak bone mass and bone metabolism is genetically determined. Non-genetic factors such as nutrition, exercise, a normal hormone milieu and social habits are other important determinants of peak bone mass.

Bone loss occurs as a result of an increased rate of bone remodeling and an imbalance between the activity of osteoclasts and osteoblasts. The major causes of bone loss relate to estrogen deficiency and aging, where the intricate molecular interactions at the cellular level between estrogen, androgens, parathyroid hormone, vitamin D, growth factors such as insulin-like growth factor-I and transforming growth factor-β, and cytokines such as the interleukins, tumor necrosis factor-α, osteoprotegerin and bone morphogenic proteins, are likely to feature. Other accelerating factors such as drugs, illnesses and toxins also contribute to bone loss.

The normal skeletal architecture is disrupted, with fewer and thinner bony spicules, perforations and loss of trabecular connectivity, especially horizontally, microfractures and fatigue damage, all serving to undermine the structural integrity of the bone.

Clinical Manifestations

Osteoporosis is usually asymptomatic unless the patient experiences pain from a fracture. The fracture suggests skeletal fragility; non-specific body aches and pains should generally not be attributed to osteoporosis unless there is progressive deformity or radiological documentation of fractures. In the absence of fractures, the description of "radiographic osteopenia"

might subjectively suggest the presence of osteoporosis, but this needs to be quantified objectively.

Consequent to fractures, patients may experience pain, deformity, prolonged disability, physical deconditioning due to inactivity, changes in self-image and increased mortality in the case of hip and vertebral fractures. Vertebral compression fractures may also result in loss of height, kyphosis (dowager's hump), crowding of internal organs, and chronic back pain.

Etiology

In several large epidemiological studies, including recent studies from Asia, many risk factors have been found to be associated with osteoporosis, falls and fractures (Table 1).

Table 1 Risk Factors for Osteoporotic Fractures

Risk factors for low bone mass and fracture
- Personal history of previous fracture as an adult
- History of fracture in a first degree relative (especially maternal)
- Low body weight & older age
- Early natural or surgical menopause before age 45 years, or prolonged pre-menopausal amenorrhea lasting >1 year
- Drugs, e.g. corticosteroids (equivalent to prednisolone >7.5 mg/day for more than 6 months), excess thyroxine, anti-convulsants
- Ongoing disease conditions, e.g. hypogonadism, hyperthyroidism, hyperparathyroidism, Cushing's syndrome, chronic obstructive airways disease, liver disease, malabsorption, chronic renal failure, rheumatoid arthritis, organ transplantation and anorexia nervosa
- Prolonged immobilization, poor health or frailty
- Current cigarette smoking
- Alcohol abuse (stronger data in men)
- Lifelong low calcium intake (<500 mg/day among Asians)
- Lack of regular physical activity

Other risk factors for fracture independent of bone mass
- Taller individuals (>1.6 m among Asian women, >1.7 m among Asian men)
- One or more previous falls in the past year
- Strokes, poor balance, weak quadriceps muscle strength
- Impaired eyesight despite adequate correction
- Drugs, e.g. sedatives, polypharmacy
- Environmental factors, e.g. slippery floors, inadequate lighting

Diagnostic Approach

Osteoporosis is defined as a skeletal disorder characterized by compromised bone strength predisposing a person to an increased risk of fracture. Bone strength primarily reflects the integration of bone density and bone quality. This conceptual definition captures the notion that bone mineral density (BMD) is central to the diagnosis of osteoporosis in the absence of a fracture, and that an individual experiencing a fragility fracture (fracture due to minimal trauma, such as after a fall from standing height or less) also has osteoporosis, regardless of bone density. Bone quality refers to architecture, turnover, damage accumulation and mineralization.

Operationally, osteoporosis is defined as a value for BMD that is 2.5 standard deviations or more below the mean value in young adults (T-score $\leqslant -2.5$ SD). This allows standardization of the diagnosis, and has been widely used in epidemiological and clinical intervention trials, and in clinical practice.

In recent years, several techniques have been available for the measurement of BMD, but the most well-validated technique of choice for diagnosing osteoporosis remains dual-energy X-ray absorptiometry (DXA) of the hip. The lumbar spine DXA (antero-posterior projection) is the best site for monitoring therapeutic response.

Secondary causes of osteoporosis should be considered and evaluated for (Table 2), since part of the management of osteoporosis would be to reverse these causes as far as possible. There is no role for biochemical markers of bone turnover in the diagnosis of osteoporosis although there may be a role as a risk factor in predicting fracture risk or in monitoring response to anti-resorptive therapy. Patients should not be treated with drugs for osteoporosis until a definitive diagnosis of osteoporosis is made.

Selection of Individuals for BMD

Because BMD measurements are relatively costly, clinical evaluation tools have been developed to assist in proper selection of individuals for diagnostic BMD testing, guided by the presence of risk factors for osteoporosis.

The US National Osteoporosis Foundation (NOF) uses an *age* criterion in recommending BMD testing in all women over the age of 65 years. In *postmenopausal* women under 65 years of age, BMD is recommended in those with risk factors for hip fracture independent of bone density, including history of previous fracture as an adult, family history of

Table 2 Laboratory Investigations to Consider

Routine
- Bone mineral density (BMD)
  Indications for measuring bone density
  - Previous fragility fracture
  - Individuals at high risk for osteoporosis
  - Radiological evidence of osteopenia or vertebral deformity
  - Women who are considering therapy for osteoporosis, if BMD testing would facilitate the decision
  - Monitoring of treatment
- Relevant radiographs to document fractures
- Full blood count, sedimentation rate
- Creatinine, urinalysis, calcium, phosphate, intact parathyroid hormone (iPTH), liver function tests

Optional
- 24 hr urinary calcium
- Bone turnover markers, e.g. osteocalcin, C-telopeptide, N-telopeptide
- Endocrine tests: Free thyroxine (fT_4), thyroid stimulating hormone (TSH), 25-hydroxyvitamin D (25-OHD), follicle stimulating hormone (FSH), luteinizing hormone (LH), prolactin, urinary free cortisol, dexamethasone suppression test
- Testosterone levels in men
- Tumor markers, myeloma screen, bone marrow examination
- Scintigraphic bone studies

fracture, body weight < 57 kg or current cigarette smoking. It is not yet clear if these criteria are applicable to Asians.

The Osteoporosis Self-assessment Tool for Asians (OSTA), which is based on *age* and *weight*, categorizes postmenopausal Asian women into high, moderate and low risk of having osteoporosis on subsequent BMD measurement. BMD is recommended for women in the high-risk category (Table 3), and in the moderate-risk category if they had risk factors for low bone mass, particularly if multiple risk factors were present (Table 1). In the low-risk category, the prevalence of osteoporosis is sufficiently low to defer BMD unless the woman has a past fragility fracture or has a condition known to be associated with osteoporosis.

Management

Prevention of osteoporosis

The goal of strategies to prevent osteoporosis would be to optimize skeletal development during childhood and adolescence, maximize peak bone

Table 3 Osteoporosis Self-assessment Tool for Asians (OSTA)

Age (Yr)	Weight (kg)							
	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79
45–49								
50–54								
55–59						Low Risk		
60–64								
65–69			Moderate Risk					
70–74								
75–79	High Risk							
80–84								
85–89								

Copyright 2001, MSD, Whitehouse Station, NJ, USA. All rights reserved. Reproduced with permission.

mass at skeletal maturity, and preserve the structural integrity of the skeleton by preventing age-related and secondary causes of bone loss, thereby preventing fracture.

A population-based strategy addresses lifestyle issues such as good general nutrition, and in particular, adequate dietary calcium and vitamin D intake (Tables 4, 12 and 13), regular weight-bearing exercise, and discouraging cigarette smoking and alcohol abuse in an attempt to improve the overall health of the population. The risk of falling, or the degree of injury from falls, could be reduced by measures such as modifying the environment. However, there remains scant evidence about the effect of these changes on fracture risk.

The use of pharmacological agents in patients with osteopenia (BMD T-score between −1 and −2.5 SD) but without fractures for 'prevention', to maintain bone density above the osteoporotic threshold, is debatable. It must be borne in mind that the most important measure of the effectiveness of any osteoporosis therapy is anti-fracture efficacy. Apart from hormone replacement therapy, which prevents fractures in early postmenopausal women, other therapies, when initiated in osteopenic patients over periods of 2–4 years, have resulted in increased BMD, but have generally not been shown to reduce fracture risk. These drugs include alendronate, cyclical etidronate, risedronate, raloxifene, calcitonin and tibolone (a synthetic steroid primarily used to treat climacteric

Table 4 Recommended Dietary Allowance of Calcium and Vitamin D

Category	Calcium (mg)	Vitamin D (IU)
Adolescents 10–18 years old	1000	400
All adults 19–50 years old	800	400
All adults above 50 years	1000	400–800
Pregnancy & lactation	1000	400

symptoms). These might perhaps be considered if the BMD approaches the osteoporotic range, or if the patient is at high risk of fracture.

Treatment of osteoporosis

The goal of strategies to treat osteoporosis would be to prevent fractures, stabilize or increase bone mass, relieve symptoms of fractures, retard skeletal deformity and maximize physical function.

Patients with established osteoporosis, either with BMD T-score $\leqslant -2.5$ SD or with fragility fractures, should be considered for treatment with drugs shown to reduce the risk of the first fracture or further fractures. Treatment of patients with baseline fragility fractures (e.g. vertebral fractures) is probably more cost-effective because incidence of subsequent fracture is higher compared to patients without fracture. Unfortunately, evidence suggests that the majority of patients who have had fragility fractures do not appear to subsequently receive osteoporosis therapy, which has been shown to reduce the risk of second fractures.

Early surgical management of hip fractures is essential to decrease mortality rate and to improve the significant perioperative morbidity, especially in the frail elderly. In patients with major pain related to a crushed vertebra, vertebroplasty, involving injection of bone cement into the vertebral body, has proven fairly effective in controlling symptoms.

The decision to institute medical therapy should be guided by an integration of fracture risk, considering the patient's age, BMD, and the presence of other risk factors for fracture, falls or bone loss.

The drugs with the most rigorous experimental data showing fracture reduction in osteoporotic individuals with or without prevalent fractures include the anti-resorptive agents, alendronate, risedronate and raloxifene, and the anabolic agent, parathyroid hormone (PTH). Alendronate and risedronate are bisphophonates. Daily alendronate

increases bone density and consistently reduces the risk of fractures, including those in the spine and hip. A more convenient, once-weekly dosing regimen of alendronate is as effective in increasing bone density. Risedronate increases bone density and reduces vertebral, non-vertebral and hip fractures. A once-weekly dosing regimen is also available. Raloxifene is a selective estrogen receptor modulator (SERM), which prevents bone loss and reduces vertebral fracture risk, but not non-vertebral fracture risk. Raloxifene appears to decrease the risk of breast cancer and cardiovascular events in women with high cardiovascular risk, but is associated with an increased risk of thromboembolism. Intermittent recombinant human PTH (1–34) therapy has a marked anabolic effect on bones and has recently been demonstrated to reduce vertebral as well as non-vertebral fracture risk.

Other drugs with some evidence for anti-fracture efficacy, mostly in women with prevalent vertebral fractures, include hormone replacement therapy (HRT), intranasal calcitonin, other bisphosphonates (cyclical etidronate and oral pamidronate), and strontium ranelate (an anabolic agent). The recently reported Women's Health Initiative showed that HRT decreases vertebral, non-vertebral and hip fractures. However, the overall risk-benefit ratio in terms of endpoints such as coronary, cerebrovascular and thromboembolic disease and total cancer was unfavorable. Calcitonin has analgesic properties in the setting of acute vertebral crush fractures. Oral pamidronate and strontium ranelate have recently been shown to increase bone density and reduce vertebral fractures.

The effectiveness of vitamin D analogs, such as calcitriol and alphacalcidol, and fluoride remain doubtful as both reductions and increases in vertebral fractures have been demonstrated in different trials. Fluoride is anabolic to bone while the vitamin D analogs have inconsistently decreased bone loss. No protective effect has been demonstrated at the hip.

Calcium and vitamin D remain as supplementary to other forms of treatment for osteoporosis, except in housebound individuals, where calcium and vitamin D deficiency are likely. In such individuals, calcium and vitamin D supplementation alone has resulted in fracture reduction.

The recommended doses of available osteoporosis drugs (calcium doses refer to elemental calcium) are:

- Oral *alendronate* 10 mg daily with calcium 500 mg daily or 70 mg once weekly with calcium 500 mg daily.

- Oral *risedronate* 5 mg daily with calcium 500 mg daily or 35 mg once weekly with calcium 500 mg daily.
- Oral *raloxifene* 60 mg daily with calcium 500 mg daily.
- *Hormone replacement therapy* (cyclical or continuous combined) with calcium 500 mg daily. Prescribe the lowest effective dose of estrogen and titrate upwards until the desired therapeutic effect is achieved.
- Intranasal *calcitonin* 200 IU daily with calcium 500 mg daily.
- Cyclical oral *etidronate* 400 mg daily for 2 weeks followed by calcium 500 mg daily for 11 weeks, then repeat cycle.
- Oral *clodronate* 800 mg daily with calcium 500 mg daily.
- Oral *alfacalcidol* 0.5–1 μg daily usually without calcium supplements.
- Oral *calcitriol* 0.5 μg daily usually without calcium supplements.
- Intramuscular *nandrolone* 25–50 mg once every 3 weeks for up to 6 months.
- Oral *tibolone* 2.5 mg daily.

Monitoring of treatment

As a reduction in fractures is difficult to document, the most important surrogate marker of therapeutic response is the follow-up bone density measurement in comparison with the baseline, which is usually performed at an interval in excess of one year, the lumbar spine site usually showing the most marked changes.

In patients who fail to respond, the following should be considered: non-compliance to therapy, incorrect administration of drug (e.g. failure to follow specific instructions on how to take the drug, interactions with other drugs), ongoing, undiagnosed pathology that accelerates bone loss, imprecision of the BMD measurement technique, and true treatment failure, in which case an alternative therapy might be considered.

The occurrence of a fracture, however, does not necessarily imply treatment failure, particularly if bone density has risen, and the osteoporosis drug should be continued. If the patient has frequent multiple fractures, combination therapy to enhance bone density further may be attempted, despite the lack of data on additive anti-fracture efficacy.

Male Osteoporosis

Worldwide, and in Singapore, approximately 20% of symptomatic vertebral fractures and 30% of hip fractures occur in men, the incidence rates of

hip fractures having risen over time. Epidemiological data suggests that all major fractures are associated with increased mortality, especially in men.

Although the data is scanty, the relationship between BMD and relative risk of hip fracture appears similar in men and women. This has led to some controversy with regard to the diagnostic threshold for osteoporosis in men. There has been a suggestion to use the BMD value (in g/cm^2), which corresponds to the female femoral neck T-score of -2.5 SD, as opposed to the male T-score of -2.5 SD, as the male threshold value. It would be necessary to obtain this female threshold value from the center providing the BMD result since this value is specific to each measurement device.

Secondary causes of osteoporosis are more commonly found among men. Investigations for hypogonadism, high alcohol intake, corticosteroid therapy, and medical disorders associated with secondary osteoporosis and muscular instability, should always be considered. Other risk factors for fractures in men include past fracture from age 50 years, physical inactivity, recent falls, sedative use, low body mass index and smoking.

Established treatment options for osteoporosis in men are fewer than for women. Alendronate has been shown to increase bone density and reduce vertebral fracture risk in men. Another therapy which has shown similar benefits is low-dose intermittent monofluorophosphate (3 months on, 1 month off), but caution on its use is recommended in view of the inconsistent reduction of fractures seen with this drug among women. There is evidence that calcium and vitamin D supplementation may be useful in preserving bone density and reducing non-vertebral fractures in men. Other drugs which have been shown to be beneficial in increasing or preserving bone density, but without data on fracture reduction, include intermittent PTH therapy, cyclical etidronate, calcitonin, and testosterone.

Conclusion

As awareness of the importance and impact of osteoporosis increases, continued research and the evolution in knowledge have allowed this problem to be addressed on many fronts. The challenge remains to manage the majority of individuals suffering from this worldwide epidemic, by improving on the effectiveness of risk evaluation, diagnostics and therapeutics, and translating these into patient care, through the most cost-efficient strategies.

HYPERCALCEMIA

Hypercalcemia is a relatively common clinical problem with the widespread use of routine biochemical screening. Population studies have suggested a prevalence of 3% in women and <1% in men above the age of 60 years. The clinical presentation varies from an asymptomatic biochemical abnormality, which is increasingly common, through symptomatic disease, to a life-threatening medical emergency. The presentation determines to a large extent the therapeutic measures that might be utilized in managing this problem.

Pathophysiology

Calcium has 2 major roles: It provides structural integrity in the skeleton, and it is critically important in the maintenance and control of major biochemical processes. The concentration of calcium in extra- and intracellular fluids is maintained with great accuracy. Calcium is distributed in 3 major fractions: The ionized portion (~50%), which is physiologically important; the protein-bound portion (~40%), most of which is bound to albumin; and a complexed portion (~10%) linked mainly to citrate, or sulphate. Under normal circumstances, the plasma calcium concentration reflects a balance between the flux of calcium between the extracellular fluid and the intracellular fluid, skeleton, gastrointestinal tract, kidney and skin. This calcium flux is regulated by PTH, 1,25-dihydroxyvitamin D ($1,25(OH)_2D$) and calcitonin. The normal serum total calcium level ranges between 2.1–2.6 mmol/L.

In response to hypercalcemia, the secretion of PTH decreases, and this impacts on several target organs. Calcium and phosphate mobilization from the bone decreases, as do calcium reabsorption from the distal renal tubules, urinary phosphate excretion and synthesis of $1,25(OH)_2D$. The latter in turn decreases the efficiency of calcium and phosphorus absorption in the intestine. Whether the putative effects of calcitonin in counter-regulating PTH are of pathophysiologic importance in humans remains unclear. PTH-mediated distal tubular calcium reabsorption and osteoclastic bone resorption are the major control points in minute-to-minute serum calcium homeostasis, with the former being quantitatively more important. Maximal adjustments to the rate of calcium absorption

in the intestine via the PTH−1,25$(OH)_2$D axis require 24–48 hours to become fully operative.

Hypercalcemia results when the entry of calcium into the circulation exceeds the excretion of calcium into the urine or deposition in bone. The limiting factor in the system's defence against hypercalcemia is the capacity of the kidneys to excrete calcium.

'Factitious' hypercalcemia may occur with severe dehydration and volume contraction, where there is relative hyperalbuminemia. As about half of the calcium in serum is bound to protein, principally albumin, an elevated serum total calcium concentration without any rise in the serum ionized calcium concentration may occur, the patient actually being eucalcemic.

Clinical Manifestations

Most patients now present with an asymptomatic biochemical abnormality or minimally symptomatic disease. Although hypercalcemia can affect a variety of organ systems, the most common symptoms are relatively non-specific (Table 5), seen in 30–75% of patients, which often resolve

Table 5 Clinical Features of Hypercalcemia

Asymptomatic
Non-specific fatigue, lethargy, muscle weakness, anorexia, depression
Gastrointestinal constipation, vague abdominal pain, nausea, vomiting, thirst, dry mouth, polydipsia, dehydration, peptic ulcer disease, pancreatitis
Genitourinary polyuria, nocturia, renal tubular dysfunction, nephrolithiasis, nephrocalcinosis, acute and chronic renal insufficiency, uremia
Neuropsychiatric cognitive dysfunction, depression, anxiety, irritability, psychosis, dementia, headache, blurred vision, drowsiness, confusion, stupor, coma, hyporeflexia, corneal or conjunctival calcification
Musculoskeletal gout, pseudogout, chondrocalcinosis, arthralgia, bone pain, osteoporosis
Cardiovascular hypertension, calcification in aortic and mitral valves, and myocardium, ECG changes: Short QT, arrhythmia, cardiac arrest

with the reversal of hypercalcemia. In symptomatic individuals, the extent of symptoms may broadly be attributed to both the degree, and the rate of onset, of the elevation in the serum calcium concentration, or to the underlying disease. Thus, a serum calcium of 3–3.5 mmol/L may be well-tolerated chronically, while an acute rise to these concentrations may cause marked changes in sensorium.

There are usually no specific physical findings of hypercalcemia other than those related to an underlying disease such as malignancy. Band keratopathy is a very rare finding in patients with hypercalcemia. The subepithelial calcium phosphate deposits extend as a horizontal band across the cornea in the area that is exposed between the eyelids, and is usually detected by slit-lamp examination.

Etiology

Hyperparathyroidism and malignancy account for 80–90% of all hypercalcemic patients (Table 6). In ambulatory patients, hyperparathyroidism is the most common cause (60% of cases), as other manifestations of

Table 6 Differential Diagnosis of Hypercalcemia

More common
Primary hyperparathyroidism: Solitary adenomas (80%); hyperplasia (15–20%); double or multiple adenomas (2%); carcinoma (<1%)
Malignancy-associated: Humoral hypercalcemia, especially parathyroid hormone-related protein (PTH-rP) (80%); localized osteolysis (20%); calcitriol production by tumor (1%)
Associated with chronic renal failure (multifactorial)
Vitamin D analog or vitamin D therapy
Less common
Miscellaneous: Immobilization, dehydration, total peripheral nutrition
Granulomatous diseases: Sarcoidosis
Drugs & minerals: Milk-alkali syndrome, thiazides, lithium, aminophylline, aluminium toxicity, vitamin A intoxication
Endocrine disorders: Thyrotoxicosis
Rare
Endocrine disorders: Pheochromocytoma, acromegaly, acute adrenal insufficiency
Granulomatous diseases: Tuberculosis, histoplasmosis, leprosy
Miscellaneous: Hypophosphatasia, familial benign hypocalciuric hypercalcemia (FBHH)

underlying malignancies are usually clinically apparent when hypercalcemia is first noted. Among hospitalized patients, cancer is the most common cause (about 65% of cases), while hyperparathyroidism accounts for another 25%. Occasionally, an occult tumor gives rise to hypercalcemia, and the clinical picture may resemble primary hyperparathyroidism. The more chronic the hypercalcemia, such as over the period of a year, the less likely malignancy is the cause.

Renal failure *per se* is usually not associated with hypercalcemia because of the calcium-lowering effects of concurrent hyperphosphatemia and decreased calcitriol synthesis. However, problems which contribute to hypercalcemia in chronic renal failure patients include the administration of calcium compounds to bind dietary phosphate, vitamin D analogs to reverse both hypocalcemia and secondary hyperparathyroidism, immobilization, dehydration, tertiary hyperparathyroidism and aluminium intoxication. As many as 50% (81 of 162) of hospitalized patients with hypercalcemia were reportedly associated with chronic renal failure.

A high calcium intake alone rarely results in hypercalcemia, because the initial elevation in serum calcium concentration usually inhibits both the release of PTH and the synthesis of calcitriol. However, in susceptible individuals, excessive intake of calcium carbonate to treat osteoporosis and dyspepsia may result in hypercalcemia from the milk-alkali syndrome.

Diagnostic Approach

For clinical purposes, the total calcium concentration is the most widely evaluated index of calcium status. The ionized calcium concentration is technically more difficult to perform and less widely used, although it provides a more precise estimate in situations such as critically ill patients with decreased serum protein levels, acid-base disturbances or those who have received large volumes of citrated blood products.

Normal values for serum total calcium vary somewhat among clinical laboratories, but in general, range from 2.1–2.6 mmol/L. For routine clinical interpretation, the total serum calcium concentration (in mmol/L) is corrected for albumin concentration (in g/L) using: *corrected serum calcium = measured calcium + 0.02(40 − serum albumin)*.

A single raised serum calcium value should be repeated to confirm the diagnosis, together with serum phosphate and intact PTH (iPTH) levels. Because primary hyperparathyroidism and malignancy account

Table 7 Laboratory Investigations to Consider

- serum calcium, phosphate, albumin, ionized calcium
- serum iPTH

Likely problem	Ca^{2+}	PO_4	iPTH
PTH-mediated	↑ / ↑↑	↓ / N	↑ / N
Malignancy: PTHrP-mediated	↑↑	↓	↓
Malignancy: Local osteolysis	↑↑	N	↓
Vitamin D excess (supplements, granulomas)	↑	↑ / N	↓
Thyrotoxicosis; milk-alkali syndrome; thiazides	↑	N / ↑	↓
FBHH (rare)	↑	N	↑ / N
Tertiary hyperparathyroidism	↑	↑	↑

Several problems may superimpose on calcium and phosphate levels: In dehydration, Ca ↑ and PO_4 N/↑; in immobilization, Ca ↑ and PO_4 N; in poor nutrition states, Ca ↓ and PO_4 ↓; in renal dysfunction, Ca ↓/N and PO_4 ↑.

- urea & electrolytes
- liver function test
- myeloma screen
- chest radiograph, skeletal radiology, where indicated
- bone scan, parathyroid scan
- bone turnover markers
- 25(OH)D
- tumor markers
- endoscopy and biopsy of nasopharynx, gastrointestinal or genito-urinary systems, where indicated
- intraabdominal imaging, where indicated
- [PTH-rP & $1,25(OH)_2D$: Both currently unavailable locally]

for the majority of cases, more involved diagnostic investigations should be undertaken only after these have been excluded (Table 7).

The degree of hypercalcemia in primary hyperparathyroidism is often milder than that of malignancy. In patients with primary hyperparathyroidism, 10–20% have a serum iPTH concentration in the upper end of the normal range, which is inappropriate in the presence of hypercalcemia.

Treatment

The decision to institute therapy and the choice of therapy would depend on several factors such as the level of serum calcium, the presence of symptoms and clinical setting (acute and life-threatening vs. chronic and asymptomatic), and the underlying cause.

Table 8 Therapy for Acute Hypercalcemia

Treatment (*mechanism*)	Usual Dose	Onset of Action	Duration (days)	Response (%)
Increase urinary calcium excretion				
Saline	~2.5–4 L/d	2–4 hr	1–2	0–10
Saline + loop diuretic e.g. frusemide	20–40 mg q2–12 hr	2–4 hr	1–2	0–10
Decrease bone resorption				
Clodronate	300 mg/d × 5 d or 1500 mg	1–2 d	10–14	~ 80
Pamidronate	90 mg	1–2 d	11–28	70–90
Ibandronate	4–6 mg	1–2 d	18–26	70–80
Zoledronate	4–8 mg	1–2 d	32–43	~ 90
Calcitonin	4–8 IU/kg q6–12 hr	4–6 hr	2–3	60–70
[Mithramycin]	25 mg/kg	1–2 d	5–7	40–60
[Gallium nitrate]	200 mg/m^2/d	2–3 d	10–14	70–80
Decreased intestinal calcium absorption				
Corticosteroids	40–60 mg/d	2–5 d	—	—
Exchange with extracorporeal calcium pool				
Dialysis	—	hr–d	—	—

General measures such as hydration and calciuresis may suffice in patients with mild to moderate, asymptomatic hypercalcemia (serum calcium usually < 3.0 mmol/L). Specific measures such as bisphosphonates, calcitonin or glucocorticoids (Table 8) may be considered in patients with symptoms and serum calcium > 3.0–3.5 mmol/L. The decision on measures employed may be tempered by the presence of advanced malignancy with no definitive treatment plans and grave prognosis.

The basic goals of therapy for hypercalcemia are to correct dehydration, to lower the serum calcium concentration (by enhancing renal excretion of calcium and inhibiting accelerated bone resorption) and, where possible, to reverse the underlying disease.

General Measures

Hydration

The first step in management of severe hypercalcemia (serum calcium > 3.0 mmol/L) is volume re-expansion. This enhances renal calcium clearance by increasing filtration of calcium, reducing renal tubular

sodium and calcium reabsorption, and causing obligatory calciuresis. Serum calcium may decline by ~0.5 mmol/L. The rate of administration of isotonic saline should be based on the severity of the hypercalcemia, the extent of dehydration, and the tolerance of the cardiovascular system for volume expansion, e.g. in patients with heart failure or renal failure.

Calciuresis

A loop diuretic (e.g. IV or oral frusemide) can be added once hypervolemia-induced natriuresis has been induced, and this may help guard against volume overload. The patient should be monitored carefully to prevent recurrent hypovolemia, which will limit the calciuresis.

Specific Measures

General measures may reduce serum calcium concentrations appreciably, but do not impact on excessive mobilization of calcium from bone, which is the pathogenetic process in the majority of cases. If serum calcium remains above 3.0 mmol/L despite adequate saline/loop diuretic therapy, specific therapy should be implemented. There is currently no longer any role for intravenous phosphate in the management of acute hypercalcemia.

Bisphosphonates

The bisphosphonates are relatively non-toxic compounds that are more potent than intravenous saline and are very effective in patients with moderately severe hypercalcemia. Their maximum effect occurs in 2–4 days. Clodronate, pamidronate, ibandronate and zoledronate are currently available for the treatment of hypercalcemia.

Clodronate is given intravenously initially to achieve normocalcemia. The patient may then be maintained on oral clodronate 800–1600 mg bd. *Pamidronate* is more potent than clodronate with longer-lasting hypocalcemic effects, and is generally well-tolerated. The dose could be repeated after 2–4 weeks or more, depending on the recurrence of hypercalcemia. Although pamidronate is excreted by the kidney, it appears to be safe and effective for the treatment of hypercalcemia in dialysis patients induced by the combination of calcium carbonate and calcitriol, and hypercalcemia

related to renal insufficiency and immobilization. *Ibandronate* is a newer bisphosphonate with a similar calcium-lowering profile to pamidronate. *Zoledronate* is another promising bisphosphonate with a longer duration and greater hypocalcemic effect than pamidronate.

Calcitonin

Salmon calcitonin is safe and nontoxic, and is given intramuscularly or subcutaneously. It lowers serum calcium concentration rapidly, but usually only modestly, and transiently. Simultaneous administration of glucocorticoids may enhance and prolong the effectiveness of calcitonin. There is an additive effect when calcitonin is given with a bisphosphonate.

Mithramycin and gallium nitrate

These drugs are invariably mentioned in most articles on this subject, but are seldom used locally as they are not readily available and are toxic. In addition, extensive experience with them is lacking.

Corticosteroids

Corticosteroids, e.g. prednisolone, are the therapy of choice for patients with vitamin D excess such as in granulomatous diseases, vitamin D toxicity, and in occasional patients with malignant lymphomas. They decrease calcitriol production by the activated mononuclear cells in the lung and lymph nodes. If this is ineffective or not tolerated, chloroquine, hydroxychloroquine or ketoconazole, which similarly reduce calcitriol production, could be given.

Dialysis

Hemodialysis should be considered, in addition to the above treatments, in patients who have serum calcium concentrations in the range of 4.5–5 mmol/L and neurologic symptoms but a stable circulation. Hemodialysis with little or no calcium in the dialysis fluid and peritoneal dialysis (though it is slower) are both very effective modes of therapy for hypercalcemia. Dialysis is particularly useful in patients with renal insufficiency or congestive heart failure who cannot safely be given intravenous saline.

Correction of Underlying Disease

After managing the episode of acute hypercalcemia, it is important to determine the underlying disease, and treat it to prevent recurrence of hypercalcemia.

Primary hyperparathyroidism

To date, parathyroid surgery remains the only effective treatment for primary hyperparathyroidism. Hyperparathyroid patients should be evaluated for indications for parathyroid surgery. The need for preoperative localization of parathyroid tissue vs. surgical exploration, and type and extent of parathyroid surgery, are somewhat controversial. The basis for deferring surgery in asymptomatic individuals with hyperparathyroidism is the lack of progression and generally benign long-term sequelae, although some recent epidemiological data seems to contradict this.

Conservatively managed patients should be evaluated clinically, with 6–12 monthly measurements of serum calcium, creatinine and parathyroid hormone, and annual cortical and trabecular bone density measurements. Dehydration should be avoided, and calcium intake kept at or below 1000 mg per day. Most long-term medical therapies appear unable to lower serum calcium levels significantly, although several agents such as estrogen, alendronate, and raloxifene may do so marginally, in addition to improving low bone density, which may be the only manifestation of mild primary hyperparathyroidism. Their effects on fracture reduction remain unknown.

Hypercalcemia of malignancy

The control of the underlying malignancy with anti-tumor therapy is paramount in achieving long-term control of hypercalcemia. Without this, hypercalcemia, which may be temporized initially by the measures outlined above, invariably recurs and eventually becomes resistant to therapy. More recently, it has been reported that bisphosphonates appear to prevent hypercalcemia in patients with breast cancer and multiple myeloma, in addition to lowering the incidence of other skeletal complications and symptoms.

Potential Future Approaches

Osteoprotegerin (OPG) specifically inhibits osteoclast differentiation by acting as a decoy receptor for OPG ligand (OPG-L or RANKL). OPG-L is believed to be the final common pathway by which multiple hormones and cytokines regulate osteoclast formation, the biological activity being neutralized by binding to OPG. In a murine model of humoral hypercalcemia of malignancy, OPG prevented and reversed hypercalcemia. Studies on humans are ongoing.

Conclusion

Hypercalcemia is a relatively common clinical problem. The therapeutic approach to hypercalcemia should be tailored to the clinical setting, with the aims of controlling hypercalcemia, and finding and instituting therapy for the underlying cause.

HYPOCALCEMIA

Hypocalcemia is encountered commonly in medical practice. Hypocalcemia, like hypercalcemia, varies in its clinical presentation from an asymptomatic biochemical abnormality to a severe life-threatening condition.

Pathophysiology

The physiology of calcium balance has been discussed in the previous section. Hypocalcemia results from either increased loss of ionized calcium from the circulation (deposition in tissue, loss in urine, or increased binding of calcium in serum) or decreased entry of calcium into the circulation (gastrointestinal malabsorption, decreased bone resorption). The major determinants of serum calcium are the serum phosphate concentration (acutely) and serum PTH and vitamin D metabolite concentrations (chronically).

Clinical Manifestations

Symptoms and signs of hypocalcemia are related to the severity and chronicity of the hypocalcemia, as well as the underlying cause. The hallmark of acute hypocalcemia is tetany (Table 9), which is uncommon unless the serum total calcium concentration falls below 1.8–1.9 mmol/L.

Table 9 Clinical Features of Hypocalcemia

Asymptomatic
Non-specific fatigue, hyperirritability, anxiety, depression
Neuromuscular acute: Tetany (circumoral numbness, paresthesias of the hands and feet, muscle cramps, carpopedal spasm, laryngospasm), seizures (grand mal, petit mal, focal), autonomic (diaphoresis, bronchospasm, dysphagia, biliary colic) chronic: Extrapyramidal (parkinsonism, dystonia, choreoathetosis), basal ganglia calcifications, ocular (cataracts, papilloedema), intellectual (mental retardation), psychiatric (confusional state, psychosis)
Cardiovascular bradycardia, hypotension, impaired cardiac contractility, congestive heart failure, digitalis insensitivity ECG: Prolongation of the QT interval and ST-segment, T-wave abnormalities
Ectodermal dry skin, coarse and sparse hair, brittle nails, alopecia, mucocutaneous candidiasis (in idiopathic hypoparathyroidism)
Dental delayed dentition, dental caries, enamel hypoplasia
Skeletal rickets, osteomalacia (in vitamin D deficiency and hypophosphatemia) Albright's hereditary osteodystrophy, osteitis fibrosa cystica (in pseudohypoparathyroidism)
Gastrointestinal steatorrhea

Other factors that adversely influence the frequency and severity of symptoms include acid-base status (alkalosis), hypomagnesemia, hypokalemia and increased levels of epinephrine. Patients in whom the onset of hypocalcemia is gradual tend to have fewer symptoms.

The classic physical findings in patients with latent tetany are Trousseau's and Chvostek's signs. Trousseau's sign is the induction of carpal spasm by inflation of a sphygmomanometer above systolic blood pressure for 3 minutes. Chvostek's sign is the contraction of the ipsilateral facial muscles (corner of the mouth, nose and eye) elicited by tapping the facial nerve just anterior to the ear. Contraction of the corner of the mouth alone occurs in about 25% of normal subjects. Although Trousseau's sign is more specific than Chvostek's sign, both may be negative in patients with hypocalcemia, even in those with hypocalcemic seizures.

Etiology of Hypocalcemia in Adults

Hypocalcemia is seen relatively frequently in hospitalized patients, the most common cause being apparent hypocalcemia, in relation to hypoalbuminemia (Table 10). Truly low serum calcium concentrations are caused mainly by disorders of PTH or vitamin D, as well as sepsis, and disorders that result in decreased serum ionized calcium concentration through binding of calcium within the vascular space or calcium deposition in tissues.

Table 10 Differential Diagnosis of Hypocalcemia

Apparent Hypocalcemia
hypoalbuminemia

True Hypocalcemia
Parathyroid disorders
Hypoparathyroidism (reduced PTH secretion)
- Post-surgical, after parathyroid, thyroid or radical neck surgery
- Autoimmune: Isolated, end-organ deficiency syndrome
- Other causes: Irradiation, infiltrative diseases (hemochromatosis, Wilson's disease, granulomas, metastatic cancer)
- Congenital: DiGeorge's syndrome, calcium sensor mutation
- Severe magnesium deficiency, severe sepsis

Resistance to PTH action
- Pseudohypoparathyroidism
- Severe magnesium deficiency, severe sepsis

Vitamin D disorders
Inadequate intake or malabsorption of vitamin D
Inadequate 25(OH)D formation: Severe liver disease
Inadequate $1,25(OH)_2D$ formation: Renal insufficiency, 1-α hydroxylase deficiency (vitamin-D dependent rickets type I), severe sepsis
Resistance to $1,25(OH)_2D$: Vitamin-D dependent rickets type II, severe sepsis

Loss of calcium from the circulation
Extravascular deposition: Hyperphosphatemia (secondary hyperparathyroidism in renal failure, rhabdomyolysis, tumor lysis, phosphate administration), acute pancreatitis, osteoblastic metastases (breast or prostate cancer), post-parathyroidectomy "hungry bone" syndrome
Intravascular binding: Alkalosis, massive blood transfusion, lactic acidosis due to shock or sepsis, chelating agents (citrate, lactate, foscarnet, EDTA)

Drugs
bisphosphonates, loop diuretic, chemotherapy (cisplatinum, asparaginase, cytosine arabinoside, doxorubicin, 5-fluorouracil-elucovorin)

Diagnostic Approach

Firstly, the diagnosis of hypocalcemia should be confirmed by repeated measurement of serum calcium with serum albumin, verifying with measurements of serum ionized calcium if the presence of hypocalcemia remains in doubt. Serum calcium (mmol/L) is corrected for serum albumin (g/L) using: *corrected serum calcium = measured calcium + 0.02(40 − serum albumin)*. Secondly, the cause should be determined. Causes may be obvious from the patient's history, e.g. chronic renal failure and post-surgical hypoparathyroidism. When the cause is not obvious, further investigations should be considered, based on clinical suspicion (see Table 11). Occasionally, symptomatic hypocalcemia can occur in the absence of a fall in the serum total calcium concentration, as with increased protein-binding of ionized calcium due to acute respiratory alkalosis.

Treatment

The decision to institute therapy and the choice of therapy would depend on several factors such as the level of serum calcium, rapidity of hypocalcemia, presence of symptoms and signs, and the underlying cause.

Table 11 Laboratory Investigations to Consider

Likely problem	Ca^{2+}	PO_4	iPTH
• **serum calcium, phosphate, albumin, ionised calcium, iPTH**			
PTH-deficiency	↓	↑	↓
Vitamin D-deficiency;	↓	↓	↑
Secondary hyperparathyroidism, hyperphosphatemia	↓	↑	↑
Pseudohypoparathyroidism	↓	↑	↑
Hypomagnesemia	↓	N/↓	↑/N/↓

- Serum creatinine, magnesium, liver function studies
- arterial blood gases, lactate
- 25(OH)D, [1,25$(OH)_2$D currently unavailable locally]
- amylase, muscle enzymes
- hormone panel: Thyroid, gonadal, adrenal
- radiographs of hands, feet, skull or other relevant areas
- chromosomal studies

Mild asymptomatic hypocalcemia

Patients with mild hypocalcemia (serum total calcium concentration between 1.9–2.1 mmol/L) are usually asymptomatic, and can often be treated adequately by increasing oral calcium intake by 1000–2000 mg/day through supplements or diet.

Acute symptomatic hypocalcemia

Patients with serum total calcium concentrations less than 1.8 mmol/L, or with hypocalcemic symptoms such as paresthesia, tetany or seizures at serum calcium levels above this, require parenteral calcium therapy.

Intravenous calcium gluconate 10% 1–2 ampules (2.25–4.5 mmol or 90–180 mg elemental calcium per 10 ml-ampule) diluted in 50–100 ml of 5% dextrose or saline should be infused slowly over 5–10 minutes because of the risk of cardiac dysfunction, including systolic arrest. This procedure could be repeated as necessary to control symptomatic hypocalcemia. However, such infusions do not raise the serum calcium concentration for more than 2–3 hours, and therefore should be followed by a slow infusion of elemental calcium at 0.02–0.04 mmol/kg/hr. This translates approximately into 0.5–1 ampule of calcium gluconate per hour for a 60 kg individual.

An alternative preparation is calcium chloride 10% (9 mmol or 360 mg per 10 ml ampule). However, calcium chloride is more likely to cause tissue necrosis than calcium gluconate if extravasated. Calcium should be diluted in dextrose, water or saline, because concentrated calcium solutions, e.g. more than 2 ampules of calcium gluconate/100 ml, are irritating to veins. The calcium solution should not contain bicarbonate or phosphate, which can form insoluble calcium salts; if these anions are needed, another intravenous line should be used.

Intravenous calcium should be continued until the patient is receiving an effective regimen of oral calcium (elemental calcium 1–2 g/day) and vitamin D (calcitriol 0.25–1.0 μg/day).

Hypomagnesemia is not an uncommon cause of hypocalcemia. If the serum magnesium concentration is not known or is low, and if renal function is normal (which allows excretion of excess magnesium), magnesium sulphate 50% 2–4 ml (each 5 ml-ampule contains 10 mmol or 20 mEq or 2.5 g elemental Mg), diluted in 20–50 ml of saline, should be infused over

10 minutes, followed by 12 ml (24 mmol or 48 mEq or 6 g) infused intravenously per day.

Some post-parathyroidectomy patients may require prolonged, massive calcium therapy due to hungry bone disease. Patients with renal failure who have symptomatic hypocalcemia can be treated by adding calcium to the dialysis fluid. Treatment for patients with hypocalcemia and severe acute hyperphosphatemia due to a hypercatabolic state, such as the tumor lysis syndrome, should be aimed at correcting the hyperphosphatemia (e.g. by hemodialysis in patients with impaired renal function).

Chronic hypocalcemia

These patients should be treated initially with an oral calcium preparation (Table 12). Patients with documented achlorhydria and normal renal function should ideally receive calcium citrate.

The goals are to relieve symptoms, and to raise and maintain the serum calcium concentration in the low-normal range, e.g. 2.0–2.1 mmol/L. Attainment of higher values is not necessary and is usually limited by the development of hypercalciuria. The initial dose of oral calcium should be 1.5–2 g of elemental calcium. The diet also should contain liberal amounts of calcium, and be limited in phosphate to minimize hyperphosphatemia. Treatment with any calcium salt also lowers the serum phosphate concentrations in these patients.

Table 12 Available Oral Calcium Preparations

Calcium Salt	Preparation	% Ca	Score (mg)	Elemental Ca	
				(mg)	(mmol)
Ca carbonate	Oscal 250, Vitacal	40	625	250	6.25
	Oscal 500	40	1250	500	12.5
	Caltrate, Vitacal	40	1500	600	15
Ca citrate	Citracal tab	21	950	200	5
	Citracal caplet		1500	315	7.9
	Citracal liquitab		2376	500	12.5
Ca lactate-gluconate,	Ca-sandoz forte	13	2940	380 + 120	12.5
Ca carbonate		40	300	= 500	
Ca lactate-gluconate,	Ca-C 1000 sandoz	13	1000	130 + 130	6.5
Ca carbonate		40	327	= 260	
Ca acetate		25	667	169	4.2

If there is an insufficient response to these doses of elemental calcium, a vitamin D preparation should be added, the usual initial dose being calciferol 50 000 units/day (1.25 mg/day) or calcitriol 0.25 μg/day (Table 13). Subsequently, the dose of calcium or vitamin D should be adjusted according to the patient's symptoms and serum calcium values. Requirements vary considerably from patient to patient, and the correct dose in any given patient must be determined by trial and error. Patients with hypoparathyroidism need much more vitamin D than patients with vitamin D deficiency. Patients with renal failure are usually treated with pulsed vitamin D analogs to control secondary hyperparathyroidism and renal bone disease.

Patients with hypoparathyroidism excrete more calcium than normal subjects at the same serum calcium concentration, and are at risk of hypercalciuria as their serum calcium concentrations are increased toward normal. This may lead to nephrolithiasis, nephrocalcinosis, and chronic renal insufficiency. To prevent these complications, urinary calcium excretion and serum creatinine should be measured periodically, and the dose of calcium reduced if these are elevated. A thiazide diuretic, with or without dietary sodium restriction, may be used to decrease urinary calcium excretion.

Table 13 Available Oral Vitamin D Preparations

Preparation	Vitamin D Content	Physiologic Dose*	Pharmacologic Dose**	t 1/2 (days)	Onset (days)
Ergocalciferol [D2]	1.25 mg (50 000 IU)	5–20 μg (200–800 IU)	1.25–2.5 mg (50 000–100 000 IU)	30	10–14
Cholecalciferol [D3] (Calcium/multi-vitamin combinations)	2.5–10 μg (100–400 IU)	5–20 μg (200–800 IU)	generally not used	30	10–14
One-alpha [1α (OH)D]	0.25, 1 μg	0.25–1 μg	0.5–3 μg	2	1–2
Rocaltriol [1,25$(OH)_2$D]	0.25, 0.5 μg	0.25–0.5 μg	0.25–2 μg	1	1–2

*Physiologic dose for treatment of nutritional deficiency.
**Pharmacologic dose for treatment of hypoparathyroidism.

Potential future approaches

A possible future treatment for patients with hypoparathyroidism is the administration of PTH. Recent reports have shown that twice-daily subcutaneous administration of synthetic PTH can control hypocalcemia with a lower risk of hypercalciuria.

Conclusion

Hypocalcemia is a relatively common clinical problem. The therapeutic approach to hypocalcemia should be tailored to the clinical setting, with the aims of controlling hypocalcemia, and finding and instituting therapy for the underlying cause.

REFERENCES

Osteoporosis

1. NIH Consensus Development Panel on Osteoporosis prevention, diagnosis, and therapy, *JAMA* **285**:785–795, 2001.
2. National Osteoporosis Foundation, Osteoporosis: review of the evidence for prevention, diagnosis and treatment and cost-effectiveness analysis, *Osteoporos Int* **8**(Suppl 4): S1–S80, 1998.
3. Osteoporosis, in *Clinical Practice Guidelines*, Ministry of Health, Singapore, 2002.
4. Koh LKH, Saw SM, Lee JJM, Leong KH, Lee J, Hip fracture incidence rates in Singapore 1991–1998, in *Osteoporos Int* **12**:311–318, 2001.
5. Koh LKH, Ben Sedrine WB, Torralba TP, Kung A, Fujiwara S, Chan SP, *et al.*, A simple tool to identify Asian women at increased risk of osteoporosis, *Osteoporos Int* **12**:699–705, 2001.
6. Kanis JA, Diagnosis of osteoporosis and assessment of fracture risk, *Lancet* **359**:1929–1936, 2002.

Hypercalcemia

1. Bilezikian JP, Clinical review 51: Management of hypercalcemia, *J Clin Endocrinol Metab* **77**:1445–1449, 1993.
2. Bilezikian JP, Management of acute hypercalcemia, *N Engl J Med* **326**:1196–1203, 1992.
3. Shane E, Hypercalcaemia: pathogenesis, clinical manifestations, differential diagnosis and management, in Favus MJ (ed.), *Primer on the Metabolic Bone Diseases and Disorders*

of Mineral Metabolism, 4th ed, Philadelphia, Lippincott Williams & Wilkins, pp. 183–187, 1999.

4. Koh LKH, The diagnosis and management of hypercalcaemia, *Ann Acad Med Singapore* **32**:129–139, 2003.

5. Bilezikian JP, Potts JT, El–Hajj Fuleihan G, Kleerekoper M, Neer R, Peacock M, *et al*., Summary statement from a workshop on asymptomatic primary hyperparathyroidism: a perspective for the 21st century, *J Clin Endocrinol Metab* **87**:5353–5361, 2002.

Hypocalcemia

1. Tohme JF, Bilezikian JP, Hypocalcaemic emergencies, *Endocrinol Metab Clin North Am* **22**:363–375, 1993.

2. Reber PM, Heath H, Hypocalcemic emergencies, *Med Clin North Am* **79**:93–106, 1995.

3. Shane E, Hypocalcaemia: pathogenesis, differential diagnosis, and management, in Favus MJ (ed.), *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism, 4th ed*, Philadelphia, Lippincott Williams & Wilkins, pp. 223–226, 1999.

20

Endocrine Emergencies

Tai E Shyong

The endocrine system is the means by which different organ systems communicate with each other. As such, disorders of the endocrine system have the potential to alter the function of multiple organ systems. Severe dysfunction can result in endocrine emergencies that, by virtue of the multi-system involvement, can be life-threatening. Often, the 'crisis' is precipitated by other concurrent illnesses in the setting of a disease that is relatively benign when well-controlled, such as diabetes mellitus (DM) or thyrotoxicosis.

This chapter will deal with: 1) hyperglycemic emergencies, specifically, diabetic ketoacidosis (DKA) and hyperglycemic hyperosmolar states (HHS). 2) thyroid disorders, specifically, thyroid storm; and 3) adrenal dysfunction.

DIABETIC KETOACIDOSIS (DKA) AND HYPEROSMOLAR HYPERGLYCEMIC STATES (HHS)

Background

Hyperglycemic states are the most commonly encountered endocrine emergencies. They are associated with considerable morbidity and mortality (3–5% for DKA and ~15% for HHS).[1]

DKA comprises the triad of hyperglycemia, ketosis and acidosis. It is classically associated with insulin deficiency in Type 1 DM. In these patients, omission of insulin is a common cause of DKA, and social or psychological factors that predispose to non-compliance to therapy may predispose individuals to recurrent episodes.[2,3] However, DKA also occurs in patients with Type 2 DM. In these patients, it is often precipitated by concurrent illness, most commonly infection.

HHS occurs most commonly in elderly patients with Type 2 DM and, like DKA, is often associated with precipitating factors. The term HHS better reflects the mild to moderate ketosis that may occur in association with hyperglycemia and hyperosmolarity than the older term — hyperglycemic, hyperosmolar, non-ketotic coma. Furthermore, significant hyperosmolarity can be present without the degree of mental obtundation one would associate with the term coma.

Diagnosis

The diagnosis of DKA and HHS should be considered in any patient presenting with severe dehydration and/or mental obtundation, especially if it seems out of proportion to any concurrent illness. It should be remembered that up to 20% of patients with DKA will not give a history of preexisting DM[1] and that the absence of a history of DM should not be used to exclude the diagnosis. A simple bedside test for blood glucose should form a routine part of the examination of ill patients and significant hyperglycemia should alert the physician as to the need to exclude either DKA or HHS.

Dehydration is a key component of both DKA and HHS, and the clinical signs of severe dehydration are often present. The signs described include tachycardia, hypotension (especially postural hypotension), and sunken eyes. However, I have found that decreased skin turgor and dry mucus membranes, especially of the tongue, are the most useful signs at the bedside to get a clue as to the presence of severe dehydration.

The technical definitions of DKA and HHS are shown in Table 1. Therefore, in addition to the blood glucose, other investigations that are critical to establishing the diagnosis should include an arterial blood gas, a urea and electrolytes panel, and tests for ketones in the serum or urine.

Table 1 Diagnostic Criteria for DKA and HHS

	DKA	HHS
pH	< 7.3	> 7.3
Plasma glucose	> 250 mg/dL (13.9 mmol/L)	> 600 mg/dL (33.3 mmol/L)
Serum bicarbonate	< 15 mmol/L	> 15 mmol/L
Ketosis	Moderate to severe ketonemia/ketonuria	Absence of severe ketosis. Mild to moderate ketosis are compatible with a diagnosis of HHS
Serum total osmolarity	—	> 330 mOsm/kg

Two calculations based on the serum electrolytes are useful. Serum total osmolarity can be calculated using the formula:

$$2 \times ([Na^+] + [K^+]) + [\text{glucose}] + [\text{blood urea nitrogen}],$$

where [] signifies the measured serum concentration of the particular electrolyte in mmol/L. This is important to establish or refute the diagnosis of HHS. In addition, the anion gap can be calculated from the formula: $([Na^+] + [K^+]) - ([Cl^-] + [HCO_3^-])$. A high anion gap metabolic acidosis is present in a large proportion of patients with DKA. Others may have a mixed picture of high anion gap acidosis and hypercholeremic/normal anion gap acidosis or, in about 10% of patients, a pure hypercholeremic/normal anion gap acidosis.

The presence of significant quantities of ketones in the urine of patients is a good indication of ketonemia. Alternatively, the nitroprusside reaction (which is the basis of the tests such as the ketostix® test) can be applied to serum. One thing that has to be remembered is that the nitroprusside reaction detects acetoacetate but not β-hydroxybutyrate. The latter is the main ketoacid produced in DKA. As such, failure to detect ketones in the urine or serum of the patient does not exclude the possibility of diabetic ketoacidosis. β-hydroxybutyrate can also contribute to prolonged acidosis in the absence of detectable ketones in the serum or the urine and should be considered in the situation where acidosis fails to resolve despite adequate hydration. For these instances, although not yet in common use, simple bedside tests have become available to assay β-hydroxybutyrate and some laboratories measure this routinely.

Other concomitant conditions can also alter the metabolic picture. For example, alcoholism is associated with relatively normal blood glucose,

possibly related to the depletion of nutritional stores in these patients prior to the onset of relative insulin deficiency. Other acute illnesses that may precipitate DKA could also cause a raised anion gap metabolic acidosis, for example, lactic acidosis.

Acute Treatment

In any medical emergency, ensuring an adequate and patent airway, especially in patients with mental obtundation, is a critical first step. Subsequently, one has to deal rapidly with the cardiovascular status of the patient and this may require urgent fluid resuscitation in the patient who has circulatory collapse due to the severity of the metabolic disorder or dehydration. The priority at this stage should be to restore intravascular volume and the initial resuscitation fluid should be isotonic saline. Once the initial resuscitation has been performed, attention can turn to the specific management of the hyperglycemic state. In general, management of hyperglycemic state following diagnosis follows 4 main principles.

1) **Rehydration.** Almost without exception, patients with DKA or HHS have a large fluid deficit at the time of presentation. This can be 4–6 L in the case of DKA while with HHS, it can be as high as 8–10 L.[4,5] Adequate hydration is a cornerstone of treatment of both these conditions allowing the restoration of circulating plasma volume and the reduction of the levels of glucose counter-regulatory hormones in the plasma. In fact, inadequate fluid replacement is a common cause for the failure of the plasma glucose to respond to insulin therapy and the failure of acidosis to resolve with treatment. An acceptable regimen might be 1 L of intravenous fluids in the first hour, followed by 1 L in the subsequent 2 hours and then 1 L in the next 4 hours. Even in the elderly and in others in whom fluid overload is a real concern, aggressive fluid resuscitation is essential until the patient is hemodynamically stable. Subsequently, a more conservative fluid regimen may be appropriate. Where necessary, more care may be required and invasive monitoring with a central venous catheter or a pulmonary artery wedge catheter may be required. Once intravascular volume is replete and the patient is hemodynamically stable, the use of hypotonic solutions can be considered in the patient with effective total osmolarity > 340 mosmol/kg to restore both intracellular and extracellular

volume. However, it should be noted that the timing and the type of hypotonic solutions used remain controversial.

2) **Insulin therapy.** Insulin deficiency, whether absolute or relative, is a critical part of the pathogenesis of DKA and HHS. As such, the use of insulin therapy is an essential component in the management of these conditions. The goal of therapy is to achieve a reduction in the blood glucose by 75 to 100 mg per dL (4.2 to 5.6 mmol per L) per hour. Two aspects of insulin therapy remain hotly debated. The first concerns the mode of delivery. Some clinicians advocate an intravenous insulin infusion for treatment of this condition, whereas, others feel that subcutaneous or intramuscular injections of insulin may be sufficient. In the literature, the only clear advantage of an intravenous infusion over other modalities is a more rapid response of the blood glucose. In patients who are severely dehydrated with peripheral vascular shutdown, intravenous infusions may provide a more reliable, rapid lowering of the blood glucose. On the other hand, for example, in individuals with mild DKA, it may be appropriate to use a subcutaneous or intramuscular injection to simplify the treatment regimen. The second area of contention lies in the use of an intravenous bolus prior to the institution of a continuous intravenous infusion. While an intravenous bolus of insulin clearly raises the serum concentrations of insulin rapidly to levels many times in excess physiological levels, studies have shown that high physiological levels of serum insulin rapidly corrects the hyperglycemia.[6] These levels are readily achieved by an intravenous infusion of 0.1 U/kg/h (5–7 U per hour) without a preceding bolus. In addition, this may help avoid the dangerous hypokalemia that is associated with therapy because the effect on serum potassium with an intravenous infusion and no bolus is more gradual than if a bolus is administered.

3) **Solute replacement.** As part of the osmotic diuresis which occurs with DKA and HHS, the patient often loses large amounts of solutes in the urine, including potassium, magnesium and phosphate. The most critical of these is potassium. Even though the serum levels of potassium may be high at the time of presentation due to insulin deficiency and acidosis, this usually represents a shift in this cation from the intracellular to the extracellular compartment due to acidosis, dehydration and insulin deficiency. Invariably, total body potassium is low and this requires supplementation to avoid the dangerous

hypokalemia that can occur as a consequence of rehydration and the institution of insulin therapy. Potassium replacement should begin as soon as the serum potassium is within the normal range and there are no abnormalities on the ECG to suggest the presence of hyperkalemia. Total body magnesium and phosphate are often also depleted. However, the benefits of routinely replacing these electrolytes as part of the treatment of DKA or HHS are currently unclear. Except in severe instances, for example, when tetany or cardiac arrythmias occur, replacement of magnesium and phosphate has not been recommended as routine at this time. Another electrolyte to consider is bicarbonate. Although it may seem rational to give bicarbonate to patients with severe acidosis, this can cause worsening of intracranial pH due to the inability of bicarbonate to cross the blood brain barrier. Studies have shown no advantage in the administration of bicarbonate to patients with DKA and a pH of >6.8.[7] As such, some physicians recommend withholding bicarbonate in patients irrespective of the pH as long as they are responding to other therapy. However, in the setting of shock or cardiovascular compromise and severe acidosis, bicarbonate replacement may be required.

4) **Treatment of the underlying condition.** The most common precipitating factor for DKA and HHS are infections. These should be aggressively sought and treated. Cultures of the blood as well as a chest X-ray and urine microscopy and culture should form part of the standard work-up for these conditions since pneumonia and urinary tract infections form the 2 most common sites of infection-precipitating DKA. Appropriate empirical therapy may be considered based on local data on infective agents involved and antibiotic sensitivities. In addition, other common precipitants include cardiovascular disease (such as myocardial infarction) or other medical illnesses. These should therefore be actively sought and investigated according to the clinical findings. An electrocardiogram should form part of the initial work-up both to exclude concomitant cardiovascular pathology and to look for changes associated with hyperkalemia.

Once the patient's blood glucose comes down to 14 mmol/L (250 mg/dL) or below, the patient should be started on a glucose infusion. This provides calories and also prevents the catabolism of muscle and fat in the presence of insulin, allowing clearance of the metabolic acidosis. Giving the insulin and glucose containing solutions separately gives the

greatest flexibility for maintaining the blood glucose in the normal range as the rate of each can be adjusted independantly. However, it has been suggested that it is a good idea to use the same iv cannula in order to avoid the situation of severe hypoglycemia in the event that the cannula for the glucose containing solution becomes blocked. An important aspect of insulin therapy in DKA or HHS revolves around the conversion of patients from intravenous infusions of insulin to other modalities of drug delivery such as subcutaneous injections following acute treatment. The recurrence of acidosis or hyperglycemia often results from cessation of intravenous insulin infusions before adequate absorption can occur from other modes of drug delivery. As such, some overlap (30–45 min) between a s/c injection and the cessation of intravenous therapy should be allowed to avoid this. Feeding should be re-established as soon as possible. This provides the calories that dextrose containing iv solutions cannot provide sufficiently. Once patients are able to tolerate oral feeding and are converted to subcutaneous insulin injections, they can actually be discharged provided the underlying precipitant of DKA has been treated.[8] For patients with Type 1 DM who develop mild DKA due to omission of their insulin, they can be treated as outpatients if they are able to take adequate fluids/food orally and respond to subcutaneous insulin. However, because of the greater mortality associated with HHS, hospitalization and inpatient stabilization is almost always required.

Long-Term Treatment

The management of the patient following an episode of DKA or HHS should focus on the avoidance of recurrence of these events. In patients with Type 1 DM, the most common cause of DKA is omission of insulin. Ironically, the increasing use of insulin pumps for intensive glycemic control may lead to an increase in the frequency of DKA, particularly among those with Type 1 DM. The DCCT showed that intensive control using an insulin pump was associated with a 2-fold increased risk of DKA compared to subjects who used multiple daily injections to achieve the same degree of control.[9] This may relate to the use of short-acting insulin alone, which leaves the patient without a reservoir of insulin in the event of pump failure. It is uncertain whether the use of such devices would be associated with increased risk of diabetes ketoacidosis in those with Type 2 DM.

Not all patients with an episode of DKA require insulin. Indeed, among adult patients presenting with DKA, insulin therapy can be discontinued in one-third of patients and they can be maintained on oral hypoglycemic agents. Significant predictors of the need for insulin include older age at the time of presentation, shorter duration of DM, higher BMI, higher serum osmolarity during DKA and lower insulin dose following recovery.[10]

THYROID STORM

Background

Thyroid storm describes a severe exacerbation of thyrotoxicosis that occurs most commonly in patients with Graves' disease. However, it can occur in patients with other causes of thyrotoxicosis. It is a life-threatening condition that is fortunately rare. In the past, it was most often seen in patients undergoing surgery for the treatment of thyrotoxicosis. However, with the use of anti-thyroid therapy pre-operatively, to ensure the patient is euthyroid prior to surgery, it is now most often precipitated by other medical illnesses. These include surgery, parturition, radioiodine therapy, iodinated contrast materials, stroke, diabetic ketoacidosis, infection, and withdrawal or discontinuation of anti-thyroid medications.[11]

The pathogenesis of thyroid storm is unclear but may involve increased levels of free thyroid hormones due to release from the thyroid gland, decreases in thyroid binding or increased tissue sensitivity to the effects of thyroid hormones.

Diagnosis

The clinical features of thyroid storm are those of thyrotoxicosis, except they are more exaggerated. These include pyrexia, tachycardia, a widened pulse pressure, diarrhea, nausea and vomiting, jaundice, and motor and psychic restlessness. However, since many of these features are non-specific and can be the result of, for example, severe infection, the diagnosis can be difficult. Certainly, a preceding history of thyrotoxicosis should lead one to suspect the diagnosis as should clinical signs associated specifically with thyrotoxicosis, such as goiter or exopthalmos, which may be associated with Graves' disease. However, it should be borne in mind that thyroid storm can occur in patients with few or no signs of thyrotoxicosis

such as those with apathetic thyrotoxicosis, a condition that is more common in the elderly and can lead to delayed diagnosis.

There are no firm diagnostic criteria for the diagnosis of thyroid storm. Biochemical tests should confirm the presence of thyrotoxicosis in the form of high levels of T4 and T3 with suppressed TSH. Although the levels of free thyroid hormone in the plasma of patients with thyroid storm are thought to be higher than those in patients who just have thyrotoxicosis, no diagnostic criteria exist. In general, the diagnosis is clinical. A scoring system has been suggested for the diagnosis of thyroid storm in patients with severe thyrotoxicosis (see Table 2). Even so, since many of the symptoms are in common with underlying medical conditions precipitating the crisis, the proponents of the scoring system recommend that the highest score for any of the clinical features be applied to the patient so as to err on the side of empirical treatment rather than the withholding of treatment.

Acute Treatment

Treatment involves 3 major components:

1) **Correcting the hyperthyroidism.** This can be achieved in a number of ways and often these are used simultaneously.[13] These are summarized in Fig. 1. Anti-thyroid medication (such as the thionamides propylthiouracil or methimazole) can be used to prevent synthesis of additional thyroid hormone. Since there are no parenteral preparations of these drugs available, these are most often given orally or via a nasogastric tube. In situations where oral ingestion is not possible, such as in patients with surgical acute abdomens of bowel obstruction, rectal administration of the same dose of these drugs as part of a retention enema has been shown to achieve similar serum levels to those following oral dosing.[14,15] Propylthiouracil has the added advantage of being able to block peripheral conversion of T4 to T3, the active form of thyroid hormone. An initial dose of propylthiouracil of 600–1200 mg can be followed by 200 mg 6-hourly. To prevent the release of pre-formed thyroid hormone from the thyroid gland, we often use iodide, either orally (SSKI 5 drops 6-hourly, lugol's iodine 30 drops per day in 3–4 divided doses or as an intravenous infusion of sodium iodide 1 g 12-hourly). Alternatively, the radio-contrast labels ipodate and iopanoate also block release of

Table 2 Diagnostic Criteria for Thyroid Storm-adapted from Burch *et al.*[12] A Score of 45 or Greater is Highly Suggestive of Thyroid Storm, a Score of 25–44 Is Suggestive of Impending Storm and a Score of below 25 is Unlikely to Represent Thyroid Storm

Thermoregulatory Dysfunction			**Gastrointestinal/Hepatic Dysfunction**		
		Score			Score
Temperature	99–99.9	5	Absent		0
	100–100.9	10			
	101–101.9	15			
	102–102.9	20			
	103–103.9	25			
	⩾104	30			
Central nervous system effects			Moderate	Diarrhea	10
				Nausea/vomiting	
				Abdominal pain	
Absent	0		Severe	Unexplained jaundice	20
Mild agitation		10	**Cardiovascular dysfunction**		
Moderate	Delirium		Tachycardia	99–109	5
	Psychosis	20		110–119	10
	Extreme lethargy			120–129	15
				130–139	20
				⩾140	25
Severe	Seizure		Congestive cardiac failure	Absent	0
		30		Pedal edemia	5
	Coma			Bilateral basal	10
				Pulmonary edema	15
			Atrial	Absent	0
				Present	10
			Precipitant history		
			Negative		0
			Positive		10

pre-formed thyroid hormone and organification of inorganic iodide. One has to bear in mind that high doses of iodide can actually precipitate thyroid storm and pre-treatment with anti-thyroid medication to block the incorporation of iodide into thyroglobulin should be undertaken prior to the initiation of iodide therapy. In practice, we often administer PTU or methimazole 1–2 hours prior to starting iodide therapy. Other agents that block thyroid hormone release include

glucocorticoids (e.g. iv dexamethasone 2 mg 6 hourly) and lithium (300 mg 6-hourly orally or via NG tube), which may be used concurrently. These agents also block the peripheral conversion of T4 to T3. Beta-blockers also block T4 to T3 conversion. Since activation of the adrenergic system is responsible for many of the signs and symptoms of thyrotoxicosis and can result in hemodynamic compromise, beta-blockers also reverse some of the effects of severe thyrotoxicosis on the cardiovascular system and forms a critical part of the treatment of thyroid storm. Other agents that can be useful include the anti-arrythmic agent, amiodarone, which can be used to treat any arrythmias that arise as a consequence of severe thyrotoxicosis. Amiodarone also contains large quantities of iodide. These treatments will usually rapidly correct the serum levels of thyroid hormone. Serum levels of T3 will begin to correct within hours and often return to normal in 48–72 hours. Serum levels of T4 take longer and begin to decline after approximately 3 days.

2) Clearly, any **other supportive therapy** required should be administered concurrently and these would include fluids for rehydration, nutrition, especially given the hypercatabolic state often associated with thyrotoxicosis. Wernicke's encephalopathy has also been associated with treatment of thyroid storm and some would advise the use of vitamin supplements.
3) **Search and treatment of precipitating event.** As with the hyperglycemic emergencies, it is important to look for and treat any precipitating event and therapy should be directed according to the clinical findings.

Long-term therapy needs to be considered for the treatment of underlying thyrotoxicosis and the prevention of subsequent events. These may include surgery, long-term thionamide therapy or radioiodine, although the latter may have to be delayed since the large doses of iodide used in the treatment of thyroid storm will inhibit uptake of the radio-isotope.

The prevention of thyroid storm remains controversial. Since RAI therapy has been associated with thyroid storm, it has been suggested that pre-treatment with anti-thyroid medications should be carried out to prevent this. However, at least 1 study has shown that pre-treatment does not prevent the radioiodine induced rise in thyroid hormone

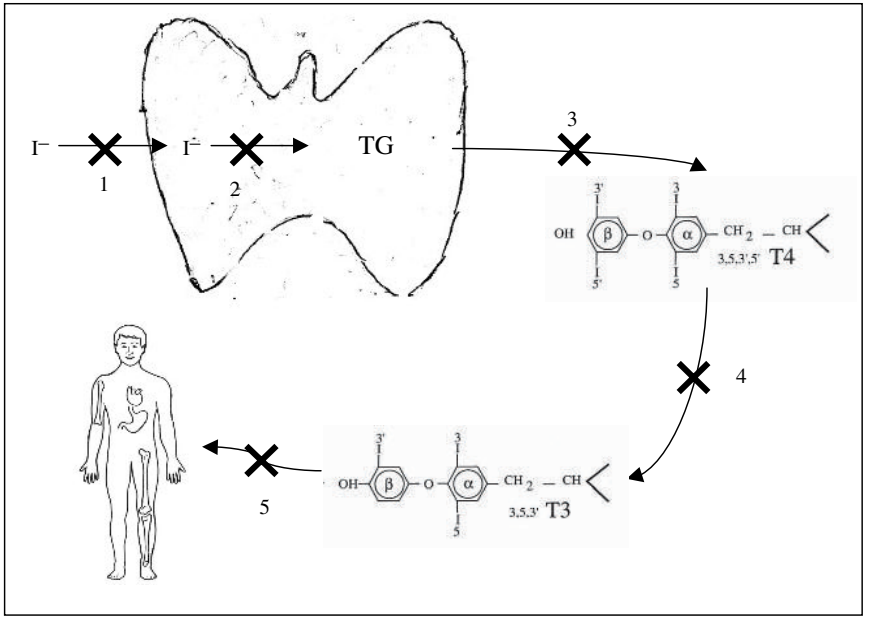


Fig. 1 Correction of hyperthyroidism in thyroid storm.

Steps in Thyroid Hormone Production and Action	**Drugs that Correct Hyperthyroidism**
1) Iodide uptake	
2) Organification of iodide	Thionamides (Propylthiouracil, carbimazole, methimazole)
3) Thyroid hormone release	Iodide (administered after blockage of step 2) Lithium Amiodarone Ipodate and iopanoate Glucocorticoids
4) Peripheral conversion of thyroxine (T4) to triiodothyronine (T3)	Propylthiouracil Glucocorticoids Beta-blockers (propanolol) Ipodate and iopanoate
5) Tissue effects of thyroid hormone	Beta-blockers

levels and in fact, the rise in thyroid hormone levels in these patients was more associated with a withdrawal of anti-thyroid medication than with RAI therapy *per se*.[16] With regards to surgery, most clinicians still recommend pre-treatment to normalize thyroid function prior to any elective surgery.

ADRENAL INSUFFICIENCY

Background

Primary adrenal insufficiency results from destruction of the adrenal glands, most commonly due to autoimmune disease, infection or hemorrhage. The most common cause of secondary adrenal insufficiency is the use of corticosteroids. In Singapore, it is important to remember that some traditional treatments may contain corticosteroids. Secondary adrenal insufficiency can also be due to pituitary disease and failure of production of ACTH. Once again, in an Asian context, it is important to remember the high incidence of nasopharyngeal carcinoma treated with cranial irradiation and associated hypopituitarism.[17] In critically ill patients with hypotension, especially pressor-refractory hypotension, acute adrenal insufficiency always should be considered in the differential diagnosis.

Diagnosis

Although specific features such as skin pigmentation or vitiligo may point to a diagnosis of primary autoimmune adrenal failure, this is a rare cause of adrenal insufficiency. Acute destruction of the adrenal glands by infection or hemorrhage is too acute to manifest as clinical signs and the diagnosis should be suspected in the appropriate setting, such as patients with meningococcal septicemia who develop refractory hypotension.

The diagnosis of adrenal insufficiency is confirmed on measurement of serum cortisol and a synacthen test (250 mcg of iv synacthen). Unfortunately, there is much controversy regarding the level of cortisol that is considered an adequate stress response. Many textbooks and published manuscripts state that the normal cortisol response to stress is a level $>$ 500–550 nmol/l (18–20 mcg/dL). However, physiological stress such as septic shock is often associated with much higher serum levels of cortisol and some physicians use a much higher cut-off in the region of 700 nmol/L (25 mcg/dL).[18] Levels below this should raise suspicion of adrenal insufficiency and lead to empiric therapy and further testing. The ACTH stimulation test remains the study of choice. A baseline cortisol sample is obtained, and 250 μg of synthetic ACTH (cosyntropin) is then administered intravenously. Additional cortisol samples are then drawn 30 and 60 minutes later. Traditionally, a plasma cortisol concentration

greater than 18–20 μg/dL at any time excludes the diagnosis of primary adrenal insufficiency, as well as long-standing secondary insufficiency. Again, this is controversial and there are authors who suggest that the dose of ACTH used in these tests is supraphysiological and that a 1 μg synacthen test or a CRH test better differentiates normal adrenal reserve from adrenal insufficiency.[18] However, until such time as these tests become routine, we generally rely on the 250 mcg synacthen test.

It is becoming increasingly recognized that, in patients with critical illness such as septicemic shock, a state of relative adrenal insufficiency may exist that may predict mortality and be amenable to therapy by corticosteroid replacement. 23.8% of patients in one series showed evidence of adrenal insufficiency on synacthen testing.[19] It has been reported that an increase in the serum cortisol of less than 9 ng/dL (180 nmol/L) after 250 mcg of cosyntropin given intravenously predicts higher mortality in patients with septicemic shock.[20] It is uncertain currently where the source of this insufficiency lies. Data seems to suggest that, in addition to an impaired ability of the adrenal to respond to ACTH,[20] there may also be an inability of the pituitary to responsd to CRH.[21] Several mechanisms may be involved in the development of adrenal insufficiency in septic shock. Insufficient blood flow to the adrenal cortex, the presence of circulating substances, which inhibit adrenal function or ACTH action, or diminished pituitary function must all be considered.

Treatment

The treatment of adrenal insufficiency is the replacement of adrenal steroids. This often begins with intravenous hydrocortisone, and in the acute stages, a dose of 100 mg 6-hourly would be appropriate. Treatment should begin as soon as the corticotropin stimulation tests is performed. If the corticotropin stimulation test can not be performed immediately, treatment should not be delayed. Instead, dexamethasone (2 mg) should be administered, and the test should be performed within the next 12 hours. Dexamethasone does not significantly cross-react with cortisol in the assay for cortisol and can be given to patients pending the results of adrenal testing. In patients who have a normal response to corticotropin, treatment can subsequently be discontinued. Some physicians would recommend the continuation of glucocorticoid replacement in patients who show a clinical response to treatment even if the test is normal.

Once their conditions are stable, patients should be tapered to maintenance doses of hydrocortisone appropriate to their levels of stress. Doses of mineralocorticoid are not usually required acutely because higher doses of hydrocortisone have sufficient mineralocorticoid activity. However, in primary adrenal insufficiency, once the hydrocortisone taper reaches less than 50 mg/day, supplementation with 9-alpha-fludrocortisone (0.05–0.2 mg/day) may be required.

As for relative adrenal insufficiency in septicemic shock, there is limited data regarding the use of corticosteroid therapy in those who fail to respond to a short iv synacthen test adequately. In a randomized placebo-controlled trial, Annane *et al.* found that a dose of iv hydrocortisone 50 mg 6-hourly and oral fludrocortisone 50 mcg per day for 7 days reduces mortality and improves hemodynamic status in patients with septic shock and evidence of relative adrenal insufficiency.[22]

CONCLUSION

With the exception of hyperglycemic emergencies, endocrine emergencies are rare. Although some of the symptoms are specific to classical endocrine disorders, many, such as nausea, vomiting or fever, are sufficiently general that the diagnosis is missed without a sufficiently high index of suspicion. Consequently, appropriate treatment is not accorded to the patient. Prompt recognition and diagnosis of the endocrine problem avoids unnecessary delay in the administration of appropriate therapy, and can reduce morbidity and mortality from these conditions.

REFERENCES

1. Kitabchi AE, Umpierrez GE, Murphy MB, *et al.,* Management of hyperglycemic crises in patients with diabetes, *Diabetes Care* **24**(1):131–153, 2001.

2. Polonsky WH, Anderson BJ, Lohrer PA, Aponte JE, Jacobson AM, Cole CF, Insulin omission in women with IDDM, *Diabetes Care* **17**(10):1178–1185, 1994.

3. Musey VC, Lee JK, Crawford R, Klatka MA, McAdams D, Phillips LS, Diabetes in urban African-Americans. I. Cessation of insulin therapy is the major precipitating cause of diabetic ketoacidosis, *Diabetes Care* **18**(4):483–489, 1995.

4. Kreisberg RA, Diabetic ketoacidosis: new concepts and trends in pathogenesis and treatment, *Ann Intern Med* **88**(5):681–695, 1978.

5. Ennis E, Stahl E, RA K, The hyperosmolar hyperglycemic syndrome, *Diabetes Rev* **2**:115–126, 1994.

6. Schade DS, Eaton RP, Dose response to insulin in man: differential effects on glucose and ketone body regulation, *J Clin Endocrinol Metab* **44**(6):1038–1053, 1977.

7. Morris LR, Murphy MB, Kitabchi AE, Bicarbonate therapy in severe diabetic ketoacidosis, *Ann Intern Med* **105**(6):836–840, 1986.

8. Carroll MF, Schade DS, Ten pivotal questions about diabetic ketoacidosis. Answers that clarify new concepts in treatment, *Postgrad Med* **110**(5):89–92, 95, 2001.

9. Implementation of treatment protocols in the Diabetes Control and Complications Trial, *Diabetes Care* **18**(3):361–376, 1995.

10. Hsin Yu E, Guo HR, Wu TJ, Factors associated with discontinuing insulin therapy after diabetic ketoacidosis in adult diabetic patients, *Diabet Med* **18**(11):895–899, 2001.

11. Ringel MD, Management of hypothyroidism and hyperthyroidism in the intensive care unit, *Crit Care Clin* **17**(1):59–74, 2001.

12. Burch HB, Wartofsky L, Life-threatening thyrotoxicosis. Thyroid storm, *Endocrinol Metab Clin North Am* **22**(2):263–277, 1993.

13. Tietgens ST, Leinung MC, Thyroid storm, *Med Clin North Am* **79**(1):169–184, 1995.

14. Nabil N, Miner DJ, Amatruda JM, Methimazole: an alternative route of administration, *J Clin Endocrinol Metab* **54**(1):180–181, 1982.

15. Yeung SC, Go R, Balasubramanyam A, Rectal administration of iodide and propylthiouracil in the treatment of thyroid storm, *Thyroid* **5**(5):403–405, 1995.

16. Burch HB, Solomon BL, Cooper DS, Ferguson P, Walpert N, Howard R, The effect of antithyroid drug pretreatment on acute changes in thyroid hormone levels after (131) I ablation for Graves' disease, *J Clin Endocrinol Metab* **86**(7):3016–3021, 2001.

17. Lam KS, Tse VK, Wang C, Yeung RT, Ho JH, Effects of cranial irradiation on hypothalamic-pituitary function — a 5-year longitudinal study in patients with nasopharyngeal carcinoma, *Q J Med* **78**(286):165–176, 1991.

18. Zaloga GP, Marik P, Hypothalamic-pituitary-adrenal insufficiency, *Crit Care Clin* **17**(1):25–41, 2001.

19. Soni A, Pepper GM, Wyrwinski PM, *et al*., Adrenal insufficiency occurring during septic shock: incidence, outcome, and relationship to peripheral cytokine levels, *Am J Med* **98**(3):266–271, 1995.

20. Annane D, Bellissant E, Prognostic value of cortisol response in septic shock, *JAMA* **284**(3):308–309, 2000.

21. Schroeder S, Wichers M, Klingmuller D, *et al*., The hypothalamic-pituitary-adrenal axis of patients with severe sepsis: altered response to corticotropin-releasing hormone, *Crit Care Med* **29**(2):310–316, 2001.

22. Annane D, Sebille V, Charpentier C, *et al*., Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock, *JAMA* **288**(7):862–871, 2002.

21

Adrenal Diseases

Rajasoorya C and Lim Ling Choo

INTRODUCTION

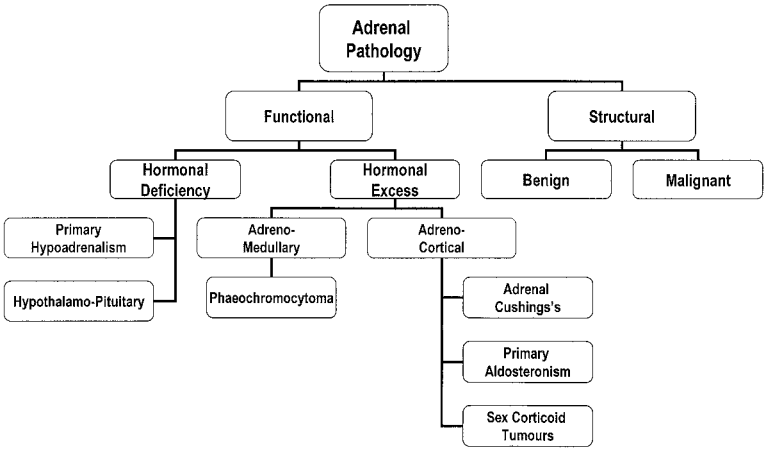
The widespread availability and utility of measurements of the different hormones produced by or regulating the adrenal glands as well as imaging modalities have contributed to a much better understanding of adrenal diseases. Advances in imaging provided by Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) as well as the selective utilization of other imaging modalities like Scintigraphy, and procedures like angiography, adrenal venous sampling and percutaneous fine needle aspiration cytology (FNAC) have allowed for better clinical-functional correlation and enabled the clinician to better delineate the need for non-intervention or intervention as the individual case may warrant. This chapter covers the current approaches to diagnostic evaluation as well as understanding of the more common adrenal diseases. It will not discuss issues specifically related to the pediatric adrenal disease.

ANATOMY AND PHYSIOLOGY

The adrenal glands occupy the retroperitoneal space at the superior poles of both renal glands and as such the left adrenal is at a higher level than the right. Each adrenal gland is made up of 2 limbs that are referred to as the lateral and medial crura. Due to the distinct 3-dimensional structure of the adrenal glands, the right adrenal gland appears on CT section as a linear structure or an inverted **V**, while the left adrenal gland appears as an inverted **V** or **Y**, or occasionally, as a triangle. While the size of the adrenal gland varies, a good clinical rule to judge enlargement on CT scans is to compare it with the ipsilateral crus of the diaphragm. Neither limb nor crura of the adrenal gland should be wider than the ipsilateral crus. The venous drainage of the left adrenal gland leads to the left renal vein while the right adrenal gland drains directly into the inferior vena cava in most instances. The adrenal gland consists of 2 distinct functional halves — the cortex and the medulla. The cortex is further subdivided into 3 anatomic zones comprising of the zona **g**lomerulosa, **f**asciculata as well as the **r**eticularis. The zona glomerulosa is responsible for the production of aldosterone and is primarily regulated by the renin-angiotensin axis. In contrast, the zona fasciculata and reticularis are primarily adrenocorticotrophic hormone (ACTH) dependent and synthesize cortisol and adrenal androgens respectively. The adrenal medulla is functionally related to the sympathetic nervous system and has its origins in the enterochromaffin cells that migrated and invaded the adrenal glands during fetal life and secrete adrenaline and noradrenaline. These anatomical and physiologic differences have implications on the functional categorization and radiological identification of the adrenal glands as well as the interpretation of specialized studies like adrenal venous sampling.

ADRENAL PATHOLOGY — AN OVERVIEW

The diagram below provides a non-exhaustive, simplified framework for the understanding of adrenal pathology. The diagram is based on the principle that adrenal disorders, just like any other endocrine gland problems may cause anatomical or structural problems and functional or physiological disturbances. Physiological disturbance may be related to excess or deficient hormone secretion.



ENDOCRINE CLINICAL SYNDROMES

Cushing's Syndrome

Incidence and causes

Cushing's syndrome remains one of the most challenging diagnostic problems in clinical endocrinology. The annual incidence of Cushing's disease has been estimated as 0.1–1.0/100 000 and the rest of the other causes of Cushing' syndrome is 5–6 times less. In most series, Cushing's disease accounts for about 70% of all cases whereas ectopic ACTH syndrome represents about 12% of cases. Among the ACTH-independent causes, adrenal adenoma makes up about 10% and adrenal carcinoma, 8% of cases.[1] The most common cause, however, is still exogenous Cushing's syndrome (factitious or iatrogenic). Table 1 below summarizes the modern classification of the endogenous causes of Cushing's syndrome based on ACTH dependence or independence. Amongst the ACTH-dependent causes, most are of eutopic origin (i.e. **ACTH** arises from the expected normal site of the pituitary) and approximately 20% arise from ectopic sites, which include **A**denomas, **C**arcinomas or **C**arcinoids, **T**hymomas, medullary **T**hyroid Carcinoma and **H**ematological malignancies, amongst others. Very rarely, an Adrenal Pheochromocytoma may secrete ACTH or even corticotrophic releasing hormone (CRH), causing confusion in the evaluation of Cushing's syndrome.

Table 1 Classification of the Causes of Endogenous Cushing's Syndrome

ACTH-dependent	Eutopic	Cushing's disease — corticotroph adenoma
	Ectopic	Ectopic ACTH secretion
		Ectopic CRH secretion(rare)
ACTH-independent	Unilateral	Adenoma
		Carcinoma
	Unilateral or Bilateral (Rare)	Gastro Inhibitory Peptide (GIP)/ Food–induced
	Bilateral (Rare)	Primary Pigmented Micro-nodular Hyperplasia (Carney's Complex)
		Macronodular Adrenal Hyperplasia
		McCune–Albright Syndrome

Table 2 Common Symptoms and Signs in Cushing's Syndrome

Disturbance	**Clinical Features**
Fat metabolism	Patients characteristically present with centripetally distributed fat seen as dorso-cervical fat hump, submental and supraclavicular fat projections.
Carbohydrate metabolism	Due to the effect of cortisol on carbohydrate metabolism, patients with Cushing's syndrome present with clinical or subclinical diabetes and its related complications. They may also present with the insulin-resistance and its associated features like hypertension and acanthosis nigricans.
Protein metabolism	Patients with Cushing's may have disturbances related to collagen formation and hence have a thin skin, striae that are progressive and present at unusual sites. Proximal myopathy is a common feature that may be easily tested by asking a patient to get up unaided from the squatting position.
Fluid and electrolytes	Disturbances may be related to sodium retention and hence water retention and hypertension, hypokalemia (reflective of the degree of hypercortisolemia rather than etiology as previously thought).
Mineral metabolism	Calciuresis and calculous nephropathy. Osteoporosis is common in patients with Cushing's syndrome for a variety of reasons, including decreased gut absorption and calciuresis.
Miscellaneous	Patients with Cushing's syndrome do have manifestations related to susceptibility to atypical infections at unusual and uncommon sites, including tuberculosis, *Pneumocystis carinni*, growth disturbances in childhood and adolescence, peptic ulcer, cataracts and avascular necrosis of the femora.

Clinical presentation

While the incidence of clinical symptoms and signs vary in the populations studied, there are 2 important considerations when discussing clinical presentation. The first consideration being the symptoms and signs of patients with Cushing's syndrome are very common amongst the non-Cushingnoid population. Secondly, the presentation features can be easily grouped by its causation of metabolic disturbances (Table 2). The former dilemma has partly been addressed by looking for signs and symptoms of high discriminant value.[1] Signs that are reflective of disturbance in protein metabolism have much higher discriminatory value in the diagnosis of Cushing's syndrome and these include findings of a papery thin skin, broad voilaceous striae and proximal myopathy. These are helpful stigmata in adults while growth retardation is usually seen in children. A high index of suspicion should arise particularly in patients with recent weight gain, history of impaired glucose intolerance, diabetes mellitus or hypertension, especially if these are resistant to conventional treatment.[2]

Diagnosis of Cushing's syndrome

The diagnostic strategy involves 6 main steps:

1) **Exclusion** of pseudo-Cushing's states;
2) **Screening** for possible Cushing's syndrome;
3) **Confirmation** of the diagnosis of Cushing's syndrome;
4) **Endocrine Localization** of primary source of hormone excess;
5) **Radiological localization** of the tumor; and
6) **Careful follow-up** to confirm cure.

1) While a good clinical history and examination is most useful for **Exclusion** of pseudo-Cushing's states like severe obesity, depression and alcoholism, it may not always be possible and easy, particularly in a subtle case, to exclude a pseudo-Cushing's state on clinical or even biochemical grounds. Indiscriminate use of radiology may allow one to be trapped into the fallacy of assuming that an incidental tumor is the cause of Cushing's syndrome. Exogenous intake of corticosteroids should always be excluded. Some patients may have episodic or periodic increases in cortisol secretion and this has to be considered during biochemical screening.

2) **Screening** for Cushing's syndrome is usually performed with the 24-hour Urinary Free Cortisol (UFC) and/or a post 1 mg overnight dexamethasone cortisol test. The 24-hour UFC is a sensitive and accurate screening procedure for Cushing's syndrome with a sensitivity of 95% and specificity of 94–98% if the urine collection is adequate. It is an integrated index of the free circulating cortisol and is therefore not affected by factors that influence the corticosteroid-binding globulin (CBG) levels.[3] Elevations of 24-hour UFC to values 4-fold greater than the upper limit of normal ranges can be considered diagnostic for Cushing's syndrome.[2] One main disadvantage of this test include the difficulties in obtaining an accurate 24-hour sample for UFC and, in women, in particular the simultaneous occurrence of menses poses practical collection problems. In approximate 4% of obese patients, the cortisol production rate may be increased, leading to an erroneous diagnosis of Cushing's syndrome.[4] Conversely, those with proven Cushing's syndrome may have a normal 24-hour UFC. About 10–15% of patients with Cushing's syndrome have at least one of four 24-hour determinations of UFC within normal ranges. With the possibility of intermittent hypercortisolism, up to three 24-hour urine collections have been advocated.[2] Recent extensive reviews assessing the utility of overnight 1 mg dexamethasone suppression test have suggested that suppression of the 8 am cortisol level of < 50 nmol/L excludes Cushing's syndrome. It has a sensitivity of about 98% but low specificity of 70–80%. The false positives can arise in pseudo-Cushing's states as well as any conditions that result in decreased delivery of oral dexamethasone to the circulation such as decreased absorption, increased metabolism of dexamethasone or interaction with drug inducers such as anti-epileptics. Conversely, false-negative results can come about due to chronic renal failure with decreased creatinine clearance and hypometabolism of dexamethasone as in hepatic failures. The current much lowered criterion of suppression of 8 am cortisol to less than 50 nmol/L will enhance the sensitivity of this test, particularly in those with mild hypercortisolism. However, the false positive rates will increase, thereby decreasing the overall diagnostic utility of the test.

A recent development in the diagnosis of Cushing's syndrome has been the measurement of the late night salivary cortisol. The concentration of cortisol in saliva is an excellent indication of biologically free cortisol. Cortisol concentrations in saliva are also independent of salivary

flow rate and reflect the serum cortisol levels. It takes advantage of the ease of collection of saliva, stability of the sample for analysis and has practical convenience for the patient. The advantages are that salivary samples are easy to collect and multiple samples can be sent. The majority of normal patients will have 11 pm salivary cortisol < 3 nmol/L. Late night salivary cortisol levels consistently above 6 nmol/L are found to be diagnostic of Cushing's syndrome.[5] Further studies will be needed before late-night salivary cortisol can be considered to substitute the 24-hour UFC or 1 mg overnight dexamethasone suppression test as a first-line screening test.

3) **Confirmation** of Cushing's syndrome has been traditionally achieved with the low-dose dexamethasone suppression tests. Some subtle cases of Cushing's syndrome may require the demonstration of loss of diurnal variation in the cortisol levels.

   Low-dose Dexamethasone Suppression Test (0.5 mg dexamethasone 6 hourly for 8 doses). At the end of the 48-hour test, a cortisol suppression to < 50 nmol/L is considered as almost certainly excluding Cushing's syndrome. Alternatively, measurement of the 24-hour UFC and 17-hydroxy- steroids during the last 24 hours of the test, and demonstration of failure to suppress 17-hydroxysteroids or 24-hour UFC is considered diagnostic for Cushing's syndrome. A decrease of UFC to less than 27 nmol/L per 24 hours on the second day of dexamethasone administration was considered a normal response.[2]

4) The **Endocrine Localization** of the etiology of Cushing's syndrome has traditionally relied on the high-dose dexamethasone suppression test as well as ACTH measurements. Due to the lack of easy availability of metyrapone, this test is rarely done locally. In recent years, a whole series of newer tests have been proposed including the CRH stimulation test, combined Dexamethasone-CRH test, desmopressin test, Adrenal Venous Sampling and Petrosal Sinus Sampling. These more elaborate tests are generally not required in the diagnosis of Adrenal Cushing's.

   ACTH levels are subnormal (usually < 5 pg/ml) in patients with ACTH-independent Cushing's syndrome where adrenal autonomy is suggested and imaging of the adrenal glands should be undertaken. Normal or elevated ACTH levels are indicative of ACTH-secreting neoplasm although those with ectopic ACTH syndrome generally have much higher levels of ACTH compared to the pituitary causes.

The ACTH-secreting pheochromocytoma is responsible for the paradox of an adrenal Cushing's having high ACTH levels.

Determination of ACTH levels requires use of prechilled EDTA tube for blood collection with placement in ice and rapid delivery to the laboratory. This is to prevent degradation of ACTH by plasma proteases, which may give rise to erroneous result.

High-dose Dexamethasone Suppression Test has been traditionally used to distinguish pituitary from non-pituitary sources of ACTH hypersecretion as Cushing's disease has been known to suppress with high-dose dexamethasone. However, the value of this test has been questioned in some studies, particularly looking at the level of suppression with high-dose dexamethasone in patients with ectopic ACTH and Cushing's disease. There was a significant overlap of the suppression of cortisol in both groups, thereby posing a question regarding the diagnostic value of this test in distinguishing Cushing's disease from ectopic ACTH.[6] There is an approximate 20% error in false localization with this test, prompting some to suggest that this test should be aborted for more elaborate and accurate diagnostic tests like sampling through radiological procedures.

5) **Radiological Localization**. The classically recommended strategy has been to attempt endocrine localization before radiological localization, although this may not always be practical or convenient. Newer strategic evaluations have proposed the more aggressive use of radiological sampling techniques wherever doubts exist and to avoid unexpected rarities.

In the detection of adrenal masses, the CT can detect masses as small as 0.5 cm and the use of intravenous contrast can help to distinguish adrenals from adjacent vessels. In comparison, the MRI has better resolution compared to the CT and is able to detect adrenal lesions as small as 0.5–1.0 cm. Paramagnetic contrast medium can be given to distinguish solid from cystic lesions and determine the vascularity of the lesion. However, for most cases of adrenal Cushing's, the lesions are large enough to be clearly demonstrated on the CT scan and MRI has not been shown to confer any added advantage. A pituitary MRI with gadolinium enhancement will reveal a discrete pituitary adenoma in up to 60% of patients.[1] It is also important to note that pituitary incidentalomas occur in about 10% of general population although these incidentalomas are usually small and less than 5 mm

in diameter. One may have to resort to more specialized test such as the bilateral inferior petrosal sinus sampling (BIPSS) if and when clinical, biochemical or radiological investigations in Cushing's syndrome are discordant or equivocal.

6) **Careful Follow-up**. Follow-up should ensure the recognition of the transient hypoadrenal state that accompanies cure after surgery. It is not uncommon to see the need for steroid replacement in gradually tailed down dose even for a year or more. There should be both clinical and biochemical resolution of symptoms and abnormalities. The criterion of "cure" has been defined in a recent consensus review. A low serum cortisol of below 50 nmol/L within 2 weeks after surgery is thought to be a good index of remission. The cortisol level is usually measured 5 to 14 days after surgery and at least 24 hours after the last dose of hydrocortisone.[2] If these do not occur, the diagnostic evaluation should then take into consideration a wrong initial localization or less commonly encountered entities like Carney's Complex as described below.

Rare causes and associations of adrenal Cushing's syndrome

Although the diagnosis of adrenal Cushing's often does not pose problems to the degree seen in ACTH-dependent Cushing's syndrome, some recently identified subsets of adrenal ACTH-independent Cushing's have heightened the need for careful evaluation prior to surgery.

A bilateral adrenal cause of Cushing's that is ACTH-independent has been described by Carney based on its characteristic histological presence of primary pigmented nodular adrenocortical disease (PPNAD). **Carney's Complex** comprises the association of PPNAD, spotty skin pigmentation together with a variety of tumors. The tumor encountered included cardiac and cutaneous myxomas, mammary myxoid fibroadenomas, Sertoli cell tumors of the testes, growth hormone pituitary adenomas, psammomatous melanotic Schwannomas, and possibly other neoplasms including adrenocortical and thyroid follicular carcinomas and ovarian cysts.[7]

Gastroinhibitory Peptide (GIP) secretory or **Food-dependent Cushing's** has been identified recently wherein most have either unilateral adrenal adenoma or bilateral macronodular adrenal hyperplasia with food-dependent hormonogenesis.[8] There is ectopic over-expression of gastroinhibitory polypeptide receptor but the precise mechanism of this aberrant expression is not known.

Subtle hypercortisolism can occur in patients with adrenal incidentalomas where there is abnormal response to at least 2 standard tests of the hypothalmo-pituitary axis function in the absence of clinical signs of Cushing's syndrome. A better terminology will be **Subclinical autonomous glucocorticoid excess**. There is still controversy over whether this entity is associated with long-term morbidity and whether treatment to reverse the subtle glucocorticoid excess is truly beneficial, although numerous case reports in the literature suggest the transient occurrence of cortisol deficiency upon surgical excision of these incidentalomas.

Normocortisolemic Cushing's has been previously described in literature with the presence of classical Cushingnoid features despite being eucortisolemic. Patients were found to have markedly elevated lymphocyte glucocorticoid receptors, speculated to be the cause of this phenotype. The exact cause is unknown.[9]

Cushing's syndrome may occur in association with **Multiple Endocrine Neoplasia I**.

Treatment

Therapy in Cushing's syndrome related to an adrenal etiology has to be adrenalectomy. This is currently performed increasingly with laparoscopic surgery. If the disease is one of the rare bilateral ACTH-independent adrenal hyperplasia, bilateral adrenalectomy with adequate steroid and fludrocortisone replacement is mandatory. Patients who have had Cushing's syndrome of longstanding duration will require steroid replacement after adrenalectomy and this should be carefully tailed down over the next 6–12 months. This steroid replacement will be necessary until the residually suppressed adrenal tissue fully recovers. In Cushing's disease, transsphenoidal surgery is the first line of treatment with the aim of complete resection of the pituitary adenoma with correction of hypercortisolism without inducing permanent pituitary deficiencies.

Primary Hyperaldosteronism

Prevalence and causes

Primary hyperaldosteronism often referred to as Conn's Syndrome, is a common cause for endocrine hypertension. Originally described by Conn as an adenoma, it is characterized by over-secretion of aldosterone and

hypertension with suppressed renin. The prevalence rates of 0.5–2.0% have been reported in unselected patients with essential hypertension. Among the causes, benign cortical adenoma is found in 79% of the cases, being 3 times more common in women than in men. Bilateral adrenal hyperplasia is found in 20% of the cases. Other rarer causes include adrenocortical carcinoma or glucocorticoid suppressible hyperaldosteronism.

Presentation

Symptoms can occur as a result of hypokalemia and are usually nonspecific, e.g. tiredness, muscle weakness, thirst, polyuria, and nocturia. Often, hypertension accompanies the hypokalemia and it can be resistant to control despite pharmacological therapy. The incidence of periodic paralysis as a presenting feature appears to be higher amongst the Asian populations.[10] Characteristically, patients with primary aldosteronism present with the biochemical features of

- **spontaneous** (unprovoked by diuretics) **hypokalemia** (usually potassium < 3.5 mmol/L);
- **inappropriate kalliuresis** (urinary potassium > 20 mmol/L);
- **mild metabolic alkalosis** (serum bicarbonate > 31 mmol/L); and
- **relative hypernatremia** (serum sodium > 142 mmol/L).

Up to 40% patients with surgically confirmed primary aldosteronism have been found to have normokalemia.

A recent retrospective study across five centers, has estimated that only between 9–37% of patients with proven hyperaldosteronism had hypokalemia.[11]

Screening

In patients with suspected (particularly those with hypertension and hypokalemia) primary hyperaldosteronism, the simple screening can be accomplished by measuring the paired random ambulatory plasma aldosterone (PAC) to plasma renin activity (PRA). Young *et al*.[12] found a ratio of PAC (in ng/dL) to PRA (in ng/ml/hr) of more than 20 to be indicative of probable primary aldosteronism. The PAC/PRA ratio will help identify patients who may have primary aldosteronism and yet normal plasma aldosterone levels. The demonstration of simultaneous high PAC and PRA would also help identify patients who have renovascular tumors, diuretic-induced hypokalemia and renin-secreting tumors, which

may cause a secondary aldosteronism. Congenital adrenal hyperplasia, Liddle's syndrome, Liquorice ingestion and exogenous mineralocorticoid treatment may often have a low PRA with a low PAC.

Confirmatory test

Primary aldosteronism can be confirmed by showing subnormal supine and erect renin activity with elevated or normal PAC. Failure of aldosterone suppression with salt loading (either orally for 3 days or with intravenous sodium chloride, before PAC is measured) is another feature of primary aldosteronism. This test, however, must be attempted with caution in those with cardiac failure. Adequate potassium repletion is important before testing, because of the stimulatory effect of aldosterone by hypokalemia. Patients may need to discontinue anti-hypertensive agents about 3 weeks prior to the test. Diuretics, angiotensin-converting enzyme inhibitors and angiotensin receptor blockers should preferably be stopped. Beta-blockers have a tendency to inhibit renin activity. However, if severe hypertension prevents the cessation of anti-hypertensives during the diagnostic period, use of alpha-blockers such as prazosin or doxazosin will be least interfering with the renin-angiotensin-aldosterone axis. Alpha-blockers and calcium channel blockers represent the best choices for anti-hypertensives that will have the least interfering effect on the measurements of renin and aldosterone. Discontinuation of anti-hypertensives should not put the patient at risk of the consequences of untreated hypertension.

Common subtype differentiation

Table 3 below summarizes the expected changes in response in PRA and PAC in adrenal adenoma as compared with hyperplasia. The postural test with the guiding interpretation that adenomas are ACTH-dependent and hyperplasias are angiotensin-driven has an approximate 20% error rate in the diagnosis of adenomas or hyperplasias.

The distinction of adenoma from hyperplasia on biochemical and radiological grounds is fraught with numerous confounding elements, some of which are highlighted under the rare causes of hyperaldosteronism.

Imaging

On unenhanced CT scans, adenomas are of low attenuation (<10 Hounsfield Units, HU) as they have high cytoplasmic lipid content. On

Table 3 Characteristics of Adenoma vs. Hyperplasia in Primary Aldosteronism

Subtypes	Time	Posture	PRA	PAC	Plasma Cortisol
Adenoma (ACTH-driven)	8 am	Supine	↓	↑	The conclusion is valid on the assumption that plasma ACTH or cortisol shows a higher 8 am value than 12 noon value, consistent with the diurnal variation.
	12 noon	Upright	↓	↓	
Hyperplasia (Angiotensin-driven)	12 noon	Upright	↓	N/↑	

MRI, most adenomas appear slightly hypodense or isodense relative to liver on T1 weighted images and hyperdense on T2 weighted images. Most adrenal adenomas are smaller than 1 cm in diameter.[12] Adrenal venous sampling is reserved for cases where adrenals appear normal and/or there is disagreement between imaging and biochemistry. It is accurate for preoperative localization of the source of abnormal hormone secretion, when findings are equivocal on cross-sectional imaging. However, this procedure is invasive, technically difficult and requires fluoroscopy guidance. The potential complications include adrenal infarction, hemorrhage, adrenal vein thrombosis, hypotensive crisis and adrenal insufficiency.

Rare causes and associations of primary aldosteronism

When primary aldosteronism leads to kidney damage and advanced arterial hypertension, a reversal of renin suppression may occur and this may obscure diagnosis of the underlying disorder. **A normal or even high-normal plasma renin** does not exclude the diagnosis of primary hyperaldosteronism. Severe arterial hypertension caused by primary hyperaldosteronism may lead to arteriolosclerotic kidney damage that counteracts renin suppression and accelerates the course of hypertension.

Glucocorticoid — remediable hyperaldosteronism, is suspected when there is a family history of hypertension with hypokalemia, early onset of hypertension, and a higher incidence of cerebrovascular event. Inherited autosomal dominantly, there is aberrancy in the expression of a chimeric gene in the section of the isozyme CPY11B2 that usually helps to synthesize aldosterone. As a result, the mineralocorticoid that is produced is under the feedback of ACTH. Chronic administration of

low-dose glucocorticoid can suppress the aldosterone excess, cause natriuresis and thus controlling the hypertension.

Unilateral hyperplasia of the adrenal gland has been rarely described, causing confusion in the diagnosis and evaluation. Similarly, **aldosterone-producing carcinomas** have been reported. Most of the latter are associated with concomitant hypersecretion of other adrenal hormones, but few may be only aldosterone-producing. Aldosteronomas have also been described in association with acromegaly, primary hyperparathyroidism, the multiple endocrine neoplasia 1 syndrome, neurofibromatosis, familial adenomatous polyposis, coexistent non-functional contralateral adenoma, angiomyolipomas and type IV renal tubular acidosis (hyperkalemic variety) thus causing confusion in the biochemical and radiological evaluation of patients.

Treatment

In aldosterone-producing adenomas, laparoscopic unilateral adrenalectomy is still an option of choice associated with a shorter hospital stay and less morbidity. With idiopathic bilateral adrenal hyperplasia, the cure rate with surgery is not as good. The medical option is feasible, and usually spironolactone is used coupled with other anti-hypertensives. It is not uncommon to see a spironolactone dose requirement exceeding 200 mg and in the authors' experience, the dosage may be reduced at a later stage. Pre-operative spironolactone responsiveness, in instances of adrenal adenoma usually predicts good post-operative control of hypertension. Whether the patient requires any post-operative anti-hypertensive medication would depend on the chronicity of the aldosteronism and hypertension, as well as any resulting complications including atherosclerosis.

Pheochromocytoma

Prevalence/Definition

Pheochromocytomas are rare chromaffin catecholamine-producing tumors, which arise from anywhere in the autonomic nervous system and are frequently sought for but rarely found. The prevalence estimates range from 0.01–0.1% of the hypertensive population with no sex preponderance and are seen usually from the third to fifth decade. If the lesion is located in the adrenal medulla it is known as a pheochromocytoma, and a paraganglioma if it is found in the sympathetic ganglia.

Classically, it has been known as the "10% tumor" (i.e. 10% are extra-adrenal in origin, 10% associated with the neuroectodermal syndrome, 10% multiple, 10% malignant, 10% in children, 10% with family history, 10% with recurrence and 10% as incidental findings). A very recent review challenged[13] the notion of this "10% rule" based on a study that on screening a large cohort of patients with sporadic pheochromocytoma with negative family history, as much as 25% had germline mutations of one of the four susceptibility genes for pheochromocytoma-, proto-oncogene RET, tumor suppressor gene VHL, SDHD (succinate dehydrogenase subunit D) and SDHB (succinate dehydrogenase subunit B). Each of these is associated with multiple endocrine neoplasia 2(MEN2), Von Hippel–Lindau and paragangliomas respectively. This study needs to be repeated as questions have been raised as to the selective bias involved in the populations studied.

Presentation

The clinical manifestations of pheochromocytoma are related to cardiovascular or metabolic consequences and the associated features, and are summarized below.

Table 4 Clinical Manifestations in Pheochromocytoma

Cardiovascular	Hypertension may be **p**aroxysmal, and associated with **p**ain (headaches), **p**erspiration, and **p**alpitations
	Arrhythmias or palpitations may be the initial or the only clinical findings
	Crisis states — myocardial infarction, pulmonary edema, hypertensive or hypotensive Crisis
Metabolic	Lactic acidosis
	Excess hormonal states — ACTH, parathyroid hormone, vasoactive intestinal peptides
	Glucose intolerance
Associated syndromes	Neurofibromatosis
	Multiple Endocrine Neoplasia Type II
	Von Hippel–Lindau Disease
Unsuspected	Autopsy findings
	Incidentalomas on CT
	Intraoperative finding during surgery performed for other reasons

Screening for pheochromocytoma

The 24-hour urinary excretion of catecholamines and the metabolites are tests of choice for screening of pheochromocytoma. 24-hour vanillylmandelic acid (VMA), metanephrines/normetanephrines and catecholamines such as adrenaline, noradrenalin, and dopamine may be secreted in excess. These tests are mostly done by High Performance Liquid Chromatography (HPLC).

VMA is least specific with a false-positive rate of 15% presumably due to dietary factors, which may interfere with the result. Drugs that may cause false-positive results in 24-hour urine metanephrines and/or catecholamines include tricyclic anti-depressants, labetolol, levodopa, ethanol, methyldopa, and benzodiazepines. The concept of periodic hormone excess must be borne in mind as normal 24-hour collection of catecholamines or metanephrines on one day may represent a trough between episodes of hypersecretion. Fractionated plasma metanephrine is highly sensitive but not specific, with 15% false-positivity. It may be reserved for high-risk patients such as those with positive family history and a previous resected pheochromocytoma. Hence repeat values during a spell/symptomatic period would be useful.

Localization of pheochromocytoma

CT and MRI can locate adrenal pheochromocytomas larger than 5–10 mm with greater than 95% sensitivity. The radiological appearances seen on CT are characteristically one of a rounded mass of similar density to the surrounding soft tissue structures, often with some central necrosis. Speckled calcification occurs in 12% of the tumors. Intense enhancement occurs after intravenous contrast media. MRI appearance of pheochromocytomas has been typically described as a "glow-bulb" on T2 weighted image, although this appearance is not pathognomonic as was once thought.

More specialized evaluation includes the utility of metaiodobenzylguanidine (MIBG) and Octreotide Scintigraphy. These are selectively used in instances where a pheochromocytoma is diagnosed but not localized or where multiple pheochromocytomas are suspected. A newer and novel approach for localizing "occult" pheochromocytoma involves employing vena caval blood sampling with measurements of plasma concentrations of normetanephrine and metanephrine, the

O-methylated metabolites of noradrenalin and adrenaline. Other approaches described involve positron emission tomographic (PET) scanning after intravenous injection of the sympathoneural-imaging agent, 6-[^{18}F] fluorodopamine.

Treatment of pheochromocytomas

Treatment involves

1) **Control of Blood Pressure** — This is usually accomplished by the use of alpha-blockers (e.g. phenoxybenzamine, prazosin, tetrazosin) before the introduction of beta-blockers. Calcium channel blockers have also been utilized successfully in some centers for the pre-operative control of blood pressure.
2) **Management of Hypertensive Crisis** — Phentolamine and sodium nitroprusside have been used in the treatment of emergencies and the intravenous dose is titrated carefully with blood pressure control.
3) **Management of Arrhythmias** — Beta-blockers like propanolol.
4) **Adequate Pre-operative Preparation** — This is mandatory to prevent intraoperative hypertensive crisis as well as immediate post-operative hypotension. Adequate alpha- and beta-blockade, volume replacement, intraoperative monitoring, and readily available pressor agents (noradrenaline and phenylephrine) as well as experienced anesthetic cover are mandatory.
5) **Surgery** — The common adage in surgery is to dissect the patient away from the tumor rather than the tumor from the patient, highlighting the need for experience as well as careful dissection of the tumor.

Adrenal Insufficiency

Definition

Addison first described adrenal insufficiency in 1855 as a "constitutional and local effect of disease of the suprarenal capsules". It can be classified as primary or secondary adrenal insufficiency. In primary adrenal insufficiency, destruction of adrenal gland tissue occurs with consequent reduction in production of either or both of cortisol and aldosterone. Compensatory elevation of ACTH and melanocyte-stimulating hormone

(MSH) occurs. In secondary adrenal insufficiency, the locus of failure is hypothalmo-pituitary axis with a decrease in production of ACTH and therefore cortisol. Aldosterone production usually remains intact, as it is dependent upon the renin-angiotensin axis. Functional adrenal insufficiency occurs when administration of exogenous corticosteroids leads to depressed ACTH. When these exogenous steroids are discontinued, a clinical picture like secondary adrenal insufficiency ensues.

Causes

Table 5 Etiology of Adrenal Insufficiency

Primary Adrenal Insufficiency	Autoimmune adrenalitis (Addison's disease) Tuberculosis Aids-related Cytomegalovirus infection, histoplasmosis Vascular hemorrhage (Waterhouse Friderichsen) Thrombosis, including Lupus anti-coagulant syndrome Metastatic cancer: Lung, gastric, breast Granulomatous disease: Sarcoidosis, Amyloidosis Hemochromatosis Adrenoleukodystrophy (Schilder's disease) ACTH or Glucocorticoid resistance
Secondary Adrenal Insufficiency	Secondary to hypothalamo-pituitary disease of whatever etiology affecting either CRH and/or ACTH
Functional Adrenal Insufficiency	Long-term glucocorticoid treatment may result in suppressed adrenal function, which may never always recover

Presentation in adrenal crisis

The classical description of an acute crisis is a profoundly ill patient who presents with nausea, vomiting and severe abdominal pain with fever. There may be confusion and disorientation with lethargy. Volume depletion occurs with resultant hypotension and shock. Chronic presentation may be subtle with chronic fatigue, lethargy with an occasional development of hyper pigmentation in a hypoadrenal patient. Mechanism of hypotension stems from cortisol deficiency, which may depress myocardial contractility; responsiveness to catecholamines may be decreased, resulting in poor response to pressors. Aldosterone deficiency may coexist.

Biochemical manifestations of adrenal insufficiency

Table 6 Biochemical Manifestations of Adrenal Insufficiency

Biochemical	Mechanism
Hypoglycemia	decreased gluconeogenesis due to cortisol deficit and increased peripheral glucose use secondary to lipolysis
Hyponatremia	elevated ADH levels due to decreased circulating blood volume, aldosterone deficiency may coexist
Hypercalcemia	related to increased protein binding or volume depletion
Azotemia	volume depletion
Elevated hematocrit levels	volume depletion
Mild metabolic acidosis	aldosterone deficiency

Diagnosis and investigations

Screening Tests: A random plasma cortisol level of >500–$525\,\text{nmol/L}$ generally rules out adrenal insufficiency. Levels $<100\,\text{nmol/L}$ (unless taken during the nadir of the diurnal variation at 10 pm to 12 midnight) are highly indicative of adrenal insufficiency. Levels in the mid-range of >100–$500\,\text{nmol/L}$ would require dynamic testing for adrenal insufficiency.

An ACTH level of $\geqslant 100\,\text{pg/ml}$ or $22\,\text{pmol/L}$ is consistent with primary adrenal insufficiency.

Dynamic Tests: ***Corticotrophin Stimulation Test*** or ***Synacthen Test*** involves the administration of $250\,\mu\text{g}$ corticotrophin intravenously or intramuscularly. Blood for cortisol levels are taken at 0, 30, 60 minutes thereafter. A maximal level of cortisol ($\geqslant 500$–$550\,\text{nmol/L}$) rules out adrenal insufficiency whereas a level of <500–$550\,\text{nmol/L}$ is suggestive of adrenal insufficiency. The role of $1\,\mu\text{g}$ corticotrophin test in the evaluation of hypothalmo-pituitary adrenal (HPA) function has been suggested and discussed. Assuming the "gold standard" and 100% accuracy of insulin tolerance test (ITT), $1\,\mu\text{g}$ low-dose ACTH test yielded sensitivity of 71% and specificity of 93%. The low-dose ACTH test is a safe and inexpensive tool to assess the HPA function as outpatient. If the peak cortisol response is $>750\,\text{nmol/L}$, it strongly predicts normal HPA function and there may not be a need to subject to ITT. If basal cortisol is of a very low level, this may be indicative of overt hypoadrenalism,[14] and subjecting to

hypoglycemic stress may not be necessary. ITT is still mandatory between the grey zone limits. There seems to be very little consensus on the advantages of the 1 μg vs. the 250 μg ACTH stimulation tests and normal ranges for different centers are not easily available. More recent reviews have highlighted the need to look at not only the basal cortisol levels but the change with corticotrophin stimulation and the free plasma cortisol level (rather than total which is usually measured) in critical illness.

Treatment

Replacement with glucocorticoids

Maintenance: Glucocorticoid dosages should be individualized with a starting dose of 25 mg hydrocortisone (usually 15/10 mg or 10/10/5 mg) or an equivalent dose of 37.5 mg cortisone. Longer-acting glucocorticoid preparations such as dexamethasone (0.5 mg/day) or prednisolone (5 mg/day) have been advocated by some to avoid the high peak levels and inadequate trough coverage. However, these compounds have variable metabolism and higher incidence of over-replacement. The dose is titrated according to sense of well-being and physical strength while avoiding weight gain, hypertension, hyperglycemia and osteoporosis.

Stress Dosing Glucocorticoid: Replacement with corticosteroids such as intravenous dexamethasone 6–10 mg 6 h that is 100 times more potent than cortisol and does not interfere with cortisol levels unlike hydrocortisone can be given if the synacthen test needs to be performed. Alternatively, intravenous hydrocortisone 100 mg 6–8 h can be used. Cortisone acetate 100 mg can be given intramuscularly 6–8 h but absorption is erratic and not reliable. Different requirements are required according to varying degrees of surgical stress.

Degrees of Surgical Stress

Minor	25 mg hydrocortisone day of procedure
Moderate	50–75 mg hydrocortisone day of procedure, rapid taper in 2 days
Severe	100–150 mg hydrocortisone day of procedure, rapid taper in 2 days
Critically ill	100 mg hydrocortisone iv bolus followed by 50–100 mg hydrocortisone iv every 6–8 hours with 0.05 mg/day fludrocortisone until shock resolves

Replacement with mineralocorticoids

Mineralocorticoid replacement, mainly fludrocortisone, is usually replaced as a single daily dose of 0.05–0.20 mg/day. This is titrated to salt craving and postural hypotension together with serum potassium and upper-range renin.

Dosing and Relative Potency

Table 7 shows the relative potency and mineralocorticoid activity[15] of the common clinically used form of steroids.

Table 7 Dosing and Relative Potency

Glucocorticoids	Approximate Equivalent Dose (mg)	Relative Mineralocorticoid Potency
Cortisone	25	2
Hydrocortisone	20	2
Prednisolone	5	1
Dexamethasone	0.75	0

Adrenal Incidentalomas

Definition

These are defined[16] as clinically inapparent adrenal masses discovered inadvertently during the course of evaluation and treatment for other clinical conditions not related to suspicion of adrenal diseases. Therefore, adrenal masses picked up during imaging procedures as part of staging and work-ups of cancer are excluded. The prevalence increases with age. Based on autopsy studies, adrenal masses are among the most common tumors in humans: at autopsy, an adrenal mass occurs in at least 3% of persons over age 50.

Functional assessment

Incidentalomas present a dilemma for physicians because while many of these masses are harmless, a few can progress or become silent or apparently silent carcinomas or hormonally active tumors. Thus, the evaluation of the functional status of adrenal incidentalomas is necessary. It is important to exclude silent pheochromocytoma, subclinical Cushing's syndrome, primary aldosteronsim and virilizing or feminizing tumors, all of which

may be associated with morbidity if left untreated. A simple screen should thus include a 1 mg dexamethasone suppressed cortisol, evaluation of the renin-aldosterone ratio with electrolytes, measurement of testosterone/estradiol levels, and urinary metanephrines.

Evaluation for possible malignancy

Size has been considered a significant factor in determining whether an adrenal lesion is benign or malignant. The general consensus is that most benign adrenal lesions are less than 3 cm, whereas malignant lesions tend to be larger than 6 cm. The nature of those lesions between 3–6 cm remains uncertain. It is difficult to just rely on size alone to determine the nature of the lesion. Other than gross anatomical size, other clinically important factors include a previous history of cancer and the extent of vascularization of the mass.

Beyond size,[17] further information can be gathered from other diagnostic techniques. CT and MRI of the adrenal gland can ascertain whether the mass is rich or poor in lipids. Lipid-rich masses tend to be benign whereas lipid-poor masses have a higher probability of being malignant. With un-enhanced CT, an attenuation of less than +10 HU is consistent with adenoma and non-adenomas do usually have un-enhanced attenuation of greater than +30HU.With contrast CT, benign lesions have a greater than 50% washout of contrasts compared to malignant lesions which have a slower washout phase. Chemical shift MR shows decrease in signal intensity in adenomas whereas non-adenomas show no significant change in relative signal intensity. Malignant lesions also are inhomogeneous due to presence of areas of necrosis compared to benign tumors. Scintigraphy with ^{131}I-^{6}beta-iodomethylnorcholesterol has been used. Concordant images are generally benign whereas discordant images are not.

Fine Needle Aspiration Cytology (FNAC): In patients with history of malignancy whereby no outward clinical sign of metastasis is present, a FNAC of those having a heterogeneous adrenal mass with a high attenuation value (>20HU) can have a role. However, FNA still has a false-negative rate and it cannot totally exclude malignancy. In patients with no history of malignancy with an adrenal incidentalomas, FNA cannot be a recommended standard procedure. Functional assessment, particularly the exclusion of pheochromocytoma, is important before any FNA of the

adrenal is performed. The documentation of a functional disorder may negate the need for a FNA.

Surgical option in adrenal incidentalomas

Recommendations for surgery based on tumor size are derived from studies not standardized for inclusion criteria, length of follow-ups, or methods of estimating the risk of carcinoma. For patients with evidence of biochemical hormonal excess in absence of clinical symptoms, surgery should still be the considered option particularly in those with pheochromocytoma and hyperaldosteronism. More data on long-term morbidity in those with subclinical Cushing's will be needed as adrenalectomy or careful observation have both been suggested as management options. In nonfunctional incidentalomas, the determination of malignant and benign tumors will guide managements. It is generally accepted that lesions greater than 6 cm should be removed. Lesions lesser than 4 cm and with low risk of malignancy from criteria discussed above can be monitored. For lesions between 4–6 cm, either close follow-up or surgery can be reasonable approaches. In summary, surgery is reserved for patients with large tumors, imaging characteristics suggestive of malignancy, recent severe hypertension, evidence of increase in tumor size or overt hypersecretion during follow-up until the discrepancy between clear improvement in clinical and endocrine features in surgically treated patients and apparent stability of those treated medically is elucidated.

ADRENAL ANDROGENS AND ESTROGENS

Adrenal androgen- or estrogen-producing tumors are less common than the other clinical syndromes described and may be benign or malignant. The production of excess androgens, particularly in large adrenal tumors should raise the possible concern of an underlying adrenal carcinoma. Androgen-producing tumors may cause virilization in women. Similarly, estrogen-producing tumors may produce feminizing features in men.

Dehydroepiandrosterone (DHEA) and its sulphate form DHEAs are steroid precursors of androgens and estrogens that are secreted in large amounts by the adrenal glands. Neglected as hormones of significance for many years, they have received widespread interest and publicity in the past decade. The alleged benefits include enhancement of cognition,

generalized well-being, reversal of depression and obesity, enhancement of immune responsiveness, prevention of malignancies and reversing of aging. DHEA circulates in high levels in fetus, falls in the neonate and remains relatively low for 5–7 years, then rises during adrenarche and peaks at the age of 20–30. Serum levels then fall progressively at the rate of 1–2% per year throughout adulthood. The ratio of DHEA/DHEAs remains unchanged with age. Given the paucity of information available on the pharmacologic effects of DHEA and the limited data on the beneficial effects in humans, the risk: Benefit ratio favors caution in the use of DHEA. When given orally or transdermally DHEA is converted to androgens and estrogens. Thus, in men, the potential for prostatic hypertrophy, prostate cancer, or gynecomastia should be guarded against. In women, potential risks include virilization, dyslipidemia and reduction of HDL.

REFERENCES

1. Boscaro M, Barzon L, Fallo F, Sonino N, Cushing's syndrome, *Lancet* **357**: 783–791, 2001.
2. Arnaldi G, Angeli A, Atkinson AB, Bertagna X, Cavagnini F, Chrousos GP, Fava GA, Findling JW, Gaillard RC, Grossman AB, Kola B, Lacroix A, Mancini T, Mantero F, Newell-Price J, Nieman LK, Sonino N, Vance ML, Giustina A, Boscaro M, Diagnosis and complications of Cushing's syndrome: A consensus statement, *J Clin Endocrinol Metab* **88**:5593–5602, 2003.
3. Boscaro M, Barzon L, Sonino N, The diagnosis of Cushing's syndrome: atypical presentations and laboratory shortcomings, *Arch Intern Med* **160**:3045– 3053, 2000.
4. Castro M, Elias PCL, Quidute ARP, Halah FPB, Moreira AC, Out-patient screening for Cushing's syndrome: The sensitivity of combination circadian rhythm and overnight dexamethasone suppression salivary cortisol test, *J Clin Endocrinol Metab* **84**:878–882, 1999.
5. Raff H, Raff JL, Findling JF, Late night salivary cortisol as a screening tool for Cushing's syndrome, *J Clin Endocrinol Metab* **83**:2681–2686, 1998.
6. Aron DC, Raff H, Findling JW, Effectiveness versus efficacy: The limited value in clinical practice of high dose dexamethasone suppression testing in the differential diagnosis of adrenocorticotrophin-dependent Cushing's syndrome, *J Clin Endocrinol Metab* **82**:1780–1785, 1997.
7. Carney A, Young WF, Primary pigmented nodular adrenocortical disease and its associated conditions, *The Endocrinologist* **2**:6–21, 1992.
8. Reznik, *et al*., Food-dependent Cushing's syndrome mediated by aberrant adrenal sensitivity to gastric inhibitory polypeptide, *New Eng J Med* **327**: 981–986, 1992.

9. Newfield RS, Kalaitzoglou G, Licholai T, Chilton D, Ashraf J, Thompson EB, New MI, Normocortisolemic Cushing's syndrome initially presenting with increased glucocorticoid receptor numbers, *J Clin Endocrinol Metab* **85**:14–21, 2000.

10. Ma JT, Wang C, Lam KS, Yeung RT, Chan FL, Boey J, Cheung PS, Coghlan JP, Scoggins BA, Stockigt JR, Fifty cases of primary hyperaldosteronism in Hong Kong Chinese with a high frequency of periodic paralysis, Evaluation of techniques for tumor localization, *Q J Med* **61**(235):1021–1037, 1986.

11. Mulatero P, Stowasser M, Loh KC, Fardella CE, Gordon RD, Mosso L, Gomez–Sanchez CE, Veglio F, Young WF. Increased diagnosis of primary aldosteronism, including surgically correctable forms, in centers from five continents, *J Clin Endocrinol Metab* **89**:1045–1050, 1994.

12. Young WF, Hogan MJ, Klee GG, Grant CS, Van Heerden JA, Primary aldosteronism: Diagnosis and treatment, *Mayo Clin Proc* **65**:96–110, 1990.

13. Rossi GP, Sacchetto A, Chiesura-Corona M, De Toni R, Gallina M, Feltrin GP, Pessina AC, Identification of primary aldosteronism with adrenal vein sampling in patients with equivocal computed tomography and magnetic resonance findings, *J Clin Endocrinol Metab* **86**:1039–1045, 2001.

14. Dluhy RG, Phaeochromocytoma — death of an axiom, *New Eng J Med* **346**:1486–1488, 2002.

15. Ambrosi B, Burbetta L, Re T, Passini E, Faglia G, The one microgram adrenocorticotrophin test in the assessment of hypothalmic-pituitary-adrenal function, *Eur J Endocrinol* **139**:575–579, 1998.

16. Lacy, *et al., Drug Information Handbook,* 8th ed., Lexi–Comp Inc. Hudson (Cleveland) pp. 1343, 2000.

17. National Institutes of Health State-of-the-Science Conference Statement, Management of the Clinically Inapparent Adrenal Mass (Incidentaloma), February 4–6, 2002, Final Statement July 16, 2002.

18. Schteingart DE, Endocrine incidentalomas — Management approaches to adrenal incidentalomas, *Endocrinol Metab Clinics* **29**:127–139, 2000.

This page intentionally left blank

Gastroenterology

This page intentionally left blank

22

Practical Approaches to GI Disease

Ng Han Seong

Gastrointestinal (GI) complaints such as abdominal pain, diarrhea, constipation and blood in the stools, are common presenting symptoms of gastrointestinal diseases. They form a significant proportion of medical problems/cases seen by the medical practitioner. Most of these symptoms are due to GI diseases. However, complaints like diarrhea/constipation, abdominal pain, difficulty in swallowing, and even weight loss, may be due to systemic diseases that secondarily involve the GI tract, e.g. systemic lupus and thyrotoxicosis. Yet another big problem is that of "functional bowel" disease — the non-ulcer dyspepsia and irritable bowel syndrome, whose symptoms and presentation may mimic organic GI diseases, e.g. peptic ulcer, inflammatory bowel diseases or cancers of the stomach and colon.

In order for the medical practitioner to select the appropriate diagnostic tests, and to arrive at a definitive diagnosis, and subsequently the treatment, a careful and detailed history and physical examination are essential.

GASTROINTESTINAL DISEASES IN SINGAPORE

GI functional diseases are very common in Singapore. Fifty percent of GI outpatients have functional disease — non-ulcer dyspepsia or irritable bowel syndrome. Among community subjects surveyed in Singapore, 8% and 2% reported having dyspepsia (6 times/year) and irritable bowel syndrome, respectively. It is also important to note that local Chinese and Malays are more likely to be lactase deficient compared to Westerners.

Among the organic GI diseases, peptic ulcer disease is a common cause of chronic upper abdominal (epigastric) pain. Most particular diseases are associated with *Helicobacter pylori* infection, with non-steroidal anti-inflammatory drugs (NSAIDs) accounting for the rest. The incidence of peptic ulcer disease is higher in the Chinese (5 times more than Malays). This is reflected in the *Helicobacter pylori* infection (prevalence of the *Helicobacter pylori* serology), which is low among Malays. Colon, stomach, liver and esophageal cancers are among the top cancers seen in Singapore, with colon, stomach and liver making it to the top 5. These cancers are also more prevalent among the Chinese. The incidence of reflux esophagitis seem to be increasing, although it is still relatively uncommon compared to the West. In a study from Singapore, 4.5% of patients undergoing upper GI endoscopy have reflux esophagitis compared to 9–22% from the West. There was a slight preponderance of this disease among the Indians. Inflammatory bowel disease (IBD), another relatively uncommon GI disease in Singapore, is seen among all the races (Chinese, Malays, Indians). Its incidence is also increasing, and again, with slight preponderance among the Indians. Colonic diverticular disease is a disease of the elderly. In the West, sigmoid colon is most commonly affected. In Asia, as well as in Singapore, diverticular disease is usually right-sided, often asymptomatic, and tends to be an incidental finding in barium enema or colonoscopy.

Table 1 Common GI Diseases in Singapore

Non-ulcer dyspepsia
Irritable bowel syndrome
Peptic ulcer disease
Cancers — colon, stomach, esophagus
Infection — gastroenteritis
** GI diseases with increasing incidence —
Gastroesophageal reflux disease
Inflammatory bowel disease

History

Common GI complaints are abdominal pain, and change in bowel habit as in constipation and diarrhea, dyspepsia, dysphagia, heartburn, bloatedness/wind and blood in the stools. Full analysis of these complaints requires further details with regards to — character, location, temporal sequence, aggravating/relieving factors (including self-treatment, prescribed medicine) and other related/associated symptoms.

Abdominal Pain

Abdominal pain is one of the more important presenting symptoms. Pain may be described as "indigestion" or "discomfort." Abdominal pain is usually due to stretching or contracture, ischemia, inflammation or neoplastic disease. Location of the pain may be useful, as listed in Table 2.

Typically, pain from duodenal ulcer occurs during "hunger," relieved by food, and wakes the patient up from sleep at night. Substernal burning sensation (heartburn) associated with acid reflux are quite specific for gastroesophageal reflux disease. Colicky pain in the right hypochondrium/epigastrium is suggestive of biliary stones. Lower abdomen pain with passing of small amount of stools with blood/mucus are quite characteristic of colonic inflammation or neoplasm. Patients with vasculitis involving the small intestine may present with abdominal pain and blood in the stools. Another condition that is not a GI disease and which presents with lower abdominal pain in young women is endometriosis.

Dyspepsia

Dyspepsia is another very common GI complaint. Dyspepsia may be defined as episodic or persistent upper abdominal pain. Other symptoms

Table 2 Abdominal Pain — Location and Organ Involved

Retrosternal	Esophagus (angina!)
Epigastric	Stomach, duodenum, pancreas (angina!)
Right hypochondrium	Liver, gallbladder, duodenum
Periumbilical	Small intestine (ischemia)
Right iliac fossa	Cecum, appendix
Lower abdomen	Colon
Suprapubic	Colon, urinary bladder

Table 3 Common Causes of Organic Dyspepsia

Gastric and duodenal ulcer
Gastric cancer and liver cancer
Gastritis (*Helicobacter pylori* +ve)
Drugs — NSAIDs and Aspirin
Biliary stone disease

may also be present — bloatedness/wind, early satiety/fullness, anorexia, nausea and vomiting. It should be present for some duration, perhaps, 3 months. Dyspepsia may be organic (see Table 3) or functional.

Functional dyspepsia may be considered in young patients (less than 35 years of age) and without an alarm "symptoms/signs." In Singapore, where the incidences of stomach and liver cancers are high, one should be always on the alert to exclude such GI diseases.

Heartburn

Heartburn is a "burning" or "hot" sensation felt over the lower sternum or near the xiphoid process. Smoking, drinking alcohol and overeating may aggravate the symptom. A patient with heartburn has the symptom intermittently. Heartburn is often suggestive of gastroesophageal reflux disease. It may also be present in patients with mechanical obstruction, e.g. esophageal cancer, peptic stricture, or gastric outlet obstruction. Its association with respiratory problems, e.g. asthma, recurrent pneumonias; or ENT problems, e.g. hoarseness of voice and recurrent laryngitis, should lead the medical practitioner to diagnose gastroesophageal reflux. It is also important that angina pectoris be excluded.

Dysphagia

Dysphagia is the perception of difficulty in swallowing. It is often due to neuromuscular disorders or structural abnormalities, e.g. cancer or stricture of esophagus, causing obstruction. Dysphagia from cancer in the esophagus is progressive, and is worse with solids than liquids. Patients with oropharyngeal dysphagia from neuromuscular disorders have difficulty swallowing liquids compared with solids. They often have difficulties in initiating swallowing, and there may be associated choking and coughing on swallowing.

Painful swallowing (odynophagia) associated with dysphagia suggests inflammation/ulceration in the esophagus — e.g. severe erosive gastroesophageal reflux disease; HIV infection, e.g. cytomegalovirus, candidiasis; or from medications lodged in the esophagus eg. tetracycline. Dysphagia may be a manifestation of a systemic disease, e.g. scleroderma, which had affected esophageal motility.

Nausea and Vomiting

Nausea is the sensation one gets prior to vomiting. Vomiting is the expulsion of gastric content through the mouth. It may be preceded by retching. Nausea and vomiting may be acute or chronic. It is important to note that these symptoms may be related to neurological disorders rather than the GI. Labyrinthitis/vestibular neuronitis may present with nausea and vomiting. Raised intracranial pressure from brain tumor, for example, may be a cause of chronic vomiting/nausea. One must also not forget about pregnancy! Chronic renal failure may also manifest or present with nausea/vomiting.

As for GI diseases, acute symptoms may be due to infective gastroenteritis, acute infection/inflammation, e.g. acute cholecystitis or pancreatitis, and bowel obstruction. Gastric outlet obstruction or gastroparesis may result in chronic vomiting. Emotional disturbances and pyschogenic disorders may also present as chronic vomiting/nausea, e.g. eating disorders and psychogenic vomiting.

Diarrhea

Diarrhea is increased fluidity and frequency of stools. There may be increased stool volume/weight as well. In evaluating diarrhea, it is important to verify the patient's complaint, as there can be much disparity. The complaint could well be due to increased frequency yet stools are formed, urgency to defecate especially after a meal (gastrocolic reflex) or incontinence.

The nature, frequency, timing and duration of the diarrhea are useful information. The patient should describe the characteristics of the stool — color, any blood/mucus, consistency, and buoyancy. Nocturnal diarrhea strongly suggests an organic cause. In a diabetic, it suggests autonomic dysfunction. The presence of blood with/without mucus suggests an

inflammatory or a neoplastic process. In the elderly, an ischemic process may be important. Abrupt onset (acute) may be due to infection, while chronic diarrhea (more than 3 weeks) would indicate malabsorption, inflammatory bowel disease, chronic infection, e.g. TB, or cancers.

In the diagnostic workup of diarrhea, a careful inquiry into the presence of associated symptoms will help in sharpening the focus and selection of investigation. Weight loss may be significant, and may suggest malabsorption from chronic inflammatory or neoplastic process. Abdominal pain may indicate an inflammatory, a neoplastic or an ischemic process. Tenesmus indicates rectal involvement. Fever, fatigue and malaise may suggest a chronic infection/inflammatory or a neoplastic process. Systemic diseases, e.g. thyrotoxicosis may also cause diarrhea.

Medication, laxatives, abdominal operation (cholecystectomy, gastrectomy, ileal resection) are important history in the evaluation of diarrhea.

Constipation

Constipation is another common GI complaint. It may be due to an organic disease like colonic cancer; metabolic, e.g. hypercalcemia, hypothyroidism; pelvic floor dysfunction; or a functional disorder, e.g. irritable bowel syndrome.

It may be defined as less than three bowel movements per week associated with straining at more than 25% of bowel movements.

A detailed history of bowel habit should be elicited — frequency, stool consistency, straining, pain upon defecation, feeling of incomplete defecation, and use of laxatives/enemas/manual evacuation. Medication may cause constipation. Dietary habits may be relevant. The patient should be questioned regarding weight loss, rectal bleeding, abdominal pain and distension. Duration of the complaint is also relevant, and the patient may be constipated for years! but would report to the doctor, just because a close friend died of colonic cancer recently. A short history should raise the possibility of colorectal malignancy. Anal pain and pain on defecation suggest a local cause such as perianal abscess or fissure.

Rectal bleeding

Rectal bleeding is another common GI complaint. Most bleeding are mild/minor, and are from anal disease and hemorrhoids. About 25% may have significant disease, e.g. colonic/rectal cancers, diverticular disease,

chronic inflammatory bowel disease, angiodysplasia and ischemic colitis. The presence of tenesmus may suggest rectal disease. Painless bleeding, sometime massive, may be from diverticular disease. Abdominal pain and bloody stools, especially in the elderly, is highly suggestive of ischemic colitis. In massive upper GI bleeding, malena may be passed mixed with fresh blood.

PHYSICAL EXAMINATION

Physical examination forms an integral part of the GI evaluation. In some instances, the examination may be normal. Nonetheless, a normal examination is an important step towards the diagnosis of functional disorders, e.g. non-ulcer dyspepsia and irritable bowel syndrome.

The initial assessment should include a thorough general examination. It provides information regarding the patient's general well being/health and nutritional status. It is also able to provide valuable information on the diagnosis and management of GI diseases. The presence of dehydration and hypotension would require immediate attention in a patient with diarrhea. In the assessment of pallor, a per rectal examination which reveals malena or blood would lead to a full assessment for GI bleeding. Abdominal pain may be due to small intestine vasculitis as in systemic lupus; or ischemia in the older patient with cardiac problems and atherosclerosis; or in a case of hemolysis. Endometriosis and other gynecological problems present as lower abdominal pain. Steroid withdrawal may be the cause of abdominal pain in a patient with Cushingnoid features. The presence of an enlarged thyroid and peripheral signs of thyrotoxicosis would explain the diarrhea and weight loss.

Examination of the abdomen is largely influenced by the presenting GI complaints. The abdomen is examined for tenderness, masses, visceromegaly, ascites, and if acute surgical problems are suspected — bowel sounds and rebound tenderness, and other signs of peritonitis. The location of tenderness may reflect the involvement of the underlying organ, as listed in Table 2. Auscultation for bruits and rubs is often omitted. They, however, provide useful information if present, as a hepatic rub or bruit may suggest hepatic malignancy. Per rectal examination is often carried out for lower abdominal pain and for bleeding per rectum, malena, and for detecting piles/lower rectal cancer and anal canal disease.

Nutritional assessment

One of the primary function of the GI tract is digestion and assimilation of nutrients. GI diseases can give rise to nutrient deficiency through decreased intake of food, decreased absorption, increased loss or increased requirement of nutrients (e.g. in sepsis). Evaluation of nutrient deficiency, and how this state of deficiency affects the body's function, form part of the GI assessment. Nutrient repletion and support are important aspects of treatment in GI diseases, e.g. inflammatory bowel disease.

Information regarding the patient's height, usual weight, unintentional weight loss, coexisting medical problems, medications, past surgery (cancer operation, gastrectomy, gut resection), food preference, taking of supplements, dentition and access to food, are important.

Physical examination includes looking out for angular stomatitis/cheilosis, glossitis, gingivitis, bruising of skin, perifollicular petechiae, scaly dermatitis, pitting ankle edema, and loss of cutaneous fat. Body mass index (BMI) and anthropometric measurements, e.g. tricep skin fold thickness, mid-arm muscle circumferences, waist to hip ratio, may be taken to assist in the evaluation of the nutritional status of a patient.

Simple investigation of "nutritional significance" includes full blood count (hemoglobin, red cell morphology, MCV, lymphocyte count), serum iron/TIBC, serum albumin and Vit B12/folate.

Psychological assessment

Psychological disturbances are related to GI disorders. Stress may be a precipitating or aggravating factor for organic diseases, e.g. inflammatory bowel disease, and functional disorders like non-ulcer dyspepsia and irritable bowel syndrome. Disorders such as depression, anxiety, cancer phobia, and disordered cognitive function may affect manifestation/presentation of GI disorders. Referrals to a psychiatrist may be necessary to provide additional support and care of such patients.

IMAGING IN GASTROENTEROLOGY

Diagnostic imaging of the GI tract and its related structures has made tremendous progress over the past decade. It has increasingly become more complex. There are also more options to choose from. In addition, interventional radiology also plays an important role in the non-surgical

treatment of many GI conditions. In order to tailor the studies to each GI problem, clinical information and discussion with the radiologist are essential. Pregnancy is contraindicated in X-ray studies.

Plain X-rays

Chest X-ray (CXR) and **Abdominal X-ray (AXR)** are still being used in the evaluation of GI patients. The erect **CXR** has been used frequently for detecting free air under the diaphragm to confirm a ruptured viscus, e.g. perforated gastric ulcer. Mediastinal air and/or pleural effusion may indicate esophageal perforation following instrumentation. **AXR** is helpful for detecting intestinal obstruction, although the site of obstruction may not be identifiable. Dilated loops of bowel with fluid levels typically signify obstruction. Bowel wall thickening may suggest ischemia or inflammation. The presence of air in the biliary tree may suggest a recent passage of gallstone or past sphincterotomy or biliary enteric anastomosis. Calcification may be seen, and can be due to gallstones, renal stones or calcification in the pancreas.

Intestinal transit time may be studied using ingested radio-opaque markers, and **AXR** taken serially daily for 5 days. This is a simple investigation for chronic constipation.

Contrast radiography

Oral cholecystography had in the past been the investigation for suspected cholelithiasis as well as gallbladder function. Presently it is replaced by ultrasound. It may be used occasionally in very obese patients where ultrasound examination has failed.

Barium swallow and **barium meal** are still being used to assess/diagnose upper GI diseases. They have certain limitations, e.g. no tissue biopsy, problem of aspiration, and contraindicated in perforation (unless water-soluble contract medium is used), and in small intestine obstruction and bleeding GI. Barium swallow can be used to assess dysphagia, and it may demonstrate tumors or strictures. It may also show up esophageal spasm and achalasia. It is not sensitive for the diagnosis of erosions and varices. Barium meal may be used to diagnose peptic ulceration, pyloric stenosis or gastric cancer. With the advent of upper GI endoscopy, the role of barium swallow and meal has diminished, as endoscopy allows tissue biopsy for histology, culture and testing for *Helicobacter pylori.*

Barium meal follow through/small bowel enema are performed to study the small bowel. They may demonstrate strictures, tumors or thickened folds. Some of these abnormalities may be localized or more widespread.

Barium enema may be used to evaluate patients with change in bowel habit or lower abdominal pain. Its main limitation is again that it does not allow for histology and for the removal of polyps. It may also miss small polyps. It can demonstrate strictures, diverticular, cancers and big polyps. Ulcerative colitis is better diagnosed with the colonoscopy, although barium enema may demonstrate ulcerations associated with colitis. CT scan of the abdomen is better for demonstrating diverticulitis and inflammation in the pericolic fat. Colonoscopy occasionally failed to reach the cecum (those with long and tortuous colon), and barium enema is ordered to complete the assessment.

Computed tomography

Computed tomography (CT) consistently produces good images of the GI tract, gall bladder, liver, pancreas and abdominal lymph nodes. It is increasingly used in the diagnosis of perforated viscus, intestinal obstruction, bowel ischemia, intraabdominal abscesses associated with appendicitis or diverticulitis, right iliac fossa mass, thickening of stomach/bowel, and in cancer of the esophagus, pancreas and metastatic cancers. Biopsies and drainage of cyst can be performed under CT guidance.

Ultrasound and Magnetic Resonance Imaging found better applications in the evaluation/diagnosis of hepatobiliary and pancreatic diseases. Doppler ultrasound allows the assessment of blood flow, and may be helpful in ischemic bowel. CT and angiogram are better tests for bowel ischemia.

Gastrointestinal endoscopy

Practically, all endoscopy are performed using video endoscopy technique. Endoscopy allows the direct inspection of the lining of the esophagus, stomach and proximal duodenum, ampullar of Vater, terminal ileum and colon/rectum. It is, therefore, possible to view directly and diagnose erosions, ulceration, polyps and tumors. Biopsies can be taken for histological diagnosis and other tests, e.g. culture for bacteria, and rapid urea test (clotest) for *Helicobacter pylori* (gastric biopsies). Therapeutic

endoscopy allows for resection of mucosal lesions, superficial cancers, polypectomy, dilatation and stenting of strictures, and treatment of focal bleeding with injection, banding, hemoclips or even laser.

For details on endoscopy, please refer to Chapter 27 on Endoscopy.

Manometry, Motility and pH Measurement

The GI tract is a hollow tube with two layers of smooth muscle. Contents are moved by a propulsive movement. Disordered movement (motility) gives rise to GI complaints or disorders, e.g. constipation, diarrhea, gastroesophageal reflux. Manometry (measurement of pressure within the gut lumen) and motility studies help to identify patients according to specific abnormalities, e.g. achalasia of the esophagus, gastroesophageal reflux disease, ano-rectum for incontinence or constipation. Information obtained from the studies contribute to assessment and appropriate treatment.

Esophageal pH measurement and manometry are established adjunct in the management of patients with esophageal motility dysfunction; in patients with gastroesophageal reflux with atypical presentation; in those not responding to standard treatment; and in those being considered for surgery.

REFERENCES

1. *Sleisenger & Fordtran's Gastrointestinal and Liver Disease Pathophysiology/Diagnosis/Management*, Feldman M, Friedman LS, Sleisenger MH, Saunders, 2002.
2. Eastwood GL, Avunduk C, *Manual of Gastroenterology*, Little Brown and Company, 1994.
3. Guan R, Kang JY, Ng HS, Merican I (eds.), *Management of Common Gastroenterological Problems, A Malaysia & Singapore Perspective*, Havas MediMedia, 2000.
4. Kang JY, *Gastroenterology in The East and West*, Guan R, Kang JY, Ng HS, Merican I (eds.), in *Management of Common Gastroenterological Problems, A Malaysia & Singapore Perspective*, Havas MediMedia, 2000.
5. Kang JY, Yap ILE, Gwee KA, The pattern of functional and organic disorders in an Asian Gastroenterological clinic, *J Gastroenterol Hepatol* **9**:124–127, 1994.

This page intentionally left blank

23

Functional Gastrointestinal Disorders

Lui Hock Foong

INTRODUCTION

The term functional gastrointestinal disorders refers to a group of conditions manifesting gastrointestinal symptoms arising from disturbance of bowel function for which an underlying structural or organic pathology is absent. In this context, the term functional does not imply a psychiatric or psychological basis for the symptoms; functional bowel disorder should not be regarded as a "condition of the mind" or "supratentorial," as it is often misconstrued. Psychosocial factors, however, may influence the healthcare seeking behavior of an individual experiencing these symptoms.

Functional gastrointestinal disorders can be further subdivided according to the section of the GIT primarily affected (Table 1).[1] They share in common some features of epidemiology, pathophysiology and approach to clinical management.

This chapter will discuss two of the most commonly seen types of functional gastrointestinal disorders, non-ulcer dyspepsia and irritable bowel syndrome.

Table 1 Rome Classification of Functional Gastrointestinal Disorders

Disorder
Functional bowel disorders
Irritable bowel syndrome
Functional abdominal bloating
Functional constipation
Functional diarrhea
Functional gastroduodenal disorders
Functional (non-ulcer) dyspepsia
Aerophagia
Functional esophageal disorders
Non-cardiac chest pain
Rumination syndrome
Globus
Functional abdominal pain
Functional biliary pain (biliary dyskinesia)
Functional anorectal disorders
Functional incontinence
Functional anorectal pain
Levator syndrome
Proctalgia fugax
Pelvic floor dyssynergia

Epidemiology

Functional gastrointestinal disorder is the most common gastrointestinal condition seen, both at the primary care level and also by gastroenterologists. It makes up to 20% of a typical general practitioner's workload and over 40% of a gastroenterologist's cases. It is seen across all age groups, ranging from school-going children to the elderly. Most patients present between the ages of 20 and 40 years and this condition becomes less common above the age of 50.[2]

Gender differences are present. In the West, women sufferers of functional bowel outnumber men in the order of 4:1. In some other cultures, the gender differences are reversed; the ratio of women to men presenting with functional gastrointestinal disorder in the Middle Eastern countries and South Asia stands around 1:4. The gender ratio of women to men in Singapore with this condition is 1.2:1.

Pathophysiology

The cause of functional gastrointestinal disorder is not clearly understood. However, many studies into functional gastrointestinal disorders have produced several consistent observations. These observations point to dysfunction of the "brain-gut axis." The brain-gut axis refers to the complex interplay between the central nervous system and the digestive tract; it involves components of higher centers (the brainstem reticular formation, limbic system, cerebral cortex), the autonomic nervous system, afferent and efferent neurons of the GIT, mechanoreceptors and mediating neurotransmitters. All forms of functional gastrointestinal disorders have in common several features involving the brain-gut axis.

1) **Abnormal motor activity**. Patients with functional gastrointestinal disorders have been shown to have disorders of motor function. The affected section of the bowel may have abnormal baseline motility or abnormal motility only when exposed to various stimuli such as meals, drugs, GI hormones and stress.
2) **Heightened visceral sensitivity**. Studies have consistently showed that patients with functional gastrointestinal disorder, compared with healthy counterparts, felt discomfort or pain at lower thresholds of experimental visceral stimulation, e.g. balloon distension in the rectum, stomach, small intestine and bladder, compared with normal individuals. The reason for this is unclear. Both inflammation and central mechanisms, such as emotional stress, have been hypothesized as potential causes of this increased sensitivity.
3) **Psychosocial factors**. Some patients with functional gastrointestinal disorders display abnormal personality profiles, health beliefs, illness behaviors and illness coping mechanisms. While these features are thought not to be primarily responsible for functional gastrointestinal symptoms, they are thought to be important determinants in their healthcare seeking behavior. Individuals in the community who have functional gastrointestinal symptoms but have not sought medical attention do not have a higher prevalence of abnormal psychological traits. Psychological abnormalities, however, can be found in up to 50% of those who see their GPs and in up to 90% who go to the specialists. Psychosocial factors may also act through higher centers in the brain-gut axis to modulate sensory thresholds, causing these individuals to have heightened visceral sensation.

Clinical approach

The clinical approach to a patient with functional gastrointestinal disorder takes into consideration the following premises:

1) Functional gastrointestinal disorders and organic pathology of the GIT share common symptoms; it is often difficult to separate the two based purely on the characteristics of the GI symptoms. The diagnosis of functional gastrointestinal disorders is based on symptoms and exclusion of organic pathology. There are no specific tests for functional gastrointestinal disorders.
2) A significant proportion of patients who seek medical attention have psychosocial problems, fear of organic diseases, especially cancers.
3) A good patient-doctor relationship is especially important in ensuring favorable outcomes — reassured and satisfied patient, avoiding excessive and unnecessary investigations, minimizing "doctor-hopping."

The following components are thought to be important in the successful management of functional gastrointestinal disorders.

1) Non-judgemental, careful and complete history taking and physical examination.
2) Adequate emphasis on exploring psychological factors and reasons for consultation.
3) Judicious use of investigations. These have value in providing patients with reassurance but should not be excessive to avoid the impression of diagnostic uncertainty and consequent erosion of patient's confidence.
4) Convey a positive diagnosis to the patient, i.e. that the condition is real and "not in the mind"; reassure and explain the non-threatening natural history of the condition; educate patient on pathophysiology of symptoms.
5) Manage patient's expectations from early on; set realistic treatment targets.
6) Select drug treatment appropriately. Many patients prefer to be without medication after reassurance. Consider all available types of treatment options, including psychotherapy, etc.
7) Ensure follow-up to monitor symptom response and revisit the diagnosis if necessary, especially if alarm features appear.

NON-ULCER (FUNCTIONAL) DYSPEPSIA (NUD)

Introduction

Dyspepsia generally refers to a discomfort centered around the upper abdomen. The term "dyspepsia" encompasses bloating, ache, early satiety, postprandial fullness, nausea, anorexia, heartburn, regurgitation, and burping or belching and generally implies a problem with the upper GI tract. Many conditions involving the organs in the vicinity of the upper abdomen can give rise to dyspepsia, ranging from gastric dysmotility, peptic ulcer disease and gastric malignancy to biliary pathology and coronary heart disease.

Definition/Diagnosis

Non-ulcer dyspepsia refers to dyspepsia for which investigations have yielded no abnormalities to explain the symptom and is thought to arise from a disturbed function of the stomach. The diagnosis of NUD has been refined over the years by panels of experts, with the latest major revision called the Rome II criteria (Table 1).[1]

Table 2 Rome II Classification of Dyspepsia

I)	**Functional Dyspepsia** At least 12 weeks, which need not be consecutive, in the preceding 12 months of: 1) Persistent or recurrent symptoms (pain or discomfort centered in the upper abdomen); 2) No evidence of organic disease (including at upper endoscopy) that is likely to explain the symptoms; *and* 3) No evidence that dyspepsia is exclusively relieved by defecation or associated with the onset of a change in stool frequency or stool form (i.e., not irritable bowel).
IIa)	**Ulcer-like Dyspepsia** Pain centered in the upper abdomen is the predominant (most bothersome) symptom.
IIb)	**Dysmotility-like Dyspepsia** An unpleasant or troublesome nonpainful sensation (discomfort) centered in the upper abdomen is the predominant symptom; this sensation may be characterized by or associated with upper abdominal fullness, early satiety, bloating, or nausea.
IIc)	**Unspecified (Nonspecific) Dyspepsia** Symptomatic patients whose symptoms do not fulfill the criteria for ulcer-like or dysmotility-like dyspepsia.

Epidemiology

NUD is common. It can be found in up to 25% of the general population. Together with irritable bowel syndrome, they form the bulk of functional gastrointestinal disorders seen in the primary care setting (making up 15% of cases seen by GPs) and the specialist setting (making up 40–70% of cases seen by gastroenterologists). The gender and age characteristics follow the pattern seen in functional gastrointestinal disease (see above).

Pathophysiology

Many patients with NUD have abnormal gastric motility — delayed emptying, impaired gastric accommodation, myoelectrical abnormalities. Visceral hypersensitivity is present. The role of *Helicobacter pylori* in NUD has been extensively studied. *H. pylori* is not more common in NUD patients than a matched population. Results to date have not implicated *H. pylori* in the pathogenesis of NUD. Meta-analysis of large trials of *H. pylori* treatment in NUD has also not shown the benefit of *H. pylori* eradication on NUD.

Studies from Western countries have shown that patients with NUD who seek medical help have more problems related to personality, illness coping mechanisms and increased anxiety and neuroticism. These factors are thought to be important in influencing their reaction to the symptoms experienced and are not thought to be primarily responsible for the symptoms *per se*.

Clinical approach

The diagnosis of NUD is usually arrived at after evaluation of the patient presenting with the symptom of dyspepsia. Figure 1 provides a suggested approach to evaluating this symptom.[3] In the initial evaluation, differential diagnoses of peptic ulcer disease, gastric malignancies, biliary disorders, cardiac disease, medical conditions with gastroparesis and NSAIDs-induced dyspepsia have to be considered. The next consideration is whether to subject the patient to an upper GI endoscopy, principally to exclude malignancy. This decision hinges on the patient's age and the presence of "alarm" symptoms. The Asia-Pacific consensus suggested a cut-off age of 35 years, above which endoscopy is recommended. Alarm symptoms refer to constitutional symptoms (weight loss, anorexia, fever),

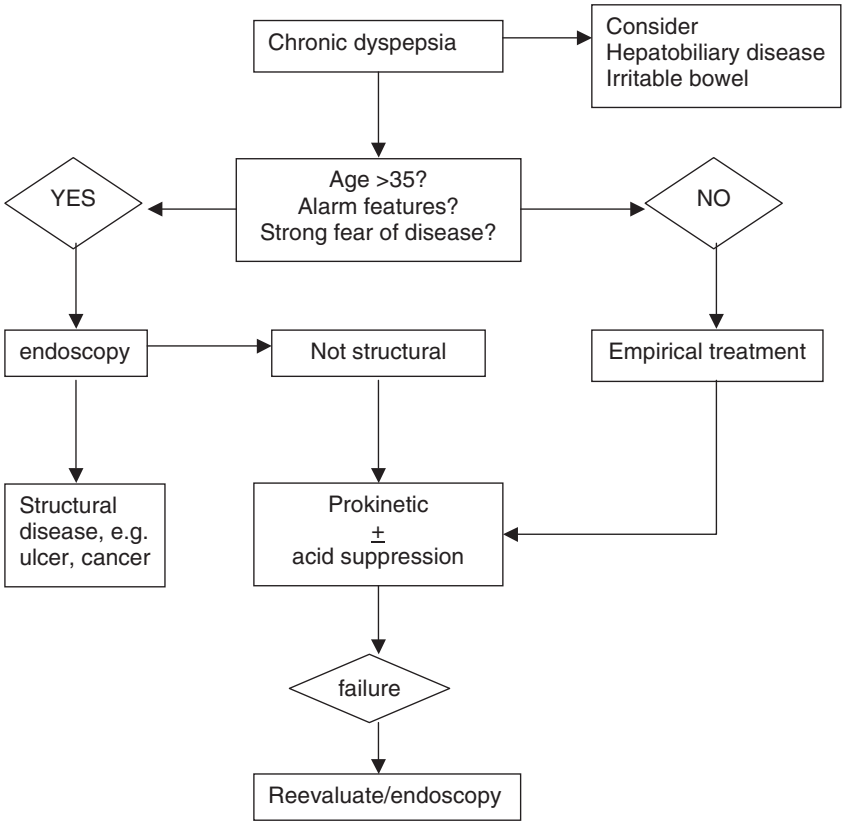


Fig. 1 Approach to uninvestigated dyspepsia.

dysphagia, severe pain, protracted vomiting, and investigation indicating positive stool occult blood, or anemia.

Management

The management of NUD follows the general principles described above for functional gastrointestinal disorders. The establishment of a good patient-doctor relationship is crucial to the success of managing NUD. This forms the foundation upon which patient trust and confidence can be built and also serves as a good platform for educating and reassuring the patient. The patient has to be given a positive diagnosis and not told "nothing is wrong with you." In addition, the patient should be reassured

of the benign natural history of NUD. The fact that cancer and ulcers are not consequences should be emphasized. In the case of NUD, dietary advice is important. Avoidance of food which may delay or alter gastric motility (coffee, oily foods, caffeine, alcohol) should be given a trial. Often with these, the patient is satisfied and there is no need to progress to pharmacotherapy, which should be reserved for those who do not improve after reassurance and lifestyle changes.

The mainstays of pharmacotherapy in NUD are acid-suppression therapy and prokinetics. Overall, the benefit of acid suppression using wither histamine-2 receptor antagonists (H2-blockers) or proton pump inhibitors (PPI) have only shown marginal benefit. The subgroup with significant pain and reflux fares better. Prokinetic agents (cisapride, domperidone and metoclopramide) promote gastric emptying and improve symptoms, especially in patients with dysmotility symptoms, when compared with placebo. Cisapride was shown to produce the most impressive results but its use came to a premature halt on account of the adverse effects of QT prolongation and arrhythmias. Metoclopramide carries significant CNS and extrapyramidal side-effects in long-term use and should be avoided in NUD.

One of the management issues of dyspepsia is whether to perform early endoscopy in all patients with dyspepsia without alarm symptoms or whether empiric testing for *H. pylori* and treatment (acid-suppression therapy + *H. Pylori* eradication) should be first employed, with endoscopy reserved for those who do not improve or experience relapse. The rationale for empiric therapy is to achieve cost-savings by avoiding endoscopy. Early endoscopy, however, may result in better patient satisfaction and reassurance. The psychological patient profile and natural history of NUD of chronic symptoms are such that empirical treatment may only be delaying the inevitable endoscopy, whilst prolonging the anxiety of the patient. Current data neither support the hypothesis that empiric therapy is more cost-effective nor show that early endoscopy results in improved outcome. The choice of approach is perhaps that which both the patient and physician are most comfortable with.

IRRITABLE BOWEL SYNDROME

Irritable bowel syndrome (IBS) is one part of the continuum of functional gastrointestinal disorder, which is characterized by the presence of

chronic, or recurrent abdominal pain occurring in relation to altered bowel habit in the absence of any organic or structural cause.

Definition/diagnosis

The diagnosis of IBS is clinical; there is no specific diagnostic test. An expert panel of clinical investigators have agreed upon a set of symptoms upon which the diagnosis is made. This has undergone revisions, the latest being the Rome II consensus criteria (Table 3).[1]

Epidemiology

The epidemiology of IBS follows that of functional gastrointestinal disorders. It presents across all age groups but is less common above the age of 50 years. It is estimated that only one-third of individuals with IBS seek medical attention; still, IBS makes up 12% of primary care visits. The gender and cultural differences follow that described for functional gastrointestinal disorders (see above).

Pathophysiology

IBS, like other forms of functional gastrointestinal disorders, have several pathophysiologic features in common. Patients with IBS do not have different basal motility of the colon compared with non-sufferers. However,

Table 3 Diagnostic Criteria for Irritable Bowel Syndrome

At least 12 weeks or more, which need not be consecutive, in the preceding 12 months of abdominal discomfort or pain that has 2 of the following 3 features:
1) Relief by defecation
2) Onset associated with change in frequency of stool
3) Onset associated with change in form (appearance) of stool
Symptoms that cumulatively support the diagnosis of IBS
1) Abnormal stool frequency (>3 bowel movements per day or <3 per week)
2) Abnormal stool form (lumpy/hard or loose/watery)
3) Abnormal stool passage (straining, urgency, or feeling of incomplete evacuation)
4) Passage of mucus
5) Bloating or feeling of abdominal distension

they have different and exaggerated colonic motility response when subjected to various forms of stimuli such as food, hormones, stress and drugs. In addition, the gastrocolic reflex (motility of distal colon after meals) is amplified in IBS.

IBS sufferers exhibit greater sensitivity to distension of the rectum and other parts of the gut. They also show more symptoms during the transit of a meal. This hypersensitivity may be responsible for the heightened motility response described above.

IBS sufferers who seek medical attention are more likely to have abnormal personality profiles and higher levels of anxiety about health and disease, and are more likely to seek medical attention for minor complaints such as colds and headaches and display a history of a certain pattern of illness behavior — report of poor health in childhood, greater parental attention to illness, rewards during illness, school absences — which in turn may influence subsequent health seeking patterns.[4]

Clinical approach

IBS should be suspected in any individual whose symptoms fit into the Rome II criteria unless alarm symptoms are present. Further features that support the diagnosis of IBS are the presence of a long history in the presence of good general health, coexistence of depression and anxiety and exacerbations related to difficult life events or situations. The presence of alarm symptoms should alert the clinician to carry out investigations to exclude more sinister pathology, in particular colon cancer, inflammatory bowel disease, diverticular disease and in the female, pelvic pathology. A reasonable set of initial investigations are full blood count, erythrocyte sedimentation rate (ESR) and sigmoidoscopy. Alarm symptoms are: 1) presentation above the age of 50 years; 2) recent onset of symptoms, especially in the older patient; 3) persistent pain that features a nocturnal component; 4) bleeding per rectum and weight loss; and 5) family history of colon cancer.

Management[5]

Once the diagnosis of IBS is established, it is important that the patient be informed, educated and reassured of the natural history and prognosis of this condition. Particular fears or misconceptions the patient may hold

should be identified and addressed. IBS is likely to be a recurring problem — 30% of patients have resolution of symptoms eventually — and the response to medication is unpredictable and often unsatisfactory. As such, it is important to manage the patient's expectations. For all of the above to be successful, a healthy patient-doctor relationship is paramount.

Specific therapy of IBS can be divided into dietary modification, drug therapy and psychotherapy. One study established that Singaporeans consume 13 g of fiber a day, compared with the recommended 30 g. An increase in intake of fiber should be advised. This can be achieved through dietary fiber or bulking agents such as Ispaghula or Psyllium products. Drugs are used to alleviate the main symptoms experienced by the patient (Table 4). In using drugs, the clinician must keep in mind that the placebo response in IBS is 40–70%. A better understanding of the

Table 4 Treatment Options in Irritable Bowel Syndrome

Increase Fiber	Dietary fiber to 30 g/day Bulking agents, e.g. Isphagula husk, methylcellulose
General medication	Librax Trimebutine
Anti-spasmodic (Pain)	Mebeverine hydrochloride 135 mg bd 20 min before meals Alverine citrate 60 mg tds before meals Enteric coated peppermint oil 1–2 capsules before meals
Anti-diarrheals	Loperamide 2 mg om-tds Diphenoxylate (more side-effects) Cholestyramine (if bile acid malabsorption suspected) 1–3 pkts/day 30 min before meals
Newer agents	
$5HT_3$ antagonist	Alosetron — for diarrhea-IBS (caution-cases of ischemic colitis reported)
$5HT_4$ agonist	Prucalopride, Tegaserod Effective for constipation-IBS; relieve pain, bloating
Anti-depressants	Amitriptyline 10–100 mg/day Fluoxetine 20–60 mg/day Paroxetine 20–50 mg/day
Psychotherapy	Hypnotherapy Relaxation/stress management Cognitive behavioral treatment Psychodynamic therapy

physiology of the brain-gut axis, and in particular of the mediating neurotransmitters, has led to the development of newer treatment options. Early data show promise in some of these drugs.

Some patients with IBS have symptoms that are refractory. In this group, the diagnosis has to be reevaluated if alarm symptoms develop over time. Otherwise, exploration of psychosocial factors should be conducted with a view of commencing anti-depressants or psychotherapy. In this respect, alternative and complementary remedies may have value (even if the response is placebo), especially in patients who hold health beliefs favoring "non-Western" medicine.[6]

In summary, the management of IBS requires that the clinician spends adequate time with the IBS sufferer, to establish rapport and trust, identify unvoiced concerns, educate and reassure the patient, and communicate the treatment plan.

REFERENCES

1. Rome II: A multinational consensus document on Functional Gastrointestinal Disorders, *Gut* **45** (Suppl II), 1999.

2. Heading R, Prevalence of upper gastrointestinal symptoms in the general population: A systematic review, *Scand J Gastroenterol* **34**:3, 1999.

3. Talley NJ, Lam SK, Goh KL, Fock KM, Management guidelines for uninvestigated and functional dyspepsia in the Asian Pacific region: First Asian Pacific working party on functional yspepsia, *J Gastroenterol Hepatol* **13**:335–353, 1998.

4. Drossman DA, McKee DC, Sandler RS, *et al.*, Psychosocial factors in the irritable bowel syndrome: A multivariate study of patients and non-patients with irritable bowel syndrome, *Gastroenterology* **95**:701–708, 1998.

5. Jones J, Boorman J, Cann P, *et al.*, British Society of Gastroenterology Guidelines for the Management of Functional Bowel Disorders with Special Reference to the Irritable Bowel Syndrome, *Gut* **47** (suppl II): 1–19, 2000.

6. Read NW, Harnessing the patient's powers of recovery: The role of psychotherapies in the irritable bowel syndrome, *Clin Gastroenterol* **13**:473–487, 1999.

24

Management of Peptic Ulcer Disease

Widjaja Luman

INTRODUCTION

A peptic ulcer is a focal mucosal defect with inflammatory cell infiltration and coagulation necrosis extending through the muscularis mucosae. Endoscopically, an ulcer is a loss of the mucosal surface that has depth and is at least 5 mm in diameter. It should be distinguished from erosion, which is a superficial focal lesion less than 5 mm in diameter. Peptic ulcers tend to occur in parts of the gastrointestinal tract that is in contact with gastric juice containing acid and pepsin. They are not usually observed beyond the ligament of Treitz except in patients suffering from the Zollinger–Ellsion syndrome, or in post-surgical conditions where free acid can reach that section of the intestine. Occasionally, peptic ulcers are found within the acid-secreting mucosa of a Meckel's diverticulum.

PATHOGENESIS

Two etiological factors cause the majority of peptic ulcers seen in clinical practice: *Helicobacter pylori* (*H. pylori*) infection and non-steroidal anti-inflammatory drugs (NSAIDs).

H. pylori

Fifty percent of the world's population has been estimated to be infected with *H. pylori*. In developing countries, much higher prevalence of up to 80% has been reported, with most of the population infected by the age of 10 years.[1] The prevalence is much lower in developed countries, especially in the younger cohorts of the population. It is now believed that infection occurs mostly in childhood and with the gradual improvement in living standards, there is a reduced transmission of the infection, thus accounting for the lower prevalence in developed countries.[2] In Singapore, the sero-prevalence has been estimated to be slightly above 30%, with much higher prevalence among the Chinese than the Malays.[3,4]

H. pylori has been associated with the pathogenesis of chronic active gastritis, peptic ulcer disease[5] and gastric carcinoma.[6] Three lines of evidence support the hypothesis that *H. pylori* infection is associated with peptic ulcer disease. First, with the exception of patients with gastrinoma and those taking NSAIDs, more than 95% of patients with duodenal ulcer and more than 80% of patients with gastric ulcers are infected with *H. pylori*.[7] Second, duodenal ulcers develop far more frequently in persons infected with *H. pylori* than in non-infected persons.[8] Third, the recurrence of peptic ulcers is markedly reduced after successful eradication of the organism.[5,9]

Once the organism colonizes the human gastric mucosa, it induces inflammation and has the ability to persist for the lifetime of the individuals. The effect of the infection on gastric acid secretion is related to the distribution of the gastritis within the stomach, in particular to the extent to which it involves the antrum or the body of the stomach or both of these regions. The antral mucosa produces the hormone gastrin which circulates and stimulates the parietal cells in the body region of the stomach to secrete acid. In antrum-predominant *H. pylori* gastritis, there is increased gastrin release and consequently increased acid secretion. This is the predominant gastritis seen in patients with duodenal ulcer when both basal- and meal-stimulated acid outputs have been reported to be elevated.[10] After eradication, these acid levels fall significantly, probably reflecting a decrease in the trophic drive due to hypergastrinemia as basal gastrin levels are restored, thus restoring a normal gastric mucosal environment. In contrast, patients with gastric ulcers tend to have gastritis towards the corpus or body of the stomachs. These patients tend to get

normal or even lower basal and stimulated acid secretions. They are more at risk of developing atrophic gastritis and gastric ulcers rather than duodenal ulcers.

There is increasing evidence that the clinical outcome of *H. pylori* gastritis is associated with differences in virulence among *H. pylori* strains as only 20% of infected persons will develop peptic ulcers in their lifetime.[8] Strains with *cagA* and *vacA* genes are associated with increased inflammatory cytokines production and higher prevalence of peptic ulcers. Urease released from *H. pylori* can lead to activation of inflammatory cells and release of cytokines such as tumor necrosis alpha (TNF α), and interleukin 1 and 6 (IL-1 and IL-6). The chronic inflammation induced can lead to progressive loss in epithelial barrier function and changes in glycoprotein composition of mucus.

NSAIDs

NSAID-induced gastroduodenal mucosal injury is due to topical and systemic effects. NSAIDs cause topical mucosal damage by diminishing the hydrophobicity of gastric mucus, thereby allowing endogenous gastric acid and pepsin to injure the surface epithelium. Furthermore, NSAIDs stimulate neutrophil adhesion and migration with resultant free radical production leading to gastric mucosal damage. The systemic effect of NSAIDs in the pathogenesis of gastroduodenal disease is largely mediated through their effect on reduction in mucosal prostaglandin. Prostaglandins play a key role in mucosal protection through their influence on mucus synthesis, mucosal blood flow, secretion of bicarbonate, epithelial proliferation and inhibition of leucocyte adherence.[11] Prostaglandins are products of metabolism of arachidonic acid by cyclooxygenase 1 and 2 (COX 1 and 2). COX 1 is expressed constitutionally and is involved in the production of mucosal prostaglandins. In contrast, COX2 is inducible by inflammatory cytokines and produces prostaglandins and leukotrienes which are more inflammatory and are involved in mediation of pain in inflamed tissues. NSAIDs inhibit activity of both cyclooxygenases, thereby delivering its analgesic effects and reducing production of mucosal prostaglandins at the same time. The newer COX2-specific NSAIDs are an effective anti-inflammatory and analgesic agent as they inhibit only the inflammatory pathway of COX2. They do not inhibit the constitutive COX1 and reduce the level of

mucosal prostaglandins, thereby sparing the gastroduodenal damages that are so well associated with the non-specific NSAIDs.

Non-H. pylori, non-NSAID ulcers

Over the past decade, the incidence of non-*H. pylori*, non-NSAID ulcers appears to be increasing. A recent study from Hong Kong reported that 4.1% of patients with bleeding ulcers have non-*H. pylori*, non-NSAID ulcers.[12] Unreported NSAID use must first be considered, especially given the large number of over-the-counter NSAID-containing medications available. Illicit contamination of traditional medicine with NSAID is another possibility. The pathogenesis of ulcers in patients with non-*H. pylori*, non-NSAID ulcers is not known. Smoking and psychological stress have been thought to be possible etiological factors, but the causal links of these factors remain to be defined. Alendronate, a biphosphonate, has recently been reported to be ulcerogenic in the stomach apart from its well documented erosive effects in the esophagus.[13] Other conditions to be considered are Crohn's disease, Zollinger–Allison's syndrome and lymphoma.

HISTORY AND PHYSICAL EXAMINATION

Peptic ulcer disease is associated with epigastric discomfort or pain which can be burning or gnawing in nature, commonly occurring 2 hours or more after a meal. It may wake the patients up at night or appear when they get up in the morning. For patients with duodenal ulcers, the pain is typically relieved by ingestion of food or a dose of antacids. As a result, duodenal ulcer patients tend to snack and gain weight. In contrast, the pain of gastric ulcer tends to occur 5–10 minutes after food. These patients tend to avoid food and lose weight. The other major characteristic of ulcer pain is periodicity, with the symptoms occurring at intervals in weeks or months. However, this symptomatology has low specificity as many patients with these symptoms do not have peptic ulcer disease and suffer from what is commonly referred to as functional or non-ulcer dyspepsia. General examination in uncomplicated peptic ulcer disease is usually unremarkable. A marked increase in the severity and intensity of the pain with spread to the whole abdomen indicate that the ulcer has perforated. This is usually followed by rebound tenderness with guarding and absent bowel sounds.

INVESTIGATIONS

Since the discovery of the role of *H. pylori* in peptic ulcer disease, a number of diagnostic tests have been developed for the detection of the organism. They can be grouped as invasive, i.e. requiring endoscopy, and non-invasive, i.e. not requiring endoscopy.

H. pylori serology

Serological assessment utilizing laboratory-based ELISA assays measuring IgG antibodies are both sensitive and specific for the diagnosis of *H. pylori* infection before as well as after eradication therapy.[14] A 50% decrease in titer at 6 months after therapy has high specificity for successful eradication. However, serological testings using near-patient kits have not been shown to be reliable and are not recommended.[15]

Stool antigen tests for H. pylori

H. pylori stool antigen has recently been introduced for the detection of the organism.[16] Most studies have shown this test to have sensitivity and specificity of close to 90%. It is one of the two tests recommended for primary care physicians by the European Helicobacter pylori study group.[17] The other recommended test is the Urea Breath tests.

Urea breath tests

Urea, the substrate of the *H. pylori* urease enzyme, can be labeled with either ^{14}C or ^{13}C. After hydrolysis of the urea by *H. pylori* urease, the labeled carbon dioxide is exhaled in the breath and an increase in the excretion of labeled CO_2 can be used as an indicator of colonization by *H. pylori*. After ingestion of the radiolabeled urea, breath samples are collected for up to 2 hours or a single sample can be collected at 40 minutes, by exhaling into a CO_2-trapping agent. For ^{14}C, the radioactivity of each sample can be measured by a scintillation counter. For ^{13}C, the CO_2 can be detected by a mass spectrometer as it is non-radioactive. The sensitivity and specificity of the test are in the range of above 95%. The test can be used to screen patients prior to endoscopy and as a follow-up to determine the effect of treatment. It has been reported to be unreliable in

patients taking ranitidine or proton pump inhibitors and both medications should be discontinued for one week prior to testing.

Barium meal

Barium meal is well tolerated and cheaper than gastroscopy. Compared with gastroscopy, however, it suffers from a lack of sensitivity in diagnosing superficial mucosal lesions, lack of histologic information and lack of therapeutic potential in bleeding patients. If a duodenal ulcer or a scarred duodenal bulb is identified on barium meal, therapy for *H. pylori* eradication should be prescribed due to the strong associations with the organism and a low likelihood of malignancies. Further endoscopic investigation is not indicated. In contrast, if the barium meal shows gastric ulcer, the patient should be referred for gastroscopy in order to obtain histologic biopsies for exclusion of malignancies.

Gastroscopy

Barium meal has given way to gastroscopy, which is a sensitive, specific and safe method for the diagnosis of gastroduodenal lesions. Biopsy can be performed from gastric ulcers to exclude malignancy and from antrum to establish the presence of *H. pylori.* This can be determined directly by the detection of bacterial urease by a dye test for urea (e.g. CLO-test), through specific histologic staining of the biopsy material for the presence of *H. pylori* or by culture under special microaerobic conditions. In patients with upper gastrointestinal hemorrhage, gastroscopy is of particular value for diagnosing peptic ulcers as the bleeding source and it offers therapeutic measures to control the bleeding.

COMPLICATIONS

Despite the advances made in the treatment of peptic ulcers, the frequency of complications such as gastric hemorrhage and perforations has not declined. This is largely attributed to increased use of NSAIDs, especially in the elderly population.

Hemorrhage

Hemorrhage is the most common complication and patients present with either hemetemesis or melena. The presence of both these symptoms

suggests that there has been brisk bleeding. The author would like to stress that negative gastric aspirate with nasogastric tubes does not exclude the possibility of upper gastrointestinal bleeding, as blood from bleeding duodenal ulcers may not reflux back into the stomach. Bleeding peptic ulcers in elderly patients may prove to be fatal despite advances made in endoscopic therapy to control the hemorrhage. However, the cause of death is usually from cardiopulmonary events rather than from exsanguination.

Perforation

Perforation occurs much less frequently than bleeding. Patients present with sudden and severe abdominal pain. Abdominal examination reveals tender abdomen with guarding and classically, erect chest radiograph should show free gas under the diaphragm.

Other complications are pyloric stenosis with gastric outlet obstruction as a result of chronic recurrent duodenal ulcers. This will lead to postprandial vomiting and fullness, and may eventually lead to hypokalemia and alkalosis.

THERAPY

Empirical therapy

Dyspepsia is a common complaint in the community and most guidelines recommend initial empirical therapy for young patients.[17] For patients who do not have alarm symptoms, the cut-off age for referral for endoscopy is above 45 years based on guidelines from the European countries. As gastric cancer commonly occurs at a much younger age in the Asia-Pacific region, the cut-off age has been taken at 35 years in Singapore.[15] Alarm symptoms are presence of anemia, severe and persistent pain, odynophagia, dysphagia, anorexia, and weight loss.

Young patients below 35-years-old and who do not have these alarm symptoms can be treated empirically with H_2 blockers or proton pump inhibitors. Patients who do not fulfill these criteria should be referred for further investigations. If the symptom relapses, further investigations are warranted as it is impossible for clinicians to distinguish patients with peptic ulcer disease from the much larger group of patients with non-ulcer or functional dyspepsia. However, the type of investigations recommended for primary care clinicians are rapidly changing and controversial. One approach might be to perform serologic antibodies or Urea Breath tests for

H. pylori for patients who fail to respond to or relapse after empirical therapy. The rationale of this approach is that *H. pylori* negative patients are unlikely to have serious pathology and could safely be treated with empirical therapy. Patients who are *H. pylori* infected should either receive therapy for *H. pylori* eradication or be referred for gastroscopy to exclude carcinomas. The other extreme approach is to refer all dyspeptic patients for gastroscopy, but this is an expensive approach with low yield for revealing serious pathology.

H. pylori eradication

H. pylori is difficult to eradicate and none of the drug regimens currently used achieves 100% eradication rate. Luminal acidity affects the effectiveness of some drugs against *H. pylori*. Raising the gastric pH from 3.5 to 5.5 increases the effectiveness of amoxycillin and erythromycin more than 10-fold. This increased activity at higher pH explains the effectiveness of regimens that combine potent inhibition of gastric acid secretion with proton pump inhibitors (PPI).[18] Furthermore, PPIs have been shown to suppress the growth of the organism.[18] Current guidelines recommend a one-week duration of triple therapy with PPIs in standard dose and two antibiotics. The regime is usually in the form of **(PPI in standard dose BD + clarithromycin 500 mg BD + amoxicillin 1 g BD)** for one week. For individuals who are allergic to penicillin, the amoxicillin can be replaced with metronidazole, though eradication rate with this combination is lower as shown in the MACH 2 trial.[19] In the event of treatment failure with these regimes, second-line quadruple therapy can be attempted with the following combinations **(PPI in standard dose BD + bismuth subcitrate 120 mg QDS + tetracycline 500 mg QDS + metronodazole 400 mg TDS)** for a minimum of seven days, preferably for two weeks.[17] When eradication of *H. pylori* infection has been documented at least three months after treatment, subsequent reinfection is rare.

H. pylori eradication is strongly recommended for patients with peptic ulcer disease, be it complicated or uncomplicated. For uncomplicated duodenal ulcers, one week of triple therapy is sufficient to heal ulcers and bring about symptomatic relief. For gastric ulcers or complicated ulcers, one week of triple therapy should be followed by another 1–2 months of either H_2 blockers or PPIs. For patients with non-ulcer dyspepsia, the chances of symptomatic improvement are about 25% as shown by a

recent meta-analysis.[17] Therapy is justifiable for patients from areas with low rate of reinfection, as successful eradication can improve symptoms and reduce ulcer occurrence.

Approach to healing NSAID-induced ulcers

PPIs have been shown to be more effective than misoprostol or ranitidine for healing of NSAID ulcers if NSAID is required to be continued.[20,21] In general, it is safer to stop the NSAIDs especially if there has been presentation with ulcer complications, such as bleeding or perforation, and look for an alternative analgesic agent.

Approach to preventing NSAID-induced ulcers

Endoscopic ulcers develop in up to 45% of patients taking NSAIDs for a duration of six months. The strategy is to choose NSAIDs from the less damaging end of the spectrum such as ibuprufen or naproxen and use the lowest effective dose. Co-prescription with antacids may reduce NSAID-induced dyspeptic symptoms but not ulcerations and their complications. Double-blind controlled trials show that the rate of ulcer formation can be significantly decreased with misoprostol,[20] proton pump inhibitors[22] and COX2-specific inhibitors (coxibs).[23] It is very common for H_2 blockers to be co-prescribed with NSAIDs, but it has so far been shown to be protective against duodenal but not gastric ulcers when used at the standard dose. Misoprostol is effective against NSAIDs ulcers but it is not commonly used due to its side-effects, most commonly abdominal cramps and diarrhea. The results from maintenance therapy with omeprazole is superior to either misoprostol or ranitidine.[20,22] Coxibs have been shown to decrease ulcer complications in double-blind trials, with relative risk reductions of 50–60%, but this risk reduction is lost if patients were on low dose aspirin for cardioprotection.[23]

Use of coxibs or maintenance therapy with omeprazole is clearly indicated for high-risk patients. These include elderly patients above 65 years of age; patients with prior dyspeptic symptoms or a history of peptic ulcer disease; patients on aspirin/steroids/anti-coagulants; or patients who need high-dose or multiple NSAIDs. If the patients are already on low-dose aspirin, it may be more sensible to prescribe co-therapy with PPI rather than using coxib.

Role of H. pylori eradication in long-term NSAID users

The relationship between *H. pylori* and NSAIDs is complex. Each agent causes peptic ulcers by distinct pathophysiological mechanisms. It is not known whether they act independently or synergistically in the development of peptic ulcers. *H. pylori* infection induces gastric mucosal prostaglandin production through increased COX2 expression. Theoretically, *H. pylori* infection can ameliorate NSAID-induced reduction in mucosal prostaglandins and therefore is protective against NSAID-induced gastroduodenal damages. In both the Astronaut[22] and OMNIUM[20] trials, the presence of *H. pylori* infection was associated with greater likelihood of ulcer healing and maintenance of remission. In comparison to patients who were maintained on long-term PPIs, eradication of *H. pylori* has not been shown to reduce incidence of ulcer complications in patients who were kept on either aspirin[24] or NSAIDs[25] without co-therapy with PPIs. It is possible that the *H. pylori* might, in some way, protect against the deleterious effect of NSAID therapy.

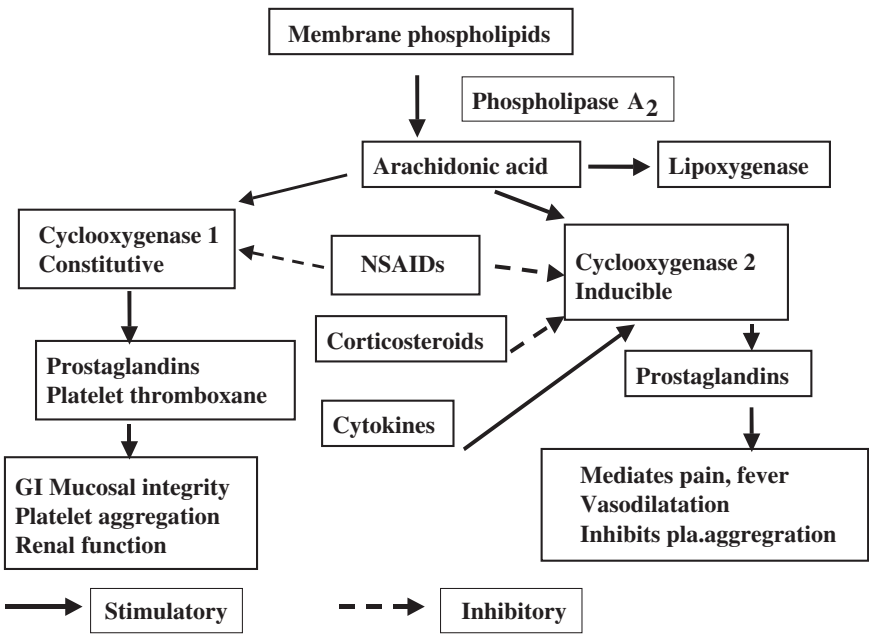


Fig. 1 Pathways of cyclooxygenases.

REFERENCES

1. Perez Perez G, Taylor D, Bodhidatta L, Wongsrichanalai J, Baze W, Dunn B *et al.*, Seroprevalence of *Helicobacer pylori* infectionsin Thailand, *J Infect Dis* **161**:1237–1241, 1990.

2. Cullen D, Collins B, Christiansen K, Epis J, Warren J, Surveyor I *et al.*, When is *Helicobacter pylori* infection acquired? *Gut* **34**: 1681–1682, 1993.

3. Kang J, Yeoh K, Ho Kea, Racial differences in Helicobacter pylori seroprevalence in Singapore: Correlation with differences in peptic ulcer frequency, *J Gastroenterol Hepatol* **12**:655–659, 1997.

4. Fock KM, Helicobacter pylori infection — current status in Singapore, *Ann Acad Med Singapore* **26**:637–641, 1997.

5. Graham DY, Lew GM, Klein PD, Dolores GE, Evans DJ, Saeed ZA *et al.*, Effect of treatment of *Helicobacter pylori* infection on the long-term recurrence of gastric or duodenal ulcer, *Ann Intern Med* **116**:705–708, 1992.

6. Logan RPH, *Helicobacter pylori* and gastric cancer, *Lancet* **344**:1078–1079, 1994.

7. Peterson WL, *Helicobacter pylori* and peptic ulcer disease, *N Engl J Med* **324**:1043–1048, 1991.

8. Sipponen P, Seppala K, Aaryynen M, Helske T, Kettunen P, Long term risk of peptic ulcer disease in people with Helicobacter pylori infection — community based study, *Gut* **30**:922–929, 1989.

9. Rauws EAJ, Tytgat GNJ, Cure of duodenal ulcer associated with eradication of *Helicobacter pylori*, *Lancet* **335**:1233–1235, 1990.

10. El-Omar E, Penman I, Dorrian Cea, Eradicating *Helicobacter pylori* infection lowers agstrin mediated acid secretion by two thirds in patients with duodenal ulcer, *Gut* **34**:1060–1065, 1993.

11. Cryer B, Mucosal defense and repair: role of protaglandins in the stomach and duodenum, *Gastroenterol Clin North America* **30**(4): 877–894, 2001.

12. Chan H, Wu J, Chan F, Choi C, Ching J, Lee Y *et al.*, Is non-*Helicobacter pylori*, non-NSAID peptic ulcer a common cause of upper GI bleeding? A prospective study of 977 patients, *Gastrointest Endosc* **53**:438–442, 2001.

13. Marshall J, Rainsford K, James C, Hunt R, A randomised controlled trial to assess alendronate-associated injury of the upper gastrointestinal tract, *Aliment Pharmacol Ther* **14**:1451–1457, 2000.

14. Kosunen T, Seppala K, Sarna Sea, Diagnostic value of decreasing IgG, IgA and IgM antibody titres after eradication of *Helicobacter pylori*, *Lancet* **339**:893–895, 1992.

15. MOH Clinical Practice Guidelines, *Management of Helicobacter pylori infection*, 1998.

16. Gisbert JP, Pajares JM, Diagnosis of Helicobacter pylori infection by stool antigen determination: A systematic review, *Am J Gastroenterol* **96**:2829–2838, 2001.

17. Malfertheiner P, Megraud F, O'Morain C, Hungin APS, Axon ATR, Graham DY *et al*., Current concepts in the management of *Helicobacter pylori* infection — The Maastricht 2-2000 Consensus Report, *Aliment Pharmacol Ther* **16**:167–180, 2002.

18. Walsh J, Peterson W, The treatment of *Helicobacter pylori* infection in the management of of peptic ulcer disease, *N Engl J Med* **333**(15): 984–991, 1995.

19. Lind T, Megraud F, Unge P, Bayerdorffer E, O'Morain C, Spiller RC *et al*., The MACH2 study: Role of omeprazole in eradication of Helicobacter pylori with 1-week tripe therapies, *Gastroenterology* **116**: 248–253, 1999.

20. Hawkey CJ, Kurrasch J, Szczepanski Lea, Omeprazole compared with misoprostol for ulcers associated with non-steroidal anti-inflammatory drugs. Omeprazole versus misoprostol for NSAID induced ulcer management study group, *N Engl J Med* **338**:727–734, 1998.

21. Brignoli R, Watkins P, Halter P, The Omega Project: a comparison of two diagnostic strategies for risk- and cost-oriented management of dyspepsia, *Eu J Gastroenterol Hepatol* **9**:337–343, 1997.

22. Yeomans N, Tulassay T, Jujasz Lea, A comparison of omeprazole with ranitidine for ulcers associated with non-steroidal anti-inflammatory drugs, *N Engl J Med* **338**:719–726, 1998.

23. Silverstein F, Faich G, Goldstein J, Simon L, Pincus T, Whelton A *et al*., Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis, *JAMA* **284**:1247–1255, 2000.

24. Lai K, Lam S, Chu K, Wong B, Hui W, Hu W *et al*., Lansoprazole for the prevention of recurrence of ulcer complications from long-term low-dose aspirin use, *N Engl J Med* **346**:2033–2038, 2002.

25. Chan FKL, Chung S, Suen B, Lee Y, Leung W, Leung VKS *et al*., Preventing recurrent upper gastrointestinal bleeding in patients with *Helicobacter pylori* infections who are taking low dose aspirin or naproxen, *N Engl J Med* **344**:967–973, 2001.

25

Gastrointestinal and Liver Cancer

Tan Chee Kiat

Gastrointestinal (GI) cancer is one of the most common cancers in the Asia-Pacific region. In Singapore, three of the top five cancers are from the GI system (colorectal, liver and stomach).[1] In most of the rest of Asia, liver cancer ranks as the top cancer in terms of incidence and mortality.[2]

GI CANCERS

Predisposing factors (Table 1)

Certain cancers of the GI tract are clearly associated with environmental risk factors. Cancer of the esophagus is associated with smoking. This is not surprising as, just like lung cancer, esophagus cancer is typically squamous cell in origin and both the lungs and upper GI tract are exposed to carcinogens in inhaled cigarette smoke. *Helicobacter pylori* infection has been established as a class I carcinogen for stomach cancer.

Besides environmental factors, other factors associated with GI cancers are chronic mucosal inflammation and genetic predisposition.

Table 1 Local Epidemiology and Summary of Treatment Modalities

GI Organ	Rank	Sex Dominance	Predisposing Conditions	Curative Therapy			
				Surgery	Chemo	Radiation	Others
Esophagus	3	Male (3X)	tobacco alcohol long-stading achalasia tylosis	yes*	yes (combination with radiation)	yes (combination with chemo)	
Stomach	2	Male (2X)	tobacco *Helicobacter pylori* infection diet high in salted/preserved foods	yes*	no	no	EMR PDT
Small bowel	5	None	diet high in smoked/cured foods FAP celiac disease Crohn's disease	yes	no	no	
Large bowel	1	None	genetic (FAP) high intake of fat long-standing ulcerative colitis	yes	no	no	
Pancreas	4	None	tobacco high intake of meat/fat herditary pancreatitis	yes	no	no	

*treatment of choice.
EMR = endoscopic mucosal resection; PDT = photodynamic therapy.

Patients with hereditary pancreatitis have a more than 50-fold increased risk of pancreatic cancer.[3] Chronic inflammatory bowel disease such as Crohn's disease predisposes to subsequent development of lymphoma, while ulcerative colitis predisposes to colonic cancer. Colorectal cancer is a classic example of the role of genetic predisposition in cancer. Familial adenomatous polyposis (FAP) and the Lynch syndrome (also known as hereditary non-polyposis colorectal cancer or HNPCC) both have strong genetic bases.

Epidemiology and treatment (Table 1)

Our local epidemiological features of the common GI cancers are summarized in Table 1. Established predisposing features are also indicated.

As evident from the table, the only curative therapy for all GI cancers is surgery. The only controversy is GI lymphoma for which some believe that chemotherapy should be the treatment of choice.

Diagnosis

Although barium studies may suggest a diagnosis of luminal cancer, it may also be mistaken as a benign condition. Hence, it is best to obtain histological confirmation. Pancreatic cancers can often be confidently diagnosed on MRI, but even in these cases, a histological assessment is important in the event that it is a pancreatic lymphoma rather than the more common adenocarcinoma which has better treatment prospects and hence a better prognosis.

Screening

Screening of the general population has been unequivocally shown to be useful only for colorectal cancer.[4] As the disease pattern in Singapore is similar to that in the West, these screening guidelines are likely to be applicable to us even though they were established based on Western populations. In individuals with Barrett's esophagitis, there is much data to support screening for development of esophageal adenocarcinoma, but the optimal frequency of screening is still unknown.[5] Although the Japanese are ardent proponents of screening for gastric cancer, there is little acceptance elsewhere.

Specific cancers

Esophagus

Most patients present with dysphagia with anorexia and loss of weight. The classical history is that of progressive dysphagia from solids to semi-solids to liquid. There may be odynophagia if the cancer ulcerates. Chest pain is a late symptom and may denote local invasion. Rarely, a patient may present catastrophically with massive hematemesis from invasion of the cancer into the adjacent aorta.

Diagnosis can be rapidly and easily made from esophagoscopy or barium swallow. CT scan of the thorax is mandatory for staging, although endoscopic ultrasonography is also useful.

Most esophageal cancers are of squamous cell origin. However, adenocarcinoma is seen in those that develop from pre-existing Barrett's esophagus.

The treatment of choice is surgery. Chemotherapy and radiotherapy are useful for inoperable cases of squamous cell carcinoma.

Stomach

Early gastric cancer is asymptomatic. Symptomatic presentation is usually late. Symptoms then include pain, loss of appetite and weight. There may be delayed vomiting after a meal if there is malignant obstruction of the gastric outlet. Characteristic metastases may occur to the ovaries causing massive ascites (also known as Krukenberg's tumor).

The two main histological types of gastric cancer are adenocarcinoma and lymphoma. Mucosa-associated lymphoid tissue (MALT) lymphoma is distinctive in that it is associated with *Helicobacter pylori* infection and is potentially curable by eradication of the *Helicobacter pylori* with a simple triple therapy regimen.[6] The macroscopoic appearance on gastroscopy is nondescript and hence it is important to do biopsy of the areas of inflammation in the stomach.

The treatment of choice for adenocarcinoma is surgery. As for lymphoma of the stomach, the trend is towards chemotherapy with or without radiotherapy as the first line as results have been as good as resection.[7]

Pancreas

Most patients with pancreatic cancers present with pain and loss of weight and appetite. Unresectable pancreatic adenocarcinomas have a very poor prognosis because of the lack of efficacious chemotherapeutic agents. On the other hand, pancreatic lymphomas, albeit rare, are responsive to the usual chemotherapy for lymphomas. Hence the importance of tissue diagnosis in cases of pancreatic cancers.

Small intestine

Cancers arising from the small intestine most often present with abdominal pain or occult bleeding. Small bowel obstruction is rare because unlike benign lesions, malignant lesions are not likely to remain clinically silent until they reach a large enough size to obstruct.

Other than the usual adenocarcinoma, carcinoid is the other main malignancy occurring in the small intestine. It may present cryptically with flushing and diarrhea (carcinoid syndrome).

Surgical resection is the only treatment for adenocarcinoma of the small intestine. Treatment outlook for carcinoids is better than adenocarcinoma because of the possibility of using somatostatin analogs. Small intestinal lymphomas are treated as an extranodal lymphoma with standard chemotherapy.

Large intestine

Large intestinal cancer is rapidly becoming the most common cancer in Singapore. Fortunately, there are proven protocols for screening individuals for large intestinal cancer.[4]

There are two distinct familial risks for large intestinal cancer — familial adenomatous polyps (FAP) and the Lynch syndrome. Polyps in individuals with FAP may turn malignant not only in the colon, but also in the stomach and periampullary region of the duodenum. Thus, screening for cancers in individuals with FAP involves panendoscopy. The Lynch syndrome is actually a constellation of cancers and large intestinal cancer is just one of the associated sites.

Large intestinal cancers commonly present with occult anemia (right-sided lesions), intestinal obstruction or bleeding per rectum (left-sided

lesions). Colonoscopy for histological confirmation of malignancy should be mandatory, for even typical applecore lesions seen on barium enema may turn out to be artefactual.

Surgical resection is necessary if there is impending obstruction, even though there may be extensive metastases. Curative resection is to be followed by long-term surveillance for recurrence of local disease as well as for development of hepatic metastases, as these conditions are amenable to surgical treatment.

HEPATOBILIARY CANCERS

Hepatocellular carcinoma (HCC) (Fig. 1) and cholangiocarcinoma (CCA) (Fig. 2) are by far the largest groups of hepatobiliary cancer locally, constituting 83.6% and 4.4% during the period 1993–1997.[8] It is often difficult to differentiate them clinically, although HCC is more likely if the liver is cirrhotic and the converse for CCA. Tumor markers (AFP for HCC versus CEA and CA 19-9 for CCA) may also be useful in some cases in differentiating between the two.

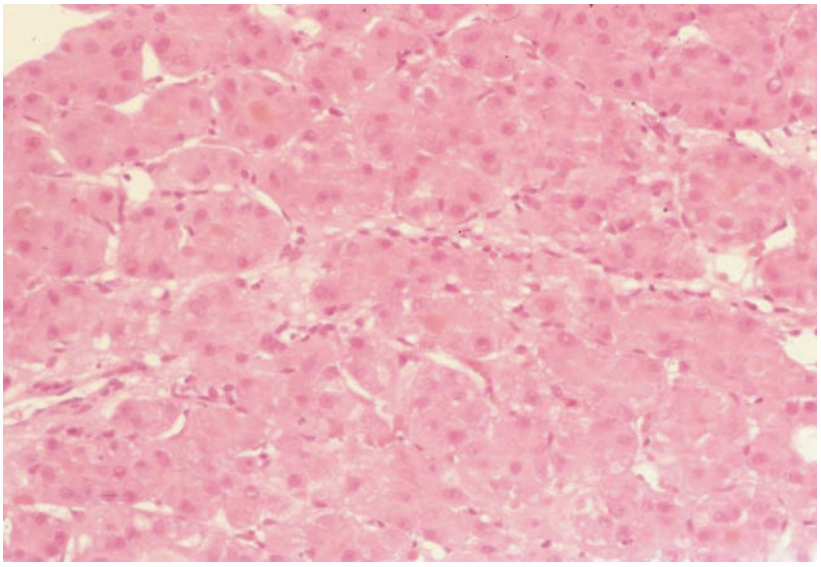


Fig. 1 Hepatocellular carcinoma. Large malignant hepatocytes with enlarged hyperchromatic nuclei and multiple layers of cords instead of the normal single layer.

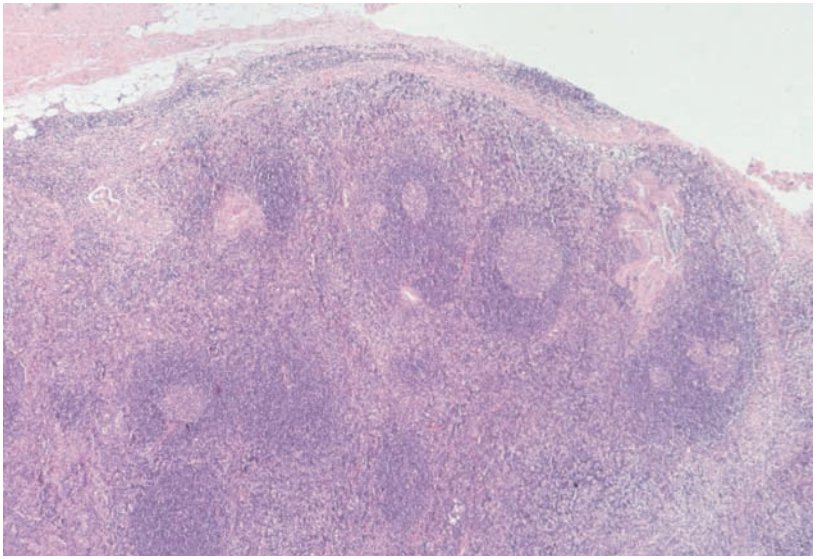


Fig. 2 Cholangiocarcinoma. Typical features of adenocarcinoma.

Table 2 Predisposing Factors for Hepatocellular Carcinoma and Cholangiocarcinoma

Hepatocellular Carcinoma	Cholangiocarcinoma
Chronic hepatitis B	Primary sclerosing cholangitis
Chronic hepatitis C	Liver fluke infestation
Excessive alcohol ingestion	Choledochol cysts
Other causes of cirrhosis	
Anabolic steroids	

Predisposing factors (Table 2)

The predisposing factors for the two cancers are very different. For HCC, the risk factors are those that lead to cirrhosis (chronic liver parenchyma damage and regeneration) whereas for CCA, conditions that cause chronic damage to the biliary tree are important. A special predisposition to CCA is found in congenital choledochol cysts, where the malignant transformation rate has been reported to be as high as 7%.[9]

Epidemiology

HCC is more common in males, whereas the incidence of CCA is almost similar in males and females. Data from our local cancer registry shows a male preponderance of 79.5% in HCC and only 53.5% in CCA. The incidence of HCC is slowly declining in Singapore for reasons that are not apparent. The decline cannot be attributed to our national hepatitis B vaccination program yet, as the program was started only in 1987 and hence the first cohort of vaccinees attains only 17 years of age this year, way before the peak age incidence of HCC.

Diagnosis

Unlike the luminal cancers, diagnosis of hepatobiliary cancers does not usually require histological confirmation. HCC can be confidently diagnosed by a typical triphasic CT appearance in a cirrhotic liver with an elevated serum alphafetoprotein level. CCA also has a typical appearance on endoscopic retrograde cholangiopancreatogram (ERCP) and magentic resonance cholangiopancreatography (MRCP). The latter is often the diagnostic modality of choice at present as it is non-invasive compared with ERCP.

Screening

The screening of HCC has always been debatable. Whilst screening of population-at-risk, namely individuals with chronic hepatitis B infection, has been conclusively shown to discover HCCs at an earlier stage, the survival of screened patients has not been shown to be significantly altered, a key criterion to justify the screening of a disease. Our institution's experience shows that patients who undergo screening has a significantly better curative resection rate and hence better long-term outlook.[10]

Treatment (Table 3)

Other than surgical therapy (resection or liver transplantation), there is no proven curative treatment for HCC. However, resection is besieged by the problem of frequent postoperative recurrence. A recent report of adjuvant treatment with ^{131}I delivered transarterially after a curative

Table 3 Treatment of Hepatocellular Carcinoma

		Disease Extent		
		Curative	**Local**	**Extrahepatic**
Liver Reserves	Good	Resection	(as below if not resectable)	Systemic chemotherapy
	Poor	Transplantation	• Percutaneous ethanol injection • Radiofrequency ablation • Transhepatic arterial chemoembolization	Supportive

resection has been shown to reduce the rate of recurrence.[11] All other treatments for HCC are considered palliative, of which the most promising are percutaneous ethanol injection therapy and radiofrequency ablation for local disease. Treatment for systemic disease is still protocol-based, with no outstanding results except for combination therapy of interferon, doxorubicin, 5-fluorouracil and cisplatin.[12]

TUMOR MARKERS IN GI AND HEPATOBILIARY CANCERS

There are 3 tumor markers pertinent to the GI tract, namely, alphafetoprotein (AFP), carcinoembryonic antigen (CEA) and carbohydrate antigen 19–9 (CA 19–9). These tumor markers should be employed in the proper clinical context rather than being used indiscriminately on every patient.

AFP

AFP is normally produced by a developing fetus and is usually not elevated in normal adults except in pregnant women. An elevated AFP strongly suggests either HCC or germ cell cancer. However, an elevated AFP must be assessed in conjunction with the degree of necroinflammation reflected by the serum alanine transaminase (ALT) level, as the AFP level can also be transiently elevated in the face of hepatitis B flare. Most ominous is a rising AFP with a normal serum ALT level. AFP can also be

subfractionated into a *Lens culinaris* agglutinin-reactive fraction (AFP-L3) that is more specific for HCC.[13]

CEA

The main use of CEA is in the follow-up of colorectal cancer. It is not to be used for screening because of its very low specificity for malignancy. Simple factors like age, and lifestyle habits like smoking, can elevate CEA.

CA 19–9

CA 19–9 was originally found in colorectal cancer. Its main utility at present is for the diagnosis of pancreatic cancer and CCA. However, one must be aware that CA 19–9 is secreted only by individuals who are secretors of blood group Lewis antigen. Thus, an individual who is a non-Lewis antigen secretor can have a normal CA 19–9 even in the presence of advanced pancreatic cancer.

CONCLUSION

As in all cancers, prevention is a more effective strategy than treatment advances. Most of the GI cancers have known predisposing factors and therefore there is some role for prevention. For individuals with a genetic predisposition, surveillance is the only option. Screening protocols are well-developed for some GI cancers, e.g. colorectal, hepatocellular carcinoma, and we look forward to screening strategies for the other cancers.

REFERENCES

1. Chia KS, Seow A, Lee HP, Shanmugaratnam K, Cancer incidence in Singapore 1993–1997, *Singapore Cancer Registry* **14**, 2000.

2. Parkin DM, Bray F, Ferlay J, Pisani P, Estimating the world cancer burden: Globcan 2000, *Int J Cancer* **94**:153–156, 2001.

3. Lowenfels AB, Maisonneuve P, DiMagno EP, Elitsur Y, Gates LK Jr, Perrault J, *et al*., Hereditary pancreatitis and the risk of pancreatic cancer. International hereditary pancreatitis study group, *J Natl Cancer Inst* **89**:442–446, 1997.

4. U.S. Preventive Services Task Force, Screening for colorectal cancer: recommendation and rationale, *Ann Intern Med* **137**:129–131, 2002.

5. Falk GW, Barrett's oesophagus, *Gastroenterology* **122**:1569–1591, 2002.

6. Fischbach W, Dragosics B, Kolve–Goebeler M, Ohmann C, Greiner A, Yang Q, Bohm S, Verreet P, Horstmann O, Busch M, Duhmke E, Primary gastric B-cell lymphoma: Results of a prospective multicenter study, *Gastroenterology* **119**:1191–1202, 2000.

7. Liu H, Hsu C, Chen C, *et al*., Chemotherapy alone versus surgery followed by chemotherapy for stage I/IIE large-cell lymphoma of the stomach, *Am J Hematol* **64**:175–179, 2000.

8. Chia KS, Seow A, Lee HP, Shanmugaratnam K, Cancer incidence in Singapore 1993–1997, *Singapore Cancer Registry* **82**, 2000.

9. Bloustein PA, Association of carcinoma with congenital cystic conditions of the liver and bile ducts, *Am J Gastroenterol* **67**:40–46, 1977.

10. Tan CK, The problem of hepatitis B and C viruses — and does screening for hepatocellular carcinoma help? *SGH Proc* **9**:126–129, 2000.

11. Lau WY, Leung TW, Ho SK, Chan M, Machin D, Lau J, *et al*., Adjuvant intra-arterial iodine-131-labelled lipiodol for resectable hepatocellular carcinoma: a prospective randomised trial, *Lancet* **353**:797–801, 1999.

12. Leung TWT, Patt YZ, Lau WY, Ho SKW, Chan ATC, Mok TSK, Yeo W, Liew CT, Leung NWY, Tang AMY, Johnson PJ, Complete pathological remission is possible with systemic combination chemotherapy for inoperable hepatocellular carcinoma, *Clin Cancer Res* **5**:1676–1681, 1999.

13. Oka H, Saito A, Ito K, Kumada T, Satomura S, Kasugai H, Osaki Y, Seki T, Kudo M, Tanaka M, Multicenter prospective analysis of newly diagnosed hepatocellular carcinoma with respect to the percentage of *Lens culinaris* agglutinin-reactive α-foetoprotein, *J Gastroenterol Hepatol* **16**:1378–1383, 2001.

This page intentionally left blank

26

Inflammatory Bowel Disease

Ooi Choon Jin

INTRODUCTION

The two main forms of idiopathic inflammatory bowel diseases (IBD) are ulcerative colitis (UC) and Crohn's disease (CD). Although these diseases are rare in Singapore, the numbers have been increasing. In the West, the prevalence of UC ranges from 37–121 per 100 000, while that of CD ranges from 6–104 per 100 000. In Singapore, the estimated minimum prevalence for UC is 10–15 per 100 000 and that of CD is per 6–10 per 100 000. No clear etiology has been identified for IBD, but it is well accepted that these diseases arise in genetically susceptible individuals exposed to an antigenic trigger that results in a dysregulated immune system. Lending credence to current thinking that genetic predisposition plays a part in disease causation, the NOD2 gene on chromosome 16[1] and cytokine gene cluster on chromosome 5[2] have recently been discovered and linked to CD patients. Some racial groups are more prone than others. In Singapore, Indians are over-represented in UC and CD.[3,4] These diseases occur mainly in the younger age groups with peak incidences in ages 20–39.[3,4]

DIAGNOSIS

Ulcerative Colitis

History and physical examination

UC usually presents with symptoms of rectal bleeding, diarrhea, urgency, tenesmus, abdominal discomfort, decreased appetite and loss of weight. The severity of the disease usually parallels the extent and intensity of colonic involvement. Although these symptoms present gradually in most cases, UC can sometimes present abruptly even in patients with limited disease.

Severity of disease dictates the clinical presentation. While mild disease may not manifest any abnormal physical findings, pallor, fever, tachycardia, lethargy and a distended tender abdomen may reflect severe disease. Common extragastrointestinal manifestations include ocular disease (iritis, uveitis, episcleritis); skin problems (erythema nodosum, pyoderma gangrenosum); and arthralgia/arthritis (peripheral and axial arthropathies). The peripheral arthropathies usually mirror the course of the bowel disease, unlike central arthropathies which tend to run a course independent of gastrointestinal manifestations. Primary sclerosing cholangitis is a rare association with UC (less than 1% locally). Hypercoagulable states presenting as venous thromboembolism and arterial thrombosis have been described, albeit infrequently.

Investigations

Endoscopic evaluation is recommended at initial presentation to establish the diagnosis and document the extent of the disease. Endoscopy is also employed to evaluate for exacerbations and dysplasia surveillance, the latter for patients whose disease is long-standing. The rectum is invariably involved, with the disease extending proximally in a symmetric fashion. In some patients, the disease may involve the whole colon, i.e. pancolitis. Histologic findings of UC include alteration of crypt architecture with crypt distortion, crypt abscesses, goblet cell depletion, erosions and superficial mucosal ulcers. The pattern of acute on chronic inflammatory changes is classically confined to the mucosa and superficial submucosa, with the deeper tissue layers unaffected except in fulminant disease.

Table 1 Laboratory Tests in Patients with Inflammatory Bowel Disease

Test	Comments
Full blood count	Anemia, leucocytosis and thrombocytosis in severe disease
ESR	Elevated in severe disease
C-reactive protein	Elevated in severe disease
Electrolytes	Depleted in marked diarrrhea
Serum albumin	Reduced in active disease; function as a negative phase reactant as well as a reflection of protein loss and undernutrition
Stool for leucocytes	Indicate active intestinal inflammation
Stool cultures for enteric organisms	Exclude infection as cause for relapse
Stool for *Clostridium difficle* toxin	Exclude *Clostridium difficle* infection
Serologic tests for pANCA and ASCA	To help differentiate between UC and CD
Hepatitis B surface antigen	If the patient is a carrier, be aware of hepatits B flares and to exhibit caution when tailing down prednisolone

Laboratory tests (see Table 1) also play a role in the clinical evaluation of UC. They confirm the severity of disease (hemoglobin levels, leukocyte count, platelet count, erythrocyte sedimentation rate and C-reactive protein). Low serum albumin indicates a chronic disease, and electrolyte levels may be depleted with profuse diarrhea.

Given the role of microbial agents in UC exacerbation, stool cultures for pathogens, including *Salmonella, Shigella, Yersinia enterocolitica* and *Campylobacter* should be done. A search for ova, cysts and parasites, and *Clostridium difficle* toxin would also be justified in most cases. In patients on immunosuppresant therapy (steroids and/or azathioprine), superimposed CMV colitis should be assiduously excluded through colonic biopsies. Serologic measurements for pANCA (perinuclear anti-neutrophilic cytoplasmic antibodies)[5] and ASCA (antibodies to *Saccharomyces cerevisiae* mannan)[6] may be useful to differentiate UC from CD. Plain abdominal radiographs are useful for detection of colonic dilatation. Colonic diameter above 5 cm or thumb printing sign on radiographs may indicate severe disease. Barium enemas are not routinely used as a means of follow-up once the diagnosis of UC has been made. However, it is useful as a diagnostic tool with evidence of loss of haustral pattern seen in burnt out disease (see Fig. 1).

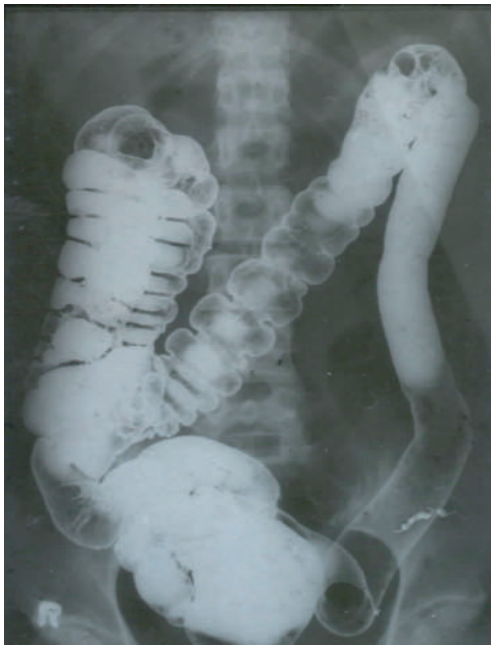


Fig. 1 Barium enema of a patient with chronic UC showing loss of haustral pattern in left colon.

Classification by disease activity

One of the most established criteria used is that of Truelove and Witts.[7] It is based on symptoms, physical findings and laboratory tests (see Table 2).

Crohn's Disease

History and physical examination

CD has a more heterogeneous presentation. Its manifestation largely depends on the location, intensity of the inflammation and presence of specific intestinal or extragastrointestinal complications. Unlike UC, which essentially involves the colon, CD can affect any part of the gastrointestinal tract.

In ileocecal CD (seen in a third of our patients), the main symptoms include diarrhea, colicky abdominal pains and weight loss. Any one of the three symptoms can be predominant, unlike UC in which diarrhea is

Table 2 Truelove and Witts Criteria for Evaluating the Severity of UC

Variable	Mild Disease	Severe Disease	Fulminant Disease
Stools	< 4 per day	> 6 per day	> 10 per day
Blood in stool	Intermittent	Frequent	Continuous
Temperature	Normal	> 37.5°C	> 37.5°C
Pulse	Normal	> 90 per min	> 90 per min
Hemoglobin	Normal	< 75 of normal value	Transfusion required
ESR	< 30 mm/hr	> 30 mm/hr	> 30 mm/hr
Colonic features on AXR	Normal	Air, edematous wall, thumbprinting	Dilatation of colon (> 5 cm diameter)
Clinical signs	Minimal tenderness	Abdominal tenderness	Increased abdominal distension, tenderness

the most frequent and prominent complaint. Low grade temperature is not uncommon. When there is concomitant fibrostenotic disease, the patients may present with small bowel obstruction.

Small bowel disease (seen in about 20% of patients locally) may present with anorexia, malabsorption and weight loss. Colonic disease alone (seen in a third of our patients) usually presents with diarrhea, sometimes with blood. Perianal disease is commonly seen in 10–15% of our patients with Crohn's colitis. CD affecting the esophagus and stomach alone are rare. They are usually seen in conjunction with distal disease.

Extraintestinal manifestations are largely similar to those of UC with the exception of oral apthous ulcers which occur more frequently in CD. Episcleritis also tends to be more common in CD than UC. Renal and genitourinary manifestations occur exclusively in CD and take the form of uric acid or oxalate stones. Rarely, glomerulonephritis and renal amyloidosis may occur. Clinical examination reveals low grade fever, some degree of pallor, undernutrition, lethargy and abdominal tenderness as the common denominator for most of the different CD types. In ileocecal disease, an inflammatory mass may be palpable in the right iliac fossa. Nutritional deficiencies are more marked in small bowel disease. Colonic disease may present with inflammatory masses and strictures resulting in large bowel obstruction. The manifestations of perianal disease include anorectal fistula, anal fissures, perirectal abscess, anal stricture, large hemorrhoidal tags and rarely, anal incontinence.

Investigations

Gastroscopy and colonoscopy with ileoscopy are commonly used to assess the extent and severity of upper gastrointestinal and ileocolonic CD. Classic endoscopic features include apthous ulcers and serpiginous ulcerations in the long axis of the bowel with intervening normal mucosa, rectal sparing and strictures.

Diagnostic imaging is of paramount use in small intestine CD. Enteroclysis or small bowel enema, a procedure that requires intubation, yields excellent views of structural lesions in the areas of study. Alternatively, barium meal follow through, which do not require any intubation, is easier to perform and may yield as good a result (see Fig. 2). Recently, capsule endoscopy has gained increased prominence in its utility in the diagnosis of small bowel CD. CT scanning of the abdomen and pelvis are useful to exclude intraabdominal abscesses.

Biochemical tests used are similar to those used in UC. The utility of these tests has been described in Table 1.

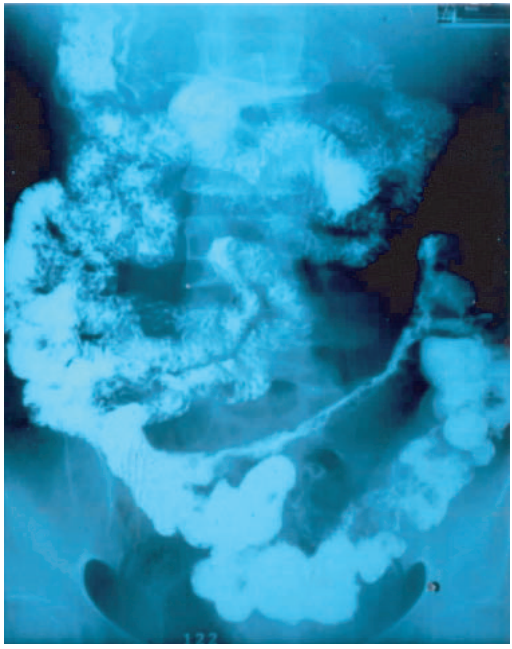


Fig. 2 Barium meal and follow through in a patient with a long CD stricture of the terminal ileum.

Tissues obtained from endoscopic biopsies or resected specimens are essential for the diagnosis of CD. Characteristic features of CD include focal inflammation, transmural involvement and microgranulomas. Analysis of resected specimens reveals clefts, deep fissures with sinus tracts or fistula formation. The presence of non-caseating granulomas suggests the diagnosis of CD. However in CD, just under 20% of patients have granulomas on their biopsies. As tuberculosis is endemic in Singapore, it is important to consider the diagnosis of tuberculosis of the gut when granulomas are observed. TB granulomas tend to be well formed, bigger in size and more confluent with central caseation. Upon applying Ziehl–Neelsen stains, at least 50% of tissues from TB gut are positive for acid fast bacilli. Where indicated, tissues should also be sent for mycobacterial culture as well as PCR analysis.

Classification by disease activity

Disease activity indices developed for CD are more prediction rules for objective measurements of disease activity. They are used in clinical trials. This is unlike the Truelove and Witts criteria which is used for everyday clinical UC management. The most established index for CD is the Crohn's Disease Activity Index (CDAI).[8] This index takes into consideration a 7-day course of symptoms (liquid stools, abdominal pain, sense of general-well being), hematocrit, physical findings (body weight, abdominal mass, extraintestinal manifestations) and the use of imodium and/or opiates for diarrhea (see Table 3). Scores less than 150 usually indicate a quiescent disease.

Table 3 Components of Crohn's Disease Activity Index (CDAI)

Variable (Taken for the Preceeding 1 Week)
Number of liquid or very soft stools
Daily ratings for abdominal pain
General well-being
Number of extragastrointestinal manifestations
Opiates for diarrhea
Presence of abdominal mass
Hematocrit relative to normal values
Body weight relative to normal values

Table 4 Aminosalicylate Preparations

Preparation	Formulation	Site of Action
Oral Therapy		
Azo-bond		
Sulphasalazine	Sulphapyridine carrier	Colon
Olsalazine	5-ASA dimer with dipentum bond	Colon
Balsalazide	Aminobenzoyl alanine carrier	Colon
Delayed release		
Asacol	Eudragit S (pH 7)	Distal ileum–colon
Salofalk	Eudragit L (pH 6)	Ileum–colon
Sustained release		
Pentasa	Ethylcellulose granules	Duodenum–colon
Rectal Therapy		
Mesalamine suppository		Rectum — sigmoid
Mesalamine enema		Rectum — splenic flexure

TREATMENT

Medical Therapy for Ulcerative Colitis and Crohn's Disease

The short- and long-term management of IBD are governed by the type, severity and disease location (see Table 5).

Supportive therapies

There is a role for the treatment that ameliorates a patient's symptoms without affecting disease activity. These include anti-spasmodics, anti-diarrheals, analgesics and even anxiolytics or anti-depressant therapies. Most gastroenterologists will allow for some anti-spasmodics (hyoscine or clidinium-based compounds like Librax) and anti-diarrheal preparations in a mild to moderate disease. It is still prudent to avoid these compounds in a severe disease, so as not to depress colonic muscle tone further and potentially contribute to the advent of a toxic megacolon.

1) Aminosalicylates
   Sulphasalazine, a compound derived by linking sulphapyridine and 5-aminosalicylic acid (5-ASA) with an azo bond, has been used in

Table 5 Suggested Therapy for UC and CD

	Distal Limited UC	Extensive UC	CD
Mild disease	Oral or topical 5-ASA Topical steroids	Oral 5-ASA	Oral 5-ASA
Moderate disease	Oral or topical 5-ASA Topical steroids	Oral 5-ASA	Oral steroids Azathioprine/6-MP
Severe disease	Oral/Parenteral steroids Topical steroids	Oral/Parenteral steroids IV cyclosporine	Oral/Parenteral steroids Parenteral MTX IV Infliximab
Perianal disease	—	—	Oral antibiotics IV Infliximab Azathioprine/6-MP
Quiescent disease	Oral or topical 5-ASA Azathioprine/6-MP	Oral 5-ASA Azathioprine/6-MP	Oral 5-ASA Azathioprine/6-MP

mild to moderate UC and CD alone with reasonable success. Since most of the side-effects are related to the sulphapyridine moiety, many new 5-ASA compounds have become available (see Table 4). Higher doses are usually required for induction of remission as opposed to maintenance of quiescent disease, and as in the case of sulphasalazine, the recommended dose is 3–6 g/day for induction and 2–4 g/day for maintenance. For distal disease, topical formulations for distal disease are useful in treating active disease and for maintenance therapy. Mesalamine suppositories are suitable for distal proctosigmoiditis. The enema preparations are efficacious for disease up to the splenic flexure. Aminosalicylates alone are not adequate for treatment of severe disease.

2) Corticosteroids

Topical therapy in the form of hydrocortisone foam in the treatment of mild to moderate proctosigmoiditis have been successful and can be used as an alternative to 5-ASA-based compounds. Topical steroids can also be used as an adjunct to parenteral steroid therapy in severe colitis.

Oral corticosteroids are indicated for the treatment of patients with moderately severe disease exacerbation. Prednisolone doses prescribed usually range from 0.5–1.0 mg/kg body weight. Although there is a documented dose response between 20 and 60 mg per day, the modest benefits of a higher dose are offset by the increasing side-effects. Some of these side-effects can be partially minimized by

careful management (early management of steroid-induced osteoporosis by supplemental calcium, vitamin D and biphosphonates and regular monitoring of blood pressure and blood glucose levels for possible steroid-induced hypertension and diabetes). In some patients with CD of the right colon and terminal ileum, controlled ileal-release preparations of novel steroids like budesonide can be useful. The decrease in steroid side-effects of budesonide is based on a high first-pass metabolism that eliminates the body of 90 % of the drug. However, this is offset by the fact that budesonide has an affinity for the glucocorticoid receptor that is 50–100 times that of prednisolone. Oral corticosteroids are not useful in maintenance therapy in that studies employing up to 15 mg per day have failed to maintain remissions. It is important to allow for a complete response before oral steroids are tapered as this may precipitate a rapid clinical deterioration. Parenteral corticosteroids are the mainstay of therapy for patients with severe disease requiring hospitalization.

Immunomodulators

1) Azathioprine and 6-Mercaptopurine (6-MP)
   Azathioprine and its active metabolite, 6-MP may require 2–3 months before they are completely effective. These drugs are useful maintenance therapies in subsets of IBD patients who are unable to taper off steroids. These drugs are usually initiated at 50 mg doses and are gradually increased to 2.5 mg/kg body weight for azathioprine and 1.5 mg/kg body weight for 6-MP. Given that the main side-effects include leucopenia, hepatitis and pancreatitis in addition to nausea and dyspepsia, monitoring should include full blood count, serum amylase and liver enzymes. The earlier concern about an increased risk for lymphoma remains unresolved, owing in part to a slight underlying increase in the risk of lymphoma in patients with inflammatory bowel disease.
2) Cyclosporine
   Intravenous cyclosporine has benefited short-term management of severe refractory UC in 50–80% of patients. In a small, selected group of patients with fistulous disease in CD, intravenous cyclosporine has been shown to yield rapid closure of fistulas. Conversion to oral dosing tends to cause relapses in these subsets of UC and CD

patients. Toxicities like nephrotoxicity (hypertension, elevated urea and creatinine, decreased glomerular filtration rate) and opportunistic infections (*Pneumocystis carinii* pneumonia) limit its use.

3) Methotrexate
   Methotrexate is effective in steroid-dependent active CD and in maintaining remission in CD. It has no role in UC. This drug is usually administered parenterally (25 mg per week subcutaneously or intramuscularly). Its main side-effects are related to its immunosuppression, interstitial pneumonitis and hepatic fibrosis (dose related). Its response time is measured in weeks.
4) Antibiotics
   Antibiotics have no proven value in UC. The data, however, points to the beneficial role of metronidazole in CD patients with perianal disease. The fluoroquinolone, ciprofloxacin, is usually used as an adjunctive therapy to metronidazole in perianal disease and in the unproven setting of active luminal CD.
5) Anti-tumor necrosis factor (infliximab)
   This chimeric monoclonal antibody that binds to soluble tumor necrosis factor (TNF) and precursor cell-surface TNF has been shown in pivotal trials[9,10] to be efficacious in both luminal CD and perianal fistulous CD. Response is quite prompt, usually within two weeks. The efficacy may last from a few weeks to more than six months. Given the nature of this biologic agent, its side-effects may include serum-like sickness. This side effect tends to occur in individuals who have received the treatment and resumed treatment after a hiatus. The duration of response after infusions may also become progressively shorter. Other complications include reactivation of tuberculosis and lupus like syndromes. Above all, the drug is expensive. The utility of this agent for ulcerative colitis remains uncertain, although increasing anecdotal reports involving small number of patients have been encouraging.

Investigational therapies

New UC therapies being studied include nicotine, short chain fatty acids, fish oils, heparin and biologic therapies. Investigational therapies for CD include growth hormone, thalidomide, elemental diet, mycophenolate

mofetil, interleukin-11 and probiotics. As yet, many have not achieved the results desired for general acceptance or regulatory approval.

Surgical therapy

The main indications for surgery in UC include: 1) uncontrollable hemorrhage; 2) toxic megacolon with impending or frank perforation; 3) steroid or cyclosporine resistant fulminant active disease; 4) dysplasia or cancer; and 5) intractable disease. The previous popular surgical procedure of choice — proctocolectomy and ileostomy — has now given way to the more acceptable ileoanal anastomosis.

Indications for surgical intervention in CD include: 1) uncontrollable hemorrhage; 2) fulminant disease; 3) dysplasia and cancer; 4) perforations/fistulas/inflammatory mass; 5) obstructions from stricures; and 6) intractable disease. The most common types of surgery performed in CD include intestinal resection with or without anastomosis, bypass procedures and stricturoplasty. Procedures done for perianal CD include examination under anaesthesia, incision and drainage for perirectal abscesses, fistulotomy and Seton insertions.

Although many physicians, in general, tend to regard surgery in IBD as a "last resort," surgery can be a form of effective (in most instances) treatment. Surgery, in selected cases, does afford the patient a chance to get well and proceed to achieve a decent quality of life.

REFERENCES

1. Ogura Y *et al.*, A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease, *Nature* **411**:603–606.
2. Rioux JD *et al.*, Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn's disease, *Nature Genetics* **29**:223–228.
3. Ling KL, Ooi CJ, Luman W, Cheong WK, Seow–Choen F, Ng HS, Clinical characteristics of ulcerative colitis in Singapore, a multiracial city-state, *J Clin Gastroenterol* **35**:144–148.
4. Data from Singapore General Hospital IBD registry (unpublished).
5. Ooi CJ, Lim BL, Cheong WK, Ling AE, Ng HS, Antineutrophil cytoplasmic Antibodies (ANCAs) in patients with inflammatory bowel disease show no corrrelation with proteinase 3, lactoferrin, myeloperoxidase, elastase, cathepsin G and lysozyme: a Singapore study, *Ann Acad Med Singapore* **29**:704–707.

6. Vermeire S *et al.*, Comparative study of ASCA (anti-Saccharomyces cerevisiae Antibody) assays in inflammatory bowel disease, *Gastroenterology* **120**:827–833.

7. Truelove SC, Witts LS, Cortisone in ulcerative colitis, *Br Med J* **2**:1041.

8. Best WR, Becktel JM, Singleton JW. Rederived values of the eight coefficients of Crohn's disease activity index (CDAI), *Gastroenterology* **77**:843.

9. Targan SR *et al.*, A short term study of chimeric monoclonal antibody cA2 to tumour necrosis factor α for Crohn's disease, *N Engl J Med* **337**:1029–1035.

10. Present DH *et al.*, Infliximab for the treatment of fistulas in patients with crohn's disease, *N Engl J Med* **340**:1398–1405.

This page intentionally left blank

27

Gastrointestinal Endoscopy

Yap Chin Kong

INTRODUCTION

Advances in technology have lead to the development of flexible endoscopes that can be inserted into various parts of the digestive tract to visualize intra-luminal pathology and treat diseases. Endoscopy can be regarded as an extension of the clinical examination to diagnose gastrointestinal (GI) symptoms. Like the physical examination, a systematic approach, a meticulous technique and a trained eye, are needed to detect abnormalities. Adequate visualization is the key. A provisional diagnosis can be made based on morphological features in most cases, but the definitive diagnosis relies on histology. With accurate diagnosis, appropriate therapy can then be applied via the endoscope. Endoscopic ultrasonography (EUS) may be used in further evaluation of mucosal cancers, submucosal lesions and biliary-pancreatic diseases. The use of endoscopy for the more important and established situations will be discussed. The principles and some specifics are outlined for the benefit of the general physician.

PRE-PROCEDURE MANAGEMENT

A cardio-respiratory assessment is performed and consent is taken for the procedure. Antibiotic prophylaxis for endocarditis is considered for patients with valvular heart disease (rheumatic valve disease, mitral valve prolapse with regurgitation) undergoing invasive therapeutic endoscopy (injection sclerotherapy, dilation, percutaneous gastrostomy, ERCP). Prophylactic antibiotics are recommended for patients undergoing ERCP for biliary obstruction and percutaneous gastrostomy. Anti-platelet agents and warfarin should be discontinued for a week before invasive procedures like polypectomy, endoscopic mucosal resection, dilation of stenoses, endoscopic sphincterotomy and percutaneous gastrostomy. Routine cross-matching of blood is not necessary unless the patient is actively bleeding. Immediate bleeding during endoscopy can usually be treated successfully by endoscopy techniques. A 6-hour fast is desirable to enable proper visualization in elective cases. Contraindications to endoscopy are suspected viscus perforation and cardio-respiratory instability, which cannot be corrected by supportive measures. Emergency endoscopy can be performed in the intensive-care setting under close monitoring.

SEDATION AND ENDOSCOPY PROCEDURE

Diagnostic upper endoscopy and colonoscopy are performed on an ambulatory basis while more invasive procedures are treated as inpatient cases. The patient is positioned in the lateral, supine or prone position according to the procedure to be performed. An assessment of the anesthetic risks helps to decide on the use and type of sedation required. Very ill patients (e.g. variceal bleeding or cholangitis with neurological depression or cardio-respiratory instability) may require endotracheal intubation for airway protection before proceeding with sedation and endoscopy. In patients requiring sedation, pulse oximetry monitoring is obligatory and oxygen is delivered via nasal cannulae if oxygen saturation falls below 90%. Medications used for sedation include short-acting Dormicum (midazolam), Valium (diazepam) or Propofol, depending on the anticipated length of procedure. Short-acting Fentanyl or Pethidine can be added, depending on the degree of analgesia required. Buscopan or glucagon is used to reduce peristaltic contractions when required. Side-effects of sedation include paradoxical agitation with midazolam

and cardio-respiratory depression. Flumazenil should be available to reverse the effects of midazolam or diazepam. Complications of sedation include aspiration pneumonia and cardio-respiratory depression.

UPPER GASTROINTESTINAL ENDOSCOPY

1 Diagnostic Endoscopy

The indications for diagnostic esophago-gastroduodenoscopy (OGD) are chronic dyspepsia, gastro-esophageal reflux, dysphagia, iron-deficiency anemia, hemetemesis or malena, evaluation of abnormal barium studies and malabsorption syndromes. During endoscopy, biopsies can be obtained from abnormal areas for tissue diagnosis. Rapid urease test for *Helicobacter pylori* can be performed where appropriate. Blind areas at OGD include the posterior wall, lesser curve and duodenal angle. A cap-fitted technique helps to improve visualization and targeting of lesions at these areas during diagnostic and therapeutic OGD.

The image obtained at OGD is in two-dimensions and subtle variations in mucosal contour is difficult to appreciate. Chromoendoscopy is a simple technique using dyes (indigo-carmine, methylene blue) that fill in the subtle depressions in the mucosa and accentuates the contours better, improving the diagnosis of small lesions. It accentuates the morphological features of small mucosal cancers that may otherwise be dismissed as normal. In Japan, a country with a high prevalence of gastric carcinoma, screening for early cancer is recommended for high-risk groups: age $>$ 50 years, *Helicobacter pylori* positive currently or previously, intestinal metaplasia and positive family history. The risk for early gastric cancer in *Helicobacter pylori* positive patients is estimated at 0.5% per year in one Japanese study. Most of early gastric cancers are detected incidentally ("opportunistic" screening) during OGD while a formal screening detects only 10% of early gastric cancers. In Western countries where gastro-esophageal reflux disease is prevalent, screening for high grade dysplasia and cancer is recommended for long segment ($>$ 3 cm) Barrett's esophagus. The risk for esophageal adenocarcinoma in long segment Barrett's esophagus is estimated at 0.5% per year.

2 Therapeutic Endoscopy

Therapeutic endoscopy is indicated for upper gastrointestinal bleeding, dilation of benign and malignant stenoses, insertion of stents for malignant

stenoses of the esophagus, stomach and duodenum, removal of mucosal lesions (by polypectomy or endoscopic mucosal resection), removal of foreign bodies, enteral nutrition support and treatment of gastro-esophageal reflux disease.

Bleeding from esophageal varices can be treated by endoscopic variceal ligation (EVL) using rubber bands loaded onto a device that can be mounted on the tip of the endoscope. EVL has effectively replaced injection with sclerosants because of efficacy, lower complications and a short learning curve. The risk-benefit profile of EVL is safe enough to allow primary prophylactic ligation of moderate sized (grade 3) varices that have not bled. Gastric varices can be treated by injection of cyanoacrylate, a synthetic polymer that acts as a glue to achieve obliteration. Endoscopic treatment has largely replaced emergency surgery for variceal bleeding. Complications after EVL include transient mild chest pain and odynophagia, superficial ulcerations after the rubber bands are dislodged and, rarely, bleeding from the ulcers. Injection of cyanoacrylate can be complicated by embolization to the lungs via the azygous venous system, or rarely systemically via a patent foramen ovale. The polymer cast that forms after cyanoacrylate injection is extruded several weeks or months later and may be accompanied by a small bleeding episode.

Bleeding peptic ulcers can be treated by injection of diluted adrenaline, hypertonic dextrose, sclerosants, application of metallic clips and thermal coagulation using a heat-probe, BICAP or argon plasma coagulation. The advantages of endoscopic treatment are a reduction in mortality and need for emergency surgery. Surgery is performed if endoscopic treatment fails to control bleeding, for high-risk lesions located over large arteries (posterior duodenal wall, lesser curve of stomach) and if re-bleeding occurs after endoscopic treatment. Urgent therapeutic endoscopy can be performed relatively safely in actively bleeding patients after resuscitation with close monitoring. Patients with coronary artery syndromes and active bleeding should have endoscopy under continuous monitoring in an intensive-care environment. Complications include perforation if excessive coagulation is used. Arrhythmias can occur if excessive adrenaline is absorbed.

A variety of lesions that bleed can be treated endoscopically. These include: vascular malformation, Dieulafoy lesion and large polyps. Argon plasma coagulation is used to coagulate vascular malformations. Metal clips or rubber bands are used to treat dieulafoy lesions, and polypectomy is performed to remove bleeding polyps.

Benign strictures due to reflux esophagitis, ingestion of corrosives or post-operative stricture in the esophagus can be dilated using polyvinyl dilators or pneumatic balloons. Symptomatic recurrences are treated with repeated dilations. Achalasia can be similarly dilated using a large diameter (30–40 mm) pneumatic balloon for symptom control. Alternatively, botulin toxin can be injected into the lower esophageal sphincter. Complications include bleeding and perforation, the majority of which are small tears that can be managed conservatively with nil orally and antibiotics.

Malignant strictures in the esophagus, stomach and duodenum may cause dysphagia and gastric outlet obstruction. Palliative relief of dysphagia, tracheo-esophageal fistula or gastric outlet obstruction is effective by stent insertion. Alternative options are periodic dilation, alcohol injection or laser ablation. However, other problems such as pain, anorexia, bleeding and hypomotility do not respond to stenting. Self-expandable metal stents (SEMS), which require a lesser degree of dilation, are easier to deploy, though more expensive compared to plastic stents. SEMS with a covering membrane ("covered" stents) reduce the incidence of tumor ingrowth and can be used to treat tracheo-esophageal fistulae. Complications include: perforation associated with the dilation process and is uncommon with self-expandable metal stents. Stent-related problems include migration, reflux, tumor ingrowth and overgrowth.

Endoscopic mucosal resection (EMR), or strip biopsy is a technique used to remove early cancers of the stomach for pathological examination. Early gastric cancers are deemed to be resected with curative intent if the resected specimen has the following features: < 2 cm, of well or moderate cytological differentiation, no ulceration, tumor-free margins, no submucosal, lymphatic or venous infiltration. Areas of 1–2 cm diameter can be resected with low complications. The first step is injection of a large volume of saline to separate the mucosa and submucosa from the muscularis propria. The elevated mucosa and submucosa, together with the lesion, can then be safety resected using electrocautery and a polypectomy snare. Alternatively, a cap-suction technique can be employed for EMR. An en-bloc resection is preferred to piece-meal resections because of higher local recurrence in the latter. EMR is also used to treat patients with early cancers of the esophagus (squamous and Barrett's adenocarcinoma) who are high anesthetic risks. Complications of EMR include pain, which is common and is managed conservatively. Bleeding can be

treated endoscopically, and small perforations can be managed using metal clips. Large perforations require surgery.

Foreign bodies can be removed from the upper digestive tract using a variety of snares, forceps and a protective sheath to prevent mucosal laceration during extraction.

Percutaneous endoscopic gastrostomy (PEG) can be established with endoscopy. PEG is commonly indicated in patients with established (>30 days) impaired swallowing due to neurologic conditions such as strokes, motor neurone disease and malignancy of the oropharynx, larynx and esophagus. The aim of PEG is to maintain nutritional status and improve the quality of life. Predicting life expectancy and judging quality of life can be difficult in such patients. As PEG is an invasive procedure and does not prevent aspiration of gastric contents, the benefits of PEG compared to naso-gastric feeding tube needs careful consideration in each individual patient. A PEG extended into the jejunum (PEG-J) or a direct percutaneous endoscopic jejunostomy (PEJ) are more difficult procedures that hope to reduce the risk of aspiration. Endoscopy can also assist in the placement of short-term naso-enteric tubes into the small bowel with the hope of reducing aspiration risks. Complications related to the procedure include pain, perforation of adjacent viscera, bleeding, infection and tube dislodgement. Pneumoperitoneum is common and does not require treatment unless there is peritonitis. The PEG tube can be left *in situ* until problems occur which can be many months or even years, or be changed to a button PEG after a percutaneous tract has formed.

Endoscopic treatment of mild gastro-esophageal reflux disease (GERD), an essentially benign disease, is a new and rapidly evolving technology. The aim is to cause a physical narrowing of the cardio-esophageal junction using a variety of techniques to form a barrier to acid reflux. Suturing devices, radio-frequency treatment and injections of a biologically inert material have been tried with fair success, durability and occasional mortality. Long-term studies regarding the safety and role of various endoscopic treatments for GERD need to be performed, given the availability of safe and effective pharmacological therapies.

3 Endoscopic Ultrasonography

Endoscopic ultrasonography (EUS) is a technique that uses an endoscope with an ultrasound scanner mounted at the tip. The wall of the digestive

tract is a 5-layered structure comprising mucosa, submucosa, muscularis propria and adventitia. Outside these layers, the surrounding lymph nodes, vessels and various organs (heart, liver, pancreas, spleen and kidneys) can be visualized on ultrasound to varying degrees depending on the penetration of the ultrasonic waves (7.5 or 12 MHz). The image obtained using the common "circular scan" mode, results in a cross-sectional image. Alternatively, small diameter ultrasonic probes (20 MHz) can be inserted through the instrument channel of the endoscope to target lesions for EUS with greater resolution but less penetration. A "blind probe" that is thin, yet with good ultrasound penetration can be inserted over a guide-wire to stage stenotic esophageal cancer. Fine-needle aspiration cytology (FNAC) requires a "linear scan" mode that can direct the cytology needle into the lesion of interest using real-time ultrasound.

The indications for EUS include diagnosis of submucosal lesions, staging of cancers of the esophagus, stomach, pancreas and periampullary area, evaluation of a dilated bile duct for stones, detection of neuroendocrine tumors in the pancreas and chronic pancreatitis.

Submucosal lesions are usually suspected after endoscopy or barium studies. EUS is then applied to differentiate a true submucosal lesion from extra-luminal compression by other structures. EUS is very useful in diagnosing the cause of thickened gastric folds. It can differentiate varices from infiltrative cancer, lymphoma and Menetrier's disease. Large particle biopsy can be obtained if EUS shows suspicious thickening of the wall structure but the initial endoscopic biopsy is negative. Although EUS provides clues to the nature of a submucosal lesion, judged by the layer of origin, size, shape, echo pattern, and margins, it cannot substitute for histopathology in differentiating benign from malignant lesions. Inter-observer agreement is good for lipoma, leiomyoma, vascular structures and extrinsic compressions.

Cancer of the digestive tract can be staged for the depth of involvement (T-stage) and lymph node involvement (N-stage). Accuracy of T-stage varies from 70–90%, depending on the under- or over-staging phenomenon due to microscopic infiltration or peri-tumor inflammation. The N-stage can be confirmed pathologically with fine-needle aspiration cytology (FNAC). In esophagus cancer, infiltration into the aorta (T4) and celiac lymph nodes can be assessed. In pancreas cancer, infiltration into the spleno-portal confluence can be assessed for resectability. The influence of EUS on cancer management depends on the use of stage-dependent

protocols for decision-making and treatment. For example, if a "three-field" radical surgery for esophageal cancer is the strategy adopted to clear the cervical, mediastinal and celiac nodes, EUS is irrelevant with regards to N-staging. If resection is the strategy adopted for cancer, identifying the T4 stage (infiltration into surrounding structures) will help to decide on non-resectable cases. Tumor stenosis prevents the passage of the EUS scope and may require prior dilation that increases the complication rate. A "blind probe" is specially designed for this situation to avoid dilation.

Bile duct stones can be difficult to visualize on percutaneous ultrasound due to bowel gas, and stones can be isodense with bile on CT scan. EUS can visualize the bile duct and periampullary region well and diagnose small stones with high sensitivity. This is useful to select patients with common bile duct stones for therapeutic ERCP and hence avoid diagnostic ERCP, with small but real risks.

In the appropriate setting of hypoglycemia and high insulin levels, small insulinoma tumors in the pancreas, particularly in the head region can be localized using EUS.

The diagnosis of chronic pancreatitis relies on morphological definition of pancreatic ductal abnormalities and calcifications based on CT, ERCP or MRCP. Recently, parenchymal features of chronic pancreatitis on EUS have been described.

LOWER GASTROINTESTINAL ENDOSCOPY

Diagnostic Colonoscopy

The indications for colonoscopy include alteration in bowel habit, positive fecal occult blood test, per rectal bleeding, chronic or bloody diarrhea, iron deficiency anemia, and cancer screening in high-risk groups. Chronic abdominal pain without other features, has a low yield of colonic pathology.

Colonoscopy is performed on an ambulatory basis and is fairly well tolerated. Bowel preparation with osmotic purgatives is necessary to allow proper visualization. This is usually done one day before, or even on the same day. A "clean" colon and a meticulous examination help to reduce the chance of missing lesions behind folds and anatomical flexures that are potentially blind areas. It is considered the "gold standard" for examination of the colon and terminal ileum, compared to barium enema

and CT virtual colonoscopy. However, miss rates for polyps > 1 cm was reported in 6%, using back-to-back colonoscopy by experts. Sigmoidoscopy is the insertion of the endoscope up to the sigmoid-descending colon junction. This junction is commonly a difficult region to negotiate, especially in patients with a thin habitus or women who had a hysterectomy. The combination of sigmoidoscopy plus barium enema is a reasonable alternative to total colonoscopy in difficult cases. Chromoendoscopy in the colon allows the detection of "flat" or depressed early cancers. In addition, magnifying colonoscopy can identify various pit-patterns on the surface of lesions that correlate with the glandular architecture on histology and predict the likelihood of malignancy and submucosal infiltration.

Alterations in bowel habit and abdominal discomfort can be due to irritable bowel syndrome (IBS). This can be accurately and safely diagnosed based on Manning's or the Rome criteria for functional digestive disorders. A sigmoidoscopy or colonoscopy is indicated to exclude pathology that can mimic IBS, especially in older patients or if there is anemia, fever, weight loss or bloody stools.

A positive fecal occult blood test requires total colonoscopy for evaluation. Cancer, large polyps, vascular malformations, hemorrhoids, aspirin, NSAID ingestion, bleeding gums and peptic ulcer need to be considered.

Bleeding from the colon, or hematochezia, can be due to hemorrhoids, cancer, large polyps, diverticulae, NSAIDs or vascular malformation. Colonoscopy can be performed during active bleeding to localize the source and render appropriate endoscopic treatment.

The evaluation of chronic (> 2 weeks) or bloody diarrhea includes stool cultures for infectious agents. Total colonoscopy and ileoscopy is indicated to diagnose inflammatory bowel disease (IBD). Crohn's disease needs to be differentiated from tuberculosis in our population. Biopsies should be taken from inflammed and normal-looking mucosa for microscopic involvement in Crohn's disease. The morphology of lesions, pattern of distribution of inflammation is important in the differential diagnosis of ulcerative colitis, ischemic colitis and infectious colitis. Bacterial infections, clostridium difficile, cytomegalovirus and herpes simplex can be superimposed on or aggravate IBD.

Colonoscopy for cancer surveillance is indicated for patients with a personal history of adenomatous polyp, first-degree family member with colon cancer or adenomatous polyp, genetic syndromes (familial

adenomatous polyposis, FAP or hereditary non-polyposis coli, HNPCC, or Gardner's syndrome). For the "average risk" individual, screening for colorectal cancer has been advocated at age 50 years. Fecal stool occult blood testing annually has been shown in randomized controlled trials to reduce mortality from colorectal cancer. There is some evidence that colonoscopy as the initial screening procedure for cancer screening detects more cancer and significant polyps, but the impact on cancer related mortality has not been studied in randomized trials.

Polyps can be pedunculated or sessile. The range of pathology includes hyperplastic, adenomatous (tubular, villous, tubular-villous), harmartoma (juvenile polyps, Peutz–Jeghers syndrome), polypoidal and "flat" cancers. The "adenoma-cancer" model is a sequence of cumulative genetic alterations resulting in the transformation of normal mucosa to polyp and progression to dysplasia, mucosal carcinoma and invasive carcinoma over approximately 10 years. These pre-malignant adenomatous polyps can be removed with polypectomy thereby reducing the incidence of colon cancer. Cancers arising "*de novo*" have been described in Japan and the United Kingdom that account for a minority. These develop directly as cancer and invade the submucosa even as small "flat" lesions. They are difficult to detect unless chromoendoscopy is used.

Therapeutic Colonoscopy

Polypectomy is the removal of polyps using a snare and electrocautery. Pedunculated polyps with a stalk are easily removed. The polyp is assessed pathologically for malignancy (degree of differentiation), depth of penetration into the submucosa, lymphatic or venous infiltration and distance from the resection margin. An adenocarcinoma without invasion into the submucosa is deemed "cured" after polypectomy. If the tumor invades the submucosa and is well- or moderately differentiated, without lymphatic or venous infiltration and the distance from the margin is >2 mm, the chance of lymph node metastases is probably $<7\%$. The need for subsequent colectomy and lymph node clearance has to be weighed against the risk of a colectomy.

Sessile polyps are technically more difficult to remove and require EMR. The polyp is elevated with submucosal injection of saline and adrenaline. The polyp can be removed piece-meal using a snare and electrocautery.

Adrenaline injections can be used to treat bleeding from diverticula. Vascular malformations can be treated with argon plasma coagulation. Bleeding from polyps can be arrested by injection of adrenaline, application of metal clips or ligating devices. Mechanical obstruction from cancer can be treated with self-expandable metal stents (SEMS) as a temporizing measure if the patient is too ill for surgery, allowing an improvement in the condition before surgery. Pseudo-obstruction of the colon can be treated with colonoscopic insertion of colonic tube.

Complications of colonoscopy include perforation (about 1:2000 for diagnostic colonoscopy) and bleeding. This is higher if a therapeutic procedure is performed, such as polypectomy and EMR. Abdominal pain and local tenderness can follow polypectomy and this is likely due to transmural inflammation. Differentiation from perforation is important and a trial of conservative management with bowel rest, antibiotics and judicious review can be tried. Bleeding following polypectomy can be delayed up to two weeks, and can usually be controlled endoscopically by injections or metal clips.

ENDOSCOPIC RETROGRADE CHOLANGIO-PANCREATOGRAPHY (ERCP)

ERCP is a combined endoscopic and radiologic technique to image the biliary and pancreatic tract for diagnosis and therapy. A purely "diagnostic" role is rarely required and every ERCP should have a therapeutic intent. Imaging of the biliary and pancreatic ducts with MRI/MRCP and EUS has none of the potential complications of ERCP and should be considered initially if the diagnosis of biliary-pancreatic anatomy or pathology is unclear or the yield of ERCP is likely to be low.

The patient is usually in the prone position. A side-viewing duodenoscope with an elevator bridge is used to visualize the papilla and achieve cannulation of the desired duct with confirmation under fluoroscopy. A therapeutic procedure can then be performed: sphincterotomy of the biliary or pancreatic sphincter, stone extraction or fragmentation, stent insertion or stricture dilation.

The indications for ERCP include:

1) Biliary obstruction due to stone or malignancy;
2) Biliary tract injuries and benign strictures;

3) Pancreatitis, acute and chronic; and/or
4) Ampullectomy.

Stones in the biliary tract can manifest as obstructive jaundice, cholangitis, pancreatitis or Mirrizzi's syndrome, depending on the location of the stone. Stones can be classified as cholesterol, brown pigment or black pigment. Most common bile duct stones can be successfully cleared (>90%) after sphincterotomy. The failure to remove stones may be due to a large size, a small distal passage or a difficult location (Mirrizzi's syndrome, intrahepatic location). Large stones can be crushed with a mechanical lithotripter. In the other situations, a temporary biliary stent can be inserted to allow drainage and surgery can be performed later. In situations where the anesthetic risk is high, a stent can be appropriate definitive treatment. After clearance of stones from the bile duct, residual gallbladder stones may cause biliary events and require cholecystectomy in 37% and repeat ERCP in 10% of patients over two years.

Malignancy in the biliary tract may be due to pancreatic, bile duct, periampullary, gallbladder cancer or metastatic spread to the porta hepatis. Primary tumors should be assessed for resection possibilities and the role of ERCP is for palliative relief of itch. The patient should "earn" the procedure. Pre-operative drainage is not mandatory and may contribute to morbidity in potentially resectable cases. Cytological diagnosis can be obtained using brush cytology or EUS guided fine-needle aspiration for suspected malignancy. The overall sensitivity for brush cytology ranges from 40–80% and a negative test cannot reliably exclude a cancer. The anatomy of hilar strictures according to the Bismuth classification (types I, II, III and IV) should be determined because it determines procedure-related cholangitis risk and mortality. Within a year, stent "clogging" and recurrent cholangitis occur in up to 41% and 29% of patients with pancreatic cancer, using plastic and metal stents respectively, and require additional intervention. Various measures to retard the clogging phenomenon have been tried using bile acid supplements, antibiotics and various stent designs. Only stent design has made an impact on prolonging stent patency. On the average, metal stents (10 mm diameter) stay patent longer than plastic stents (3 mm diameter), for 9 months and 4 months respectively. However, stent clogging is unpredictable in the individual patient.

Bile duct injuries are uncommon but can lead to significant morbidity. This is higher in laparoscopic cholecystectomy compared to open

cholecystectomy and is related to ischemic injury from electrocoagulation. The injury can be identified during surgery or be delayed for several months. The injuries range from mild to severe, and can be classified accordingly: complete transection, leak or stricture. Minor leaks can be treated by sphincterotomy or temporary stenting. Strictures are more difficult to treat and require several sessions of dilation and stenting for a year. An overall response of 80% after two years is encouraging and favors a trial of endoscopic treatment. Complete transections or leaks isolated from the bile duct are not amenable to endoscopic treatment.

Benign strictures due to primary sclerosing cholangitis can be treated by short-term stenting. Rare cases of benign stricture due to tuberculosis can mimic cholangiocarcinoma. Sphincter of Oddi stenosis may be due to fibrosis at the level of the ampulla secondary to biliary stones. Sphincter of Oddi dysfunction is a controversial entity that has been described to cause post-cholecystectomy pain and pancreatitis. Manometric studies are advocated to justify the risks of sphincterotomy.

Pancreatitis is diagnosed based on epigastric pain, elevated amylase ($>3\times$ normal) and compatible imaging. Acute pancreatitis is commonly due to stones or sludge and is readily treated by sphincterotomy. Pancreas divisum causing pancreatitis is a rare condition and can be treated by sphincterotomy of the minor papilla. Chronic pancreatitis is diagnosed on pancreatographic features on MRCP or ERCP. Stones and strictures can occur in the pancreatic duct and these can be treated by sphincterotomy followed by stone clearance or stenting. A dilated pancreatic duct and good response to endoscopic stenting predicts successful surgical relief of pain by a pancreatico-jejunostomy.

Periampullary neoplasms that are villous adenomas or neuroendocrine tumor are amenable to endoscopic ampullectomy. The procedure is similar to polypectomy and removes the entire ampulla.

The complications of ERCP include pancreatitis, bleeding, cholangitis and perforation. Pancreatitis is usually mild (5%) and is managed conservatively by nil orally, pain relief and bowel rest. Severe pancreatitis can be life-threatening and may require intensive care. Significant bleeding (2%) can be delayed for up to two weeks after sphincterotomy and is usually successfully treated endoscopically. Cholangitis (5% in distal stricture and up to 40% in hilar stricture) indicates incomplete biliary drainage and an attempt can be made again using ERCP, or by percutaneous drainage radiologically. Perforation by the endoscope is rare, occurring in difficult

anatomy (stenosis, distortion, diverticulum, Billroth II gastrectomy) and requires surgery. Small perforations due to guidewire or catheter trauma can usually be managed conservatively with nil orally, antibiotics and percutaneous drainage of any fluid collection.

CAPSULE "ENDOSCOPY"

This is a new technology that uses a swallowable capsule with a self-contained video chip, power and light source (8 hours life-span) and radio-wave transmitter to transmit video signals to receivers mounted on the surface of the abdomen. Several thousand images are captured as the capsule is propelled by peristalsis towards the colon. These images are down-loaded to a desk top personal computer and analyzed by a trained physician. The main indication is for obscure gastrointestinal bleeding and is purely diagnostic. Small bowel enteroscopy or laparotomy may still be required to treat any lesions discovered.

CONCLUSION

The role of diagnostic endoscopy continues to evolve due to competition from other modalities such as radiological imaging (virtual colonoscopy) and capsule "endoscopy". The development of optical enhancement technologies such as fluorescent endoscopy, fluorescence and light scattering spectroscopy, optical coherent tomography and color enhancement endoscopy, are aimed at improving the diagnosis of early cancers. In the area of gastrointestinal cancer, the promise of improved diagnosis of early cancers, pre-malignant lesions and minimally invasive endoscopic treatment is resection with curative intent, organ preservation and improved quality of life.

REFERENCES

General reading

1. Silverstein FE, Tytgat GNJ, *Atlas of Gastrointestinal Endoscopy*, 2nd ed., Gower Medical Publishing, New York, 1991.
2. Cotton PB, Williams CB, *Practical Gastrointestinal Endoscopy*, 3rd ed., Blackwell Scientific, Oxford, 1983.

3. Rosch T, Classen M, *Gastroenterologic Endosonography: Textbook and Atlas,* Stuttgart, Thieme, New York, 1992.

Gastrointestinal bleeding

1. Cook DJ, Guyatt GH, Salena BJ, Laine LA, Endoscopic therapy for acute non-variceal upper gastrointestinal hemorrhage, *Gastroenterology* **102**:139–148, 1992.

2. Forrest JA, Finlayson ND, Shearman DJ, Endoscopy in gastrointestinal bleeding, *Lancet* **2**:394–397, 1974.

3. Lau JYW, Sung JJY, Lam YH, Chan ACW, Ng EKW, Lee DWH, *et al.*, Endoscopic retreatment compared with surgery in patients with recurrent bleeding after intitial endoscopic control of bleeding ulcers, *N Engl J Med* **340**:751–756, 1999.

4. Khuroo MS, Yattoo GN, Javid G, Khan BA, Shah AA, Gulzar GM. *et al.*, A comparison of omeprazole and placebo for bleeding peptic ulcer, *N Engl J Med* **336**:1054–1058, 1997.

5. D'amico G, Pagliaro L, Bosch J, The treatment of portal hypertension: a meta analytic review, *Hepatology* **22**:332–354, 1995.

6. Van Stiegmann GV, Goff S, Michaletz–Onody PA, *et al.*, Endsocopic sclerotherapy as compared with endoscopic ligation for bleeding esophageal varices, *N Engl J Med* **326**:1527–1532, 1992.

7. Sarin SK, Lamba GS, Kumar M, *et al.*, Comparison of endoscopic ligation and propranolol for the primary prevention of variceal bleeding, *N Engl J Med* **340**:988–993, 1999.

8. Lo GH, Lai KH, Cheng KS, *et al.*, Emergency band ligation versus sclerotherapy for the control of active bleeding from esophageal varices, *Hepatology* **25**:1101-1104, 1997.

9. Rossle M, Deibert P, Haag K, *et al.*, Randomized trial of transjugular-intrahepatic-portosystemic shunt versus endoscopy plus propranolol for prevention of variceal rebleeding, *Lancet* **349**:1043–1049, 1997.

10. Casado M, Bosch J, Garcia–Pagan JC, *et al.*, Clinical events after transjugular intrahepatic portosystemic shunt: correlation with hemodynamic findings, *Gastroenterology* **114**:1296–1303, 1998.

Early cancers

1. Tsukuma H, Oshima A, Narahara H, Morii T, Natural history of early gastric cancer: a non-concurrent, long term, follow up study, *Gut* **47**:618–621, 2000.

2. Yamao T, Shirao K, Ono H, Kondo H, Saito D, *et al.*, Risk of lymph node metastasis form intramucosal gastric carcinoma, *Cancer* **77**:602–606, 1996.

3. Ono H, Kondo H, Gotoda T, Shirao K, Yamaguchi H, Saito D, Hosokawa K, Shimoda T, Yoshida S, Endoscopic mucosal resection for treatment of early gastric cancer, *Gut* **48**:225–229, 2001.

4. Rembacken BJ, Gotoda T, Fuji T, Axon ATR, Endoscopic mucosal resection, *Endoscopy* **33**:709–718, 2001.

5. Rembacken BJ, Fuji T, Cairns A, Dixon MF, Yoshida S, Chalmers DM, Axon ATR, Flat and depressed colonic neoplasms: a prospective study of 1000 colonoscopies in the UK, *Lancet* **355**:1211–1214, 2000.

6. Whiting JL, Sigurdsson A, Rowlands DC, Hallissey MT, Fielding JWL, The long term results of endoscopic surveillance of pre-malignant gastric lesions, *Gut* **50**:378–381, 2002.

7. Shaheen NJ, Crosby MA, Bozymski EM, Sandler RS, Is there publication bias in the reporting of cancer risk in Barrett's esophagus? *Gastroenterology* **119**:333–339, 2000.

8. American Society for Gastrointestinal Endoscopy, The role of endoscopy in the surveillance of pre-malignant conditions of the upper gastrointestinal tract, *Gastrointest Endosc* **48**:663–668, 1998.

9. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, *et al.*, Reducing mortality from colorectal cancer by screening for fecal occult bleed. Minnesota Colon Cancer Control Study, *N Engl J Med* **328**:1365–1371, 1993.

10. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, *et al.*, Randomised controlled trial of fecal occult blood screening for colorectal cancer, *Lancet* **348**:1472–1477, 1996.

11. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O, Randomised study of screening for colorectal cancer with fecal occult blood test, *Lancet* **348**:1467–1471, 1996.

12. Lieberman DA, Weiss DG, Bond JH, Ahnen DJ, Garewal H, Chjfec G, Use of colonscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380, *N Engl J Med* **343**:162–168, 2000.

Endoscopic reterograde cholangio-pancreatography

1. Cotton PB, Evaluating ERCP is important but difficult, *Gut* **51**:287–289, 2002.

2. Boerma D, Rauws EAJ, Keulemans YCA, Janssen IMC, Bolwerk CJM, Timmer R, Boerma EJ, Obertop H, Huibregtse K, Gouma DJ, Wait-and-see policy or laparoscopic cholecystectomy after endoscopic sphincterotomy for bile-duct stones: a randomised trial, *Lancet* **360**:761–765, 2002.

3. Lai, *et al.*, Endoscopic biliary drainage for severe acute cholangitis, *N Engl J Med* **326**:1582–1586, 1992.

4. Davids, *et al.*, Randomized trial of self expanding metal stents versus polyethylene stents for distal malignant bile duct obstruction, *Lancet* **340**:1488–1492, 1992.

5. Freeman, *et al.*, Complications of endoscopic biliary sphincterotomy, *N Engl J Med* **335**:909–917, 1996.

6. Chang, *et al.*, Outcome in patients with bifurcation tumours who undergo unilateral versus bilateral hepatic duct drainage, *Gastrointest Endosc* **47**:354–362, 1998.

7. De Palma GD, Galloro G, Siciliano S, Iovino P, Catanzano C, Unilateral versus bilateral endoscopic hepatic duct drainage in patients with malignant hilar biliary obstruction: results of a prospective, randomized, and controlled study, *Gastrointest Endosc* **53**:547–553, 2001.

8. Loperfido S, *et al*., Major early complications from diagnostic and therapeutic ERCP: a prospective multicenter study, *Gastrointest Endosc* **48**:1–10, 1998.

9. Smith AC, *et al*., Randomised trial of endoscopic stenting versus surgical bypass in malignant low bile duct obstruction, *Lancet* **344**:1655–1659, 1994.

10. Bergman JJGH, *et al*., Long-term follow-up after biliary stent placement for postoperative bile duct stenosis, *Gastrointest Endosc* **54**:154–161, 2001.

11. Smits, *et al*., Endoscopic treatment of pancreatic stones in patients with chronic pancreatitis, *Gastrointest Endosc* **43**:556–560, 1996.

12. Stapfer M, *et al*., Management of duodenal perforation after endoscopic retrograde cholangiography and sphincterotomy, *Ann Surg* **232**:191–198, 1999.

13. Howard T, *et al*., Classification and management of perforations complicating endoscopic sphincterotomy, *Surgery* **126**:658–665, 1999.

Miscellaneous

1. Iddan G, Meron G, Glukhovsky A, Swain P, Wireless Capsule endoscopy, *Nature* **405**:417, 2000.

2. Yap CK, Ng HS, Cap-fitted gastroscopy improves visualization and targeting of lesions, *Gastrointest Endosc* **53**:93–95, 2001.

3. Filipi CJ, Lehman GA, Rothstein RI, Raijman I, Stiegman GV, Waring JP, Hunter JG, Gostout CJ, Edmundowicz SA, Dunne DP, Watson PA, Cornet DA, Transoral flexible endoscopic suturing for treatment of GERD: a multicenter trial, *Gastrointest Endosc* **53**:416–422, 2001.

4. Deviere J, Pastorelli A, Louis H, de Maertelaer V, Lehman G, Cicala M, Le Moine O, Silverman D, Costamagna G, Endoscopic implantation of a biopolymer in the lower esophageal sphincter for gastroesophageal reflux: a pilot study, *Gastrointest Endosc* **55**:335–341, 2002.

5. American Society for Gastrointestinal Endoscopy, The role of endoscopy in enteral feeding, *Gastrointest Endosc* **55**:794–797, 2002.

This page intentionally left blank

28

Practical Approaches to Liver Disease

Ng Han Seong

Patients with liver disease, depending on the etiology and the stage of the disease, have varied manifestations. They may range from rather seemingly trivial laboratory results (from a multiphasic screening), e.g. mildly elevated ALT or SAP; a slightly low platelet count; or vague symptoms of lethargy (as in chronic hepatitis B and C), to patients with obvious jaundice or enlarged liver. Moreover, an unwary doctor may be misled by the presence of jaundice to diagnose liver disease, not knowing that Gram-negative sepsis, hemolysis or congestive cardiac failure, may be the cause of the jaundice.

The liver has a central role in the body's metabolic and physiologic processes, e.g. synthesis of albumin, plasma proteins, clotting factors, cholesterol and lipoproteins; and maintenance of normal blood glucose. Ascites, ankle edema, bleeding tendency, somnolence and confusion, hypoglycemia, and high serum cholesterol can be the result of liver dysfunction, and these add to the varied manifestations of liver diseases.

LIVER DISEASES IN SINGAPORE

Knowing the spectrum of liver diseases in our local practice gives an additional advantage. However, one must be mindful that with the increasing numbers of foreigners working in Singapore, and that more people are traveling, "imported" cases of liver disease may be encountered.

Viral hepatitis is still the most common/important cause of acute and chronic hepatitis. One hundred and forty-three cases of acute viral hepatitis were reported in 2001, of which 55.9% was acute hepatitis B; 42%, acute hepatitis A and 2.1%, acute hepatitis E. The majority of the patients are in the 25–34 age group, with more males being affected, especially the Chinese. Chronic hepatitis B infection is also very common in Singapore, especially among the Chinese. The HBs carrier rate is about 5%, with predominance of males being mainly affected (M:F, 4:1) The majority of the hepatitis B infection was transmitted from mother to child at the time of birth (vertical transmission). Consequently, chronic hepatitis B is the major cause (80%) of liver cirrhosis and hepatocellular carcinoma in Singapore. Acute "flares" among chronic hepatitis B patients, as a result of seroconversion of HBe positive to anti-HBe positive, often present with "acute hepatitis," with anti-HBc IgM being tested negative in most cases. Hepatitis C is uncommon, and the "carrier rate" is way below 1%.

Alcoholic liver disease is a major problem in the West, but is less common in Singapore. It is more common among Indians in Singapore. Autoimmune hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis and Wilson's disease are well recognized liver diseases in Singapore, but are relatively uncommon compared with chronic hepatitis B. Hereditary hemochromatosis may be common in Northern Europe/Australia; it is extremely uncommon in Singapore. Fatty liver and non-alcoholic steatohepatitis (NASH) may be diagnosed more frequently as awareness increases. Its incidence is, however, not known.

Acute cholecystitis and cholangitis as a result of cholelithiasis is commonly seen in Singapore. "Oriental cholangiohepatitis," a condition where soft brown stones/sludge are formed primarily in the bile ducts, rather than in the gallbladder, is associated with repeated episodes of cholangitis. This condition is seen in older Chinese patients in Singapore, Hong Kong, China and Taiwan. Bile duct and gallbladder cancers are relatively uncommon. Liver abscesses associated with biliary stones are

Table 1 Common Liver Diseases in Singapore

Acute hepatitis B and A
Chronic hepatitis B
Liver cirrhosis and hepatocellular carcinoma
Obstructive jaundice from stones
Liver secondaries

more common than amoebic abscesses. The incidence of amoebic abscess has fallen, and is seen in people who had traveled to India and other parts of South-East Asia.

CLINICAL APPROACH

Most liver diseases can be diagnosed through a detailed history, thorough clinical examination, full blood count (FBC), prothrombin time (PT), liver function tests (LFTs), alpha fetoprotein (AFP), and viral markers — hepatitis A and B serology. Ultrasound examination of the hepatobiliary system is a very useful test, and often gives additional information regarding liver texture, focal lesions, liver and spleen size, bile ducts, stones in the gallbladder and bile ducts, ascites, pancreas and kidneys.

Additional blood tests such as autoimmune serological markers — anti-nuclear factor, smooth muscle antibody, anti-mitochondrial antibody, ceruloplasmin, liver biopsy and further imaging (CT, MRI, hepatic angiography) may be required to arrive at a diagnosis should one cannot be made with the initial assessment.

HISTORY

The age and perhaps sex of the patients are of diagnostic significance. Acute viral hepatitis is often diagnosed in the 25–34 age group. Males are more affected. HBs carriers and chronic hepatitis B infection are more common in males, and among the Chinese. Older patients, more than 50 years, are more likely to have obstructive jaundice from biliary stones, cancers, and drug-induced liver disease. Autoimmune hepatitis, primary biliary cirrhosis and gallbladder cancer are more prevalent in females.

Past history is also important. Past history of blood transfusion may lead one to suspect hepatitis B or C. Needle stick injury, tattooing, and intravenous drug abuse may also lead one to suspect hepatitis B or C. Past history of surgery to the liver/biliary system may alert the clinician regarding recurrence of liver cancer, retained stones, or bile duct strictures. Anesthetic agents may occasionally give rise to "hepatitis."

History of travel may lead one to suspect hepatitis A or E in a jaundice patient with "acute hepatitis". A family history of hepatitis B carriers, or the close contact with hepatitis B carriers, may lead one to the diagnosis of this condition in a patient with liver disease. Patients with a history of unprotected sex with prostitutes, multiple sex partners, and homosexuals are prone to hepatitis B.

Prodromal symptoms of nausea, anorexia, lethargy, fever followed by dark urine and jaundice, are strongly indicative of viral hepatitis. Smokers may develop an aversion to cigarettes. Pruritus, dark urine and pale stools are important symptoms of cholestasis. Pruritus may be present in the absence of jaundice. Pain in the right hypochondrium and epigastrium may lead one to suspect liver abscess or cancer; cholangitis, cholecystitis and pancreatic cancer. A tender liver may be present in drug-induced liver disease and alcoholic hepatitis. Infection should always be ruled out in patients with fever and chills.

Painless jaundice and weight loss may be ominous signs of carcinoma of the pancreas or secondaries in the liver. In hepatocellular carcinoma, jaundice is often associated with advanced disease. In young healthy adults, mild jaundice may be due to Gilbert's disease, a benign condition. Congestive cardiac failure, Gram-negative septicemia and hemolysis may present with jaundice.

A history of easy bruising, gastrointestinal bleeding, abdominal swelling (ascites) and ankle edema, sleep inversion, confusion and somnolence are indicative of portal hypertension, impaired liver function and hepatic encephalopathy.

A detailed drug history is mandatory, especially in older patients, and those who are on multiple drugs. One should not forget traditional/herbal medication, as it is quite common for patients to be on "Western" as well as "Traditional" medication at the same time. Drug- induced liver disease can mimic a whole spectrum of liver disease, and in some, "autoimmune serological markers" may be positive. Alcohol history should always be elicited, and unless the patients are questioned, this

part of the history may not be volunteered. An estimate of the amount taken per day or week should be made.

Occupational history such as exposure to aromatic chemicals should be documented. They may give rise to ALT elevations, and "fatty liver." While Wilson's and other hereditary liver diseases are uncommon in Singapore, a family history of liver disease or jaundice should be noted. In Singapore, a family history of hepatitis B is an important piece of information in the management of liver disease.

PHYSICAL EXAMINATION

General Examination

Jaundice is an important sign of liver disease. It should be looked for in the sclera and is detectable at a serum bilirubin level of 2 mg%. The depth of jaundice may provide a clue to the underlying cause. Hemolysis produces mild jaundice (less than 5 mg%), whereas in cholestatic jaundice, a greenish hue may be present in addition to the deep jaundice. Hepatocellular jaundice is often moderate, but can be deep in long-standing cases. Associated pallor may indicate hemolysis, malignancy, or a recent episode of bleeding from varices.

Signs of chronic liver disease include palmar erythema, spider naevi, clubbing, white nails, skin pigmentation, bruising, gynecomastia and testicular atrophy in males. Detection of supraclavicular lymphadenopathy, wasting and ascites may indicate malignancy in the liver or pancreas. In alcoholic patients, parotid enlargement and Dupuytren's contracture may be present. Gynecomastia, testicular atrophy, spider naevi and palmar erythema are especially prominent in alcoholics.

Pruritus, scratch marks, finger clubbing and xanthomata are usually associated with chronic cholestasis, and are classically seen in primary biliary cirrhosis. Dry eyes and mouth, as part of Sjogren's syndrome, are associated with primary biliary cirrhosis. Hemochromatosis may have increased skin pigmentation (bronze diabetes) due to increased melanin deposits.

Wilson's disease may present with hemolysis. Examination of the eyes with a slit lamp may reveal Kayser Fleischer ring in the cornea. The patient may have extrapyramidal signs.

Evaluation of the mental state and neurological function is particularly important with regard to hepatic encephalopathy. Change in

personality, mild confusional state and lethargy are subtle signs of hepatic encephalopathy. In more advanced cases, seizures, abnormal posturing, and lateralizing signs may be present. A flapping tremor (asterixis) may also be present.

Abdominal examination

Abdominal examination may reveal dilated periumbilical veins. A small liver, splenomegaly and ascites would suggest liver cirrhosis and portal hypertension. A large, hard and nodular liver suggests the presence of hepatocellular carcinoma or metastatic disease, and is especially so, if there is a bruit or rub. A small liver may indicate massive hepatic necrosis from severe hepatitis. A tender and enlarged liver is seen in congestion from cardiac failure, alcoholic hepatitis, abscess or fatty infiltration.

A palpable and enlarged gallbladder (Courvoisier's sign) would indicate extrahepatic biliary obstruction from the cancer head of the pancreas or common duct obstruction. Tenderness over the right hypochondrium in the gallbladder area (Murphy's sign) may indicate an inflammed gallbladder, as in acute cholecystitis.

INVESTIGATIONS

Full blood count and prothrombin time

Full blood count (FBC) can be a useful test in liver disease. In most cases of acute liver disease, the FBC is often normal. Slight leucopenia and atypical lymphocytes may be seen in acute viral hepatitis. Leucocytosis is uncommon in acute viral hepatitis, but may be present in alcoholic hepatitis, drug-induced hepatitis (eosinophilia may be present) and in severe cases of viral hepatitis. Bacterial coinfection often gives rise to high leucocyte count.

Hemolysis may be seen in Wilson's disease and in alcoholic hepatitis with hypertriglyceridemia (Zieve's syndrome). Target cells and acanthocytes may be seen in cirrhotics and alcoholic cirrhosis. Macrocytosis may indicate underlying alcoholic liver disease or cirrhosis. Aplastic anemia occurs infrequently in viral hepatitis B and C. Anemia in chronic liver disease may be due to bleeding from varices/portal hypertensive gastropathy, or from poor nutrition.

In cirrhotics, hypersplenism resulted in decreased red cells, white cells and platelets. A slightly low platelet count in a patient with chronic liver disease may be a sign of underlying cirrhosis.

Prothrombin together with other clotting factors (II, V, IX, X, XI, XII, XIII) and fibrinogen are synthesized in the liver. Synthesis of clotting factors II, VII, IX, X depends on the supply of vitamin K. Prothrombin time (PT) measures the rate of which prothrombin in citrated plasma is converted to thrombin in the presence of calcium, tissue thromboplastin and activated clotting factors. It is usually expressed in seconds or percentage of a standardized control. The test is abnormal in liver failure, vitamin K deficiency and in DIVC. A single injection of vitamin K 10 mg will correct the PT in vitamin K deficiency, a condition seen in cholestasis associated with bile duct obstruction.

Liver function tests

Liver function test refers to the measurement of serum total protein, albumin, bilirubin and hepatic enzymes (ALT, AST, SAP). When deranged, it indicates liver injury or dysfunction. LFT is usually carried out when one suspects a liver disease, and in those with liver disease, to help in further evaluation. Elevated ALT/AST would suggest "hepatitis," while elevated SAP with or without elevated serum bilirubin would suggest "Cholestasis." The clinician can then plan further tests — serology or imaging — to make a diagnosis, along with history and physical findings as a guide.

Total protein and albumin

Serum albumin and prothrombin time reflect synthetic function of the liver. A low albumin level is associated with poor synthetic function (exclude renal loss). It can, however, be low in severe infection of the liver, e.g. liver abscess. Serum globulin may be elevated in chronic liver disease and autoimmune hepatitis. IgM is elevated in primary biliary cirrhosis, IgG in autoimmune hepatitis, and IgA in alcoholic liver disease.

Aminotransferase (ALT & AST)

ALT (alanine aminotransferase) and AST (aspartate aminotransferase) are the two usually measured aminotransferase. The levels are associated

with liver injury/necrosis. ALT is found mainly in cytosol, and AST in mitochondria and cytosol. AST may be elevated in muscle/cardiac disease. Drugs and alcohol, which may affect the mitochondria more, AST elevation may be more than ALT. A AST/ALT ratio of more than 2 is said to be characteristic of alcoholic liver disease.

ALT and AST are markedly elevated in necroinflammatory liver disease — acute viral hepatitis, drug-induced liver disease (especially in paracetamol overdose), ischemic liver (patient in shock) and Amanita mushroom poisoning. Moderate elevation may be seen in chronic hepatitis B and C, autoimmune hepatitis. Fatty liver and obstructive jaundice are often associated with mildly elevated ALT/AST (less than 5 times). Primary biliary cirrhosis may be associated with mildly elevated ALT/AST. AST may also be elevated during hemolysis e.g. G6PD Deficiency.

Alkaline phosphatase

Serum alkaline phosphatase (SAP) is not a single enzyme, but rather a family of isoenzymes, and can come from the hepatobiliary system, bones, intestines, placenta and lung cancer. If SAP is elevated without obvious liver disease, and the other components of LFT being normal, gamma-glutamyl transferase should be checked. If elevated as well, the SAP would be of a hepatobiliary origin. If not, the SAP could have come

Table 2 Causes of Elevated Transaminases (ALT/AST)

Markedly elevated ALT (> 1000 U/I)	• Paracetamol overdose • Ischemic hepatitis • Acute viral hepatitis • Chronic hepatitis B undergoing HBe seroconversion ("flares") • Amanita mushroom poisoning
Moderately raised ALT (5–20 ULN)	• Chronic viral hepatitis B and C • Autoimmune hepatitis • Drug-induced hepatitis
Mildly elevated ALT (< 5 ULN)	• Fatty liver • Chronic viral hepatitis B and C • Cholestatic liver diseases • Liver cirrhosis (usually <3 ULN)

from the bone, kidneys, intestines and lung cancer. One must not forget that children at puberty and pregnant woman have raised SAP!

SAP is increased in cholestasis. Radiological imaging is needed to exclude obstruction in the biliary system. In "non-obstructive" cholestasis, e.g. primary biliary cirrhosis or drug-induced liver disease, serological testing for anti-mitochondrial antibody and liver biopsy would be useful to confirm diagnosis. SAP is also elevated in liver abscess, granulomatous disease (TB, Sarcoidosis) and in hepatocellular carcinoma or metastatic disease. Imaging is again useful in making a diagnosis, and liver biopsy is needed to diagnose granulomas. Mild elevation of SAP occurs in congestive cardiac failure, liver cirrhosis and fatty liver.

Bilirubin

Bilirubin is made up of unconjugated and conjugated bilirubin, with the unconjugated component making up 70% of total serum bilirubin. Serum bilirubin more than 2 mg% is detectable clinically as jaundice in the sclera. Unconjugated hyperbilirubinemia is seen in hemolysis and Gilbert's syndrome. It is important to check on the hemoglobin and reticulocyte count, and look for fragmented red cells in the peripheral blood, if one suspects hemolysis.

Conjugated hyperbilirubinemia is seen in cholestasis. It is often elevated together with SAP. In hepatitis, serum bilirubin would be elevated along with ALT/AST. SAP may be mildly elevated.

Child–Pugh classification

Child and Turcotte in 1964 devised a prognostic index for chronic liver disease based on 5 variables — serum bilirubin, serum albumin, ascites, encephalopathy and nutritional status. It was used to assess the risk of surgery in patients with liver disease. A scoring system was used to modify the grading, an example being the Pugh's modification (see Table 3). Prothrombin time was used instead of nutritional status. The Pugh's score had been useful for assessing prognosis in cirrhosis, risk of general anesthesia and operative mortality, and patients going for surgery and shunt operation. Patients who score 5–6 are considered Grade A; scores of 7–9, Grade B; and 10–15, Grade C. Patients in Grade C have high morbidity and mortality following surgery and poor 5-year survival rate.

Table 3 Points Scored for Increasing Abnormality

Clinical and Biochemical Measurements	1	2	3
Ascites	Absent	Slight	Moderate/severe
Encephalopathy	None	1 and 2	3 and 4
Bilirubin (mg/100 ml)	1–2	2–3	>3
Albumin (g/L)	>35	28–35	<28
Prothrombin (is prolonged)	1–4	4–6	>6

Total numerical score	Child–Pugh class
5–6	A
7–9	B
10–15	C

Viral hepatitis serology

Hepatitis A and B are the two common viral hepatitis seen. Hepatitis C is uncommon. As treatments for chronic hepatitis B and C are now available, a proper understanding of the viral serology is important.

Acute hepatitis A is diagnosed by detecting anti-HAV IgM, while anti-HAV IgG indicates immunity to hepatitis A (a test to do before hepatitis A vaccination).

Acute hepatitis B is diagnosed by a positive anti-HBc IgM. HBs alone is not adequate to diagnose acute hepatitis B as the HBs carrier rate is high in Singapore. HBs carriers (HBe +ve) are prone to "flares" as a result of HBe seroconversion. It may be mistaken for acute hepatitis B as HBs is positive, but anti-IgM is negative (rarely anti-HBc IgM can be positive). HBe and anti-HBe are often tested in chronic hepatitis B, as a positive HBe would indicate active viral replication. A better test would be serum HBV DNA, and its presence would certainly indicate viral replication. It is a test one would do prior to consideration for antiviral treatment of chromic hepatitis B.

Hepatitis C is diagnosed by anti-HCV, and confirmed by demonstration of HCV RNA in serum. It is important to confirm with HCV RNA, as the anti-HCV may be false-positive. It is also important to note that it may take a few weeks after acute infection for the anti-HCV to be positive.

Other viral serology may be tested if markers for viral A, B and C are negative. Acute hepatitis E can be diagnosed with anti-HEV IgM. Serology testing for EBV and CMV are also available. Delta infection is

very uncommon, and can be tested with anti-HDV IgM (for acute infection) and anti-HDV IgG.

Autoimmune markers and others

Serum IgG are elevated in chronic liver disease and in autoimmune hepatitis (AIH). Anti-nuclear factor (ANF) and smooth muscle antibody (SMA) and anti-LKM antibody are tests for diagnosis of AIH.

Anti-mitochondrial antibody (AMA) is positive in most cases of primary biliary cirrhosis. Serum copper and ceruloplasmin are low and 24-hour urinary excretion of copper is high in Wilson's disease. Serum ferritin level is markedly elevated in hemochromatosis.

Alpha fetoprotein (1–10 ng/ml) is frequently used in the screening and diagnosis of hepatocellular carcinoma. A high titer (more than 400 ng/ml) or a rising titer strongly suggests the presence of hepatocellular carcinoma. The diagnosis requires imaging, e.g. ultrasound of the abdomen or CT scan for confirmation. A low titer or fluctuating titer may be due to underlying inflammatory activity. If one is in doubt, an US examination of the liver should be carried out.

Radiology

Abdominal X-ray is of limited use. It may show up calcifications in the liver and gallstones. **Barium swallow/meal** is not as sensitive as endoscopy for the detection of esophageal and gastric varices.

Abdominal ultrasound is often the radiological study of choice. It is relatively inexpensive, with no radiation involved and is portable. It can detect focal lesions (solid or cystic), and facilitates liver biopsy, drainage of liver abscesses and localized ascites. It is also able to show changes in the texture of the liver associated with liver cirrhosis and fatty liver. It can also detect gallstones and dilated bile ducts, and is a good test in the evaluation of patients with cholestasis. With the Doppler technique, it can be used to study hepatic and portal blood flow, and aid in the diagnosis of portal vein thrombosis and Budd Chiari syndrome.

Helical CT scanning with IV contrast is more accurate compared with US abdomen in characterizing hepatic masses, especially if the liver is cirrhotic and nodular. It is, however, more costly, involves radiation and is not portable. CT after hepatic angiogram can detect lipoidol which

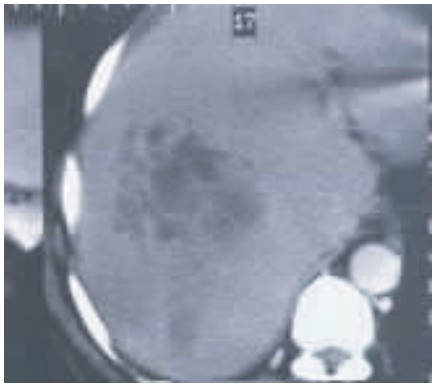


Fig. 1 CT Scan — liver abscess.

is preferentially taken up and retained by hepatocellular carcinoma. It can detect small hepatocellular carcinoma (HCC) and multicentric small HCC. It is also useful in the diagnosis of liver abscesses and cysts.

Magnetic resonance imaging (MRI) is excellent for evaluating blood flow and confirming vascular lesions, e.g. hemagioma and in differentiating regenerating nodules from HCC. **Magnetic resonance cholangiopancreatography (MRCP)** is a new, non-invasive imaging modality for bile ducts and pancreas, and provides an alternative to diagnostic ERCP.

Hepatic angiogram continues to play an important role in the diagnosis and evaluation of focal lesions, metastatic endocrine tumors and HCC. However, improvement in non-invasive imaging techniques have diminish angiographic's role in diagnosis. Nonetheless, it has remained an important option in the treatment of unresectable HCC. With superselective catherization, an interventional radiologist can perform chemotherapy/embolization of tumors, and embolotherapy for AV fistula and bleeding from traumatic injuries/liver biopsy.

Percutaneous transhepatic cholangiography (PTC), an invasive test, offers an alternative diagnostic and therapeutic procedure to ERCP in obstructive jaundice (dilated bile ducts) in whom endoscopic drainage has failed or is not possible, e.g. tight biliary structure, impacted stones, or after gastrectomy and Billroth II anastomosis. Marked ascites and irreversible coagulopathy are contraindications to this procedure.

Radioisotope scanning may be helpful in acute cholecystitis, and HIDA scanning is positive in most cases of acute acalculous cholecystitis. However, US abdomen and CT scan have largely superceded

radioisotope scanning in the diagnosis of biliary disease and focal/parenchymal liver disease.

Endoscopy

Upper GI endoscopy remains the best diagnostic test for detecting esophageal and gastric varices, and portal hypertensive gastropathy. It also provides a therapy to bleeding varices; esophageal varices can be ligated/injected, and gastric varices injected.

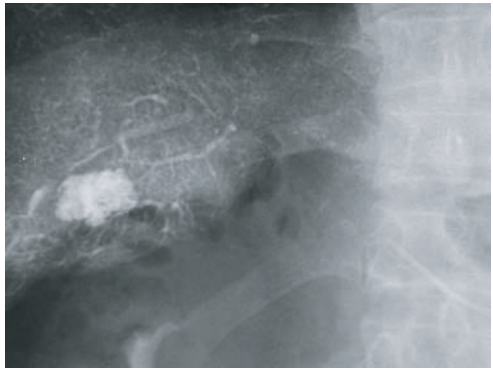


Fig. 2 Hepatic angiogram — hepatocellular carcinoma with lipoidol.

Table 4 Common Liver Diseases and Relevant Investigations

Viral Hepatitis
Hepatitis serology:
Anti-HAV IgM, HBs, Anti-HBc IgM
HBe, Anti-HBe, HBV DNA (chronic hepatitis B)
Anti-HCV (confirmed with HCV RNA)
Anti-HEV IgM, Anti-EBV IgM, Anti-CMV IgM
Chronic Liver Disease/ Cirrhosis
Hepatitis B and C serology
Autoimmune serology (ANF, SMA, AMA)
Alpha fetoprotein
Serum, urine and liver copper studies, serum ceruloplasmin
Ferritin and Iron studies
Ultrasonography
Upper GI Endoscopy (for varices)
Liver biopsy

Patients with liver cirrhosis should receive upper GI endoscopy to check for varices. Those with varices should receive propranolol as a primary prophylaxis against variceal bleeding. For those whose liver cirrhosis is not clinically evident, a platelet count of less than 100 000 may help to select patients for endoscopy and surveillance for varices.

Endoscopic retrograde cholangiopancreatography (ERCP) allows the visualization of the biliary tree and pancreatic duct. It has a definite role in the diagnosis and treatment of biliary diseases — diagnosis of primary sclerosing cholangitis and biliary stones; and treatment of strictures and biliary leak with stents; and removal of biliary stones. It is an invasive test and involve a small risk. Its diagnostic role is presently being challenged by **MRI** and **endoscopic ultrasonography.**

Liver biopsy

Liver biopsy is an invasive test, and carries with it a small risk of bleeding. It is performed for the purpose of diagnosis, e.g. fatty liver/NASH, primary biliary cirrhosis; assessment of severity of disease/inflammation, e.g. drug-induced liver disease, chronic viral hepatitis; and for prognosis, e.g. chronic viral hepatitis/cirrhosis. Liver biopsy is also performed to help make a diagnosis in medical problems such as fever of unknown origin (PUO) and in infiltrative disorders; to check on progress and effect of treatment, e.g. chronic viral hepatitis; and in liver transplanted patients to diagnose/exclude rejection, ischemia or infection.

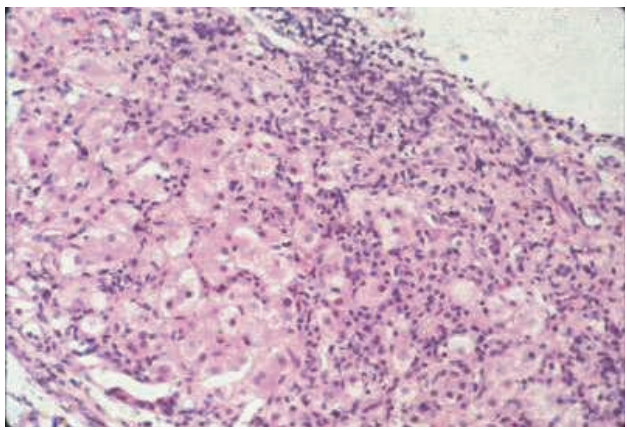


Fig. 3 Liver biopsy — chronic hepatitis B with periportal inflammation.

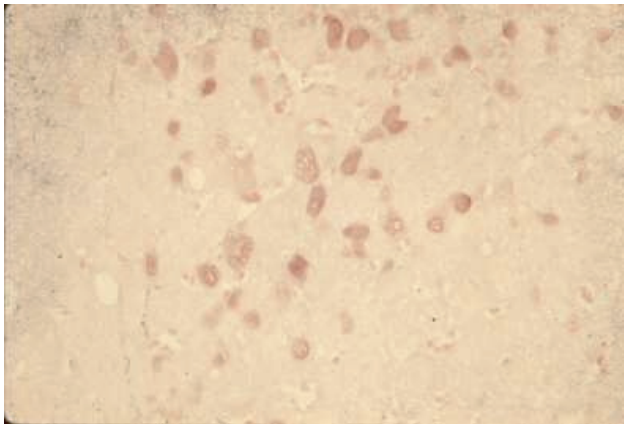


Fig. 4 Liver biopsy — chronic hepatitis B, orcein positive.

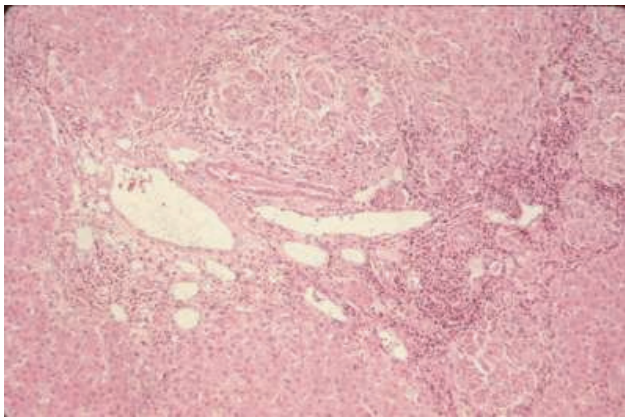


Fig. 5 Liver biopsy — primary biliary cirrhosis.

The procedure may be performed by the "blind" technique, under radiological guidance (ultrasound or CT), transjugularly or during laparoscopy/surgery. For localized disease, e.g. space occupying lesion, liver biopsy is carried out under radiological guidance. Transjugular approach is carried out in patients with coagulopathy.

Liver biopsy can be carried out safely in patients with a normal PT and a platelet count of 100 000 or more. In some institutions, a lower platelet count is acceptable. Liver biopsy is contraindicated in patients who are

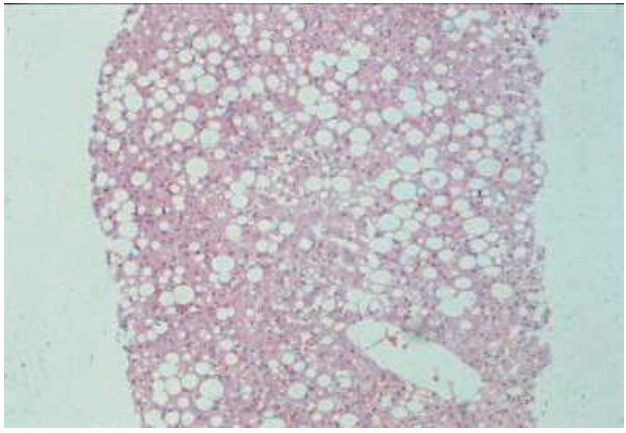


Fig. 6 Liver biopsy — fatty liver.

uncooperative; and in patients with coagulopathy, suspected hydatid disease, biliary obstruction and moderate/severe ascites.

Various grading scores (HAI score, Metavir score) are used by histopathologists to score inflammation and fibrosis in chronic hepatitis B and C. Scoring are carried out in clinical drug trials in chronic hepatitis B and C to document severity before treatment and improvement after treatment.

REFERENCES

1. *Communicable Disease Surveillance in Singapore 2001*, Quarantine and Epidemiology Department, Ministry of Environment.
2. Guan R, Kang JY, Ng HS, Merican I, *Management of Common Gastroenterological Problems — A Malaysia & Singapore Perspective*, Havas MediMedia Asia Pte Ltd, 2000.
3. McIntyre N, Benhamou J-P, Bircher J, Rizzetto M, Rodes J, *Oxford Textbook of Clinical Hepatology*, Oxford Medical Publications, 1991.
4. Zakim, Boyer, *Hepatology — A Textbook of Liver Disease*, W.B. Saunders, 1996.

29

Viral Hepatitis

Chow Wan Cheng and Ng Han Seong

INTRODUCTION

Viral hepatitis, an infection caused by a hepatotropic virus, is associated with diffuse inflammation of the liver. However, its clinical manifestation may range from an asymptomatic, subclinical, anicteric disease to an overt but self-limiting illness with non-specific systemic signs and symptoms of a viral illness and clinical features related to derangement of hepatic function, to a rapidly fatal disease. Depending on the causal viral agent, chronic hepatitis and liver cirrhosis may develop in some infected individuals.

There are at least five established etiologic agents namely, hepatitis A, B, C, D and E viruses. Selected characteristics of the five viruses are summarized in Table 1.

HEPATITIS A VIRUS

Hepatitis A virus (HAV) is a small, non-enveloped RNA picorna virus. The different strains of HAV are all of 1 serotype, making diagnosis relatively easy. The virus, spread mainly by the orofecal route, may cause sporadic infection as well as epidemic. The incubation period is 15–50 days. In

Table 1 Characteristics of Hepatitis Viruses

	HAV	HBV	HCV	HDV	HEV	HGV
Viral Type	**RNA Picorna-virus**	**DNA Hepadna-virus**	**RNA Flavi-virus**	**RNA Viriod**	**RNA Calci-virus**	**RNA Flavi-virus**
Predominant Mode of Transmission	OF	P, Sexual	P	P, Sexual	OF	P
Incubation Period (Days)	15–50	30–180	30–120	30–180	15–60	30–120
Fulminant Hepatitis	Yes	Yes	rare	Yes	Yes	?
Chronic Infection	No	Yes	Yes	Yes	No	Yes
Vaccine	Yes	Yes	No	HBV Vaccine	No	No
Passive Immunization	Yes	Yes	No	No	No	No

OF = orofaecal; P = parenteral.
? contradictory evidence.

endemic areas, it affects mainly children and young adults. Whereas the disease is usually mild and frequently anicteric in the pediatric age group, a more serious illness, including fulminant hepatitis with fatal outcome, may occur in the elderly. However, classical presentation of acute hepatitis with complete recovery and immunity to reinfection is expected in most cases.

Several unusual clinical manifestations of HAV infection have been described. In addition to a prolonged cholestatic disease, some patients may experience relapsing hepatitis A, usually 2–3 months after the initial illness. Complete recovery is the rule in these cases. HAV is also known to trigger autoimmune hepatitis in susceptible individuals.[1]

Active prophylaxis is now available with formaldehyde inactivated HAV vaccine. It is safe and effective and is recommended for travelers to endemic areas, as well as sewage workers and workers in high-risk occupations.[2] With improved living conditions and sanitation, almost all young people (age less than 20 years) in Singapore were tested negative for antibody against HAV.[3] Hence hepatitis A vaccination is recommended for patients with chronic liver disease (in view of the potentially severe clinical manifestations in this group of patients) and frequent travelers to endemic areas.

HEPATITIS B VIRUS

Hepatitis B virus (HBV) is a DNA virus, belonging to the Hepadnaviridae family. It has a circular, partially double-stranded DNA that encodes 4 major gene products: the surface antigen (HBsAg), X protein (HBx), HBV DNA polymerase and nucleocapsid proteins (HBcAg and HBeAg). HBV is transmitted parenterally, with sexual transmission and intravenous drug abuse being the predominant routes of transmission in the West. The incubation period of the infection is 30–180 days. In Asia, however, perinatal vertical transmission from HBV-infected mothers (especially those who are HBeAg positive) is probably the most important mode of HBV transmission. Once infected, more than 90% of the infected newborns will develop chronic HBV infection. While the infected newborns are usually asymptomatic till much later in life, newly infected adults usually present with signs and symptoms of acute hepatitis, with 1% or less of them progressing to develop fulminant hepatitis. The majority of these newly infected adults will recover fully from the episode and become immuned to HBV, with development of anti-HBs and anti-HBc antibodies. Only 1–3% of the adult-acquired HBV infection will progress to chronic HBV infection.

Due to selective pressure from the host's immune system, HBV mutant strains emerge in patients with chronic HBV infection. A single amino acid substitution from glycine to arginine at position 145 in the HBs gene causes a structural change of the HBs antigen, allowing the variant to escape neutralization by anti-HBs antibody. This has been described in infants born to HBV-infected mothers. The child will have chronic HBV infection despite having received hepatitis B vaccination at birth.[4] Hence, it is also known as a HBV escape mutant. Patients who receive regular monoclonal anti-HBs immunoglobulin infusion following liver transplantation are also prone to developing such a mutant.

Other clinically important HBV variants are the HBe-minus variants. They are commonly found in the Mediterranean countries and in Asia. Classically, this refers to HBV that has undergone G(glycine) to A(arginine) point mutation in the precore gene at position 1896. This results in the formation of a stop codon, thus interrupting the synthesis of HBeAg.[5] The resulting absence of HBeAg expression allows the HBV variant to escape from the HBeAg-directed host immune response, and thus viral survival is preserved. Since the description of this classical precore mutant, nucleotide mutations at other sites of the precore as well as the core

promoter genes giving rise to similar phenotypic manifestation have been described.[6,7] Fulminant hepatitis and a more severe form of chronic hepatitis B were reported in some patients infected with HBe-minus HBV variants.

The YMDD variant is another more recently described HBV variant. Mutations at the YMDD motif of the HBV polymerase gene were found in chronic hepatitis B patients who were treated for more than 6 months with lamivudine, a nucleoside analog. Its emergence is suspected when there is recurrence of HBV viremia while the patient is on lamivudine, and compliant to treatment, suggesting resistance to treatment.[8]

An effective and safe vaccine against HBV is available. It is manufactured using recombinant technique, where HBs antigen is synthesized in yeast cells. Many countries are now implementing universal vaccination against HBV. The immunization program has been very successful in Taiwan, Singapore and Malaysia. The overall prevalence of HBV infection in Singapore has dropped from 8% to 4% and, in our latest survey, practically all primary school children in Singapore are immuned to HBV.[9]

HEPATITIS C VIRUS

Hepatitis C virus (HCV) is a small, enveloped cytopathic RNA virus, belonging to the Flaviviridae family. The HCV genome has a high degree of heterogeneity. Phylogenetically, HCV can be classified into at least six major HCV genotypes and 20 subtypes. While each infected patient usually harbors only one HCV genotype, there are usually minor variations (quasispecies) within the viral population in every patient. The HCV genotype with which the patient is infected has prognostic significance as disease severity and response to treatment differ between different genotypes. For example, genotype 1b is associated with a more severe disease and poorer response to interferon treatment. Several encoded viral peptides (e.g. C22, C33, and C100) have been synthesized, using recombinant technique, and are used in current diagnostic assays. Using the ELISA technique, we can screen for ongoing HCV infection by demonstrating the presence of a combination of antibodies (anti-HCV IgG) against these viral epitopes. False positive test results may arise from screening low-risk population, e.g. healthy blood donors. Recombinant immunoblot assays, e.g. RIBA-3 are used as supplementary tests for HCV infection. Ultimately, HCV infection is confirmed by demonstration of hepatitis C viremia (HCV RNA), using polymerase chain reaction.

HCV is prevalent throughout the world, with 0.5–3.0% seroprevalence rate among healthy blood donors. In Singapore, HCV infection is relatively uncommon. Only 0.3% of our blood donors are anti-HCV IgG positive. Although sporadic cases do occur, its transmission is primarily parenteral. Intravenous drug users and those who receive multiple blood/ blood product transfusion are at high risk. Transmission through sexual contact, as opposed to HBV, is rare. Except for infants born to mothers with HCV and HIV co-infection, perinatal vertical transmission of HCV is uncommon.

The incubation period of HCV infection is 30–120 days. Acute hepatitis C is usually subclinical and fulminant hepatitis is uncommon. More than 70% of the infected patients progress to developing chronic hepatitis C, of whom, 20% will developing cirrhosis and hepatocellular carcinoma.

HEPATITIS D VIRUS

Hepatitis D virus (HDV) is a defective virus that requires HBV for its replication. It has a small genome, consisting of a single-stranded circular RNA. The viral nucleocapsid is coated by an envelope of HBs antigen. HDV is endemic in the Mediterranean countries, Middle East and parts of South America. It is uncommon in Singapore and South-East Asia. Immunity against HBV will protect against HDV infection.

HDV can give rise to acute as well as chronic infection. Two situations can occur:

1) Co-infection: Both HBV and HDV are acquired at the same time. The infection is usually self-limiting, as HDV suppresses HBV replication. Chronicity after co-infection occurs in less than 10% of patients.
2) Super-infection refers to HDV infection occurring in a patient who already has chronic HBV infection. In the presence of established hepatitis B viremia, HDV replicates rapidly. Super-infection is more likely to result in fulminant hepatitis, and if the patient survives, usually leads to chronic HDV infection.

HEPATITIS E VIRUS

Hepatitis E virus (HEV) is a small spherical non-enveloped RNA virus, of the Calciviridae family. It is a major cause of sporadic and epidemic

hepatitis in India, Pakistan, Myanmar, Central Asia, Mexico and Russia. The virus is transmitted by the orofecal route, mainly through contaminated drinking water. The incubation period is 15 to 60 days. Sporadic cases are reported in Singapore, at least half of them being found in travelers returning from endemic areas.[10]

Although there may be prolonged cholestasis, it usually causes a self-limiting illness. Fulminant hepatitis tends to occur in gravid women, especially those in their third trimester of pregnancy, with mortality up to 20%.

OTHER IMPLICATED ETIOLOGIC VIRAL AGENTS

Recent additions to the list of etiologic agents that may be responsible for some of the non-A–E hepatitis include hepatitis G virus (HGV), transfusion-transmitted virus (TTV) and SEN virus (SENV). All of these viruses are parenterally transmitted and are capable of persisting in their new host. However, evidence suggests that they do not account for most of the non-A–E acute fulminant hepatitis and "cryptogenic" liver cirrhosis.

ACUTE VIRAL HEPATITIS

Acute viral hepatitis in humans may be caused by at least any of the above mentioned five hepatitides viruses, namely HAV, HBV, HCV, HDV and HEV. The vast majority of acute viral hepatitis in Singapore are caused by HAV and HBV (see Table 2). Acute hepatitis C and delta are uncommon. Sporadic cases of acute hepatitis E has been reported; many were travelers returning from neighboring countries.

Clinical features

The clinical features of acute viral hepatitis, regardless of the causal viral agent, are similar. The course of the clinical illness can be divided into three clinical stages.

Prodromal phase

More than 90% of acute infections are either asymptomatic or associated with non-specific symptoms. The earliest symptoms are nausea, vomiting,

Table 2 Acute Viral Hepatitis in Singapore

Year	Incidence	Acute Hepatitis
1991	501	281 A, 200 B, 16 C, 4 others
1993	317	173 A, 133 B, 11 E
1995	267	131 A, 135 B, 1 E
1997	345	149 A, 179 B, 17 E

Epidemiology Department, Ministry of Environment, Singapore.[11]

anorexia and/or diarrhea. There is often malaise, fatigue, and a low grade fever. A serum sickness-like picture (rash, arthralgia, headache, and fever) may appear during this period. Clinical diagnosis may not be possible without biochemical (liver function tests) and serologic tests. Asymptomatic course is very common in HAV infection in children and in HCV infection in adults.

Icteric phase

Dark urine is usually noted a few days prior to onset of obvious jaundice. Constitutional symptoms often improve with the onset of jaundice. Moderate weight loss may occur as a result of anorexia. Pruritus may occur but passage of pale stools is rare.

Physical examination at this stage may reveal a mildly enlarged, possibly slightly tender, liver. A mildly enlarged spleen is palpable in up to 20% of patients. Severe jaundice, ascites and ankle edema may occur as a complication of severe acute hepatitis, which may progress to fulminant hepatic failure. However, this is rare and if present, consider acute-on-chronic hepatitis. Especially in hepatitis B endemic areas, acute exacerbation of underlying chronic hepatitis B may mimic acute hepatitis.

Acute fulminant hepatitis occurs in less than one percent of acute viral hepatitis. Mortality can be high (50–90%), especially in older patients. Of the five hepatitis viruses, HBV and HDV are well known to cause fulminant hepatitis. Fulminant hepatitis A is seen in elderly patients, whereas fulminant hepatitis E occurs in pregnant women. HCV, on the other hand, rarely causes fulminant hepatitis.

Prolonged and severe jaundice, despite improvement of the rest of the clinical parameters, may develop in a small proportion of patients. This is well described in elderly patients infected with HAV and patients

with acute hepatitis E. Chinese patients with glucose-6-phosphate dehydrogenase deficiency may have exacerbation of hemolysis concomitantly resulting in marked jaundice.

Convalescence

During convalescence, liver function test progressively normalizes but easy fatigability may persist for many more weeks, with full recovery occurring in 6 to 12-week time.

Recovery is complete after resolution of acute hepatitis A and E. However, persistent infection may occur with HBV (1–3% of infected adults) and HCV (80% of infected adults) infection. Hence, patients with acute hepatitis B and C need to be followed up for at least six months to ensure that they are free from chronic hepatitis virus infection.

Diagnosis

Simple laboratory tests

The serum alanine aminotransferase (ALT) is markedly elevated, and may reach peak values of more than 50–100 times normal. Serum ALT level is usually relatively higher than serum aspartate aminotransferase (AST). This is accompanied by elevation of serum bilirubin. Serum albumin and prothrombin time (PT) are usually normal at presentation. Acute severe hepatitis should be suspected if the prothrombin time becomes prolonged. Therapeutic trial of vitamin K supplement may help to distinguish prolongation of prothrombin time due to liver failure and that due to vitamin K malabsorption associated with cholestasis *per se*.

Full blood count is usually normal. Leukopenia with neutropenia and lymphopenia may be seen in the early phase of the infection, followed by and a short-lived lymphocytotic phase. There may be atypical mononuclear cells but thrombocytopenia is unusual and if present, pre-existing chronic liver disease should be considered.

Serologic tests

Serologic tests are diagnostic tests to determine the etiology of acute viral hepatitis. Anti-HAV IgM, diagnostic of acute hepatitis A, can remain positive for up to six months post-infection, whereas anti-HAV IgG,

indicative of immunity against HAV, may remain positive for decades. This may be acquired via previous viral exposure or active immunization against HAV.

Hepatitis B surface antigen (HBsAg) is often the first serologic marker which appears during acute hepatitis B. However, especially in areas with a high prevalence of chronic HBV infection, only the presence of IgM antibody to the HBV core antigen (anti-HBc IgM) is diagnostic of acute hepatitis B.

HCV infection is usually diagnosed by the detection of anti-HCV IgG antibody. Its presence cannot distinguish acute from chronic HCV infection. With the third generation EIA assays, seropositivity occurs two weeks after exposure to the virus. Assay for anti-HCV IgM is available but is seldom used due to its low sensitivity. Moreover, anti-HCV IgM is not detected much earlier than anti-HCV IgG during an acute infection.[12]

Anti-HDV IgM is positive in acute hepatitis delta. However, patients with chronic infection may also have high titer of anti-HDV IgM. With the exception of HBV and HDV co-infection, HBsAg is usually tested positive simultaneously. In the case of co-infection, anti-HBc IgM will be positive.

Anti-HEV IgM is present in acute or recent infection, and may persist up to six weeks. Anti-HEV IgG may be positive in recent or past infection, and may remain detectable for as long as 20 months.

Anti-HGV IgG is positive in a patient who has previous exposure to HGV. There is no commercially available test for anti-HGV IgM. Diagnosis of ongoing infection depends on test for HGV RNA using polymerase chain reaction.[13]

Differential diagnosis

Drugs, such as anti-hypertensives, non-steroidal anti-inflammatory drugs and anti-tuberculous agents, may cause acute hepatitis. Acetaminophen taken in excess can cause acute fulminant hepatitis. In alcoholics and malnourished patients, severe liver injury can occur with a much attenuated dose of acetaminophen. The presence of fever in association with jaundice, eosinophilia or presence of a rash (e.g. allopurinol) supports the diagnosis of drug-induced hepatitis.

While alcohol hepatitis has a similar biochemical picture as viral hepatitis, the serum aminotransferase elevation is usually much less (less

than 10 times normal), and is associated with a disproportionately higher s. AST level.

In the early phase of biliary stone disease, the serum bilirubin and alkaline phosphatase may be normal or only mildly elevated, with predominant elevation of s. transaminases mimicking hepatitis. Any significant preceding abdominal pain should raise the suspicion of biliary colic, hence, the presence of biliary stone(s). Demonstration of dilated biliary tree by ultrasonography confirms the diagnosis.

Wilson's disease and autoimmune hepatitis, less common in Singapore, may sometimes present as acute hepatitis.

Markedly elevated s. transaminases are seen in patients who sustained hepatic ischemia or congestion, following congestive cardiac failure or a hypotensive crisis.

Treatment

There is no specific treatment for acute hepatitis A, B and E, while early treatment of newly acquired HCV infection, if recognized, is associated with a high sustained response rate, and thus prevents progression to chronicity. There is, however, no consensus for the precise management of acute hepatitis C.

As a general measure, patients should be advised to abstain from alcohol and take a high caloric diet that is rich in carbohydrate. While patients with cholestasis would often find oily food unappealing, there is no necessity to enforce a low fat diet. Enforced bed rest is also not necessary.

Most patients with acute viral hepatitis can be managed as outpatients. Only those with signs and symptoms suggestive of impending hepatic failure should be admitted to hospital for monitoring and, if necessary, workup for emergency liver transplantation.

CHRONIC VIRAL HEPATITIS

Chronic hepatitis virus infection is defined as persistent infection due to a hepatitis virus lasting more than six months. Chronic viral hepatitis is the presence of such infection in association with persistent or recurrent elevation of serum transaminases. The commonest cause of this in Singapore is chronic HBV infection. Chronic HCV infection is found mainly among patients with chronic renal failure and previous intravenous drug abusers. Chronic HDV infection is almost non-existant locally.

The long-term complications of chronic viral hepatitis are cirrhosis, with its attendant problems arising mainly from progressive loss of hepatic synthetic function, portal hypertension, and hepatocellular carcinoma.

Clinical features

Symptoms and signs

The majority of chronic viral hepatitis is asymptomatic. Non-specific symptoms such as fatigue may be present. It is not uncommon that patients present with incidental finding of raised sera ALT or AST and/or a positive serologic test for hepatitis B or hepatitis C on routine blood testing. Others may only present at the onset of complications of chronic viral hepatitis, such as features of decompensated liver disease, including variceal bleeding, ascites, and encephalopathy, or hepatocellular carcinoma.

There is usually absence of physical abnormality in the early stage of chronic hepatitis. Others may have mild hepatomegaly or splenomegaly, stigmata of chronic liver disease such as spider naevi, palmar erythema, gynecomastia or testicular atrophy. Rarely, patients may present with extrahepatic manifestations e.g. cryoglobulinemia (chronic hepatitis C), polyarteritis nodosa (chronic hepatitis B), chronic glomerulonephritis, arthralgia or vasculitis.

Laboratory findings

Sera ALT and AST levels are usually elevated and tend to fluctuate. Levels as high as that of acute hepatitis may be seen during spontaneous attempts of immunoclearance of the virus (HBe seroconversion event) or during reactivation of the underlying viral infection in patients with chronic HBV infection. Marked fluctuation of serum transaminases, however, seldom occurs with chronic hepatitis C. Hyperglobulinemia and thrombocytopenia suggest presence of liver cirrhosis; and prolonged prothrombin time and hyperbilirubinemia are indicative of hepatic decompensation. Serum alpha-fetoprotein may be elevated with the development of hepatocellular carcinoma, or in association with marked hepatic necroinflammation, usually accompanied by raised serum transaminases.

Histology

Liver biopsies are done to confirm the diagnosis, determine the need for treatment (especially for chronic hepatitis C), and assess the degree of fibrosis or cirrhosis in the liver, for prognostic purposes. Characteristic histological features of chronic hepatitis B (e.g. ground glass appearance) and C (e.g. lymphoid aggregates) may be seen in these liver biopsies. Various scoring systems have been devised to objectively quantify the degree of hepatic inflammation and fibrosis.[14,15]

Management

All patients with chronic viral hepatitis should be followed-up regularly. Patients who will benefit from antiviral treatment should be identified. Patients who are already cirrhotic should also be monitored on a 4–6-monthly basis, and treated accordingly for the associated complications.

Treatment of chronic viral hepatitis

Interferon-alpha is the most established agent for the treatment of both chronic hepatitis B and C. The main aim of treatment is to arrest necroinflammatory activities in the liver, hence prevent the development of liver cirrhosis and the associated clinical complications. Unfortunately, while adverse effects associated with treatment are many, sustained responses to treatment are not seen in all treated patients. Nevertheless, the therapeutic response for chronic hepatitis C has improved significantly with the introduction of pegylated interferon, while its effects on chronic hepatitis B awaits verification. The adverse effects of interferon treatment range from the commonly seen but mild flu-like symptoms of myalgia, headache and fever, to leucopenia, thrombocytopenia, thyroid dysfunction, depression and other psychiatric manifestations.

Hepatitis B

The ultimate objective of treatment in patients with chronic hepatitis B is to achieve sustained loss of HBV DNA and HBe seroconversion (i.e. loss of HBeAg and acquisition of anti-HBe) in patients who are HBeAg positive pre-treatment. Little response to treatment with interferon is expected during the early, immunotolerant phase of the illness. Hence, it is recommended to treat patients who have high s. ALT and low HBV DNA levels, which indicate patients' pre-existing anti-HBV immunological response.

The established 24-week subcutaneous interferon-alpha (5 MU thrice weekly) treatment regimen is associated with a maximum sustained response rate of 35% in HBeAg-positive patients. Long-term follow-up has demonstrated improvement of long-term prognosis in patients treated with interferon.[16,17] Unfortunately, patients of Chinese ethnicity, pediatric age group or those with precore mutants have a much poorer response to interferon treatment.

Other therapeutic, or potential therapeutic agents available for chronic hepatitis B include various nucleoside or nucleotide analogues, thymosin alpha-1 and therapeutic hepatitis B vaccine. Lamivudine (2',3'-dideoxy-3'-thiacytidine), the first nucleoside analogue that has been available for treatment of chronic hepatitis B, acts by inhibition of HBV DNA synthesis through its interaction with the RNA-dependent DNA polymerase. It is rapidly absorbed orally with high bioavailability, and excreted unchanged through the kidneys. Dose adjustment is required for patients with renal impairment. It effectively reduces the viral load and attenuates acute exacerbation of hepatitis B, bringing about normalization of serum transaminases. Its other therapeutic advantage is its lack of adverse effects. However, its major drawback is its poor sustained response rate upon withdrawal of treatment on the one hand,[18] and the risk of emergence of lamivudine-resistant HBV mutant strains (YMDDv) with prolonged usage on the other hand. The YMDDv, however, can be countered with the recent availability of another nucleotide analogue, adefovir, for treatment of chronic hepatitis B.[19] Unfortunately, neither of these drugs are associated with high HBe seroconversion rate with up to 2 years of treatment. HBe seroconversion occurred in 17%, 27% and 40% of patients at the end of 1, 2 and 3 years of lamivudine treatment,[20,21] and in 12% and 23% of patients at the end of 1 and 1.5 years of adefovir treatment respectively.[22] None has been known to have sustained therapeutic effects on precore HBV mutants.

Subcutaneous thymosin alpha-1 was found to be of therapeutic benefits for both chronic hepatitis B and C, when used alone or in combination with interferon.[23–26] It is associated with delayed HBe seroconversion many months post-treatment.

Hepatitis C

Interferon-alpha remains the mainstay of treatment for chronic hepatitis C. Higher dose and longer treatment duration are associated with better

therapeutic outcome. The latest breakthrough is the use of pegylated interferon-alpha, together with ribavirin, which gives rise to sustained response rates of 40–50% and 80% in genotpye 1 and non-genotype 1 HCV infection, respectively.[27,28] The problem of ribavirin is its association with anemia, which may warrant dose reduction during treatment. Its other side effects include gastrointestinal symptoms (e.g. dyspepsia, nausea), dry mouth, metallic taste, insomnia, fatigue, etc.

Management of liver cirrhosis

Liver cirrhosis is generally considered as an irreversible process. Hence, other than liver transplantation, all therapeutic measures are mainly supportive in nature, aimed at providing symptomatic relief.

Ascites

Ascites and ankle edema are some of the commonest symptoms of liver cirrhosis. They are the results of both portal hypertension and hypoalbuminemia. It is associated with a hyperaldosteronemic state. Hence, other than bed-rest and restriction of salt intake, it is logical to make use of an aldosterone-inhibiting diuretic, such as spironolactone. Loop diuretics, such as frusemide, may be used temporarily at the start of treatment in combination with spironolactone to provide rapid relief, as spironolactone is slow acting. Addition of low maintenance dose of loop diuretic may be used to counteract the spironolactone-induced hyperkalemia, or if treatment with high-dose spironolactone alone is ineffective. Other options include pteridine diuretics, e.g. triamterene, amiloride, which will cause potassium-retaining natriuresis through their actions at the distal renal tubules. One should beware of prerenal azotemia and hypokalemia, which will precipitate encephalopathy, when using such therapeutic agents.

Patients who fail diuretic treatment may undergo repeated large volume paracentesis, in combination with plasma expansion using intravenous colloid infusion, as a short-term therapeutic measure. Alternatively, transjugular intrahepatic portosystemic shunting (TIPSS, see below), or peritoneovenous shunts (if the patient is not a candidate for liver transplantation) can be considered. The limitation of these shunts is the problem of shunt blockade over time.

One of the problems of massive ascites is spontaneous bacterial peritonitis. Antibiotic prophylaxis using a fluoroquinolone, Norfloxacin,[29,30]

has proven to be effective. This should be considered in patients whose ascitic fluid protein is less than 10 g/L.[31]

Varices

Non-selective beta-blockers have been proven to be effective primary and secondary prophylactic agents against variceal hemorrhage.[32]

Endoscopic therapy should be the first-line, definitive therapeutic option for active variceal hemorrhage, while vasoconstrictive therapy is a useful adjunct to endoscopic treatment. Bleeding esophageal varices may be ligated or injected with various sclerosants(sclerotherapy), whereas gastric varices are usually treated by sclerotherapy. Intravenous infusion of somatostatin, a combination of vasopressin with nitroglycerin, or terlipressin may be initiated while awaiting endoscopic therapy. Esophageal balloon tamponade is seldom employed nowadays.

Portocaval anastomosis, which reduces the portal pressure, should be considered for patients who failed endoscopic treatment. Portocaval shunts may be created surgically or transjugularly. The advantage of the transjugular technique (TIPSS, transjugular intrahepatic portosystemic shunting), done under fluoroscopic guidance, is the ease in its performance, with the avoidance of surgery and anesthesia in a probably high-risk patient. Its biggest disadvantage is the problem of shunt occlusion over time. Hence, it should be best reserved for patients who are deemed suitable candidates for liver transplantation in the near future. Otherwise, a more durable shunt will be one that is created surgically. All shunts are associated with the risk of development of hepatic encephalopathy. Among the surgical shunts, this risk is less with the distal splenorenal shunt or creation of a H-graft between the portal vein and vena cava using a prosthetic conduit.

Esophageal transection or other devascularization procedures are other surgical options for management of variceal hemorrhage.

Hepatic encephalopathy

Acute hepatic encephalopathy usually arises from some acute precipitating events such as bleeding varices, sepsis or metabolic derangements (e.g. hypokalemia, hypoglycemia). This should be identified and treated accordingly. Meanwhile, the patient should be given supportive care to prevent fecal impaction and to maintain normoglycemia.

Chronic encephalopathy usually occurs when the patient has advanced liver cirrhosis with severely decompensated hepatic function. Liver transplantation is the only definitive treatment option.

Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is one of the major complications of liver cirrhosis. Patients with cirrhosis should be monitored at least six monthly. Alpha-fetoprotein measurements and ultrasonography of the liver should be done regularly to detect early tumors.

Definitive treatment of HCC is surgical resection of the tumor. Alcohol injection of small HCC of 2 cm or less in diameter has reported good results, comparable to surgery.[33] If the patient's hepatic reserve precludes the option of curative surgical resection, liver transplantation should be considered for suitable patients.[34] Otherwise, palliative transarterial chemoembolization may be considered for patients who have patent portal vein.

REFERENCES

1. Vento S, Garofano T, Di Perri G, Dolci L, Concia E, Bassetti D, Identification of hepatitis A virus as a trigger for autoimmune chronic hepatitis type 1 in susceptible individuals, *Lancet* 1183–1187, 1991.
2. CDC Atlanta, USA, Prevention of hepatitis A thorough active or passive immunization, *MMWR* **45**(No. RR-15), 1996.
3. Yap I, Guan R, Hepatitis A sero-epidemiology in Singapore: a changing pattern, *Trans R Soc Trop Med Hyg* **87**:22–3, 1993.
4. Oon CJ, Lim GK, Ye Z, Goh KT, Tan KL, Yo SL, *et al*., Molecular epidemiology of hepatitis B virus vaccine variants in Singapore, *Vaccine* **13**:699–702, 1995.
5. Carman WF, Jacyna MR, Hadziyannis S, Karayiannis P, McGarcvey MJ, Markis A, *et al*., Mutation preventing formation of hepatitis B e antigen in pateints with chronic hepatitis B infection, *Lancet* **2**:588–591, 1989.
6. Chan HLY, Hussain M, Lok ASF, Different hepatitis B virus gentoypes are associated with different mutations in the core promoter and precore regions during hepatitis B e antigen seroconversion, *Hepatology* **29**:976–984, 1999.
7. Chan HLY, Leung NWY, Hussain M, Wong ML, Lok ASF. Hepatitis B e antigen — negative chronic hepatitis B in Hong Kong, *Hepatology* **31**:763–768, 2000.
8. Ling R, Harrison TJ, Functional analysis of mutations conferring lamivudine resistance on hepatitis B virus, *J Gen Virol* **80**:601–606, 1999.

9. James L, Fong CW, Foong BH, Wee MK, Chow A, Shum E, *et al*., Hepatitis B seroprevalence Study 1999, *Singapore Med J* **42**:420–424, 2001.

10. Chow WC, Lee ASG, Lim GK, Cheong WK, Chong R, Tan CK, *et al*., Acute viral hepatitis E: clinical and serologic studies in Singapore. *J Clin Gastroenterol* **24**:235–238, 1997.

11. The current status of viral Hepatitis in Singapore, *Epidemiol News Bull* **24**(12):67–68, 1998.

12. Kitchen AD, Barbara JAJ, The development of tests for antibody to hepatitis C virus, in *The Molecular Medicine of Viral Hepatitis*, Harrison TJ, Zuckerman AJ (eds.), West Sussex, John Wiley & Sons, pp. 141–152, 1997.

13. Chow WC, Lee ASG, Seo YC, Tan CK, Ooi CJ, Ng HS, Hepatitis G viraemic rate and seroprevalence of hepatitis G antibody (anti-HGenv) in patients with acute and chronic liver diseases in Singapore, *J Gastoenterol Hepatol* **13**(Suppl):A12, 1998.

14. Batts KP, Ludwig J, Chronic Hepatitis, An update on terminology and reporting, *Am J Path* **19**:1409–1417, 1995.

15. Poynard T, Bedossa P, Opolon P, Natural history of liver fiborsis progression in patients with chronic hepatitis C. The OBAVIRC, METAVIR, CLINIVIR, and DOSVIRC groups, *Lancet* **349**:825–830, 1997.

16. Wong DK, Cheung AM, O'Rourke K, Naylor CD, Detsky AS, Heathcote J, Effect of alpha-interferon treatment in patients with hepatitis B e antigen-positive chronic hepatitis B. A meta-analysis, *Ann Intern Med* **119**:12–23, 1993.

17. Wong JB, Koff RS, Tine F, Pauker SG, Cost-effectiveness of interferon-alpha 2b treatment for hepatitis B e antigen-positive chronic hepatitis B, *Am Intern Med* **122**:664–675, 1995.

18. Song BC, Suh DJ, Lee HC, Chung YH, Lee YS, Hepatitis B e antigen seroconversion after lamivudine therapy is not durable in pateints with chronic hepatitis B in Korea, *Hepatology* **32**:803–806, 2000.

19. Perrillo R, Hann H, Mutimer D, Willems B, Leung N, Lee WM, *et al*. Adefovir Dipivoxil added to ongoing lamivudine in chronic hepatitis B with YMDD mutant hepatitis B virus, *Gastroenterology* **126**:81–90, 2004.

20. Liaw YF, Leung NWY, Chang TT, Guan R, Tai DI, Ng KY, *et al*., Effects of extended lamivudine therapy in Asian patients with chronic hepatitis B, *Gastroenterology* **119**:172–180, 2000.

21. Leung NWY, Lai CL, Chang TT, Guan R, Lee CM, Ng KY, *et al*., Extended lamivudine treatment in patients with chronic hepatitis B enhances hepatitis B e antigen seroconversion rates: results after 3 years of therapy, *Hepatology* **33**:1527–1532, 2001.

22. Marcellin P, Chang TT, Lim SG, Tong MJ, Sievert W, Shiffman ML, *et al*., Adefovir Dipivoxil for treatment of Hepatitis B e Antigen-Positive Chronic Hepatitis B, *New Engl J Med* **348**:808–16, 2003.

23. Rasi G, Mutchnick MG, Di Virgilo D, Sinibaldi–Vallebona P, Pierimarchi P, Colella F, *et al*., Combination low-dose lymphoblastoid interferon and thymosin-alpha 1 therapy in the treatment of chronic hepatitis B, *J Viral Hep* **3**:191–196, 1996.

24. Sherman KE *et al*., Combination therapy with thymosin alpha 1 and interferon for the treatment of chronic hepatitis C infection: a randomised, placebo-controlled double-blind trial, *Hepatology* **27**:1128–1135, 1998.

25. Chen RN, Liaw YF, Chen TC, Yeh CT, Sheen IS, Efficacy of thymosin alpha 1 in patients with chronic hepatitis B: a randomised, controlled trial, *Hepatology* **27**:1383–1387, 1998.

26. Chow WC, Ng KY, Mesenas S, Law NM, Ng HS, The interim results of thymosin alpha 1 in a randomised controlled study in the treatment of chronic hepatitis B patients: Singapore experience, *Hepatology* **28**:586A, 1998.

27. Manns MP, McHutchison JG, Gordon SC, Rustgi VK, Shiffman M, Reindollar R, *et al*., Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial, *Lancet* **358**:958–965, 2001.

28. Hadziyannis S, Cheinquer H, Morgan T, Diago M, Jensen DM, Sette H Jr, *et al*., Peginterferon alfa-2a (40kd)(pegasys) in combination with ribavirin (RBV): efficacy and safety results from a phase III, randomised, double-blind, multicentre study examining effect of duration of treatment and RBV dose. Abstr: *European Association for the study of the Liver 37th Annual Meeting*, Apr 18–21, Madrid, Spain, 2002.

29. Gines P *et al*., Norfloxacin prevents spontaneous bacterial peritonitis recurrence in cirrhosis: results of a double-blind, placebo-controlled trial, *Hepatology* **12**:716–724, 1990.

30. Novella M *et al*., Continuous versus inpatient prophylaxis of the first episode of spontaneous bacterial peritonitis with norfloxacin, *Hepatology* **25**:532–536, 1997.

31. Runyon B, Low protein-concentration ascitic fluid is predisposed to spontaneous bacterial peritonitis, *Gastroenterology* **91**:1343–1346, 1986.

32. D'Amico G, Pagliaro L, Bosch J, Pharmacological treatment of portal hypertension: an evidence-based approach, *Semin Liver Dis* **19**:475–505, 1999.

33. Kotoh K, Sakai H, Sakamoto S, Nakayama S, Satoh M, Morotomi I, *et al*., The effect of percutaneous ethanol injection therapy on small solitary hepatocellular carcinoma is comparable to that of hepatectomy, *Am J Gastroenterol* **89**:194–198, 1994.

34. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirneti A, Bozzetti F, *et al*., Liver Transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis, *N Engl J Med* **334**:693–699, 1996.

30

Pancreatic Diseases

Steven J. Mesenas

The pancreas was for a long time an organ neglected by anatomists and physicians. It was not until Wirsung demonstrated the pancreatic duct in humans in 1642 that the pancreas found its way into the medical literature. Langerhans described pancreatic histology in 1869 and the concept of the pancreas as both an exocrine and endocrine organ slowly evolved thereafter. Various disorders like pancreatitis, cancers, embryological anomalies, hereditary disorders, cystic lesions and endocrine tumors are known to afflict this organ.

NORMAL ANATOMY AND PHYSIOLOGY

The pancreas is an elongated organ, 12–15 cm in length, consisting of a head, neck, body and tail. The pancreas is largely a retroperitoneal organ.

The gland consists of an exocrine portion (80%) and an endocrine portion (20%). The exocrine portion consists of subunits known as lobules which secrete pancreatic enzymes, bicarbonate in response to food intake, and is under neurohumoral control.

Hormones that stimulate pancreatic enzymes secretion are VIP (vasoactive intestinal peptide), secretin, CCK (cholecystokinin), acetylcholine, bombesin, and substance P. Inhibitors of pancreatic secretion are glucagon, somatostatin and pancreatic polypeptide (PP).

The exocrine pancreas consists of one cell type that can secrete all the digestive enzymes. The endocrine pancreas, on the other hand, is composed of four major cell types: B cells (50–80%) that produce insulin; A cells (5–20%) that secrete glucagon; PP cells (10–35%) that produce pancreatic polypeptide; and D cells (5%) that secrete somatostatin.

EMBRYOLOGICAL ANOMALIES

The pancreas develops from the fourth week of gestation from two endodermal outpouchings of the duodenum: the ventral and dorsal pancreas. The adult pancreas is formed from the fusion of the ventral and dorsal pancreas.

Failure of fusion of the ventral and dorsal pancreas results in the most common pancreatic anomaly, the pancreas divisum. It is seen in 2–7% of patients undergoing endoscopic retrograde pancreatography. Pancreas divisum is a cause of recurrent acute pancreatitis. Treatment is usually conservative. However, various operative and endoscopic methods have been employed in the management of this condition with mixed results.

Accessory pancreata and ectopic pancreatic tissue can occur anywhere in the gastrointestinal tract, but are more commonly seen in the stomach, duodenum and proximal jejunum.

ACUTE PANCREATITIS

Acute pancreatitis causes abdominal pain and elevated pancreatic enzymes. It is the most common pancreatic disease worldwide.

The two most common etiologic factors are biliary and alcoholic pancreatitis.

Acute pancreatitis can be mild or severe. Complications, both local (e.g. necrosis, infection, pseudocyst formation) or systemic may occur. The mean age of acute pancreatitis is 50–55 years, with a range of 40 to 80 years. Acute biliary pancreatitis is more common in females, in the older age group, while acute alcoholic pancreatitis mainly afflicts males between 30 and 40 years old. Biliary pancreatitis is more common in Japan, Southern Europe and South America, whereas alcoholic pancreatitis is more common in Northern/Central Europe and North America.

ETIOLOGY

Alcoholic and biliary pancreatitis account for more than 70% of all cases of acute pancreatitis. Table 1 shows the possible etiologic factors.

Biliary pancreatitis is the most common cause of pancreatitis in Singapore. The mortality of biliary pancreatitis can be as high as 12% during the first attack but falls in subsequent episodes. Gallstone pancreatitis rarely progresses to chronic pancreatitis.

Alcoholic pancreatitis occurs usually in individuals who take more than 100 g of alcohol daily. These patients may develop chronic pancreatitis and more likely afflicts men between 30 and 45 years (male: female, 3:1). "Idiopathic" pancreatitis accounts for 15–20% of cases. These tend to be due to microlithiasis in up to 70%, while sphincter of Oddi dysfunction, pancreas divisum or small ampullary tumors can account for a significant number of the remaining cases.

Hyperlipidemia due to hypertriglyceridemia is the most common metabolic cause of pancreatitis, especially if the triglyceride level exceeds 1000 mg/dL.

Hypercalcemia from hyperparathyroidism is another metabolic cause of acute pancreatitis.

The incidence of drug-induced pancreatitis is 3%, and the most commonly implicated drugs are anti-metabolities (azathioprine, 6-mercaptopurine), sulphonamides, aminosalicylactes, and metronidazoles. These cause pancreatitis via an allergic reaction. Drug abuse with amphetamines or cocaine are also known to cause pancreatitis.

Post-surgical pancreatitis can occur in 0.2–0.8% of patients post-gastrectomy or after biliary surgery, and can have a mortality of 20–50%.

Pancreatitis can occur after cardiopulmonary bypass. Post-traumatic pancreatitis is more common in children or teenagers after blunt abdominal trauma (compression by bicycle handle or steering wheel).

Post-ERCP pancreatitis can occur in up to 5% of patients undergoing ERCP. This is often related to multiple pancreatic cannulations with injection of contrast, or after sphincter of Oddi manometry.

Acute pancreatitis has been reported in pregnancy and is usually related to coexisting cholelithiasis. A penetrating posterior wall duodenal ulcer rarely progresses to overt pancreatitis.

Hereditary pancreatitis is a rare autosomal dominant disorder related to a cationic trypsinogen gene mutation at 7q35 which prevents trypsin inactivation.

Table 1 Etiologic Factors of Acute Pancreatitis

Obstruction	**Drugs**	**Infections**
Gallstones	Azathioprine, 6-mercaptopurine, estrogens, nitrofurantoin, metronidazole, furosemide, tetracycline, methyldopa, valproic acid, L-asparaginese, thiazides, sulfonamides, didanosine, aminosalicylates, pentamidine, erythromycin, H2-receptor antagonists	Viral — mumps, rubella, Coxsackie B, cytomegalovirus, Epstein–Barr virus, human immunodeficiency virus (HIV), hepatitis A, hepatitis B, hepatitis C
Pancreatic tumors		Parasitic — ascariasis, clonorchiasis
Congenital choledochal cyst		Bacterial — mycoplasma, mycobacterium, Legionella, Campylobacter jejuni, leptospirosis
Intraductal papillary mucinous tumor		
Intraluminal duodenal duplication		
Long common pancreatico-biliary channel		
Sphincter of Oddi dysfunction		
Duodenal loop obstruction		
Periampullary duodenal diverticula		
Annular pancreas		
Alcohol and toxins	**Trauma**	**Hereditary**
Ethyl alcohol	Abdominal trauma	Autosomal dominant disorder leading to chronic pancreatitis
Methyl alcohol		
Scorpion venom		
Organophosphate insecticides		
Idiopathic	**Postoperative**	**Miscellaneous**
70% of the cases attributed to microlithiasis	Following major abdominal and thoracic surgery with cardiopulmonary bypass	Crohn's disease of the duodenum
		Penetrating peptic ulcer
		Post-ERCP, endoscopic sphincterotomy, manometry of sphincter of Oddi
		Reye's syndrome
		Cystic fibrosis
Metabolic	**Vascular**	
Hypertriglyceridemia	Ischemia	
Hypercalcemia	Vasculitis (systemic lupus erythematosus, polyarteritis nodosa)	
	Atherosclerotic embolism	

End-stage renal failure patients on hemodialysis or peritoneal dialysis may develop fulminating pancreatitis.

Various infections like mumps, Coxsackie B, hepatitis A, B, C, cytomegalovirus, HIV, clonorchiasis and ascariasis can also cause pancreatitis.

PATHOPHYSIOLOGY

The activation of pancreatic enzymes is the central step in causing acute pancreatitis.

Autodigestion and release of pancreatic enzymes and the activation of the inflammatory cascade is responsible for the local and systemic effects of acute pancreatitis.

Activation of kallikrein (by chymotrypsin) and bradykinin leads to an increase in capillary permeability. Leukotrienes, tumor necrosis factor and free O_2 radicals release also result in local tissue injury and multiorgan failure. Complement activation results in white blood cell chemotaxis, accentuating the damage. Elastase causes capillary wall damage, and phospholipase A_2 activation leads to cell membrane destruction. Disseminated intravascular coagulation results from the development of microthrombi and consumption of clotting factors and complement. Adult respiratory distress syndrome occurs because of pulmonary right to left shunting from pulmonary microthrombi.

Fat necrosis and hypocalcemia is consequent to indiscriminant lipase release.

Clinical presentation

Acute pancreatitis can be mild, severe or complicated. Edematous pancreatitis can result in organ failure in 3–5%; in necrotizing pancreatitis, it is 40–60%; and in infected, necrotizing pancreatitis, 80–90%.

The manifestations include the following:

- Diffuse, dull or boring pain in the epigastrium or left upper abdomen, which radiates to the back.
- Bowel sounds are diminished or frank paralytic ileus may occur.
- The patient is usually febrile but if he or she appears septic, it would imply a complicating bacterial infection like infected pancreatic necrosis or abscess, cholangitis, pneumonitis or cholecystitis.

- Hypotension may occur in 30–40% of patients and is due to "third-space" loss from plasma exudation, accumulation in atonic intestines, vasodilatation, vomiting and hemorrhage.
- Bluish-brown discoloration of the flanks (Grey Turner's sign) or periumbilical region (Cullen's sign) due to bleeding in the subcutaneous tissue.
- Hypocalcemic tetany is a poor prognostic sign.
- Jaundice occurs in 40% of patients and could imply cholangitis from an obstructing common bile duct stone. It could also be due to extrinsic compression from an inflamed pancreas or alcoholic liver disease.
- Pulmonary manifestations include an elevated diaphragm, pleural effusion, pneumonitis and adult respiratory distress syndrome.
- Disseminated intravascular coagulation and multiorgan failure including renal and cardiovascular failure may occur.

Laboratory tests

Serum pancreatic enzymes

Elevation of serum pancreatic enzymes, namely amylase, lipase, elastase and trypsin, can help in the diagnosis of acute pancreatitis.

The most commonly used enzyme is amylase which rises 2–12 hours after symptom onset and can remain elevated for 3–5 days. However, amylase may return to normal within 36 hours. Serum lipase is elevated in 87% of patients with acute pancreatitis. It is less sensitive but more specific than serum amylase. Serum amylase elevation alone has a sensitivity of 80–90% and specificity of 70%; however, if serum amylase and lipase elevations are taken in combination, the sensitivity and specificity for acute pancreatitis is 90–95%. Additional pancreatic enzyme tests like elastase and trypsin have no further benefit.

Hematologic and biochemical tests

Leukocytosis is present in 80% of patients. Hematocrit is raised initially due to hemoconcentration.

Raised alanine transferase, bilirubin, alkaline phosphatase and g-glutamyl transpeptidase could imply biliary tract obstruction from stones or tumor.

Hypocalcemia is detected in 30% of patients and a level below 2.0 mmol/L signifies poor prognosis. Transient, mild hyperglycemia is common.

Serum levels of C-reactive protein > 120 mg/L, lactic dehydrogenase (LDH) > 270 μ/L and polymorphonuclear leukocyte-elastase > 120 μg/L in the first 3–4 days would predict progression to severe, necrotizing pancreatitis. Methemalbumin in the serum is also observed in these patients.

Radiologic investigations

The most useful radiologic tests in acute pancreatitis are the ultrasound (US) and computed tomography (CT) of the abdomen.

Ultrasonography

Ultrasonography for the diagnosis of cholelithiasis and acute cholecystitis is unsurpassed, in terms of sensitivity (93%) and specificity (96%).

Limitations include technical difficulty due to intestinal gas, operator dependence, difficulty in assessing extent and severity of pancreatitis.

CT abdomen

CT abdomen is useful in assessing the clinical course of acute pancreatitis and local complications like pancreatic necrosis, pseudocyst formation, abscess and also prior to guided needle aspiration of a pancreatic collection.

Balthazar and colleagues' have divided acute pancreatitis into grades depending on the CT features of pancreatitis. Based on these features, one is able to prognosticate the disease.

Magnetic resonance imaging (MRI)

MRI adds no extra information to a CT Abdomen but can be done if the patient has renal failure or contrast allergy. However, it is expensive and not readily available in some hospitals.

Endoscopic retrograde cholangiopancreatography (ERCP)

ERCP has both diagnostic and therapeutic uses. Emergency ERCP has been proven to reduce morbidity in patients with moderate to severe

acute biliary pancreatitis. Sphincterotomy and stone extracted is performed in these cases. ERCP is also important in evaluating patients with recurrent pancreatitis to exclude small ampullary tumors, microlithiasis, pancreatic stones or strictures, pancreatic divisum and sphincter of Oddi dysfunction. Pancreatic pseudocysts can also be drained endoscopically.

DIFFERENTIAL DIAGNOSIS

- Gastrointestinal diseases (acute cholecystitis, intestinal obstruction, perforated peptic ulcer, peritonitis and diverticultitis)
- Vascular disorders (mesenteric ischemia, mesenteric vessel occlusion, aortic aneurysm)
- Gynecological problems (ectopic pregnancy, salphingitis and ovarian tumor)

THERAPY

Treatment for acute pancreatitis includes:

- Supportive care
- Inhibition of gastric and pancreatic secretion
- Reduction of inflammation
- Management of complications including surgery.

Supportive care

This includes analgesia, replacing and maintaining intravascular volume and electrolytes. We often use intramuscular pethidine 50–75 mg 6–8 hourly. A Swan–Ganz catheter or central venous pressure line should be inserted to maintain intravascular volume.

Fluid repletion prevents shock, renal failure and pancreatic necrosis. The patient should be kept nil-by-mouth because ileus is common, but a nasogastric tube insertion with intermittent suction is only necessary in severe cases with persistent vomiting and prolonged ileus. Electrolytes should be monitored and any losses, corrected. Total parental nutrition is useful in patients with severe pancreatitis and necrotizing pancreatitis because of catabolism and increased caloric requirements. Patients with

hypotension (systolic BP < 90 mmHg), renal impairment (serum Cr > 200 μmol/L), respiratory failure ($pO_2 < 60$ mmHg) or are septic should be monitored in intensive care.

Reduction of pancreatic/gastric secretion

The most commonly used agents to achieve this goal are H_2-receptor antagonists and proton pump inhibitors.

Somatostatin and Octreotide, a long-acting Somatostatin analog, have been tried with limited success.

Reduction of inflammation

These agents target mediators of inflammation and terminate the inflammatory cascade of acute pancreatitis. Although promising, these drugs have met with limited success. Aprotinin, a protease inhibitor and gabexate, a strong protease and phosphalipase A_2 inhibitor, showed no benefit in studies.

Other methods to reduce inflammation including infusion of fresh frozen plasma, Lexipafant and antioxidants (vitamin A, C, E, acetylcysteine, selenium) have been tried with mixed results.

Management of complications and surgery

The various complications of acute pancreatitis require a multidisciplinary approach involving gastroenterologists, surgeons, radiologists and critical care physicians.

Infection

The prevention and treatment of infection is important because infection is a major determinant of mortality.

The most common organism involved is *Escherichia coli* (51% of cases), the other bacteria being *Proteus*, *Kiebsiella*, *Staphylococcus*, *Streptococcus fecalis* and *Bacteroides*.

An infection of the pancreatic fluid collection is suspected if the patient becomes toxic (raised total white cell count, septic shock).

If the CT Abdomen shows multiple fluid collections with pancreatic necrosis, the possibility of infection is high. However, confirmation requires a radiologically-guided needle aspirate of the fluid to prove infection.

If infected, surgery with removal of necrotic tissue, drainage of the infected area and closed continuous retroperitoneal lavage or open packing is the procedure of choice.

The antibiotics of choice are imipenem, ciprofloxacin and ofloxacin if infection sets in. However, prophylactic imipenem in acute necrotizing pancreatitis reduces morbidity and mortality.

Management of pseudocyst and its complications

Pseudocysts lack an epithelial lining and occur in 8–10% of patients with acute pancreatitis. The majority of cases resolve in 6–8 weeks. Medical or surgical treatment is necessary if they persist beyond this period because of the risk of complications. The other indications for treatment are persistent pain, enlargement, complications like duodenal or biliary obstruction (from pseudocyst compression), dissection into large vessels, parasplenic areas, pararenal areas or chest, and hemorrhage. Bleeding occurs either from a vessel in the pseudocyst lining or erosion into a major vessel.

The treatment of choice is endoscopic drainage. This is done by placing a stent endoscopically through the posterior stomach wall (cystgastrostomy), duodenal wall (cystduodenostomy) or transpapillary, if there is communication between the cyst and the pancreatic duct. The stent is removed 4–6 weeks later after the pseudocyst decompresses. Surgery or percutaneous drainage is rarely performed nowadays.

Biliary pancreatitis

Patients with moderate to severe acute pancreatitis, suspected of having a biliary etiology (concomitant jaundice, dilated duct and obstructing stone on ultrasonography) benefit from emergency ERCP within 24–48 hours. ERCP with splincterotomy and stone removal is the treatment of choice especially if the patient is a high-risk candidate (renal, respiratory, cardiovascular failure). If clearance of the bile duct fails, a transpapillary polyethylene stent is placed to relief the obstruction temporarily. Definitive treatment would be endoscopic removal of the common bile duct stones or open cholecystectomy and CBD exploration when the patient is more stable.

The other complications include hypocalcemia, adult respiratory distress syndrome and multiorgan failure. Hypocalcemia requires

intravenous correction with calcium gluconate. Adult respiratory distress syndrome will need intubation and ventilation. If the patient has multiorgan failure, the prognosis is uniformly poor and management is supportive. This includes hemodialysis for renal failure, fresh frozen plasma and heparin for disseminated intravascular coagulation.

PROGNOSIS

Clinical assessment and prognostication of acute pancreatitis is useful in identifying patients with potentially poor outcomes who require intensive care management.

The most commonly used assessment method are the multifactorial scoring systems which use a combination of prognostic factors. These are:

1) Ranson criteria and modified Ranson criteria for biliary pancreatitis.
2) Modified Glasgow criteria.
3) APACHE-II.

CHRONIC PANCREATITIS

Chronic pancreatitis is an uncommon problem with an incidence of 4–10 new cases per 100 000 people per year in the West.

It is a chronic inflammatory disease resulting in the loss of exocrine function, fibrosis, and in some patients, endocrine dysfunction.

The Marseilles-Rome classification differentiates three types of chronic pancreatitis:

- chronic calcifying pancreatitis;
- chronic obstructive pancreatitis; and
- chronic inflammatory pancreatitis.

Etiology

The most common etiologic factor in the West is alcohol. Gallstones rarely result in a chronic pancreatitis, unless the attacks are very frequent and no intervention is performed.

A disease entity known as tropical pancreatitis is seen in certain tropical areas, namely Southern India, Indonesia and Central/South Africa. Hereditary pancreatitis is uncommon and affects those in the younger age

group and is an autosomal dominant disorder. These patients have recurrent attacks of pancreatitis from 5–10 years old, and develop chronic pancreatitis in their teens. Hereditary pancreatitis is due to mutation in the cationic trypsinogen gene PRSS. The cumulative lifetime risk of pancreatic cancer in these individuals is 40% by age 70 years.

The other causes include hyperparathyroidism, hypercalcemia, pancreas divisum and pancreatic trauma.

Pathophysiology

The possible theories include the "stone theory" or the "necrosis fibrosis" theory.

In the "stone theory," intraductal stone formation results in ductal hypertension with subsequent ischemia and fibrosis.

The "necrosis fibrosis" theory proposes that the fibrosis and ductal stricturing in chronic pancreatitis is secondary to focal inflammation, necrosis and subsequent fibrosis.

CLINICAL PRESENTATION

Chronic pancreatitis usually develops after repeated episodes of acute pancreatitis, and this accounts for 50% of patients with chronic pancreatitis. In 35% of patients, the acute exacerbations are absent and pain develops insidiously until it becomes continuous or intermittent but of varying intensity. Interestingly, in 15% of patients, pain is not a feature or at most a minor irritation. Pain is typically in epigastrium, radiates to back and relieved by bending forward. Pain is persistent and usually only heavy analgesics can relieve it.

Nausea and vomiting occasionally occur and could imply duodenal stenosis. Jaundice is often secondary to a biliary stricture or compression of the common bile duct by a pseudocyst. Ascites rich in pancreatic enzymes with possible pleural effusions usually means a pseudocyst leak or fistula from the ruptured pancreatic duct.

Gastrointestinal hemorrhage may occur from a ruptured pseudoaneurym of the pseudocyst lining. Malabsorption resulting in steatorrhea occurs if 90% of pancreatic exocrine function is lost, but steatorrhea is less severe than in malabsorption from small intestinal disease. This develops in 70% of patients with chronic pancreatitis for eight years or more.

Diabetes mellitus develops only after 10–15 years. Malnutrition occurs in 30% of patients and the pathogenesis is multifactorial, namely fear of eating, despite a normal appetite, vomiting from duodenal stenosis, steatorrhea or uncontrolled diabetes mellitus.

LABORATORY STUDIES

Serum lipase and amylase levels are of no use in the diagnosis of chronic pancreatitis. Enzyme levels may be elevated for 2–3 days after an attack, but the levels may be normal.

The most accurate methods to diagnose chronic pancreatitis are pancreatic function tests and imaging studies.

Direct pancreatic function tests requiring aspiration and analysis of duodenal contents after injection of secretin is the method of choice. However, the secretin stimulation test has a sensitivity of 80% for diagnosis of chronic pancreatitis.

The volume and concentration of bicarbonate after secretin injection gives an idea of the severity of the pancreatic insufficiency. By injecting CCK or cerulein, the digestive enzymes (e.g. amylase, trypsin, and lipase) can also be measured.

Indirect or "tubeless" tests have been used to evaluate pancreatic function without the need for duodenal intubation. However, it can only diagnose cases of severe chronic pancreatitis where 90% of the exocrine function is lost. These tests include the bentiromide (NBT-PABA) test, and the pancreolauryl tests which measure released components in the blood or urine.

Fecal fat excretion test is used to measure fat malabsorption as a consequence of pancreatic insufficiency. It also enables one to gauge the effectiveness of therapy after administration of enzymatic supplementation. Stool fat of more than 5–7 g/day is abnormal.

Radiological studies

Only 30% of patients with chronic pancreatitis have diffused or focal pancreatic calcification on *plain abdominal radiographs*. However, when calcifications are present, the diagnosis of chronic pancreatitis is almost certain.

Dynamic *secretin magnetic resonance pancreatography (S-MRP)* is able to detect subtle ductal changes in early pancreatitis and provide information

on pancreatic exocrine function (bicarbonate excretion following secretion stimulation).

Endoscopic retrograde cholangiopancreatography (ERCP) is still the gold standard in the diagnosis of chronic pancreatitis. Although it is invasive, ERCP provides possible therapy in selected cases of chronic pancreatitis. Pancreatic strictures can be "ballooned" or "stented" endoscopically and pancreatic ductal stones can be removed. Pancreatic pseudocysts are mostly treated endoscopically nowadays. ERCP also allows for tissue acquisition.

A new and emerging radiological method for evaluation of chronic pancreatitis is *endoscopic ultrasound (EUS)*. It is sensitive in detecting typical changes of chronic pancreatitis and fine needle aspiration cytology (FNAC) enables one to biopsy a suspicious mass lesion.

MRCP (magnetic resonance cholangiocreatography) and CT are only useful in detecting biliary strictures and pseudocysts.

Management

The treatment of uncomplicated chronic pancreatitis is medical, and involves avoidance of aggravating factors (alcohol abstinence), nutritional support, relief of acute and chronic pain, treatment of diabetes mellitus and replacement of pancreatic enzymes.

Medical therapy

Medical management of chronic pain

Measures like abstention of alcohol and avoiding large meals may be useful. Non-narcotic analgesics like acetaminophen, aspirin or other non-steroidal anti-inflammatory drugs can be tried initially.

Enzyme therapy has been proposed to be beneficial because of the concept of feedback regulation which shows that intraduodenal administration of enzymes reduce the release of cholecystokinin and hence pancreatic juice secretion and induction of pain. Non-enteric-coated preparations are better than enteric-coated ones.

Malabsorption

Steatorrhea can be reduced by administration of at least eight tablets of conventional enzyme preparation (e.g. Viokase, Cotazym, or Ilozyme)

taken throughout each meal of the day. The patient should be placed on a restricted fat diet (60–80 g/day). If unsuccessful, concomitant usage of H_2-receptor antagonists or enteric-coated enzyme preparations can be employed. The former acts by reducing inactivation of pancreatic enzymes by maintaining a gastric and duodenal pH of greater than 4.0. Enteric-coated enzyme preparations only release their contents when the pH is 5.5. Medium-chain tryglyceride (MCT) administration reduces steatorrhea.

Diabetes mellitus

Management of diabetes mellitus in patients with chronic pancreatitis includes reduction of alcohol intake, treatment of malabsorption and correction of irregular food intake and malnutrition.

Hyperoxaluria and renal stones

Hyperoxaluria and renal stone formation occurs because of untreated steatorrhea. Treatment includes a low oxalate and triglyceride diet, pancreatic enzyme substitution and high calcium intake (in the form of antacids).

Endoscopic and surgical therapy

The main indication for endoscopic or surgical intervention in patients with chronic pancreatitis is often pain. This pain must be constant, disabling and must have failed treatment with conventional medical therapy.

Endoscopic therapy would include a pancreatic sphincterotomy ± biliary sphincterotomy. If there are main pancreatic duct stones, these should be extracted. Often if the stones are large, extracorporeal shock wave lithotripsy (ESWL) may be necessary to fragment the stones allowing for easier extraction. If there is a pancreatic stricture, stenting with a polyethylene stent for six weeks to dilate the stricture can relieve pain. There is often obstructive jaundice from a distal CBD stricture, which can be relieved by placement of a plastic stent. Pseudocysts are nowadays treated endoscopically by placement of a plastic stent through the gastric wall (cystogastrostomy), or duodenal wall (cystoduodostomy). If there is a disruption of the main pancreatic duct, a transpapillary stent is useful in treating the pancreatic fistula.

These stents are left in place for 4–6 weeks until pseudocysts or pancreatic fistulae have resolved, and then extracted endoscopically. Endoscopic therapy has gained success over the years because of its low complication rate and the avoidance or postponement of surgery. However, medical and endoscopic therapy does occasionally fail, and surgery is required. The most common surgical procedure would be the longitudinal pancreaticojejunostomy (Modified Puestow Procedure). It is a drainage procedure, useful if the main pancreatic duct is dilated and after 5 years, pain relief is achieved in about two-thirds of patients.

A coeliac plexus block done via CT or endoscopic ultrasound guidance can be performed for patients who have failed all other modalities of treatment and are unfit or unwilling to undergo major surgery. Pain relief rarely exceeds six months.

CARCINOMA OF THE PANCREAS

Pancreatic carcinoma is a universally fatal cancer. Ninety percent are mucinous adenocarcinomas. The incidence of pancreatic carcinomas ranges from 2.2/100 000 in Singapore, India and Kuwait to 12.5/100 000 in Sweden. In Singapore, it is uncommon in those less than 45 years old, and 80% occur in those between 60 and 80 years. The male to female ratio is between 1.5:1 and 2:1. These epidemiological findings are similar in other countries.

Etiology

The etiological factors are:

- cigarette smoking;
- alcohol consumption (particularly beer);
- gallstones;
- high animal fat diet;
- diabetes mellitus; and
- chronic pancreatitis.

Clinical findings

For the majority of patients, symptoms and signs of the disease suggest at least moderately advanced pancreatic carcinoma. Early presentation of the disease is rare.

Many patients have a vague, dull, midepigastric pain that is often constant, radiating to the back with no relieving factors. Jaundice is noted in 50% of patients and is due to invasion and obstruction of the distal common bile duct. Periampullary tumors tend to present earlier because of their propensity to cause jaundice early in the course of the illness. Often, jaundice is a late manifestation and if associated with a papable tumor mass could indicate bulky disease or hepatic invasion. Courvoisier's sign or a papable, non-tender gallbladder is indicative of an obstructing distal common bile duct tumor.

Onset of brittle diabetes (15%) or sudden worsening of pre-existing diabetes (18%) could indicate a pancreatic cancer.

Other manifestations include migratory thrombophlebitis, intense pruritis, acute pancreatitis and psychiatric disturbance (particularly depression). Vomiting is a later symptom and usually indicates tumor invasion into the duodenum or stomach. Hemetemesis or malena may also result if the tumor bleeds.

Laboratory investigations

A non-invasive, reproducible diagnostic test with high specificity and sensitivity does not exist.

Elevated CEA (carcinoembryonic antigen) and carbohydrate antigen CA 19-9 levels are often seen in pancreatic cancers. However, as screening tools for pancreatic cancer, these are inadequate.

Imaging studies

Although ERCP (endoscopic retrograde cholangiopancreatography) is the most sensitive and specific radiological investigation for the diagnosis of pancreatic carcinoma, it plays no part in staging of pancreatic cancer.

Staging requires helical computed tomography (CT), endoscopic ultrasonography (EUS) or magnetic resonance imaging (MRI).

Computed tomography (CT)

CT has a sensitivity of 80% and specificity of 95% for diagnosis of pancreatic carcinoma. Lesions larger than 2 cm in diameter are easily detectable. CT provides information on resectability (e.g. lymph nodes,

liver metastases, and major vessel involvement). Accuracy for resectability is 83% and sensitivity for vascular invasion is 64%.

However, CT is inappropriate for nodal staging, and EUS is more suitable and accurate. However, in advanced disease, lymph node involvement is usually associated with features of unresectability. CT is the first imaging modality of choice if carcinoma of the pancreas is suspected. Magnetic resonance imaging (MRI) provides little additional information.

Endoscopic retrograde cholangiopancreatography (ERCP)

ERCP is sensitive (94%) and specific (97–100%) in the diagnosis of carcinoma of the pancreas. This is especially so if the tumor is less than 2 cm and the use of other imaging modalities to detect this tumor has been unsuccessful.

The advantages of ERCP also include tissue diagnosis and palliation or temporary relief of obstructive jaundice from a distal CBD stricture.

Endoscopic ultrasonography (EUS)

EUS provides detailed imaging of the pancreas through the walls of the stomach and duodenum. EUS has a sensitivity of detection for pancreatic cancer of 97%, accuracy for resectability of 91% and sensitivity for vascular invasion of 91%. It is the most accurate modality for the detection of portal vein invasion.

EUS has the additional benefit of sampling locoregional lymph nodes like celiac nodes. A celiac plexus block can also be performed for palliation of pain in unresectable tumors.

Figure 1 shows a possible clinical algorithim for staging of pancreatic cancer.

Tissue acquisition

Tissue diagnosis for pancreatic carcinoma is not necessary in every suspected case. A patient with a lesion that is operable, who is desirous of surgery and is otherwise medically fit should be sent for surgery without tissue biopsy. Tissue acquisition is only necessary if the result of the biopsy will impact on management. This would be required if the patient is unwilling to undergo major surgery without definite histological proof of malignancy or tissue diagnosis is required prior to palliative therapy (relief of biliary, duodenal obstruction, chemotherapy and radiotherapy).

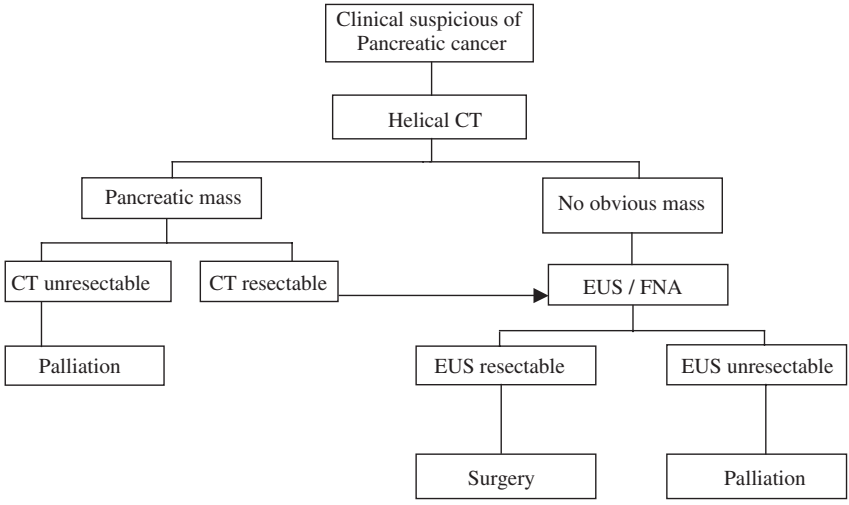


Fig. 1 Algorithm for staging of pancreatic cancer.

There are three methods whereby tissue can be acquired:

1) ERCP tissue sampling;
2) Percutaneous needle biopsy; and
3) EUS-guided FNA biopsy.

Treatment

Despite advances in diagnosis, imaging, staging and treatment of pancreatic cancer, surgery still remains the only possibility of cure. Only 20% of patients are candidates for surgical resection while the rest of the patients can only be offered palliation. Palliation includes relief of jaundice, pain and intestinal obstruction which may be surgical or non-surgical.

Medical therapy

Palliation involves drainage of the biliary system via the percutaneous route (PTC) or endoscopic route (ERCP).

ERCP and insertion of biliary endoprostheses has a lower morbidity and mortality than the percutaneous method. It also allows for tissue diagnosis and the success rate is 90% in experienced hands.

Early complications of endoscopic stenting include retroperitoneal perforation, bleeding and pancreatitis. These occur in less than 2% of

patients. The most common late complication is clogging of stent. The median time to occlusion is three to four months. Various methods have been used to prevent stent occlusion, including new stent designs and use of antibiotics. However, the only method that has proven to extend stent lifespan, is the use of self-expanding metallic stents. These metallic stents have a median time to occlusion of seven months, while plastic stents tend to last for 3–4 months. However, they are costly. Therefore metallic stents should only be used for patients expected to survive at least six months.

Palliation by endoscopic stenting is comparable to surgical bypass in terms of mean survival. Endoscopic stenting has a lower morbidity and a 30-day mortality.

Unresectable tumor may benefit from a palliative double-bypass surgical procedure at the onset. Otherwise, an enteral stent and a biliary stent may be inserted endoscopically if the patient has both duodenal and biliary obstruction from the tumor and refuses surgery or is unfit for surgery.

Adjuvant therapy includes radiotherapy and chemotherapy.

Surgery

The Whipple resection remains the surgical procedure of choice in patients with resectable pancreatic head cancers. It has an operative mortality of less than 5% in expert hands.

Total pancreatectomy with en bloc resection of the entire pancreas, duodenum, spleen, greater omentum, subtotal gastrectomy and lymph node dissection, has been suggested to be superior to the Whipple procedure, because of better tumor clearance. However, this procedure has a higher mortality (16%) and no difference in 5-year survival.

CYSTIC TUMORS OF THE PANCREAS

Cystic tumors of the pancreas are relatively uncommon, accounting for 1% of pancreatic neoplasms.

There are three main types of these cystic tumors, namely mucinous cystic neoplasms (MCN); serous cystadenoma; and intraductal papillary mucinous tumors (IPMT). These can be confused with pancreatic pseudocysts. Treatment for all pancreatic cystic neoplasms is surgical resection if feasible.

PANCREATIC ENDOCRINE TUMORS

The overall incidence of pancreatic endocrine tumors (PETs) is low, 10 per 1 million population. PETs can be functional, if the tumor produces a hormone and is associated with a clinical syndrome; or non-functional, if there is no hormone production or if the hormone produced is not associated with a clinical syndrome. The most common pancreatic endocrine tumors are insulinomas and gastrinomas. VIPomas, glucagonomas and somatostatinomas are rarer. PETs can occur sporadically with no familial associations, or it can occur in inherited disorders. The most established of these is MEN I which is characterized by tumors of the pituitary gland, parathyroid gland and endocrine tumors of the pancreas, usually gastrinomas or insulinomas.

Other inherited syndromes associated with pancreatic endocrine tumors include von Hippel–Lindau disease, von Recklinghausen's disease, and tuberous sclerosis. Only insulinomas and gastrinomas which are the most common will be dealt with.

Insulinoma

Insulinomas are the most frequently occurring pancreatic endocrine neoplasms. These tumors usually occur in females (60%) between 40 and 45 years old. Patients have symptoms of hypoglycemia, neuroglycopenia (somnolence, confusion, irritability, visual disturbances, abnormal behavior, paresthesias, seizures, drowsiness, and coma) and catecholamine release (anxiety, palpitations, tremor, and sweating).

Diagnosis is made by the presence of fasting hypoglycemia and hyperinsulinemia. A 72-hour fast is performed and plasma insulin glucose ratio of higher than 0.3 is positive for insulinoma. Measurement of C-peptide levels helps to differentiate endogenous insulin hypersecretion from surreptitious insulin usage.

Treatment includes correction of hypoglycemia by having frequent meals with complex carbohydrates or diazoxide which has potent hyperglycemic effects. Octreotide administered subcutaneously is useful for this purpose. Long-acting preparations of both lanreotide and octreotide that last 2–4 weeks are available. Surgical treatment involves localization and resection if the patient is fit.

Gastrinoma and the Zollinger–Ellison syndrome

Gastrinomas are the second most common pancreatic endocrine tumors. Sixty percent are malignant, and 30% are associated with MEN I. The clinical presentation is that of multiple ulcers in atypical sites like the esophagus, distal duodenum and upper jejunum.

Complications include perforation, pyloric stenosis, hemorrhage and gastrojejunocolic fistulas. Diarrhea and malabsorption occur because of acid-inactivation of enzymes. Diagnosis is by the demonstration of hypergastrinemia with a basal acid output > 15 mEq per hour. Medical treatment is with high-dose proton-pump inhibitors. Definitive treatment is surgical resection after tumor localization.

REFERENCES

1. Balthazar EJ, Robinson DL, Megibow AJ, Ranson JH, Acute pancreatitis: value of CT in establishing prognosis, *Radiology* **174**:331–336, 1990.
2. Steinberg WM, Tenner S, Acute pancreatitis, *N Engl J Med* **330**:1198–1210, 1994.
3. Madura J, Fiore AC, O'Connor KW, Lehman GA, Mc Cammon RL, Pancreas divisum, Detection & management, *Am Surg* **51**:353, 1985.
4. Smits ME, Rauws EAJ, Tytgat GNJ, Huitbregste K, Endoscopic treatment of pancreatic stones in patients with chronic pancreatitis, *Gastrointest Endosc* **43**:556–560, 1996.
5. Whitcomb DC, Preston RA, Aston CE *et al*., A gene for hereditary pancreatitis maps to chromosome 7q35, *Gastroenterology* **110**:1975–1980, 1996.
6. Lillemoe KD, Current management of pancreatic carcinoma, *Ann Surg* **221**:133–148, 1995.
7. DiMagno EP, Reber HA, Tempero MA, AGA technical review on the epidemiology, diagnosis, and treatment of pancreatic ductal adenocarcinoma, *Gastroenterology* **117**:1464–1484, 1999.
8. Anderson MA, Carpenter S, Thompson NW, Nostrant TT, Elta GH, Scheiman JM, Endoscopic Ultrasound is highly accurate and directs management in patients with neuroendocrine tumours of the pancreas, *Am J Gastroenterology* **95**(9):2, 271–277, 2000.
9. Mertz HR, Sechopoulos P, Delbeke D, Leach SD, EUS, PET & CT scanning for evaluation of pancreatic adenocarcinoma, *Gastrointest Endosc* **52**(3):367–371, 2000.
10. Ballinger AB, McHugh M, Catnach SM *et al*., Symptom relief and quality of life after stenting for malignant bile duct obstruction, *Gut* **35**:467–470, 1994.
11. Martin I, Hammond P, Scott J *et al*., Cystic tumours of the pancreas, *Br J Surg* **85**:1484–1486, 1998.
12. Jensen RT, Carcinoid and pancreatic endocrine tumours: Recent advances in molecular pathogenesis, localisation and treatment, *Curr Opin Oncol* **12**:368, 2000.

Hematology

This page intentionally left blank

31

An Approach to Blood Disorders

Patrick Tan and Ng Heng Joo

INTRODUCTION

Blood functions as the carrier component of the circulatory system and serves to supply oxygen, essential nutrients and other substances to all tissues in the human body. In return, the various products of these tissues, which include carbon dioxide and hormones, are carried to different organs where they perform specific metabolic functions, or are otherwise either stored or excreted. Blood forms an important part of the immune system by delivering foreign materials to sites of identification and allowing mobilized white cells of immune function to descend on sites where they are required. Antibodies, cytokines and other substances circulate within blood to serve their purpose. Within this system, blood has its own set of regulatory proteins and cells that serves to keep it in fluid phase as well as prevent loss of blood from the circulation in the event of a mechanical breach of the circulatory system.

In its physiological state and in pathology, it is useful to visualize blood and its disorders through its individual components. These may be categorized broadly into cellular and plasma components. In the normal

person, these components function in harmony to perform their designated roles. It is easy to envisage the potential for pathological states to arise in any of these components by way of quantitative or qualitative production defects, qualitative dysfunction, unregulated destruction or loss and malignant proliferation of cellular components.

BLOOD IN HEALTH

In a healthy individual, blood forms about 8% of the total body weight (about 5 liters in an average 60 kg man). Of this volume, 55% consists of plasma in which the cellular components are suspended.

Cellular Components and the Bone Marrow

The cellular elements of blood are made up of red blood cells, white blood cells and platelets. The bone marrow is the organ in which all red blood cells and platelets and most white cells are produced in healthy adults. Blood cells are formed in the liver and spleen in the fetus but such extramedullary hematopoiesis occurs only in disease states in the adult.

The bone marrow is one of the largest organs in the human body and is certainly one of the most active. The primer of all blood cells is the pluripotent stem cell, which are usually few in numbers. These cells have the potential to repopulate the whole bone marrow even when introduced in small numbers to an ablated marrow, as illustrated in the classical bone marrow transplant. Pluripotent stem cells differentiate into committed stem cells from which various cell types are then derived. White cell precursors in different stages of development outnumber red cells by an order of 3 to 1, which reflects in part, the shorter lifespan of white cells in comparison to red cells.

Red blood cells serve primarily to carry oxygen via its red pigment hemoglobin. Shaped as biconcave discs, these cells have an average lifespan of 120 days. Within each of these cells is found hemoglobin, a protein, which is a globular molecule made up of 4 subunits. Each of these subunits has a heme moeity and a globin portion. The amount of red blood cells in the body may be expressed as the red blood cell count, which is usually about 5.4 million/μl in men and 4.8 million/μl in women. The red blood series is more commonly measured via its functional entity of hemoglobin, which is usually about 16 g/dL in men and 14 g/dL in women.

Among white blood cells, granulocytes are the most numerous and may be further differentiated based on their morphology and granular contents, into neutrophils, eosinophils and basophils. Two other white cell types normally found in blood are the lymphocytes and monocytes. Granulocytes are mainly involved in inflammatory and allergic responses, and have cytoplasmic granules that contain substances mediating these responses. After leaving the bone marrow, monocytes circulate in the peripheral blood for about 72 hours before entering tissues and becoming tissue macrophages. They serve an important function in eliminating bacteria when activated and play an important role in immunity as antigen presenting cells. Lymphocytes are the pillars of the immune system with complex tasks, which ultimately culminate in an immune response to foreign antigens via activated cells or the production of antibodies.

Plasma Components

About 5% of our body weight is contributed by plasma, which amounts to about 3 liters in a 60 kg man. This fluid portion of blood contains a large number of ions, inorganic and organic substances and proteins that are being transported to various parts of the body, facilitating this transport process or performing various functions to regulate blood in its dynamic environment. Plasma clots on standing and the portion that remains after removal of the clot is called serum. This differs from plasma by lacking the coagulation factors II, V, VIII and fibrinogen.

Plasma proteins may be separated into albumin and globulin fractions. The globulin fraction may be subdivided into various components such as alpha, beta and gamma fractions. In general, within the globulin fraction of protein is found specific proteins like antibodies and clotting factors as well as carrier proteins. Albumin is the main contributor to oncotic pressure that keeps water within the vasculature. The liver is the organ in which most of the protein components like albumin and clotting factors are manufactured. Antibodies are manufactured by plasma cells.

Clotting factors play an important role in maintaining the integrity of the vascular system and are activated whenever vascular breaks occur, in order to contain and arrest bleeding. This function is carried out in concert with platelets, which are the smallest cellular component in blood. Within plasma, is found an alternative set of proteins which serve as antithrombotic and fibrinolytic agents to keep the clotting mechanism in check and ensure appropriate and controlled activation only.

Plasma proteins also serve an important function in binding to ions and hormones, thus helping in the carriage and storage of these compounds as well as prolonging their half-life in the vascular system.

BLOOD IN DISEASE

When confronted with a patient with a blood disorder, it is useful to begin by conceptualizing the origins of the disorder. Blood diseases may be disorders of its cellular or plasma component. The two components are, of course, not mutually exclusive compartments and cellular disorders can manifest as changes in plasma components and vice versa.

Viewing blood cellular disorders under broad etiological categories of ineffective, impaired or unregulated production, increased destruction or blood loss is a simple concept of the pathogenesis of blood disorders. Anemia in its most basic manifestation may be due to blood loss of any nature with no defects in the hematopoeitic system *per se*. Antibody-mediated hemolysis on the other hand will lead to anemia by way of increased destruction of blood cells. Again, the hematopoeitic potential of the bone marrow remains intact with the onset of anemic symptoms occurring only by way of its inability to keep up with the degree of destruction of red cells. An intrinsic defect of the red cells itself cannot always be absolved in all hemolytic disorders. Membrane defects like congenital spherocytosis or complement-mediated hemolysis in paroxysmal nocturnal hemoglobinuria give rise to increased red cell destruction with resultant anemia. In the absence of both blood loss and hemolysis, the possibility of ineffective or impaired production of cells by the bone marrow must then be entertained. Multiple cell line involvement provides clues to the extent of the marrow pathology. In its simplest form, marrow production of blood cells may be due to a lack of essential nutrients for blood production. This is the essence of iron deficiency anemia and the vitamin deficiencies, which contribute to ineffectual production. The inherited thalassemias are also disorders resulting from ineffectual hematopoiesis. Via an immune or infective mechanism or simply by direct toxicity from drugs including chemotherapeutic agents, marrow production of blood cells may be suppressed in a self-limited or permanent manner.

Marrow pathology may also be due to clonal disorders, which basically is the concept of an abnormal clone of cells given preferential survival benefits, overwhelming normal clones and resulting in disorders of

hematopoiesis. Dysregulated proliferation of malignant clones gives rise to the leukemias and lymphomas, myelodysplastic syndromes and myeloproliferative disorders. Non-malignant clones exist in conditions like paroxysmal nocturnal hemoglobinurias. An afterthought should also be given for non-hematological infiltrations of the bone marrow which may include malignant metastatic deposits as well as infective agents like tuberculosis. Finally, the bone marrow as an organ is not a detached entity and its capacity to produce cells may be affected by concurrent illnesses of autoimmune, infective or other etiology by indirect cytokine-mediated means.

The most common disorder in relation to the plasma component of blood is that of hemostasis which manifests clinically as either a disorder of excessive bleeding or clotting. The simplest of these disorders are due to deficiencies of specific protein factors. The hemophilias A and B are due to inherited deficiencies of factors VIII and IX respectively and give rise to bleeding tendencies. Acquired deficiencies of clotting factors are usually related to pathologies of the liver or drugs and involve multiple factor deficiencies. Platelet, a cellular component, is closely intertwined with the clotting factors in hemostasis. In a patient with a bleeding disorder, it is always useful to consider the cause as being either due to a clotting factor deficiency or defect, a low platelet level or platelet dysfunction and finally a disorder of the vasculature itself.

Deficiencies of proteins that regulate the clotting mechanism, like anti-thrombin III, protein C and S on the other hand, may lead to clotting disorders. Thrombotic tendencies are however influenced by various factors which range from disruption of the vasculature, stasis or altered blood flow, increased blood viscosity from an increase in cellular or plasma components, to disorders that affect the protein system which usually regulates and controls clot formation.

At times, the clotting and fibrinolytic system may be activated simultaneously by precipitating factors that trigger the coagulation cascade and give rise to a consumption coagulopathy. This is the basis of the syndrome of disseminated intravascular coagulation.

Changes to the plasma protein component of blood may occur in disease as well. The globulin fraction of plasma may be markedly increased in plasma cell disorders and leads to symptoms due to increase in plasma viscosity. More commonly, the albumin portion becomes low during active disease with resultant signs and symptoms secondary to this development.

THE HEMATOLOGY PATIENT

By virtue of blood's functions as a medium of transport for oxygen and other substances and as an executor of immune function, the patient with a hematological disorder presents with the consequences of defects of these functions. An anemic patient begins by complaining of lethargy, dyspnea, poor concentration and may be noted to be pale by friends and relations. Polycythemic patients on the other hand may present with symptoms of hyperviscosity or a thrombotic event. Patients with a white cell and lymphoproliferative disorder may seek medical attention because of an increased susceptibility to infection. Bleeding tendencies may be the first indication that a patient has a low platelet count. Conditions that involve more that one cell line may have any combination of the above mentioned symptoms. Thus, it is not unusual for the leukemic patient to present as an infection with a recent history of easy fatigability as well as gum bleeding.

Myelo- and lymphoproliferative conditions have a propensity to involve the liver and spleen and it is not unusual for patients to present with first symptoms arising from enlargement of these two organs. Extramedullary hematopoiesis in the liver and spleen may also lead to significant enlargement of these two organs and result in symptoms.

The patient with an acute and rapidly progressive disorder like leukemia is usually quickly diagnosed. Chronic conditions like the myeloproliferative disease may remain asymptomatic for years until an unrelated condition warrants testing of a patient's blood or when the condition progresses.

The symptoms and signs of blood disorders can at times be nonspecific and be easily missed by the unsuspecting clinician. Similarly, in systemic illness with hematological manifestations, abnormalities may at times be subtle with attention being focused on the underlying condition instead.

Thrombotic conditions may have rapid and dramatic onset. Patients with deep vein thrombosis may present with acute limb swelling and subsequently develop pulmonary embolism with circulatory collapse. On the other hand, they can also be asymptomatic if the extent of thrombosis is limited. Bleeding disorders are usually dramatic and alarming to the patient and clinician and are usually brought to rapid attention.

INVESTIGATING BLOOD DISORDERS

Rapid advances in our understanding of blood disorders have led to the availability of various diagnostic tools that allow for faster and more accurate diagnosis of blood disorders. Nonetheless, they cannot replace the most basic tool of clinical medicine, which remains as the taking of a detailed and complete medical history with careful physical examination. Presenting symptoms give clues to help narrow the possible disorders that a patient has. Duration of symptoms helps in distinguishing chronic from acute disorders. The value of history taking can hardly be overemphasized. On completion of this exercise, the clinician should be able then to order logical and relevant investigations, which will help confirm the diagnosis, detect associated conditions, determine best possible treatment options and predict outcome of the chosen treatment.

The Full Blood Count and Peripheral Blood Film

The quest for answers to explain blood cellular disorders usually begins with the full blood count and peripheral blood film. It is helpful to define the number and type of cell lineage that is involved. Pure red cell disorders are more common that those that only affect the white cell and platelets. Indices of red blood cell morphology like the mean corpuscular volume and red cell distribution width are useful in helping to differentiate red cell disorders. Isolated thrombocytopenia is commonly encountered and its causes should always be questioned along the lines of increased destruction or decreased production. More commonly, however, more than one cell line is involved, which opens up the possibility of a more generalized condition that exerts its effects on multiple cell lines. This may be seen in the example of Evan's syndrome where antibodies to both platelets and red cells are found or in aplastic anemia where all cell lines are involved.

The peripheral blood film provides a microscopic view of blood cellular component that is of immense value. Red cell morphology can clearly be described with great diagnostic value to the trained eye. Nucleated white cells give clues to the possibility of malignant disorders or infiltrative conditions. Hypersegmented polymorphs may point to a diagnosis of vitamin B12. Toxic granules may suggest recent infection. Platelet clumps in association with a low platelet count may prevent further unnecessary investigations for pseudothrombocytopenia.

Accessory investigations to exclude common deficiency disorders of essential nutrients for hematopoiesis are usually done early in the pyramid of investigations if they are suggested by indices in the full blood counts. Hemoglobin electrophoresis to exclude thalassemias are best resorted to, once iron deficiency is excluded, in order not to miss the beta thalassemia minor which may be masked by iron deficiency. Screening for systemic conditions like renal failure, autoimmune conditions, liver and endocrine disorders may provide explanation for various blood abnormalities.

Bone Marrow Studies

Bone marrow studies are usually done in the first instance when it is required to confirm suggestions of malignant or infiltrative conditions like acute leukemias, demonstrated in the peripheral blood picture. Many other conditions that can be easily defined by the full blood count and peripheral blood picture as well as accessory investigations do not deserve to have marrow studies performed. There are instances however where an underlying marrow pathology cannot be conclusively excluded from peripheral investigations, for which it must then be carried out. Bone marrow studies as commonly described, usually allow for the examination of a marrow aspirate, flow cytometric studies on marrow aspirate as well as karyotyping of marrow cells. In the same sitting a trephine biopsy may also be taken for histological examination. Marrow samples may also be taken for cultures of suspected microbiological agents that may have disseminated to the bone marrow.

An examination of the bone marrow aspirate provides a valuable insight into the morphological appearance and quantification of the precursor cells of all the cell lines. This is essential in the diagnosis of many disorders which begin in early hematopoietic phases and which may be marked by dysplastic features or frank malignant features. Non-hematological malignant cell infiltration as well as infectious microbiological agents such as Histoplasma sp. may be clearly evident under light microscopy. The leukemias and plasma cell disorders are quantified and diagnosed based on the percentage of the malignant cells counted under light microscopy. Apart from cell morphology, marrow aspirates allow estimation of the cellularity of the marrow and estimation of iron stores available in the body. A normal marrow appearance in the presence of a hematological

condition may suggest a peripheral destructive disorder or the possibility of a systemic disorder with general systemic effects.

The trephine biopsy when taken as part of the bone marrow examination, provides details of the bony architecture within which the marrow is found and allows more accurate estimation of the cellularity of the marrow as well as reticulin fibers. Lymphomatous involvement of the bone marrow is best excluded from the trephine biopsy.

Flow Cytometry, Cytogenetic and Molecular Studies

Immunophenotyping by flow cytometry provides valuable information in the reproducible diagnosis and classification of acute leukemias, chronic lymphoid leukemias and non-Hodgkin's lymphoma. Analysis can be performed on peripheral blood, bone marrow and body fluid as well as fresh specimens obtained from lymph node biopsies, spleen and extranodal tumors.

Cytogenetic and molecular analysis techniques have opened up new dimensions in hematological diagnosis by allowing for the identification of specific clonal abnormalities and their association with particular disorders. A notable example of the value of this examination is the Philadelphia chromosome, which occurs as a translocation between chromosomes 9 and 22, and is found in chronic myeloid leukemia (CML). Its molecular product is the bcr/abl protein transcript. The finding of Philadelphia chromosomes or bcr/abl protein transcripts invariably points to a diagnosis of CML. Apart from providing information on the clonal origins of particular disorders, this technique can be used in the monitoring of response to treatment of the leukemias and lymphomas, where their presence or absence in subsequent analysis can be used as an indicator of disease response as well as relapse. Cytogenetic abnormalities are also used to prognosticate the response and outcome of treatment for various hematological malignancies.

Coagulation and Bleeding Disorders

Bleeding in a patient may be due to a platelet, coagulation or vascular disorder. Initial screening tests are aimed at distinguishing a platelet disorder from a coagulation disorder and usually include a basic full blood count and prothrombin and partial thromboplastin time. Beyond this,

more specialized testing looks at platelet functions as well as deficiency of specific or multiple coagulation factors and their inhibitors. A knowledge and understanding of the coagulation cascade is essential in the logical investigation of coagulation disorders. Vascular disorders leading to bleeding are usually diagnosed on exclusion of platelet and coagulation disorder.

In the patient with thrombotic manifestations, obvious precipitating factors are often looked for initially. These include recent surgery and known malignancies that are likely to lead to hypercoagulable states. In patients with unprovoked thrombosis, an underlying acquired or inherited thombophilic state has to be excluded. Testing in these patients will include screening for antibodies associated with the anti-phospholipid syndromes, inherited conditions like protein C and S and anti-thrombin III deficiencies.

TREATING THE HEMATOLOGY PATIENT

Treating blood disorder in its most straightforward form may merely involve the replacement of deficient blood components or removal of excess components. Thus, the anemic patient may be transfused with packed red blood cells, the thrombocytopenic patient with platelets, while the patient with clotting deficiency receives factor concentrates or fresh frozen plasma. Similarly, the polycythemic patient is venesected while the patient with leucocytosis may require leucopheresis. This simple approach is adequate in instances where the deficiencies or excesses are temporary or where alternatives do not exist.

Specific therapy for blood disorders are based on an understanding of the pathophysiology of the disorder and are directed at the causative factor, aiming to control or cure the underlying condition. Autoimmune conditions that cause immune thrombocytopenia and hemolytic anemia are best treated with agents that suppress or modulate the immune system, such as steroids. This approach is used in the treatment of other conditions that are thought to have an immune component, such as the hypoplastic anemias. Chemotherapeutic agents may be used as alternative treatment in autoimmune disorders by way of modulating the immune system.

In hematological malignancies, cytotoxic agents are used with the intention of eliminating malignant clones. Most agents are myelotoxic

and will induce significant marrow suppression. Chemotherapeutic doses are usually given to maximal tolerable doses which allows reconstitution of normal cell lines within reasonable periods, that will limit infective and bleeding complications. Repeated chemotherapeutic courses are however required to achieve the best possibility of elimination of malignant clones.

Stem cell transplants are therapeutic options for suitable patients in the majority of malignant and some non-malignant hematological disorders. Classical stem cell transplant employs myeloablative doses of cytotoxic agents which eliminate the marrow's capacity to repopulate itself. Donor stem cells are then introduced to repopulate the marrow. Besides the potential of cytotoxic agents to eliminate the abnormal clone that causes the disease, the donor stem cells may exert a graft-versus-tumor effect that has the added role of eliminating further disease. The downside of this therapy are problems with rejection of the donated graft or graft-versus-host disease.

In recent years, immunotherapy with the use of monoclonal antibodies has added a new dimension to therapy with many new agents being introduced to treat both malignant and non-malignant conditions. Monoclonal antibodies deliver targeted therapy against specific cells by recognizing cell surface markers. The benefits of this therapy is the ability to act on cells that mediate disease while limiting toxicity to other tissues. Monoclonal antibodies are increasingly being used in combination with chemotherapy for a variety of malignant and non-malignant conditions.

Thrombotic disorders are treated by antagonizing the coagulation pathways to prevent the recurrence or occurrence of a thrombotic event. An important aspect of managing these patients involves assessing the risk of recurrent events, in order to determine the duration of anti-coagulation and prophylaxis against them.

Bleeding as a result of single or multiple clotting factor deficiency is treated with replacement of relevant factors with either purified factors if available or fresh frozen plasma. Factor replacement therapy has however led to the development of inhibitors in hemophiliacs. Bypassing agents are required for these patients with inhibitors during bleeding episodes. Acquired inhibitors to clotting factors, which is most commonly to factor VIII, will require treatment with immunosuppression besides bypassing agents to control active bleeding. An important component of the treatment of congenital hemophilias should also be in the prevention

and reduction of morbidities arising from complications of bleeding episodes.

CONCLUSION

There is an immense variety of blood disorders with varying complexities, which will challenge the knowledge and intellect of any clinician confronted with a hematology patient. Our understanding of blood in physiology and disease will however allow the discerning doctor to dissect each condition with the assistance of an ever-increasing array of diagnostic tools, to arrive at a diagnosis that will stand to be logical and valid. In all instances, this should be a fascinating journey.

FURTHER READING

1. Hoffbrand AV, Lewis SM, Tuddenham EGD, *Postgraduate Haematology*, 4th ed., Butterworth–Heinemann UK, 1999.
2. Beutler E *et al*., *Williams Hematology*, 6th ed., McGraw–Hill USA, 2001.

32

Nutritional Anemias

William Hwang and Charles Chuah

The nutritional anemias are thus named as they are associated with deficiencies of substances that are found in the normal diet. Nutritional anemias may be due to:

- reduced intake;
- diminished absorption;
- increased utilization; or
- increased losses.

However, in developed countries, nutritional anemia is often not due to insufficient intake but usually due to impaired absorption or excess loss/utilization.

IRON DEFICIENCY ANEMIA

It is uncommon nowadays to find patients with 'classic' clinical features of iron deficiency (e.g. atrophic glossitis, koilonychia). Patients may also present with fatigue or reduced exercise capacity; but more often than not, patients are asymptomatic (after screening blood tests). Upon

detailed enquiry, symptoms of mental deficiency impaired control of body temperature[1] and the restless legs syndrome[2] may be found.

Iron deficiency is diagnosed when the serum ferritin is decreased or when a decreased serum iron is accompanied by a normal or increased total iron binding capacity (TIBC), with the transferrin saturation being less than or equal to 15%.[3] Though serum ferritin levels appear to reflect iron stores,[4] they may be spuriously raised in certain inflammatory states like hepatitis, disseminated infections and chronic inflammatory states like rheumatoid arthritis. Also, these values become uninterpretable if blood transfusion, oral iron therapy of more than 3 weeks or desferrioxamine, are given before the serum samples for the iron status are taken.[3]

Commercial immunoassays to detect the soluble truncated form of the transferrin receptor are now available. Decreased levels are found in patients with erythroid hypoplasia (aplastic or hypoplastic anemia, chronic renal failure); while increased levels are found in patients with erythroid hyperplasia (thalassemia major, sickle cell anemia, chronic hemolytic anemia, and iron deficiency anemia). If the other causes of erythroid hyperplasia have been ruled out, an increased concentration of the soluble transferrin receptor is a sensitive, quantitative measure of iron deficiency. The ratio of the soluble transferrin receptor over the log of the plasma ferritin concentration, the "transferrin receptor-ferritin index", is the most sensitive means of distinguishing between iron deficiency and anemia of chronic illness.[5–7]

The other main differential diagnoses include thalassemia minor (see Table 2) and hypoplastic anemia/myelodysplastic syndrome (reticulocyte count usually low with erythroid hyperplasia and increased iron stain in the bone marrow).

Table 1 Laboratory Tests to Distinguish between Anemia of Chronic Illness and Iron Deficiency Anemia

	Anemia of Chronic Illness	**Iron Deficiency Anemia**
Plasma iron concentration	Reduced/normal	Reduced
Plasma transferrin concentration	Reduced/normal	Increased
Transferrin saturation	Reduced/normal	Reduced
Plasma ferritin	Normal/increased	Reduced
Soluble transferrin receptor	Normal	Increased
Transferrin receptor/log ferritin	Low (<1)	High (>4)

Table 2 Indices to Help Differentiate Between Iron Deficiency and Thalassemia Minor in Patients with Microcytosis with Mean Cell Volume (MCV) < 70

	Thalassemia Minor	Iron Deficiency
Mentze's index: MCV/RBC	< 12	> 12
Green's index: (MCV × RDW)/ (Hemoglobin (g/dL) × 100)	< 73	> 73

RBC = red blood cell count ($\times 10^9$/L), RDW = red cell distribution width.

Iron deficiency may be caused by:

1) *Blood loss*. This is the commonest cause of iron deficiency, and in females the commonest cause of blood loss is excess menstruation. Even after menses has ceased, pathological uterine and cervical bleeding may occur, but in males and postmenopausal women, the commonest cause is source of blood loss is gastrointestinal. So where possible, the patient with unexplained iron deficiency anemia should undergo investigations to exclude gastrointestinal blood loss, the causes of which are discussed elsewhere in this book. Other than from the gastrointestinal tract, blood loss may also be from:
   a) the respiratory tract, manifesting as recurrent hemoptysis or as occult blood in the stools when swallowed;
   b) the genitourinary tract, from excess uterine bleeding or urinary blood loss (from tumors, infections or inflammations of the kidneys, ureters and bladder);
   c) self-inflicted phlebotomy, these may be hard to detect as the site of blood letting may be hidden; or
   d) excess phlebotomy (iatrogenic); or
   e) extracorporeal dialysis and associated trapping of blood.
2) *Urinary iron loss*. Iron deficiency is not only due to blood loss; *urinary losses* in the form of hemoglobinuria and hemosiderinuria due to intravascular hemolysis can also lead to iron deficiencies. The causes for this include:
   a) cardiac hemolysis from mechanical heart valves, intracardiac myxomas or cardiac patches;
   b) paroxysmal nocturnal hemoglobinuria; and
   c) march hemoglobinuria (from the lysis of red cells from trauma associated with prolonged marches or long distance running).

3) *Dietary insufficiency*. This is most commonly seen with complete vegans who take only vegetables and fruits. It is less commonly seen with lacto-vegans (who take milk) and lacto-ovo-vegans (who take eggs and milk). When asking a patient whether he takes meat, it is important to ask how often he does so (once a day or once a month) and how much. Complete vegans can overcome this deficiency by taking fruits with a higher content of iron (e.g. apples). Older patients may also have diminished intake as a result of neglect, apathy and reduced appetite. Such nutritional insufficiency may be due to neglect on the part of the caregiver, but may equally well be due to apathy or reduced appetite on the part of the patient who may be suffering from some mental or physical deterioration. If allowed to go on, the iron deficiency may further aggravate the mental and physical decline and cause home accidents due to giddiness from anemia, thus causing a vicious cycle of events. A careful exploration of the nutritional status and the underlying causes of any deficiency may uncover a deeper problem requiring treatment.
4) *Malabsorption*. Iron is mainly absorbed via the duodenum and less via the jejunum. Absorption is increased in iron deficiency with its heightened erythropoietic drive. It is decreased:
    a) in patients with lesions or had prior surgery to the duodenum;
    b) when there is excess ingestion of phytates in Chapattis (an Asian delicacy) or excess drinking of teas, both of which impair iron absorption by the duodenum;

These causes should be excluded and iron supplementation increased where necessary. Ascorbic acid improves iron absorption and vitamin C supplementation should be given whenever possible.

5) *Pregnancy*. This results in diversion of iron to the fetus for erythropoiesis, blood loss at delivery, and lactation. All pregnant women should therefore receive prophylactic iron supplementation.[8]
6) *Parasitic infestation*. This may result in blood loss (imbibed by hookworms or other parasites) or malabsorption. Eosinophilia may give a clue to this cause but stool tests for ova/cysts may be falsely negative. An empirical course of anti-protozoals may be necessary.

Treatment

Routine prescription of iron is not justified unless the cause is addressed where possible, even in the elderly.[11] When the cause of iron deficiency anemia is due to insufficient intake or when the cause of the iron deficiency anemia has been treated or is untreatable, iron replacement should be undertaken to replete the body iron stores and to restore normal erythropoiesis. Iron replacement may be undertaken via ferrous sulphate or ferrous gluconate tablets, bearing in mind that even with rigorous therapy the hemoglobin would usually rise by only 1 g/dL per week. In fact many patients would require up to 6 months of therapy. 150–200 mg of elemental iron should be given daily for replacement which may be given orally in 3–4 doses 1 hour before meals.

With proper treatment a brisk reticulocytosis may be observed in a week, which settles down soon after. Hemoglobin should be near normal by 2 months. However, oral therapy may fail, and this is often due to:

1) *Insufficient intake.* The patient may be non-compliant with the prescribed therapy due to neglect or confusion with regards to drug dosing. Patients who develop constipation with oral iron therapy may become non-compliant. For this reason, prophylactic administration of stool softeners or bulk laxatives may be necessary.
2) *Malabsorption*. Iron absorption is favored by factors such as acid and reducing agents keeping the iron soluble and in the ferrous rather than ferric state. Hence iron absorption may be diminished if iron is co-administered with antacids and H2 antagonists, which may be given for the same causes that precipitated the iron deficiency in the first place (e.g. peptic ulcer, gastritis).
3) In the elderly, iron absorption may be normal but the *iron utilization* is diminished. This is due to the impaired incorporation of iron for hemoglobin production in old age.[9]
4) *Continuing blood or iron loss* will offset any benefit of iron therapy and, where possible, the cause should be elucidated and arrested. Continuing iron supplementation without addressing the primary cause is not justified unless the medical condition does not permit otherwise.[9]
5) Lastly, it is important to evaluate if there has been an *incorrect diagnosis* or if the anemia is actually *multifactorial* rather than due to just iron deficiency.

If the cause for failure of oral iron therapy cannot be rectified (e.g. malabsorption or intolerance/ insufficiency of oral iron), it may be necessary to administer iron parenterally. Iron dextran is usually given at a test dose of 0.5 ml intravenously (IV) followed by a slow IV infusion rate of 1 mL/hr. Dosage is calculated as follows:

$$\text{Iron dextran (mg)} = \text{Hb deficit} \times \text{weight (kg)} \times 2.2$$

Iron dextran may also be given intramuscularly at the same dose, but this is not always preferred as it may result in considerable pain at the site of injection. Also, the injection site may also develop a secondary malignancy or a dark stain develop lasting 1–2 years. Intravenous administration is not without its problems as patients may develop thrombophlebitis, fever, rash, arthralgia, lymphadenopathy, splenomegaly, aseptic meningitis, anaphylactic reaction (<1%), shock and even death. As such parenteral iron therapy is not a measure to be undertaken with impunity.

VITAMIN B12 DEFICIENCY

Food of animal origin is the only source of vitamin B12 in humans, with the highest amounts found in liver and kidney, while vegetables and fruits have no vitamin B12. It is not destroyed by cooking and is absorbed at the duodenum and ileum. Absorption is mediated by gastric intrinsic factor. A normal daily diet contains 5–30 μg of vitamin B12 and daily losses through urine and faeces are 1–3 μg. The body can store 2–3 mg of vitamin B12, which can last 3–4 years.

Clinical Features

Patients can be asymptomatic, with the vitamin B12 deficiency being detected during investigations of an elevated mean corpuscular volume. Otherwise the clinical features can be attributed to:

- anaemia — lethargy, anorexia, dyspnea, angina, heart failure, glossitis;
- thrombocytopenia — easy bruising and bleeding; or
- leucopenia — infections.

Vitamin B12 deficiency can also cause bilateral peripheral neuropathy, subacute degeneration of the spinal cord and psychiatric disturbances.

Laboratory Features

Vitamin B12 deficiency can lead to characteristic hematologic features. The peripheral blood will have oval macrocytes, anisocytosis and poikilocytosis; elevated mean corpuscular volume of more than 100 fl; leucopenia with hypersegmented neutrophils (more than five nuclear lobes); and thrombocytopenia. Bone marrow changes include a hypercellular marrow; dissociation between nuclear and cytoplasmic development in the erythroblasts, i.e. the nucleus retains a primitive appearance while the cytoplasm attains maturation; giant and abnormally shaped metamyelocytes; enlarged hyperpolypoid megakaryocytes; and increased iron staining. Ineffective hematopoiesis can also give rise to apoptosis of nucleated red cells resulting in elevated unconjugated bilirubin, reduced haptoglobin, raised lactate dehydrogenase and raised urine urobilinogen and urine hemosiderin.

Causes of Vitamin B12 Deficiency

1) *Inadequate dietary intake*. This occurs in vegans who do not take meat, fish, eggs and other animal products, in patients with psychiatric illness, in situations of poverty and in infants born to severely vitamin B12-deficient mothers
2) *Pernicious anemia*. This is due to severe lack of intrinsic factor arising from gastric atrophy. It is slightly more common in females and has a peak incidence at 60 years old. Pernicious anemia (PA) can be associated with various organ-specific autoimmune diseases, e.g. thyroid disorders, Addison's disease, as well as vitiligo and premature graying of hair. There is also a familial association. Male patients seem to have a higher incidence of gastric carcinoma. Two types of antibodies can be found in PA. The parietal cell antibody is present in 90% of cases but this antibody is also frequently found in other patients. The other antibody is the intrinsic factor (IF) antibody. There are 2 types: Type 1 blocks the combination of IF and vitamin B12 and is found in 55% of patients with PA;[10] and Type 2, which is found in 35% of patients, attaches to IF regardless of combination to vitamin B12, thus preventing its attachment to ileal mucosa. A gastric biopsy would show atrophy of all layers of the body and fundus, loss of glandular element, absence of parietal and chief cells, replacement of mucous

cells, a mixed inflammatory cell infiltrate and intestinal metaplasia. Gastric secretion studies will reveal absent or low level of IF (up to 250 units per hour) compared to 1000–2000 units per hour in a normal adult.

3) *Malabsorption*. This can occur in a myriad of conditions:
   a) Gastrectomy resulting in reduced IF;
   b) Intestinal stagnant loop syndrome due to the colonization of the small intestine by fecal organisms as a result of jejunal diverticulum, enter-anastomoses, intestinal stricture or fistula, anatomical blind loop due to Crohn's disease or tuberculosis;
   c) Ileal resection;
   d) Tropical sprue;
   e) Fish tape worm infestation, which can be acquired by eating raw or partly cooked fish;
   f) Gluten-induced enteropathy;
   g) Chronic pancreatitis;
   h) HIV infection;
   i) Zollinger–Ellison syndrome;
   j) Total body irradiation or radiotherapy to the ileum;
   k) Graft-versus-host disease of the small intestine; or
   l) Drugs, e.g. phenytoin, metformin, colchicine and alcohol.
4) Congenital transcobalamin II deficiency.

Diagnosis of Vitamin B12 Deficiency

The normal range of serum vitamin B12 is 180 to 1000 ng/L. In patients with megaloblastic anemia, the serum vitamin B12 level is invariably less than 100 ng/L. Levels between 100 and 200 ng/L are considered borderline and can occur in mild vitamin B12 deficiency, pregnancy and folate deficiency. Serum homocysteine and methylmalonyl coA levels can be performed as supportive evidence. The levels are raised in vitamin B12 deficient states. However, chronic renal disease, alcohol and hypothyroidism can affect the levels. The cause of vitamin B12 deficiency has to be elucidated as well. This should be done with a proper clinical history detailing the patient's diet, drug history and previous operations. A Dicopac test can also be performed, subject to availability of the required isotopes. Two isotopes of cyanocobalamin are usually labeled with ^{57}Co and ^{58}Co and are given simultaneously. ^{57}Co is attached to IF and ^{58}Co on

its own. The excretion of the 2 isotopes are then compared in 24-hour urine samples.

Treatment of Vitamin B12 Deficiency

Initial therapy consists of six 1000 μg of hydroxycobalamin given via the intramuscular route within the first month. A response, as evidenced by a reticulocytosis, will be obvious by the third day of treatment. Leucocyte and platelet counts normalize after 7 days. Maintenance therapy is given subsequently every 3 months at the same dose. Therapy should be lifelong unless there are reversible causes. Blood transfusions should be avoided or kept to a minimum unless the patient is very ill. Potassium supplements should also be given, especially during the initial therapy.

FOLATE DEFICIENCY

Folate is involved in DNA and RNA synthesis. Most foods contain folate, with highest levels found in liver, yeast and vegetables. However, unlike vitamin B12, it is easily destroyed by heat. The body stores 10 mg of folate that can last for 4 months. The daily adult requirement is 100 μg. Folate is absorbed in the upper small intestine and is lost in the urine, sweat, skin and bile.

Causes of Folate Deficiency

1) *Dietary insufficiency*. This is common in the elderly, poor, alcoholics, patients with psychiatric disorder and infants fed solely on goat's milk.
2) *Malabsorption*. This usually occurs in patients with gluten-induced enteropathy or tropical sprue, as well as the other causes of malabsorption of vitamin B12 as mentioned above.
3) *Excess utilization*. The daily folate requirement of pregnant and lactating mothers is 400 μg and therefore they may develop folate deficiency if not given folate supplements.[11] Other conditions, which utilize an excess of folate, include chronic hemolytic anemias, myelofibrosis, malignancies, chronic inflammatory diseases and long-term dialysis.
4) *Drugs*. These include phenytoin, primidone, alcohol, methotrexate, pyrimethamine and trimethoprim. The last three drugs inhibit dihydrofolate reductase.

Diagnosis of Folate Deficiency

Clinical and laboratory features are similar to that of vitamin B12 deficiency. Serum folate is a useful tool but the level can be affected by recent diet and can be increased in severe vitamin B12 deficiency, acute renal failure and liver damage. Red cell folate, on the other hand, is less affected by recent diet and can be used as an adjunct to the diagnosis of folate deficiency.

Treatment of Folate Deficiency

Folate can be given orally in doses of 5–15 mg daily for at least 4 months. It is important to exclude concomitant vitamin B12 deficiency as treatment with folate alone can result in neuropathy. Long-term folate therapy is recommended in chronic hemolytic anemias and myelofibrosis. Folate should also be given prophylactically to pregnant females and patients on long-term dialysis.

OTHER NUTRITIONAL ANEMIAS

Anemia may also result from deficiencies of other vitamins and trace minerals (see Table 3). These include vitamin A, vitamin B6, riboflavin, niacin, ascorbic acid, vitamin E and copper. The mechanisms are varied and less well-defined. Measurements of levels are also not routinely performed.

Table 3 Interpretation of the Dicopac Test

^{57}Co /^{58}Co Ratio	^{57}Co Excretion	^{58}Co Excretion	Interpretation
0.7–1.2	greater than 14%	greater than 14%	Normal
0.7–1.2	less than 7%	less than 7%	B12 malabsorption not caused by lack of intrinsic factor
0.7–1.2	7–14%	7–14%	Some deficiency in B12 absorption not caused by lack of intrinsic factor
greater than 1.2	—	—	Lack of intrinsic factor
less than 0.6	—	—	Test should be repeated

Extracted from South Bend Medical Foundation Manual Laboratory.

Table 4 Anemia due to Rarer Nutritional Deficiencies[12]

Deficiency	Clinical Presentation	Clinical Setting
Vitamin A	Low MCV and MCHC with low serum iron and increased marrow iron	Nutritional deprivation
Vitamin B6 (pyridoxine)	Hypochromic anemia	Isoniazid administration, nutritional deprivation
Riboflavin	Pure red cell aplasia	Experimental nutritional deprivation
Niacin	Pellagra	Nutritional deprivation
Vitamin E	Abnormal erythrocyte morphology, thrombocytosis, edema	Low birth weight infants, chronic fat malabsorption
Copper	Microcytic anemia unresponsive to iron, vacuolated erythroid precursors in marrow	Nutritional deprivation, parenteral alimentation, excess dietary zinc
Anemia of starvation	Mild normocytic, normochromic anemia	Response to hypometabolic state
Anemia of protein deficiency	Normocytic, normochromic anaemia, variations in red cell size and shape	Infants and children with protein-calorie malnutrition
Alcoholism	Macrocytosis	Usually with concomitant folate or iron deficiency

REFERENCES

1. Schultz BM, *et al*., Iron deficiency in the elderly, in *Baillieres Clin Haematol*, 1987.
2. O'Keeffe ST, *et al*., Iron status and restless legs syndrome in the elderly, in *Age Ageing* 1994.
3. Williams WJ, Haematology in the aged, in *William's Haematology* **8**:93–102, 2001.
4. Lipschitz DA, Mitchell CO, Thompson C, The anaemia of senescence, *Am J Hematol* **11**:47, 1981.
5. Cook JD, The measurment of serum transferrin receptor, *Am J Med Sci* **318**:269–276, 1999.
6. Ferguson BJ, Skikne BS, Simpson KM, Baynes RD, Cook JD, Serum transferrin receptor distinguishes the anemia of chronic disease from iron deficiency anemia, *J Lab Clin Med* **119**:385–390, 1992.

7. Punnonen K, Irjala K, Rajmaki A, Serum transferrin receptor and its ratio to serum ferritin in the diagnosis of iron deficiency, *Blood* **89**:1052–1057, 1997.

8. King DE, Subal J, Muncie HL Jr, *et al*., Prescribing postpartum iron supplementation: a survey of practising obstetricians, *South Med J* **79**:674, 1986.

9. Calvey HD, *et al*., Gastrointestinal investigations for anaemia in the elderly: a prospective study, *Age Ageing*, 1987.

10. Kapadia CR, Donaldson RM, Disorders of cobalamin (vitamin B12) absorption and transport, *Annu Rev Med* **36**:93, 1985.

11. Strieff RR, Little AB, Folic acid deficiency in pregnancy, *N Engl J Med* **276**:776, 1967.

12. Williams WJ, Anemia due to other nutritional deficiencies, in *William's Haematology* **39**:471–446, 2001.

FURTHER READING

1. Beutler E, Lichtman MA, Coller BS, Kipps TJ, Seligsohn U, *Williams Haematology*, 6th ed., McGraw–Hill Publishing Company, USA, pp. 271–721, 2000.

2. Hoffbrand AV, Lewis SM, Tuddenham EGD, *Postgraduate Haematology*, 4th ed., Butterworth Heinemann, UK, pp. 47–67, 1999.

33

The Chronic Anemia

Linn Yeh Ching and Charles Chuah

INTRODUCTION

The presence of anemia is based on a lower than normal hemoglobin concentration, red blood cell count or packed cell volume. The range of normal values may vary depending on an individual's age, sex or even with environmental conditions. Women of childbearing age have normal values that are 10% lower than men, while individuals living at high elevations have higher hemoglobin concentrations than people living at sea level.

Under normal circumstances, symptoms are not usually experienced unless the hemoglobin falls to less than 7–8 g/dL. It is, however, important to detect and investigate mild anemia as it may herald the presence of a potentially treatable underlying disease.

An algorithm on the approach to investigation and diagnosis of anemia is shown in Fig. 1. This goes by classifying anemia into microcytic, normocytic and macrocytic based on the mean cell volume (MCV).

The more common types of chronic anemia are discussed.

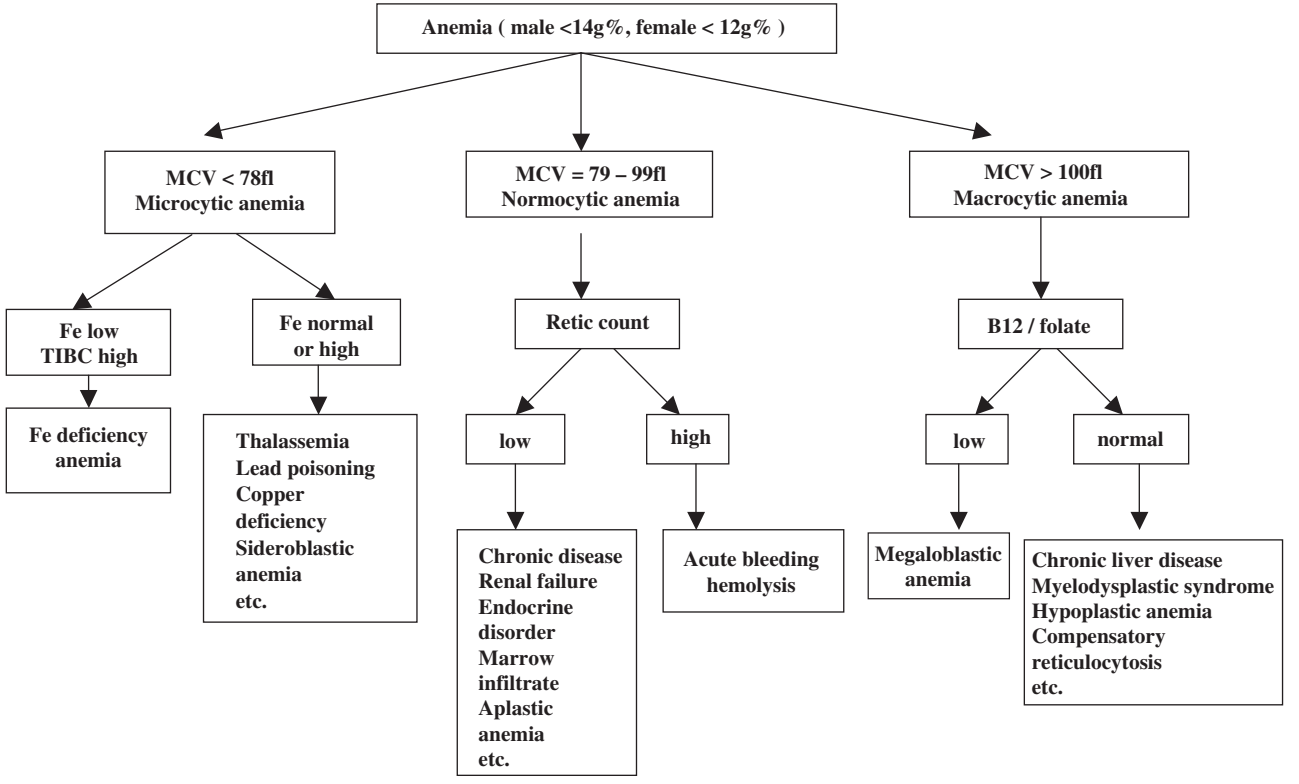


Fig. 1 Approach to investigation and diagnosis of anemia.

IRON DEFICIENCY ANEMIA

This is characterized by decreased or absent iron stores, low serum iron, low transferrin saturation together with a hypochromic and microcytic anemia. This is due to either inadequate intake or excessive loss, or both.

Causes

Dietary deficiency. In infants, this is due to insufficient iron supplement in milk diet. In older children and adults, diet may be inadequate in iron. Liver is rich in iron; beans, peas, red meat, poultry and fish to a lesser amount. With the exception of legumes, fruits and vegetables are poor in iron. The recommended daily allowance of iron for adults is 10 mg, and for menstruating females, is 18 mg.

Intestinal malabsorption. In post gastrectomy or in other malabsorption syndrome, iron can be deficient together with other nutrients.

Blood loss. In females of childbearing age, menstrual bleeding is a common cause of iron deficiency. Pregnancy at close interval is another cause. Overt or occult gastrointestinal loss is an important cause, particularly in men or postmenopausal women. Peptic ulcer, hiatus hernia, gastritis, Meckel's diverticulum, hemorrhoids, vascular anomalies and neoplasm are causes to exclude. Other sources of bleeding include intravascular hemolysis such as in paroxysmal nocturnal hemoglobinuria, prosthetic valves, loss during hemodialysis, recurrent hemoptysis, even regular venesection in blood donors or patients.

Laboratory tests to confirm the diagnosis of iron deficiency include a low serum iron together with a high total iron-binding capacity (TIBC). Iron itself may also be low in anemia of chronic disease and malignancy, but this occurs in the presence of a low TIBC. The serum ferritin is a good reflection of iron store but in many disease states, e.g. chronic illness, liver disease, infection, malignancy, it may be falsely normal or high as it is an acute phase reactant.

Iron deficiency anemia is easily treated with oral iron supplement; the more important task is to detect and treat the underlying cause of iron loss. With adequate oral replacement the hemoglobin can be expected to improve by 1 g% per week. Parenteral iron is not more effective and is associated with high hypersensitivity reaction. It is only indicated in the occasional patient who has malabsorption of iron or severe gastrointestinal intolerance to various oral iron preparations.

Thalassemia

This is due to an inherited defect in the rate of synthesis of one or more globin chain, leading to unbalanced globin chain synthesis and defective hemoglobin production and hemolysis as a result of excess globin subunit.

There is a wide variety of thalassemic genetic defects. For clinical purposes based on severity, these are classified into thalassemia major, intermedia and minor.

Thalassemia Major

This is due most commonly to absent or minimal synthesis of β globin chain as a result of mutation of both β genes. Combination of β thalassemic trait with other β gene defect, e.g. HbE, is also seen locally. Excess α chain precipitates and causes hemolysis. Severe anemia during infancy occurs together with other signs and symptoms, such as hepatosplenomegaly, thalassemic facies and tendency to fracture. Multiple transfusions are required to maintain life, and this eventually causes iron overloading leading to liver damage, endocrine failure, heart failure and susceptibility to infection. Hypertransfusion to maintain >10 g% and aggressive chelation therapy has been shown to prevent complications of end organ damage. Bone marrow transplantation has become a treatment of choice for patients with matched siblings, with long-term success as good as 80%, especially in patients without liver impairment. Screening and genetic counseling with early antenatal diagnosis and abortion to prevent the birth of a thalassemia major child is the only way to resolve this unfortunate condition.

Thalassemia Intermedia

This is clinically defined as thalassemia with a Hb between 7–10 g% which does not require regular transfusion. A variety of genetic defects include coinheritance of homozygous β thal with α thal, homozygous mild β thal, homozygous δβ thal and HbH disease. Complication is milder than thalassemia major and blood transfusion may be required e.g. during hemolytic or aplastic crisis.

Thalassemia Minor

This is a common and asymptomatic condition where the blood count shows no or only mild anemia. A variety of genetic defects, including heterozygous β thal, α thal trait with one or two gene deletion, δβ thal trait, all present as thalassemia minor. Full blood count and blood film provide the initial suspicion of thalassemia with hypochromia and microcytosis. There are formulae to differentiate between iron deficiency and thalassemia minor with efficiency of discrimination up to over 90%. Hb electrophoresis shows a raised HbA2 in the range of 4–7% in β thal trait whilst in α thal trait HbA2 might be low or normal. The definitive diagnosis of α thal trait depends on the molecular method.

HEMOLYTIC ANEMIA

There are many causes of hemolytic anemia, best classified into hereditary and acquired causes as follows:

1) Hereditary:
   a) membrane defects : e.g. spherocytosis;
   b) enzymopathy: e.g. G-6PD deficiency when under oxidant stress; and
   c) hemoglobinopathy and thalassemia
2) Acquired:
   a) immune–mediated: autoimmune hemolysis, delayed hemolysis post transfusion due to allo-antibodies;
   b) mechanical hemolysis: giant aneurysm, microangiopathic hemolysis (HUS and TTP), disseminated intravascular coagulation;
   c) infection and toxin: falciparum malaria, clostridium, snake venom;
   d) hypersplenism, and
   e) acquired membrane abnormality: paroxysmal nocturnal hemoglobinuria (PNH).

In general, the hereditary hemolytic anemia are due to intracorpuscular causes, whereas acquired conditions are due to extracorpuscular causes, with the exception of PNH.

The prominent features in hemolytic anemia, in addition to that due to low hemoglobin, share common signs such as jaundice, and often

hepatosplenomegaly, to varying degrees. All will have in common the following laboratory findings as evidence of hemolysis:

1) reticulocytosis (unless marrow is diseased too or transiently suppressed);
2) raised LDH, which is an enzyme released from red cells;
3) raised unconjugated bilirubin;
4) reduced haptoglobin, which is used to bind free hemoglobin, and
5) hemoglobinuria and hemosiderinuria in intravascular hemolysis.

DIAGNOSTIC TESTS TO DIAGNOSE THE CAUSE OF HEMOLYSIS

1) Peripheral blood film: spherocytes are seen in hereditary spherocytosis and immune–related hemolysis; fragmentation is seen in micro- or macro-angiopathic hemolysis and Heinz bodies in unstable hemoglobinopathy. Autoagglutination is seen in immune–mediated hemolysis. Polychromasia reflects reticulocytosis.
2) Enzyme assay: e.g. G-6PD level.
3) Direct Coomb's test (DCT), which will be positive for immune–mediated hemolysis; antibodies may be warm or cold. Other autoimmune markers may be positive.
4) He electrophoresis: to diagnose the various hemoglobinopathy or thalassemia.
5) Osmotic fragility test will show an increased fragility in hereditary spherocytosis and decreased fragility in thalassemia.
6) Flow cytometry for membrane markers of PNH: CD55, CD59.

AUTOIMMUNE HEMOLYTIC ANEMIA (AIHA)

This is one acquired hemolytic condition commonly seen in hemotological practice. The hallmark of AIHA is the presence of positive DCT, a test that demonstrates coating of immunoglobulin or complement on the red cell surface. Warm antibodies that are active at body temperature are IgG, often idiopathic but could be associated with lymphoid neoplasm, systemic lupus erythematosus or drugs such as methyldopa or penicillin group of drugs. Cold antibodies active at a temperature lower than body

temperature are usually IgM that binds complement. This can also arise in the background of lymphoid malignancy or MGUS, or in response to infection, classically mycoplasma and infectious mononucleosis.

The degree of hemolysis and response to treatment is variable. For warm AIHA, initial treatment is glucocorticoids, e.g. prednisolone at 1 mg/kg. Splenectomy as a second line is indicated for those who are steroid refractory or dependent on steroids, resulting in undesirable side-effects, and is effective in two-thirds of patients. Third-line treatment includes a variety of immunosuppressants, e.g. azathioprine, cyclosporine or cyclophosphamide. The underlying lymphoid neoplasm needs to be treated concomitantly. Glucocorticoids or splenectomy are usually less effective for cold AIHA.

ANEMIA OF CHRONIC DISEASE (ACD)

The anemia of chronic disease (ACD) is one of the most common forms of anemia encountered in clinical practice. This is because it is associated with a myriad of conditions which include:

1) Chronic infections, e.g. osteomyelitis, infective endocarditis, tuberculosis, abscesses, bronchiectasis, chronic urinary tract infections, AIDS;
2) Chronic inflammatory conditions, e.g. rheumatoid arthritis, polymyalgia rheumatica, systemic lupus erythematosus, inflammatory bowel diseases;
3) Malignancy, e.g. carcinoma, lymphoma, myeloma; and
4) Others, e.g. congestive cardiac failure, ischemic heart disease.

It is a normochromic, normocytic or mildly hypochromic, microcytic anemia. The hemoglobin level is usually not less than 9 g/dL unless there are other concomitant causes of anemia. Serum iron and total iron-binding capacity are both low while serum ferritin can be normal or elevated.

Pathogenesis of ACD

Many mechanisms have been implicated in the pathogenesis of ACD. There seems to be a mild decrease in red cell lifespan. There is also a disturbance of erythropoiesis due to reduced sensitivity to physiological

erythropoietic stimuli. Other possible mechanisms include decreased iron utilization and inappropriately low plasma erythropoietin levels. ACD is associated with increased plasma levels of TNF-alpha, IL-1 and IL-6 and these cytokines are known to inhibit *in vitro* erythropoiesis.

Treatment of ACD

The fundamental principle in the therapy of ACD lies in the treatment or alleviation of the underlying disease. This is especially so as the severity of anemia corresponds to the activity and severity of the chronic illness. Pharmacological doses of recombinant erythropoietin have been used successfully in rheumatoid arthritis, cancer and myeloma. However it is also essential to exclude and correct other treatable causes of anemia e.g. iron therapy in coincidental iron deficiency through blood loss or dietary insufficiency.

ANEMIA DUE TO NON-HEMATOLOGIC SYSTEMIC ILLNESS

Anemia of Malignancy

Cancer can cause anemia in the following ways:

1) Anemia of chronic disease.
2) Blood loss — gastrointestinal or gynecological neoplasms.
3) Hemolysis — autoimmune (warm or cold), which can be associated with lymphoproliferative disorders, ovarian tumors; or microangiopathic, which can be associated with mucin-secreting adenocarcinoma of stomach, breast or lung.
4) Red cell aplasia — thymoma, chronic lymphocytic leukemia or non-Hodgkin's lymphoma.
5) Leucoerythroblastic anemia — marrow infiltration by tumor or active bone marrow response to peripheral consumption, e.g. disseminated intravascular consumption, septicemia or hypersplenism.
6) Chemotherapy–induced disturbance of DNA synthesis leading to megaloblastic or myelodysplastic changes.
7) Folate deficiency due to poor appetite.
8) Vitamin B12 deficiency in pernicious anemia associated with gastric carcinoma.

Anemia of Renal Disease

Anemia is a common manifestation of chronic renal disease. It is primarily due to low endogenous erythropoietin levels but iron deficiency from blood loss and folate deficiency has to be considered as well. Treatment is with recombinant erythropoietin which can be administered by subcutaneous, intravenous or intraperitoneal routes. Initial subcutaneous doses begin at 50–75 units/kg body weight/week in 2–3 divided doses. The aim is to correct the anemia to a level of 10–12 g/dL at a rate of 1 g/dL/month. Concomitant iron therapy should be given routinely given the fact that subclinical iron deficiency and impaired mobilization of stored iron is common.

Anemia of Endocrine Disorders

The hormones released by specific endocrine organs all play a role in modulating the rate of red cell production. Therefore, dysfunction of the pituitary, thyroid, adrenal glands and gonads can give rise to anemia, and subsequent correction of the deficient hormone can restore normal hematopoiesis. Autoimmune endocrine disorders can also be associated with pernicious anemia.

Anemia due to Hematologic Diseases

Hematologic conditions can give rise to anemia through many ways. These can be due to bone marrow infiltration, e.g. myeloproliferative disorders or myelodysplastic syndromes, suppression from the underlying disease process, chemotherapy or radiotherapy, and anemia of chronic disease. Anemia can also develop through other ways in the following conditions:

1) Chronic lymphocytic leukemia — autoimmune hemolytic anemia, pure red cell aplasia.
2) Red cell aplasia — idiopathic, viral infections, e.g. parvovirus, autoimmune conditions, e.g. systemic lupus erythematosus, thymoma, exogenous erythropoietin–induced antibodies.
3) Hemoglobinopathies — aplastic crises following infections.

Treatment again is of the underlying condition, with blood transfusion and exogenous erythropoietin when indicated.

FURTHER READING

1. Beutler E *et al.*, *Williams Hematology*, 6th ed., McGraw–Hill USA, pp. 271–721, 2001.
2. Hoffbrand AV, Lewis SM, Tuddenham EGD, *Postgraduate Haematology*, 4th ed., Butterworth–Heinemann UK, pp. 68–90, 675–693, 1999.

34

The Leukemias

Goh Yeow Tee and Wong Gee Chuan

INTRODUCTION

The hematological malignancies comprise a heterogeneous group of proliferative disorders involving both the lymphoid and myeloid systems. Recent advances in the fields of cellular and molecular biology have resulted in the recently updated World Health Organization (WHO) Classification[1] of these malignancies, which should facilitate further progress in research and development of new treatment strategies in this area.

While new chemotherapeutic agents and optimal treatment schedules continue to be developed, the new millennium has seen an emergence in the role of new biologic response modifiers. Alpha interferon has been in use since the last decade. Advances in biotechnology have led to the development of monoclonal antibodies such as a chimeric anti-CD 20 monoclonal antibody (Rituximab), anti-CD 33 monoclonal antibody and the more recent use of the tyrosine kinase inhibitor (Glivec) in patients with Chronic Myeloid Leukemia.

ACUTE MYELOID LEUKEMIA (AML) IN ADULTS

Acute leukemia is a malignant disorder arising in immature hematopoietic cells. The leukemic process can begin in immature cells not committed to a particular lineage or in more mature cells committed to develop along a particular lineage pathway. Acute leukemia is broadly divided into acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL).

There has been important progress in the treatment of AML in patients under 60 years of age. A remission rate of 80% can be achieved by some schedules and 40–45% of patients will survive. Improvement in older patients have, however, been less detectable. Newer approaches include agents involved in resistance modulation and FLT 3 inhibitors.

Diagnosis

Most patients with AML present with progressive fatigue and commonly have evidence for infection or bleeding diathesis. The white blood cell (wbc) count is usually elevated but may be normal or low. It is not necessary for leukemic blast cells to be present in the peripheral blood for the diagnosis of AML to be made. Anemia is usually present and can cause cardiopulmonary symptoms. Thrombocytopenia is also common and patients may present with petechiae, ecchymosis, hematuria or gastrointestinal bleeding. Hemorrhage in the central nervous system (CNS) is rare but often fatal. It occurs most commonly in patients who present with very high wbc counts. Signs of extramedullary involvement (e.g. gingival hypertrophy, skin infiltration) or meningeal leukemia may be seen in the monocytic and myelomonocytic variants.

Peripheral blood smear may reveal dysplastic changes in the red cells, granulocytes and platelets. The bone marrow is usually hypercellular and often replaced with leukemic blast cells. The rapid turnover of leukemic blasts produces a hypermetabolic state with increased levels of lactate dehydrogenase (LDH) and uric acid.

Classification

The traditional classification of AML is based on morphologic and histochemical characterization of the peripheral blood and bone marrow cells. New technologies, including flow cytometry, cytogenetics and analyses at molecular level have also been used.

Morphologic classification

The FAB (French-American-British) classification relates the morphology of leukemia cells to their presumed hematopoietic counterparts:

- M0 — no visible myeloid features;
- M1 — undifferentiated myeloblastic leukemia;
- M2 — differentiated myeloblastic leukemia with myeloblasts and Auer rods;
- M3 — promyelocytic leukemia;
- M4 — myelomonocytic leukemia;
- M5 — monocytic leukemia;
- M6 — erythroleukemia or Di Guglielmo's Syndrome; and
- M7 — megakaryoblastic leukemia.

Cytochemical stains are used to characterize leukemic cells and aid in morphologic diagnosis: enzymatic (e.g. peroxidase, esterases) and non-enzymatic [e.g. sudan black, chloroacetate esterase and non-specific esterases and periodic acid schiff (PAS)].

Cytogenetics

Clonal cytogenetic abnormalities are present in the majority of patients with newly diagnosed AML. Certain cytogenetic abnormalities may be found in both AML and ALL, but some are found exclusively in AML, e.g. t(8;21) (q22;q22), 16q22, t(15;17), trisomy 8 and 11q23. Abnormalities of chromosomes 5 and 7 (del 5q, del 7q) are commonly seen in therapy-related AML or in elderly patients with *de novo* AML. In general, patient with a normal karyotype have a better prognosis than patients with single karyotypic abnormalities who have a better prognosis than patients with multiple cytogenetic abnormalities.

Immunophenotypic Analysis

Immunophenotyping often provides a valuable tool for detection of acute leukemia and is especially useful for detection of minimal residual disease.

Treatment

Induction therapy

Treatment consists of an anthracycline or a closely related analog, e.g. daunorubicin, idarubicin or mitoxantrone, together with cytosine arabinoside. A typical induction regimen for AML consists of daunorubicin 45–60 mg/m^2/day for 3 days, with cytosine arabinoside 100–200 mg/m^2 by continuous infusion for 7 days. Overall remission rate is approximately 65%. As many as 25–40% of patients require 2 courses of induction chemotherapy to attain complete remission. The length of hospitalization for induction chemotherapy is 4–6 weeks, during which time the patient has a 2–4 week period of absolute pancytopenia. The patient is at risk of infections and bleeding during pancytopenia. After recovery of blood counts, bone marrow studies will be done to document complete remission. Complete remission is defined as normal full blood count, with no circulating peripheral blasts, absence of signs and symptoms of disease and bone marrow aspirate showing $< 5\%$ blasts.

Post-remission therapy

Once a patient has entered a complete remission (CR), post-remission therapy is given to improve long-term survival. Several treatment modalities are now being used.

1) *Conventional consolidation* or *post-remission chemotherapy* typically consists of high doses of cytosine arabinoside with or without an anthracycline or other anti-leukemic agents.
2) *Allogenic bone marrow transplantation (BMT).* This is effective therapy in first complete remission for younger patients who have a histocompatible sibling. Relapse rate following transplantation is low (15–20%) but the peritransplant mortality, primarily from graft-versus-host disease and interstitial pneumonitis, may be as high as 25–30%.
3) *Autologous BMT.* The procedural mortality from autologous BMT is about 10%. Disadvantages include the absence of immunologic effects of graft-vs-leukemia (anti-leukemic properties) and the potential for reinfusion of leukemic cells.

The choice of post-remission therapy depends on various factors including the patient's prognosis at diagnosis, bone marrow cytogenetics and the patient's and physician's acceptance of risks associated with high-dose therapy or transplant.

Relapsed and Refractory AML

The majority of AML patients relapse. The first step is reinduction into second complete remission. Allogeneic BMT can probably cure 30–40% of patients who are transplanted in second complete remission. Patients who relapse after a very short first remission or refractory AML patients who never attained a complete remission following standard therapy (known as primary treatment failures) are best entered into experimental clinical trials to determine the efficacy of newer agents as well as combination of established agents to evaluate the best regimen.

ACUTE PROMYELOCYTIC LEUKEMIA (APL, AML M3)

Although acute promyelocytic leukemia (APL) is properly classified as a subtype of acute myeloid leukemia (AML) [FAB M3], it represents a uniquely treatable, clinical and molecular disease entity. Treated appropriately, APL is now curable in the vast majority of patients. In Asians, the relative incidence of APL may be as high as 20% to 30% of all AML. The vast majority of APL cases arise *de novo*, but secondary APL have also been reported. Unlike other subtypes of AML, APL in the elderly is highly curable. Thus, in the absence of a clinical contraindication, all patients diagnosed with APL, regardless of age, should be offered standard therapy with all-trans retinoic acid (ATRA) and chemotherapy. ATRA, together with chemotherapy, has increased the cure rate to 75% compared to 35% with chemotherapy alone.

Diagnosis of APL

The morphologic diagnosis of APL is generally straightforward. Cytogenetic or molecular documentation of the pathognomic t(15;17) will be confirmatory in most. Therapy should not be delayed pending cytogenetics in cases with diagnostic morphology. APL can also be caused by translocations other than the classic t(15;17).

The defining molecular event in APL is the disruption of the retinoic acid receptor α (RARα) gene at 17q21 and its fusion with partner genes. The PML-RARα fusion gene is the molecular counterpart of the classic t(15;17).

Treatment

Central role of ATRA in APL

The use of ATRA in the induction therapy doubles the percentage of long-term disease-free survivors. This is due to a reduction in the relapse rate. ATRA alone will induce CR in the majority of newly diagnosed patients with APL[2–5] as well as in a significant fraction of APL patients who either relapse after or are refractory to chemotherapy regimens.[4,6] However the remissions are short-lived and ATRA alone will not cure either *de novo* or relapsed APL. ATRA has a specific differentiating effect on leukemic promyelocytes, which then undergo spontaneous cell death. The majority of patients with APL should be long-term disease-free survivors when treated with the appropriate combination of ATRA and chemotherapy.

Coagulopathy

Release of procoagulant activity from APL blasts plays a major role in the pathophysiology of the APL-associated disseminated intravascular coagulation (DIC). The maturation process induced by ATRA appears to prevent or lessen the procoagulants released by APL blasts. The early death rate of APL patients treated with ATRA remains approximately 10% and many patients continue to die from bleeding. Plasma infusion, cryoprecipitate and platelet support should be instituted.

Retinoic acid syndrome (ATRA syndrome)

This is the primary life-threatening complication of ATRA induction therapy, where there is development of a distinct respiratory distress syndrome as manifested by fever, dyspnea, weight gain, clinical and radiographic signs of third space fluid accumulation, and occasionally, renal failure and hypotension. Hyperleukocytosis is common in patients who develop RAS but this can also develop in patients with normal or

low wbc counts. Treatment consists of intravenous dexamethasone 10 mg every 12 hours for at least 3 days, beginning at the first signs or symptoms of ATRA Syndrome.

Consolidation therapy

All patients who achieve CR should receive at least 2 cycles of consolidation chemotherapy with anthracycline/cytarabine-based regimens.

Maintenance therapy

Both the North American Intergroup Trial[3] and the European APL 93 trial support the use of maintenance therapy with either ATRA, chemotherapy or both. PCR monitoring for PML/RARα should be done in the first 2 years and intervention during molecular relapse is safer than treatment in clinical relapse.

Salvage therapy

Approximately 20–25% of APL patients who enter CR will relapse. The best strategy for salvage therapy in APL has not been defined. Although 20% of APL patients previously treated with ATRA appear to retain sensitivity to this agent, single agent ATRA rarely induces PCR negativity and cannot maintain remission. ATRA, combined with chemotherapy, can induce a CR rate of 95% as reported by a European group with a disease-free survival of 3 years of 54%. Other modalities include arsenic trioxide, autologous and allogeneic BMT.

ACUTE LYMPHOBLASTIC LEUKEMIA (ALL)

Acute Lymphoblastic Leukemia (ALL) is a common hematologic malignancy in childhood. In adults, there is a bimodal age distribution with a peak in early adulthood and a second peak after the fifth and sixth decade of life.

Diagnosis

Patients usually present with signs and symptoms resulting from bone marrow failure, namely pallor, tachycardia and fatigue due to anemia,

petechiae and other bleeding manifestations due to thrombocytopenia and infections due to neutropenia. Direct organ infiltration by leukemic blasts may lead to generalized lymphadenopathy, testicular swelling and hepatosplenomegaly. Central nervous system involvement may result in severe headaches and cranial nerve palsies.

Bone marrow aspiration and biopsy for morphology together with cytochemical stains is required for the initial diagnosis of lymphoid blasts. These blasts can be subtyped as L1, L2 and L3 blast according to the FAB classification, but only the subtype L3 is clinically significant and indicative of a mature B-ALL. Immunophenotyping by means of flow cytometry is useful for confirmation of morphology and further subtyping into B lineage and T lineage ALL. B lineage ALL is further subtyped according to the expression of differentiation markers into Pro-B or Pre-Pre-B ALL (expresses only HLA Dr, Tdt and CD19), common ALL (expresses CD10), Pre-B ALL (expresses cytoplasmic immunoglobulin) and mature B-ALL (express surface membrane immunoglobulin). T lineage ALL can be subdivided into Pre-T and T-phenotypes.

The common cytogenetic abnormalities in ALL are clonal translocations t(9;22), t(4;11), t(8;14), t(1;19) and t(10;14) and structural abnormalities (9p, 6q and 12p abnormalities). The presence of these abnormalities especially that of t(9;22) which is also known as Philadelphia chromosome positive ALL is indicative of a poor prognosis.

Treatment

Standard *induction therapy* for most studies usually comprise a combination of vincristine, prednisolone and an anthracycline. L-Asparaginase, cyclophosphamide and cytosine arabinoside do not improve the remission rate but may influence the length and quality of remission. Cyclophosphamide is useful for patients with T cell ALL.[7] *Intensification therapy* refers to the use of high-dose chemotherapy with or without the addition of new agents to eliminate clinically undectable residual leukemia after induction chemotherapy. Intensification schedules include the use of teniposide, etoposide, high-dose Ara-C, high-dose methotrexate amongst others. A recent randomized trial by the UK MRC demonstrated a reduction in relapse risk in patients receiving early and/or late intensification. *Maintenance therapy* in ALL usually comprises low-dose 6-mercaptopurine and methotrexate with periodic pulses of vincristine/prednisolone or

other agents. The optimal duration of maintenance therapy remains unresolved but the average duration in many protocols is 2 years. Attempts at omission of the maintenance regimen have resulted in poorer outcome in the arm without maintenance.

Supportive Measures

The majority of induction failures are due to deaths from infective complications. Rigorous hygiene procedures, reverse isolation, antibiotic and antifungal prophylaxis with prompt management of febrile neutropenia are essential measures in reducing morbidity and mortality from infective complications. The use of hematopoietic growth factors is also a valuable adjunct especially with the use of aggressive chemotherapeutic protocols. Although it has not been known that leukemia-free survival is improved with the use of growth factors, its use significantly enhances neutrophil recovery with a reduction in the duration of febrile neutropenia and infective episodes. Other general measures include adequate hydration to ensure a urine output of at least 100 ml/hr, allopurinol to reduce the risk of urate nephropathy and prophylactic platelet transfusions. Patients presenting with very high white cell counts ($> 200 \times 10^9$/L) especially those with concomitant signs of leucocostasis should be considered for urgent leucopheresis. Hydroxyurea (1–3 g stat) with/without intravenous cyclophosphamide should also be started to ensure prompt reduction of the blast counts. Central nervous system leukemia should be treated promptly and all adult ALL patients should receive intrathecal chemoprophylaxis to prevent central nervous system relapse.

Bone Marrow Transplantation

Allogenic bone marrow transplantation from a matched sibling may be considered for high-risk patients in first complete remission. These high-risk patients include any patient with a high initial total white cell count ($> 30 \times 10^9$/L) or late achievement of complete remission > 4 weeks or Pre-Pre-B and mature B phenotype or associated cytogenetic abnormalities especially Philadelphia positive ALL (t(9;22)). Patients with relapsed and refractory ALL should also be considered for allogenic transplant after salvage therapy, which usually comprises a combination of high-dose Ara-C with other agents such as idarubicin, teniposide or etoposide.

Future Directions

Advances in detection of minimal residual disease may lead to new strategies in stratification of post-remission therapy. Monoclonal antibodies such as anti-CD20 (rituximab) or anti-CD52 (alemtuzumab) have demonstratable activity in certain phenotypes of ALL and their use in induction, consolidation, eradication of minimal residual disease and relapsed ALL is being widely evaluated.

CHRONIC MYELOID LEUKEMIA (CML)

Introduction

Chronic myeloid leukemia (CML) is a malignant disease of human hematopoietic stem cells. The disease is characterized by an excessive proliferation in primarily myeloid and, to a lesser extent, the megakaryocytic and erythroid lineages. A characteristic cytogenetic abnormality, the Philadelphia (Ph) chromosome, is present in the hematopoietic stem cells in more than 90% of the cases.

Diagnosis

CML accounts for 7–15% of adult leukemias. The median age at presentation is 50–60 years, but the disease occurs in all age groups. A bimodal age distribution with a second peak during the second and third decade has been noted by some Asian investigators including the authors. In chronic phase, CML is frequently asymptomatic. Patients with symptoms usually have a gradual onset of fatigue, anorexia, weight loss, increase sweating, left upper quadrant discomfort, and early satiety because of splenomegaly. The diagnosis may not be suspected until splenomegaly or an abnormal blood count is found on routine examination.

The most common peripheral blood feature of CML is an elevated WBC count, usually above 25×10^9/L, and frequently above 100×10^9/L. The WBC differential usually shows granulocytes in all stages of maturation, from blasts to mature granulocytes, which look morphologically normal. The platelet count is elevated in 30–50% of patients, and it may be greater than 1000×10^9/L in some patients. Basophils and eosinophils are frequently elevated. Leukocyte alkaline phosphatase (LAP) activity is reduced in almost all patients in chronic phase at diagnosis.

The natural history of CML is progression of from a benign chronic phase to a rapidly fatal blast crisis within 3 to 5 years. The blast crisis is often preceded by an accelerated phase in which increasing doses of hydroxyurea and busulphan are required to lower the neutrophil count. Blastic phase CML resembles acute leukemia. Its diagnosis requires the presence of at least 30% of blasts in the bone marrow or peripheral blood. Patients in blastic phase usually die within 3 to 6 months.

Molecular Pathophysiology

In 1960, Nowell and Hungerford described the Philadelphia chromosome (Ph), which is the result of reciprocal translocation between 9 and 22 and can be found in more than 90% of patients with CML. Since the 1980s it has become clear that the Ph translocation results in the fusion of the BCR (breakpoint cluster region) gene on chromosome 22 to the ABL (Ableson leukemia virus) gene on chromosome 9. The hybrid BCR-ABL oncogene produces an abnormal 8.5 kb mRNA which encodes for a 210 kD fusion protein. This protein exhibits abnormal tyrosine kinase activity, but the precise mechanism by which the abnormal protein contributes to leukemogenesis remains uncertain, although it could be through perturbation of myeloid apoptosis.

Treatment

Chemotherapy

Busulphan, introduced into CML therapy by Galton in 1953, allows long periods of hematologic control and is inexpensive. A Medical Research Council (MRC) randomized controlled trial established the superiority of busulphan therapy over irradiation. However, hydroxyurea is now preferred.

Busulphan can cause unpredictable severe myelosuppression, which may be fatal, and delayed as well as severe idiosyncratic pulmonary reaction, interstitial fibrosis (busulphan lung), marrow fibrosis, skin pigmentation, disorder of menstruation and infertility.

Hydroxyurea, a cycle-specific inhibitor of DNA synthesis, has a better toxicity profile than busulphan. It produces rapid, but more transient, control of hematological manifestations. It is well tolerated and has few

side-effects, including nausea, anorexia, skin rashes, and occasionally, ulcers in the mucous membranes and skin rashes. Hydroxyurea may cause marked red cell macrocytosis and megaloblastic changes in the marrow.

Both hydroxyurea and busulphan can control the hematological manifestations of CML in more than 80% of patients. However, treatment with either drug results in negligible rate of cytogenetic remission and has no effect on the rate of progression to blast crisis. Treatment with these agents is therefore considered as palliative.

Interferon-Alpha (IFN-α)

In 1983, IFN-α was identified as the first agent capable of inducing significant and durable cytogenetic responses in CML without causing marrow ablation.

Various studies from single institutions and cooperative groups have confirmed the efficacy of INF-α in CML. The studies show that:

1) Cytogenetic-response rate and the survival rate were higher in patients who received long-term treatment with subcutaneous IFN-α at a dose of 5 million unit per square meter of body surface area per day than in patients who received other drugs.[8,9]
2) A complete cytogenetic response or a partial cytogenetic response to treatment with IFN-α, occurs in 20–30% of patients. Minor response occurs in an additional 10%.
3) There is a direct correlation between cytogenetic response to IFN-α and survival. But the MRC study[9] also showed a survival advantage for patients receiving IFN-α even without achieving a cytogenetic response.

Problems related to IFN-α therapy

Early **adverse effects** of IFN include fever, chills, postnasal drip and anorexia usually develop within 1 to 2 weeks. The delayed adverse effects include persistent fatigue, weight loss, neurotoxicity, depression, insomnia, alopecia, marrow hypoplasia, and infrequent immune-mediated complications. A triad of depression, fatigue, and insomnia is common and can be improved with amitriptyline at bedtime.

IFN-α toxicity increases with dose. The adverse effects may reduce enthusiasm for the use of the drug, produce difficulties in patient

compliance, and limit the dose delivered to the patient, thus resulting in ineffective use of IFN-α.

The other major problem with IFN-α therapy is **cost** as the drug is expensive often leading to suboptimal dosing which may compromise response rates.

Finally, IFN-α may cause **myelosuppression** resulting in neutropenia and thrombocytopenia thus requiring a reduction in dose.

Combined modality therapy with IFN-α

Guilhot *et al.* have shown in their French multicenter, randomized study that the combination of cytosine arabinoside (Ara-C) and IFN-α gave superior cytogenetic response and survival than IFN-α alone.[10] However, patients' compliance with IFN-α/ Ara-C combination become an important concern as not all patients can tolerate two subcutaneous injections daily.

Recently, long-acting α-interferons in the form of pegylated interferons have become commercially available. These have the advantage of once a week dosing with significantly less side-effects and comparative trials are ongoing to confirm their efficacy.

Allogenic stem cell transplantation

Allogenic bone marrow transplantation from either related or unrelated donors is the only known curative therapy for CML, and results have improved over past 5 years. Transplantation from matched sibling donors during chronic phase is associated with a 1-year survival of 60–80% and a 10-year survival of approximately 50%.[11,12]

The success of allogenic BMT in these patients is related not only to the intensive conditioning therapy, but also to the anti-leukemic properties of the donor graft called the graft-versus-leukemia (GVL) effect. Patients who relapse post-allogenic transplant can be reinduced into complete remissions with infusion of donor lymphocyte without using any conditioning chemotherapy or radiotherapy.[13]

It has been shown that the likelihood of survival is higher in those who undergo transplantation in chronic phase than those who undergo transplantation during the accelerated phase or blast crisis. It has also been shown that patients who undergo transplantation within the first year after diagnosis have a higher likelihood of survival than those who undergo

transplantation later, even if it is before progression to the accelerated phase. These results indicate that allogenic transplant should be considered as soon as possible in eligible patients with HLA-matched sibling donors.

Glivec

Glivec (imatinib mesylate) is developed as a specific inhibitor to the BCR-ABL tyrosine kinase. Since June 1998, various clinical trials were conducted and they showed that Glivec was highly effective in various phases of CML with relatively minor side effects. In a large phase 2 trial involving CML patients who are α-interferon refractory or intolerant, Glivec at a starting dose of 400 mg per day was able to achieve major cytogenetic responses in 60% of patients. This led to FDA approval for the above indication in May 2001, an unprecedented "record" time. Since then, a large multicenter phase 3 trial comparing Glivec versus α-interferon plus s/c Ara-C as first-line therapy has shown clear superiority of the former (data presented in abstracts). Various combination studies involving Glivec and PEG-interferons, Ara-C and other newer agents are also ongoing.

While Glivec is clearly highly efficacious, convenient (orally as opposed to injections) and well tolerated with relatively minor side effects, it must be emphasized that it is a very new drug and its long-term effects and efficacy remain to be proven. Resistance to Glivec is also documented and various mechanisms for this have been proposed. Finally, there is the issue of cost. A month's supply at 400 mg daily costs approximately SGD$4500 presently. A worldwide Glivec International Patient Assistance Programme (GIPAP) is available from the manufacturers. It is administered by an independent charitable organization and coordinated locally through various welfare services.

Investigational modalities

Autologous Stem Cell Transplantation and homoharringtonine are examples of modalities that are under investigation.

Approach and Conclusions

These are exciting and perhaps somewhat confusing times for the clinician and patient with CML. Advances in interferon formulations, stem cell transplantation technology making it safer and more widely available and the "story-book-like" success of Glivec means that both patient and

clinician are faced with a plethora of choices. Information on new drugs, investigations, trials and so on are constantly updated and, in this internet age, widely available through various websites and support groups. The responsible clinician must be able to sift out the established facts and advise the patient and relatives.

A reasonable approach for the management of CML for 2003 would be as follows:

Patients who have good major cytogenetic responses to IFN-α should continue IFN-α given the good prognosis and long-term data available for this group. A switch to pegylated IFN-α may be considered for those with significant side effects. Glivec should be considered for all others without a major cytogenetic response. Patients on hydroxyurea (i.e. palliative therapy) should be evaluated on an individual basis.

Newly diagnosed patients with CML should be assessed if they are transplant candidates. These would generally be younger patients (below 35 years at our institution) with fully matched sibling donors.

All others should ideally be given Glivec or IFN-α unless factors present (e.g. extreme age/comorbid conditions, social etc.) necessitate palliative therapy. Factors influencing the above choice would include patient preference, patient profile (prognostic indices such as Sokal score) and most importantly in the local context, patient affordability. This would usually involve financial counseling and assessment of eligibility for the GIPAP scheme. Periodic assessments of cytogenetic response with marrow sampling is essential. Patients with a suitable donor and who do not achieve a major cytogenetic response after 6–12 months should be considered for allogenic transplant.

CHRONIC LYMPHOCYTIC LEUKEMIA (CLL)

Chronic lymphocytic leukemia (CLL) is the most common adult leukemia in Western Europe and North America. New insights into the biology of the disease, better diagnostic and staging criteria, and the recent development of more potent therapeutic agents are changing the outlook for CLL patients.

Epidemiology

CLL accounts for approximately 30% of all adult leukemias in Western Europe and North America. The incidence of CLL varies widely in

different regions of the world, ranging from 2.5% of all leukemias in Japan and other eastern countries and Africa to 38% in Denmark. CLL is rare before the fourth decade of life but increases from 5.2 to 30.4 cases per 100 000 persons over the age of 50 and 80 years, respectively. There is a male predominance.

Etiology

The cause of CLL is unclear. There is a family occurrence two-fold to seven-fold higher than in control populations and an increased incidence of lymphomas and autoimmune diseases in patients' first-degree relatives. An increased risk is also noted in agricultural, asbestos workers. This disease is the only leukemia not associated with occupational or accidental exposure to radiation, alkylating drugs or leukemogenic chemicals.

Cytogenetics

The most frequent single chromosome abnormalities were trisomy 12 in 18% (7–29%), followed by structural aberrations of the long arms of chromosomes 13 (12–23%) and 14 (9–17%). Trisomy 12 has been associated with early, progressive or advanced disease, as well as with poor prognosis or has been correlated with atypia of blood lymphocytes.

CLINICAL FEATURES

Clinical Presentation and Laboratory Abnormalities

The diagnosis of CLL is frequently made early in an asymptomatic individual from a routine complete blood count that shows lymphocytosis. Other patients present with constitutional symptoms, an increased susceptibility to viral or bacterial infections or an episode of autoimmune hemolytic anemia (AIHA). Physical findings range from no abnormalities to lymphadenopathy and/or organomegaly.

Anemia and thrombocytopenia occur in the most advanced stages. Anemia is usually normocytic and normochromic. A hypogammaglobulinemia and monoclonal gammopathy may also occur.

Diagnosis

Minimum diagnostic criteria

According to the International Workshop on CLL (IW-CLL), a diagnosis of CLL requires the following:

- absolute lymphocyte count (ALC) greater than 10×10^9 /L sustained for at least 4 weeks; and
- either greater than 30% replacement of the marrow cellularity by these cells or clonality of blood lymphocytes as determined by phenotype.

Morphologically, the CLL cells are small, mature-appearing lymphocytes. The NCI-sponsored CLL Working Group lowered the diagnostic blood lymphocytosis requirement to greater than 5×10^9 /L when the IW-CLL marrow and clonality criteria were both met.

Demonstration of clonality

CLL cells characteristically express B cell differentiation antigens, including CD19, CD20, CD21, CD23, HLA-DR, mouse rbc-receptors, low-density surface Igs and the T cell antigen, CD5.

Staging

The modified Rai Staging System stratifies patients into low-, intermediate- and high-risk categories based on their clinical picture.

Initial Rai Staging	Risk		Survival Rates (Year)
0	low	lymphocytosis of blood and marrow	10
I, II	intermediate	lymphocytosis plus lymphadenopathy with or without splenomegaly	6
III, IV	high	lymphocytosis with anemia (Hb $<$ 11 g%) or thrombocytopenia (platelet count $< 100 \times 10^9$ /L)	2

Staging should include a thorough physical examination, complete blood count and computed tomography of chest, abdomen and pelvis.

Treatment

The treatment of CLL ranges from periodic observation to experimental therapies. Not all patients require therapy. Current indications for treatment in CLL patients include:

- stage III or IV disease
- stage 0, I or II disease with one of the following:
  - bulky lymphadenopathy (>10 cm in maximal diameter)
  - marked splenomegaly (palpable >6 cm below costal margin)
  - lymphocyte doubling time of <6 months associated with a white blood cell count $>50\times10^9$/L
  - a white cell count $>100\times10^9$/L
  - disease-related constitutional symptoms

Oral chlorambucil alone or in combination with prednisolone had been the initial treatment for active CLL. However, chlorambucil is not curative. It provides symptomatic relief and reduction of tumor bulk. In the last decade, treatment with purine nucleoside analogs (e.g. fludarabine) has shown to be effective in both previously untreated and treatment-refractory CLL patients. Possible complications of treatment with these agents include prolonged cytopenias and opportunistic infections.

Most patients will still relapse and die of their disease despite the superior quality and duration of fludarabine-induced remission. In such cases, allogenic bone marrow transplant is a reasonable consideration for patients younger than 60 years. Autologous bone marrow transplant and monoclonal antibodies treatment are other options being evaluated.

Response Criteria

The International Workshop on CLL (IW-CLL) proposed 4 possible outcomes of therapy. These are complete response, nodular partial remission, partial remission and overall remission.

Complications

Complications of CLL include cytopenias, development of a second malignancy and transformation to a diffuse large-cell lymphoma, prolymphocytic

leukemia and acute lymphoblastic leukemia. Cytopenias result from bone marrow failure associated with advanced stage CLL. CLL can also be associated with autoimmune cytopenias, which can be treated with steroids, splenectomy, immunoglobulin and danazol. Hypersplenic cytopenias can be managed with splenectomy or splenic irradiation.

CONCLUSIONS

In conclusion, despite the important progress made in the treatment of leukemias, patients are still relapsing and dying of their disease. Treatment complications remain high especially for the older patients, who also have lower response rates to treatment compared to the younger patients.

Newer agents are continually being evaluated in clinical trials in an attempt to improve on response rates and decrease treatment-related morbidity and mortality.

REFERENCES

1. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller–Hermelink HK, Vardiman J, Lister TA, Bloomfield CD, World Health Organization Classification of Neoplastic Diseases of the Hematopoietic and Lymphoid Tissues: Report of the Clinical Advisory Committee Meeting, *J Clin Oncol* **17**(12):3835–3849, 1999.

2. Fenaux P, Le Deley MC, Castaigne S *et al.*, Effect of all-trans retinoic acid in newly diagnosed acute promyelocytic leukaemia. Results of a multicenter randomized trial. European APL 91 group, *Blood* **82**:3241–3249, 1993.

3. Tallman MS, Anderson JW, Schiffer CA *et al.*, All-trans retinoic acid in acute promyelocytic leukaemia, *N Engl J Med* **337**:1021–1028, 1997.

4. Huang ME, Ye YC, Chen SR *et al.*, Use of all-trans retinoic acid in the treatment of acute promyelocytic leukaemia, *Blood* **72**:567–572, 1988.

5. Chen Z-X, Xue Y-Q, Zhang R *et al.*, A clinical and experimental study on all-trans retinoic acid treated acute promyelocytic leukaemia patients, *Blood* **78**:1413–1419, 1991.

6. Cortes JE, Kantarjian H, O'Brien S *et al.*, All-trans retinoic acid followed by chemotherapy for salvage of refractory or relapsed acute promyelocytic leukemia, *Cancer* **73**:2946–2952, 1994.

7. Larson RA *et al.*, A five-drug remission induction regimen with intensive consolidation for adults with acute lymphoblastic leukemia: Cancer and Leukemia Group B study 8811, *Blood* **85**:2025–2037, 1995.

8. Italian Cooperative Study group on Chronic Myeloid Leukemia: Interferon alfa-2a as compared with conventional chemotherapy for the treatment of chronic myeloid leukemia, *N Engl J Med* **330**:820, 1994.

9. Allan N, Richards S, Shepherd P, *et al.*, UK Medical Research Council randomised multicenter trial of interferon-αn1 for chronic myeloid leukemia: improved survival irrespective of cytogenetic response, *Lancet* **345**:1392, 1995.

10. Guilhot F, Chastang C, Michallet M, *et al.*, Interferon-alpha 2b combined with cytarabine versus interferon alone in chronic myelogenous leukemia, *N Engl J Med* **337**: 223, 1997.

11. Horowitz MM, Rowlings PA, Passweg JR, Allogenic bone marrow transplantation for CML: a report from the International Bone Marrow Transplant Registry. *Bone Marrow Transplant* **17**(3):S5–S6, 1996.

12. Van Rhee F, Szydlo RM, Hermans J, *et al.*, Long-term results after allogeneic bone marrow bone marrow transplantation for chronic myelogenous leukemia inchronic phase: a report from the Chronic Leukemia working Party of the European Group for Bllod and marrow Transplantation, *Bone Marrow Transplant* **20**:553–560, 1997.

13. Kolb HJ, Mittermuller J,Clemm C *et al.*, Donor leukocyte transfusion for treatment of recurrent chronic myelogenous leukemia in marrow transplant patients, *Blood* **76**:2462–2465, 1990.

14. Druker BJ, Talpaz M, Resta D, *et al.*, Efficacy and safety of a specific inhibitor of the Bcr-Abl tyrosine kinase in Chronic Myeloid Leukemia, *N Engl J Med* **344**:1031–1037, 2001.

15. Kantarjian HM, Sawyers CL, Hochhaus A, *et al.*, Hematologic and cytogenetic responses to Imatinib Mesylate in Chronic Myeloid Leukemia, *N Engl J Med* **346**:645–652, 2002.

35

Myeloproliferative Disorders

Wong Gee Chuan

The myeloproliferative disorders (MPD) are characterized by monoclonal myeloproliferation involving multiple lineages. Even when differentiation is predominantly to cells of a single lineage, the disorder has arisen in a multipotent myeloid stem cell or in a pluripotent stem cell capable of giving rise to cells of both myeloid and lymphoid lineages. Even though the cellular maturation of neoplastic cells is relatively normal, there still exists a propensity to undergo clonal evolution. In the case of Chronic Myeloid Leukemia (CML), transformation to blast crisis is frequent and occurs at a median interval of 2–3 years. Polycythemia rubra vera (PRV) and idiopathic myelofibrosis have a longer chronic phase and undergo acute transformation less frequently. An acute phase is least frequently seen in Essential Thrombocythemia (ET).

Early in the course of MPD, hematopoiesis is effective with overproduction of cells of at least one lineage. Dysplastic features are either absent or not prominent. Dysplastic features may appear as the disease progresses. Occasionally, it is difficult to classify the disease as a MPD or myelodysplastic syndrome (MDS) because of the presence of both myeloproliferative and myelodysplastic features.

Table 1 The WHO Classification of the Myeloproliferative and Myelodysplastic/Myeloproliferative Diseases

Myeloproliferative diseases
Chronic myelogenous leukemia, Philadelphia chromosome positive [t(9;22) (q34; q11), BCR/ABL]
Chronic neutrophilic leukemia
Chronic eosinophilic leukemia/hypereosinophilic syndrome
Chronic idiopathic myelofibrosis
Polycythemia vera
Essential thrombocythemia
Myeloproliferative disorders, unclassifiable
Myelodysplastic/myeloproliferative disorders
Chronic myelomonocytic leukemia
Atypical chronic myelogenous leukemia
Juvenile myelomonocytic leukemia
Other myelodysplastic/myeloproliferative disorder, unclassifiable

The World Health Organisation (WHO) has proposed a classification of MPD and of the overlap myelodysplastic/myeloproliferative diseases. This is shown in Table 1.[1]

This classification highlights a group of conditions with features of both myelodysplasia and myeloproliferation. We will discuss polycythemia vera, essential thrombocythemia and idiopathic myelofibrosis in this chapter. Chronic myeloid leukemia (CML) is covered in the chapter "The Leukemias".

POLYCYTHEMIA

The term "polycythemia" denotes an increased amount of blood and has been applied to conditions with an increased number of erythrocytes. Primary polycythemia, or polycythemia rubra vera (polycythemia vera), is a chronic myeloproliferative disorder in which the dominant feature is the excessive production of erythrocytes by the marrow with a resultant increase in the circulating red cell mass and venous hematocrit. Secondary polycythemia (secondary erythrocytosis) describes a group of disorders characterized by an increased red cell mass due to enhanced stimulation of red cell production. Secondary polycythemia may be subdivided into appropriate polycythemia, where the response to hypoxia is normal, and inappropriate polycythemia, where erythropoiesis is

Table 2 Classification of Polycythemia and Erythrocytoses

Polycythemia vera (primary polycythemia)
Pure erythrocytosis
Secondary polycythemia (secondary erythrocytosis)
Appropriate
- High-altitude
- Cardiopulmonary disease
- Smoker's
- Abnormal hemoglobin
- Red cell enzyme deficiencies
- Chemicals (cobalt)
Inappropriate
- Familial
- Renal disease
- Tumors — myomas, brain tumors, hepatoma, endocrine disorders
- Neonatal
Apparent (spurious)

stimulated by aberrant production of or response to erythropoietin. A classification of these disorders is shown in Table 2.[2]

Inappropriate polycythemia can result from aberrant erythropoietin production by the kidney, by certain tumors or by cobalt ingestion. Some patients do not have an increased red cell mass and their erythrocytosis is the result of a decreased plasma volume. This disorder is not a true erythrocytosis and is designated apparent, spurious or relative polycythemia.

Polycythemia Rubra Vera (PRV)

Clinical features

Polycythemia vera usually has an insidious onset, with most patients presenting between the ages of 40 and 70 years. Many of the presenting symptoms are related to hyperviscosity of the blood and to arterial or venous thromboses. They include headache, plethora, pruritus, dizziness, tinnitus, dyspnea, visual disturbances, Raynaud's phenomenon, claudication, gangrene and gastrointestinal bleeding. Some patients are diagnosed because of abnormal blood counts upon routine screening.

Thrombotic episodes are the commonest complications of PRV, occurring in about one-third of patients.[3] These include episodes of hepatic vein thrombosis (Budd Chiari syndrome), cerebrovascular accidents, myocardial infarction, deep vein thrombosis and pulmonary embolism.

Bleeding and bruising occur in about one-quarter of patients. These are usually minor; however, life-threatening hemorrhages can also occur.

Pruritus is experienced in approximately 40% of patients. Usually aggravated by bathing, its cause is unclear. This may be consequent on histamine secretion by basophils.

The incidence of peptic ulceration is 4–5 times as great as in the general population. Cardiovascular symptoms include angina, myocardial infarction and congestive heart failure. Neurological symptoms such as dizziness are very common.[4]

Laboratory features

Peripheral Blood

The blood count shows an elevation of the red cell count, hemoglobin concentration and hematocrit. Iron stores may be exhausted in some patients; hence, there may also be microcytes and hypochromia. The white cell count is usually elevated due to neutrophilia. Absolute basophilia is often present. Many patients also have a moderately elevated platelet count.

Bone Marrow Studies

The bone marrow aspirate is usually hypercellular, with marked erythroid hyperplasia and some degree of granulocytic and megakaryocytic hyperplasia. Neutrophil, eosinophil and basophil lineages may be hyperplastic. Iron stores are often absent and features of superimposed iron deficiency may be present.

On trephine biopsy, there is usually marked hypercellularity, with hemopoietic cells filling more than 90% of the marrow space. There is an increase in cells of all three hemopoietic lineages. Reticulin is mildly increased in many cases. Marrow iron stores are decreased or absent.

Up to 30% of cases of PRV have severe marrow fibrosis. This is more commonly seen in patients with marked megakaryocytic proliferation.

About a quarter of patients have karyotypic abnormalities at diagnosis and as the disease progresses, the incidence rises. Del(20)(q) is the most characteristic.

Differential diagnoses

Polycythemia vera must be distinguished from secondary polycythemia and from apparent polycythemia.

Secondary Polycythemia

Patients with secondary polycythemia have a real increase in the number of circulating erythrocytes and of the red cell mass. The increase in the red cell mass is a response to the stimulation of the marrow by erythropoietin. These patients do not have a thrombocytosis or a leukocytosis or splenomegaly. Patients with cardiac or lung disease may be clubbed. Imaging to exclude hepatoma or kidney neoplasm or cyst can be performed as well.

Spurious Polycythemia

Patients with spurious polycythemia have a decrease in plasma volume. The erythrocytosis that is observed does not represent a true increase in the red cell mass. These patients do not have leukocytosis, thrombocytosis or splenomegaly.

Polycythemia Vera (PV)

There is no single diagnostic marker for PV. The Polycythemia Vera Study Group (PVSG) used major and minor diagnostic criteria.[5] The major criteria include a red blood cell mass (RCM) of greater than 36 ml/kg in males and greater than 32 ml/kg in females, arterial oxygen saturation greater than 92% and palpable splenomegaly. Alternatively, with 2 of the major criteria, any 2 of the following minor criteria were permissible: platelet count greater than 400×10^9/L and neutrophil alkaline phosphatase (NAP) greater than 100; or vitamin B12 greater than 900 pg/ml; or unsaturated B12 binding greater than 2200 pg/ml.

The International Council for Standardization Hematology recommendations (Table 3)[6] are based on height and weight as it was felt that the use of ml/kg expression is not useful, especially in overweight individuals, as fat tissue is relatively avascular.

Exclusion of causes of secondary erythrocytosis should be guided by presence of other signs and symptoms. Absolute erythrocytosis with clonal hematopoiesis supports a diagnosis of PV. At presentation, 10–20% of PV patients have an abnormal acquired marrow karyotype.

Course and management of PV

Polycythemia vera usually remains in a "plethoric" phase for many years, after which a "spent" phase with falling blood counts and increasing splenomegaly occurs.

Table 3 Proposed Diagnostic Criteria for Polycythemia Vera

A1 — Raised Red Cell Mass (RCM): $>25\%$ above mean normal predicted value (or packed cell volume >0.60 males; >0.56 females)
A2 — Absence of cause of secondary erythrocytosis
A3 — Palpable splenomegaly
A4 — Clonality marker, i.e. acquired abnormal marrow karyotype
B1 — Thrombocytosis: platelet count $>400\times10^9/L$
B2 — Neutrophil leukocytosis: neutrophil count $>10\times10^9/L$; in smokers $>12.5\times10^9/L$
B3 — Splenomegaly demonstrated by ultrasound scanning
B4 — Characteristic burst-forming unit-erythroid (BFU-E) growth or reduced serum EPO
A1 + A2 + A3 or A4 establishes PV
A1 + A2 + two of B establishes PV

The Plethoric Phase

The treatment of the polycythemic patient in this phase is targeted at reducing the blood counts, to decrease symptoms and the risk of thrombosis or bleeding. The hematocrit can be controlled in some patients by periodic phlebotomy, while thrombocytosis and leukocytosis will require the administration of drugs to suppress marrow activity.

1) *Phlebotomy*

   Phlebotomy to a hematocrit less than 45% is the mainstay of treatment for erythrocythemia. With phlebotomy alone, the platelet count may be elevated and induce a thrombotic state. Hence, it is necessary to control both the hematocrit and platelet levels. Use of low-dose aspirin is useful, especially in patients with microvascular episodes. Phlebotomy is the treatment of choice in young patients, especially those of a childbearing age.

2) *Myelosuppression and Cytoreduction*

   Long-term treatment with radioactive phosphorus and alkylating agents to control myeloproliferation is associated with leukemogenicity and mutagenic potential of these agents. Hence, the uses of these agents are limited. In PV, an elevated hematocrit and platelet count coexist in 50% of patients. Three drugs have been efficacious in reducing the platelet count: hydroxyurea (HU), anagrelide and interferon alfa (IFN). These drugs each have a different mechanism of action and no single agent can satisfy all the needs for cytoreduction during the course of PV.

Hydroxyurea inhibits ribonucleotide reductase and is non-specific, causing generalized, dose-specific myelosuppression. However, HU is leukemogenic and this limits its use in younger populations and women of childbearing age. Anagrelide is a prostaglandin synthetase inhibitor and can prevent platelet formation associated with megakaryocytic proliferation. Adverse effects seen with anagrelide include palpitations, tachycardia, edema, fluid retention, headache and gastrointestinal side-effects. Caution should be exercised in patients with myocardial ischemia and congestive heart failure. It provides a good control of thrombocytosis in younger patients. However, because of its teratogenicity, its use in pregnant women is inappropriate. Interferon treatment has not been shown to have any teratogenic or leukemogenic potential, hence it is the agent of choice in females who wish to conceive. Side effects include flu-like symptoms, anorexia, alopecia and neuropsychiatric symptoms.

The different mechanisms of action of these drugs, together with phlebotomy, allow the physician to manage the patient with different combination therapies.

The Spent Phase

The erythrocytosis of polycythemia patients may abate after a few years or more. During this phase, marrow fibrosis becomes more marked and splenomegaly increases in size. Transfusion may be required when anemia develops. Treatment of this phase is usually symptomatic.

Prognosis of PV

Patients with PV have normal or near normal lifespan. Leukemia may develop in some patients, the incidence increased in patients that were given the more leukemogenic agents.

THROMBOCYTOSIS AND ESSENTIAL THROMBOCYTHEMIA (ET)

Thrombocytosis can be classified as 1) clonal, including essential (primary) thrombocythemia or other myeloproliferative disorders; 2) familial; 3) reactive or secondary (Table 4).[7]

Table 4 Major Causes of Thrombocytosis

1) Clonal thrombocytosis
   a) Essential (primary) thrombocythemia
   b) Other myeloproliferative disorders (polycythemia vera, chronic myelogenous leukemia, myeloid metaplasia, myelofibrosis)
2) Familial thrombocytosis
3) Reactive (secondary) thrombocytosis
   a) Acute blood loss
   b) Iron deficiency
   c) Postsplenectomy, asplenic states
   d) Recovery from thrombocytopenia ("rebound")
   e) Malignancies
   f) Chronic inflammatory and infectious diseases (inflammatory bowel disease, connective tissue disorders, temporal arteritis, tuberculosis, chronic pneumonitis)
   g) Acute inflammatory and infectious diseases
   h) Response to exercise
   i) Response to drugs (vincristine, epinephrine, all-transretinoic acid, cytokines, and growth factors)
   j) Hemolytic anemia

Familial thrombocytosis is rare and is generally inherited by autosomal dominant transmission. Specific mutations in the thrombopoietin gene result in markedly increased plasma thrombopoietin levels. Causes of reactive thrombocytosis are listed in Table 4. These have to be excluded before a diagnosis of essential (primary) thrombocythemia (ET) can be made.

Essential Thrombocythemia

Clinical features

Essential thrombocythemia (ET) is a disease resulting from the clonal proliferation of a multipotent myeloid stem cell but with predominantly an increased platelet production. It typically affects patients between the ages of 50 and 70 but more asymptomatic individuals have been incidentally discovered on routine blood screening. About two-thirds of symptomatic patients have venous or arterial thrombosis and present with headache, dizziness, visual disturbance, paresthesiae or peripheral vascular insufficiency. A third of symptomatic patients suffer from abnormal bleeding similar to those seen in platelet or vascular disorders, e.g. mucosal, gastrointestinal, cutaneous or genitourinary tract bleeding.

Splenomegaly is seen in up to 40% of patients and hepatomegaly in up to 20%.

Laboratory features

Peripheral Blood

The peripheral blood film shows an increased number of platelets with some giant platelets and agranular and hypogranular platelets. Mild leukocytosis may be present but white cell count is usually less than 20×10^9/L.

Bone Marrow Studies

The bone marrow aspirate shows an increase in megakaryocytes, which are large and well-lobulated. On trephine biopsy, the marrow is usually hypercellular, though it can be normocellular or even hypocellular. Large clusters of megakaryocytes are commonly seen and the average size of megakaryocytes is increased with increased lobulation of their nuclei.

A minority of patients with ET have a clonal cytogenetic abnormality.

Differential diagnosis

Essential thrombocythemia is a diagnosis of exclusion. Reactive thrombocytosis (Table 4), which is caused by a variety of conditions like infections, inflammatory conditions and malignancies, should be first excluded. ET should also be differentiated from other myeloproliferative disorders associated with thrombocytosis.

The modified Polycythemia Vera Study Group (PVSG) criteria for diagnosis of ET[8] are: platelet count greater than 600×10^9/L; no cause for reactive thrombocytosis; hematocrit less than 40% or normal red cell mass less than 36 ml/kg in males and less than 32 ml/kg in females; normal red cell mean corpuscular volume or serum ferritin or marrow iron stain; collagen fibrosis of marrow absent, or less than one-third of biopsy area without both splenomegaly and leukoerythroblastic reaction; no myelodysplastic changes on bone marrow smear, and no Philadelphia chromosome or bcr/abl gene rearrangement. While the minimum platelet count of 600×10^9/L to establish a diagnosis of ET has been proposed, clinicians need to recognize that thrombocythemia-related consequences may occur in individual patients with mild thrombocytosis or even platelet count in the upper normal range. Similarly, patients with

reactive thrombocytosis can have extreme thrombocytosis (platelet count $> 1{,}000 \times 10^9$/L). In general, these patients do not have thrombotic or bleeding complications and do not require treatment.

Prognosis, course and management of ET

The major causes of morbidity and mortality in ET are thrombosis and hemorrhage. Unlike the other myeloproliferative disorders, ET rarely evolves into leukemia.

Not all ET patients experience thrombotic and hemorrhage complications. Major thrombotic episodes occur in 20–30% of ET population. Established risk factors for thrombosis include prior history of thrombotic events, advanced age and duration of thrombocytosis. However, no correlation can be made between laboratory abnormalities and the development of thrombotic or bleeding complications.

Hence the decision to treat ET patients should be individualized. There is no evidence to support myelosuppressive therapy in asymptomatic patients because of high platelet count or to prevent transition to myeloid metaplasia.[9] However, retrospective reports in the literature support that in patients with a history of previous thrombotic episodes and/or with cardiovascular risk factors, the elderly, and in whom platelet counts are elevated, lowering the platelet count to near normal can prevent future complications.[9–11] The management of high-risk patients should be directed toward prevention of thrombosis, yet not incurring the increased leukemogenic risk with certain therapeutic agents.

Alkylating agents and radioactive phosphorus are effective in cytoreduction but are leukemogenic. Hydroxyurea has been used to maintain a stable platelet count and reducing the thrombohemorrhagic complications in high-risk patients. It, however, causes dose-related generalized myelosuppression. It is generally considered safe and effective in the short-term, but there has been equivocal evidence that long-term treatment with hydroxyurea alone may induce secondary leukemia. While this leukemogenic potential is less of a concern in the elderly, alternative, non-mutagenic agents such as interferon-alpha (IFN-α) and anagrelide should be considered for younger patients.

Interferon is an effective, non-leukemogenic, non-teratogenic platelet-lowering treatment for ET. Its side effects include flu-like symptoms, anorexia or neuropsychiatric symptoms. IFN use in ET is most suited to

high-risk women of childbearing age and to those who failed hydroxyurea and anagrelide. Anagrelide is the newest drug used in the treatment of myeloproliferative disorders. The common side effects of anagrelide include headache, dizziness, palpitations, tachycardia, diarrhea and fluid retention. It is effective at lowering platelet count and has not been shown to be mutagenic or leukemogenic.

Low-dose aspirin alone, or in combination with platelet-lowering agent, has been found to be effective in treating the neurological and usual symptoms commonly present in ET. The side effects of aspirin include stomach and gastrointestinal irritation. In some reports, aspirin has been suggested to unmask latent bleeding tendencies. However, other reports have justified the use of aspirin, especially in patients at high risk of thrombosis, as thrombosis is a more common complication in ET than severe bleeding. In general, low-dose aspirin is an effective, appropriate treatment for the prevention and treatment of thrombosis and symptom relief, especially in patients with no prior bleeding history.

ET is being more frequently discovered in younger, asymptomatic patients. Long-term studies are still needed to determine the complications and outcomes of treatment in this group. Conservative treatment regimens and avoidance of potential leukemogens are usually practiced. Asymptomatic patients with no cardiovascular risk factors should be observed. Platelet-lowering therapy can be given in asymptomatic patients with cardiovascular risk factors and platelet anti-aggregating agents for patients with vasomotor symptoms. Anagrelide or IFN can be used for younger asymptomatic patients.

ET is associated with increased risk of recurrent miscarriages, premature delivery and fetal growth retardation. No particular treatment is recommended for symptomatic pregnant women. Hydroxyurea has been known to cause fetal malformation. Anagrelide's safety in pregnancy has not been proven. IFN has been used successfully in pregnancy and has not shown to be teratogenic and does not cross the placenta. However, the data relating to its use is limited.

The management of a patient with ET has to be individualized, taking into consideration the age, risk of thrombosis, and the morbidity and side effects of treatment, especially that of developing secondary malignancy. In general, asymptomatic, low-risk patients can be observed with no treatment, while treatment is recommended for patients with a history of thrombosis or cardiovascular risk factors.

IDIOPATHIC MYELOFIBROSIS

Idiopathic myelofibrosis is a stem cell disease. Primary or Idiopathic Myelofibrosis is also known as myelofibrosis with myeloid metaplasia and as agnogenic myeloid metaplasia. Clonal megakaryocytes abnormally secrete growth factors, causing the proliferation of fibroblasts and overproduction of collagen, with resultant bone marrow fibrosis.

Clinical features and prognosis

Idiopathic Myelofibrosis is characterized by splenomegaly, leukoerythroblastic anemia and marrow fibrosis. The median age of diagnosis is 60 years. Extramedullary hemopoiesis is particularly observed in the spleen but can develop in the liver, spinal cord, the pleural and peritoneal cavity and other organs. Splenomegaly is almost invariable and slight or moderate hepatomegaly is also commonly seen. Median survival is 3–6 years from diagnosis. Myelofibrosis can terminate in a condition resembling chronic myeloid leukemia with striking myeloid proliferation, increasing hepatomegaly and splenomegaly. It transforms to acute leukemia in 10–20% of cases, usually a myeloblastic transformation but rarely, it is a lymphoblastic transformation. Leukemic transformation should be suspected when there is a rapid increase in splenic size or sudden development of anemia or thrombocytopenia.

Laboratory features

Peripheral Blood

Characteristic peripheral blood findings of myelofibrosis include pancytopenia with a leukoerythroblastic blood film and poikilocytes. There may be some dysplastic features in the granulocytes and platelets. In the early stages of the disease, thrombocytosis and leukocytosis may be present. As the disease progresses, leukopenia, neutropenia and thrombocytopenia become prominent. Prognosis is worse with hemoglobin concentration of less than 10 g/L, 1% or more circulating blasts, more than 10% immature granulocytes and a white blood cell count of less than 4 or more than 30×10^9/L.

Bone Marrow Studies

The aspirate usually shows very hypercellular fragments with hyperplasia of all lineages in the early stages of the disease. In the later stages of the disease, aspiration of the bone marrow may be difficult secondary to the fibrosis ("dry tap").

The biopsy is often cellular, showing granulocytic and megakaryocytic hyperplasia. There is an increase in reticulin fibers. In intensely fibrotic marrows cellularity may be decreased but megakaryocytes remain evident.

Among the patients, 60% show clonal cytogenetic abnormalities — the most common of which are del(13)(q), del(20)(q), trisomy 8 and partial trisomy of 1q.

Differential diagnosis

Other causes of bone marrow fibrosis need to be excluded before idiopathic myelofibrosis is diagnosed (Table 5).[12]

Table 5 Causes of Bone Marrow Fibrosis

Myeloid disorders
Chronic myeloproliferative diseases
Myelodysplastic syndrome
Acute myelofibrosis
Acute myeloid leukemia
Mast cell disease
Malignant histiocytosis
Lymphoid disorders
Lymphomas
Hairy cell leukemia
Multiple myeloma
Non-hematologic disorders
Metastatic cancer
Connective tissue disease
Infections
Vitamin D-deficiency rickets
Renal osteodystrophy
Gray platelet syndrome

Bone marrow fibrosis may be seen in non-hematologic and malignant conditions. Chronic myeloid leukemia (CML) should be excluded by karyotypic analysis or fluorescence *in situ* hybridization study.

Management

Allogenic bone marrow transplant may provide durable remission to young patients. Most patients are, however, treated palliatively. Anemia can be improved in a third of patients with a combination of an androgen preparation and a corticosteroid. Liver function tests should be monitored. Other patients may require regular packed cell transfusions.

Symptomatic splenomegaly can be treated by splenectomy or splenic irradiation for patients with poor surgical risk. Extramedullary hematopoiesis can occur at locations such as the spleen, liver, lymph nodes, peritoneum, pleura and paraspinal and epidural spaces. Spinal cord compression can be confirmed with spinal magnetic resonance imaging with gadolinium. Extramedullary hematopoiesis is best treated with low-dose irradiation.

REFERENCES

1. Harria NL, Jaffe ES, Diebold J, Flandrin G, Muller–Hermelink HK, Vardiman J, World Health Organisation classification of neoplastic diseases of the haematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting — Airlie House, Virginia, Nov 1997, *J Clin Oncol* **17**:3835–3849, 1999.

2. Beutler E, *Polycythaemia*, William Haematology, 6th ed. McGraw Hill, p. 690, 2001.

3. Wehmeir A, Daum I, Jamin H, Schneider W, Incidence and clinical risk factors for bleeding and thrombotic complications in myeloproliferative disorders, A retrospective analysis of 260 patients, *Ann Haematol* **63**:101, 1991.

4. Newton LK, Neurological complications of polycythaemia and their impact on therapy, *Oncology* (*Williston Park*) **4**:59, 1990.

5. Berlin NI, Diagnosis and classification of polycythaemia, *Semin Haematol* **12**:339–351, 1975.

6. Pearson TC, Guthrie DL, Simpson J, Chinn S, Barosi G, Ferrant A, Lewis SM, Najean Y, Red cell mass and plasma volume interpretation in normal adults, *Br J Haematol* **89**:748–756, 1995.

7. Schafer A, *Thrombocytosis and Essential Thrombocythaemia*, William Haematology, 6th ed. McGraw Hill, p. 1542, 2001.

8. Murphy S, Peterson P, Iland H, Laszlo J, Experience of the Polycythaemia Vera Study group with essential thrombocythaemia: A final report on diagnostic criteria, survival and leukaemia transition by treatment, *Semin Haematol* **34**:29–39, 1997.

9. Murphy S, Therapeutic dilemnas: Balancing the risks of bleeding, thrombosis and leukaemic transformation in myeloproliferative disorders (MPD), *Thromb Haemost* **78**:622–626, 1997.

10. Barbui T, Finazzi G, Dupuy E, Kiladjian JJ, Briere J, Treatment strategies in essential thrombocythaemia. A critical appraisal of various experiences in different centers, *Leuk Lymphoma* **22** (Suppl 7):149–160, 1996.

11. Van Genderen PJ, Mulder PG, Waleboer M, van de Moesdijk D, Michiels JJ, Prevention and treatment of thrombotic complications in essential thrombocythaemia: Efficacy and safety of aspirin, *Br J Haematol* **97**:179–184, 1997.

12. Ayalew Tefferi, *Primary Hematology*, New Jersey, Humana Press, p. 145, 2001.

This page intentionally left blank

36

Myelomas and other B Cell Malignancies

Lim Lay Cheng and Linn Yeh Ching

INTRODUCTION

Lymphoid cells arise from a common hematopoietic progenitor that subsequently develops into lymphoid, myeloid, erythroid, monocyte and megakaryocyte lineages. The lymphoid cell further differentiates into B and T cells. Malignancies of lymphoid cells result from aberrant growths of transformed lymphocytes. Thus B-cell neoplasms are clonal expansions of cells derived from a single B cell that in the majority of cases has productively rearranged its immunoglobulin genes. Approximately 75% of all lymphoid leukemia and 90% of all lymphoma are of B-cell origin.

CLASSIFICATION

The wide diversity of lymphoid malignancies has resulted in shifting concepts of classification and terminology of this disease over the years. For many years, the French-American-British (FAB) classification has been universally used to subtype acute lymphoblastic leukemia into L1, L2 and L3 (Burkitt's) groups. On the other hand, for lymphoma, various

Table 1 WHO Classification for B-cell neoplasms

B-cell neoplasms
Precursor B-cell neoplasm
B-cell lymphoblastic lymphoma/leukemia
Mature B-cell neoplasms
Chronic lymphocytic leukemia/small lymphocytic lymphoma
Prolymphocytic leukemia
Hairy cell leukemia
Lymphoplasmacytic lymphoma
Mantle cell lymphoma
Follicular lymphoma
Marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT) type
Nodal marginal zone lymphoma with/without monocytoid B cells
Splenic marginal zone lymphoma
Diffuse large-cell lymphoma
Mediastinal (thymic), intravascular, primary effusion
Burkitt's lymphoma
Endemic, sporadic, immunodeficiency associated, atypical
Plasmacytoma
Plasma cell myeloma

classifications have been used over the years in different countries including Kiel Classification, BNLI Classification, Working Formulation and REAL Classification.[1] However since its proposal in 1995, the World Health Organization (WHO) classification[2] for lymphoid malignancies is now increasingly adopted worldwide (Table 1). Under the WHO classification, the different subtypes are defined as far as possible on the basis of five properties: morphology (histology), immunophenotype (B or T or NK cell), genotype (cytogenetics or molecular genetics), normal cell counterpart and clinical features. In effect, morphology is the collective expression of the immunophenotype, genotype and normal cell counterpart and thus the mainstay of diagnosis. In general, the WHO system divides lymphoma into 3 main categories: Hodgkin's disease, the mature B-cell and T-/NK-cell neoplasms, which were previously collectively regarded as non-Hodgkin's lymphoma.

PRECURSOR B-CELL NEOPLASM

Under the FAB classification, acute lymphoblastic leukemia was differentiated into L1, L2 and L3 subtypes purely on the basis of morphology. With increasing understanding of the pathogenesis of leukemia, we now

recognize that L1 and L2 morphology do not predict immunophenotype, genetic abnormalities or clinical behavior of the disease. Also, L3 is the equivalent to Burkitt's lymphoma in leukemic phase and thus should be diagnosed as such. Hence the WHO classification subtypes acute lymphoblastic leukemia on the basis of their immunophenotype into precursor B cell or precursor T cell while L3 is classified under mature B cell neoplasm as Burkitt's lymphoma.

MATURE B CELL NEOPLASMS

Surface immunoglobulin and T cell receptor are the definitive markers of B cells and cells. Their production is controlled by a series of genes. During the normal differentiation of the B cell, a series of random gene rearrangements takes place that subsequently results in production of different antibodies or immunoglobulins following the processes of gene transcription and translation. In contrast, the mature B cell neoplasms are a group of diseases characterized by clonal B-lymphocyte proliferation and accompanied by the production of a monoclonal immunoglobulin.

Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma

Chronic lymphocytic leukemia is the most common adult leukemia in Western Europe and North America, though it is relatively rare in Asia and in Singapore. Diagnostic criteria for chronic lymphocytic leukemia (CLL) as defined by the CLL Working Group[3] include:

1) absolute lymphocyte count greater than 5×10^9/L sustained for at least four weeks; and
2) either a greater than 30% replacement of the marrow cellularity by these cells or clonality of blood lymphocytes as determined by the phenotype.

The diagnosis is frequently made early in asymptomatic individuals from routine blood count. Sometimes patients present with constitutional symptoms, infections or hemolytic anemia. Physical examination may be unremarkable or lymphadenopathy or organomegaly may be present. Small lymphocytic lymphoma (SLL) is similar to B-cell CLL in histology and clinical features. The distinction between the two diseases is clinical.

SLL is typically a nodal disease associated with the presence of lymphadenopathy invariably. The marrow may not be involved in SLL but if so, the pattern of involvement is typically nodular rather than interstitial or diffuse. In contrast, CLL is a leukemic disease invariably associated with blood lymphocytosis and marrow involvement.

Not all patients require treatment on diagnosis. Decision for chemotherapy should be based on the clinical stage, presence of symptoms such as hemolytic anemia, thrombocytopenia or bulky lymphadenopathy and doubling time of the lymphocyte counts. In patients where the lymphocyte counts double within 6 months, which is indicative of progressive disease, therapy is required. Two staging systems are available to predict survival (Tables 2 and 3).

Currently, there are no proven cures for this disease. Traditionally the mainstay of therapy is with chlorambucil. However, newer agents like fludarabine and the anti-CD20 monoclonal antibody have also been demonstrated to have efficacy. In more refractory cases, combination chemotherapy may be required.

Table 2 Modified Rai Staging System for CLL[4]

Initial Stage (year)	Risk	Characteristics	Survival Rates
O	Low	Lymphocytosis of blood and marrow	10
I, II	Intermediate	Lymphocytosis plus lymphadenopathy, with or without splenomegaly	6
III, IV	High	Lymphocytosis with anemia (Hb* $<$ 11 g%) or thrombocytopenia (platelet $<$ 100×10^9/L)	2

*Hemoglobin.

Table 3 Binet Staging System for CLL[5]

Binet Survival Stage	Characteristics	Median Rates (years)
A	Fewer than 3 areas of infiltration (cervical, axillary or inguinal nodes, spleen or liver)	9
B	3 or more areas of infiltration	5
C	Anemia (Hb $<$ 10 g%) or thrombocytopenia (platelet $<$ 100×10^9/L)	2

Prolymphocytic Leukemia

B cell prolymphocytic leukemia is a rare disease characterized by the presence of more than 55% of prolymphocytes in peripheral blood lymphoid population.[6] Prolymphocytes of B cell lineage are typically large lymphocytes of about 10–15 μm in diameter with moderate amount of basophilic cytoplasm and a single prominent nucleolus. Clinically the patients usually have high white cell counts in excess of 100×10^9/L and massive splenomegaly, but mild or absent hepatomegaly and lymphadenopathy. It is an aggressive disease often refractory to chemotherapy and a median survival of 2–3 years.

Hairy Cell Leukemia

This is a rare chronic lymphoproliferative condition that is typified by the presence of abnormal mononuclear cells with irregular cytoplasmic projections in blood and bone marrow. Leukemic infiltrates of these cells can also be found in the liver and spleen. The demonstration cytochemically of the presence of tartrate-resistant acid phosphatase (TRAP) in the cytoplasmic vesicles of the leukemic cells is pathognomonic of this disease.

Clinically, the patients typically present with splenomegaly and pancytopenia. Sometimes the presentation consists of a variable combination of anemia, granulocytopenia and thrombocytopenia. Infections are very common and life-threatening infections can develop in up to 80% of patients. Bone marrow biopsy is extremely important for diagnosis as it is usually involved in hairy cell leukemia. Typically, marrow aspiration yields a dry tap or is non-diagnostic. However, the trephine biopsy shows diffuse infiltrate of hairy cells with abundant, clear cytoplasm — the characteristic fried-egg pattern. Other characteristic histologic patterns include pseudosinus formation seen in the spleen and angiomatoid lesions in liver.

Treatment with pentostatin gives excellent results with prolonged remissions and is the therapy of choice currently.

LYMPHOMA

In the initial management of patients with lymphoma, histological subtyping and staging of the disease is important in determining the treatment and prognosis. The Ann Arbor staging classification[6] (Table 4) is the gold

Table 4 Ann Arbor Staging Classification for Lymphoma

Stage	Definition
I	Involvement of a single lymph node region (I) or of a single extralymphatic organ or site (IE)
II	Involvement of 2 or more lymph node regions on the same side of the diaphragm (II) Localized involvement of an extralymphatic organ or site and of one or more lymph node regions on the same side of the diaphragm (IIE)
III	Involvement of lymph node regions on both sides of the diaphragm (III), which may also be accompanied by involvement of the spleen (IIIS) or by localized involvement of an extralymphatic organ or site (IIIE) or both (IIISE)
IV	Diffuse or disseminated involvement of one or more extralymphatic organs or tissues, with or without associated lymph node involvement

The absence or presence of fever, night sweats and/or unexplained loss of 10% or more of body weight in the preceding 6 months are denoted in all cases by suffix letters A or B respectively.

Table 5 International Prognostic Index (IPI) for Non-Hodgkin's Lymphoma

Five clinical risk factors:	
Age $\geqslant 60$ years	1 point
Serum lactate dehydrogenase levels elevated	1 point
Performance status $\geqslant 2$(ECOG) or $\leqslant 70$ (Karnofsky)	1 point
Ann Arbor stage III or IV	1 point
> 1 site of extranodal involvement	1 point

The presence of constitutional symptoms such as fever in excess of 38°C, night sweats and a weight loss of more than 10% body weight in the preceding 6 months suggest "B" disease (Table 4) in lymphoma.

standard for staging lymphoma. Subsequently, the prognosis of patients can be predicted using the International Prognostic Index[7] (IPI) (Table 5) where patients are assigned a score based on the presence or absence of five adverse prognostic factors.

Staging and Investigations for Lymphoma

To confirm the disease and determine subtype, excisional lymph node biopsy should be undertaken for pathologic examination wherever possible. A detailed history and physical examination is important. Generalized pruritus may be present though it is not associated with prognostic significance.

A history of infections such as human immunodeficiency virus or hepatitis B, or any previous history of neoplasm treated with chemotherapy or radiotherapy, should be documented. Physical examination should focus on the number and size of lymph nodes enlarged, liver and spleen enlargement and Waldeyer's ring involvement.

Investigations are required to stage the disease and include radiographic studies such as chest X-ray, thoracic, abdominal and pelvic computerized tomographic (CT) scans. CT scans are the standard for evaluation of nodal disease and, in addition to defining nodal sites of involvement, are critical for response assessment post therapy. Bone marrow biopsy is important to evaluate for involvement and also to evaluate the normal marrow elements before therapy.

In addition, the presence of anemia, granulocytosis, lymphopenia and low serum albumin are important adverse prognostic factors. Assaying serum beta 2 microglobulin levels are important as they correlate with tumor burden and prognosis. Elevation of the lactate dehydrogenase levels and erythrocyte sedimentation rate are also poor prognostic indicators. Additional investigations would include uric acid, calcium and renal function studies. Also, tests to determine hepatitis B status and presence of human immunodeficiency virus are important as they affect decisions on therapy. In rare cases, patients with hepatitis B can develop acute fulminant liver failure secondary to a reactivation of the virus from the use of immunosuppressive agents, particularly in relation to steroids usage.

Lymphoplasmacytic lymphoma

This is a neoplasm of small B-lymphocytes, plasmacytoid lymphocytes and plasma cells, involving bone marrow, lymph nodes and spleen.[1] It lacks CD5 expression and is usually associated with a serum monoclonal protein with hyperviscosity or cryoglobulinemia. A monoclonal serum paraprotein of IgM type is also termed "Waldenstrom's macroglobulinemia". The median age is 63 years and most patients have bone marrow involvement. The clinical course is indolent. In patients presenting with clinical features of hyperviscosity such as headaches or acute confusional states, plasmapheresis is an effective means to reduce plasma viscosity rapidly. Chemotherapy is indicated only if patient is symptomatic due to lymph node enlargement or increased plasma viscosity. The deoxyadenosine analog, 2-chorodeoxyadenosine, has been found to be effective in treating

this disorder although chemotherapy with alkylating agent-steroid combinations can also induce responses in 50% of previously untreated patients.

Mantle cell lymphoma

This is a neoplasm of monomorphous small to medium-sized B cells with irregular nuclei, which resemble the cleaved cells of germinal centers and overexpress cyclin D1.[2] It comprises about 6% of all non-Hodgkin's lymphoma. The lymphoma is associated with a characteristic chromosomal translocation t(11;14) between the immunoglobulin heavy chain gene on chromosome 14 and the bcl-1 gene on chromosome 11.

About 70% of patients with mantle cell lymphoma have stage IV disease at presentation and some have extranodal disease, commonly gastrointestinal involvement. The disease is aggressive and chemotherapy treatments are often unsatisfactory as prolonged remissions are seldom achieved. The 5-year survival for all patients with mantle cell lymphoma is approximately 25%.

Follicular lymphoma

This is a lymphoma of follicle center B cells.[1] It is the second most common lymphoma in the United States and Western Europe, comprising 20% of all non-Hodgkin's lymphoma and up to 70% of low-grade lymphomas. Cytogenetically, follicular lymphoma is characterized by the presence of t(14;18) which occurs in 85% of cases. This chromosomal aberration represents the translocation of the immunoglobulin heavy chain gene on chromosome 14 and the oncogene bcl-2 on chromosome 18. Most patients have widespread disease at diagnosis, usually predominantly lymph nodes but also spleen, bone marrow, and occasionally peripheral blood or extranodal sites. The disease is often indolent and median survival is in excess of 8 years. However, the disease is not usually curable in spite of chemotherapy. Hence asymptomatic patients may be observed till progressive disease leading to obstructive symptoms or cytopenias occur. Treatment depends on the symptoms. For localized symptomatic adenopathy, radiation therapy can be considered while for more extensive disease single or combination chemotherapy should be used. Newer therapeutic options include treating with the anti-CD20 monoclonal antibody or radioimmunotherapy. Progression to large cell lymphoma may occur and should then be managed as such.

Marginal zone lymphoma

Three separate entities compose the marginal zone lymphomas[8] — marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT) type, nodal marginal lymphoma with/without monocytoid B cells, and splenic marginal zone lymphoma. While the first two entitities are closely related, the third is a different disease.

1) *MALT type marginal zone lymphoma*
   This is an extranodal lymphoma comprising heterogeneous small B cells. It comprises the majority of low-grade gastric lymphomas and almost 50% of all gastric lymphomas. The majority of patients present with localized stage I or II extranodal disease involving glandular epithelial tissues at various sites such as stomach, intestines, lung, thyroid, salivary gland, orbit and soft tissues. Many patients have a history of autoimmune disorders such as Sjogren's syndrome or Hashimoto's disease or a history of Helicobacter gastritis. The t(11;18) translocation has been found to be present in a high frequency of patients with indolent MALT lymphoma.

   Localized tumors of the stomach or other sites may be cured with local treatment such as surgery or radiation. If disease is more widespread, chemotherapy should be considered and can give durable remissions. Large cell transformation is clinically significant and requires treatment with combination chemotherapy.
2) *Nodal marginal zone lymphoma*
   This is a primary nodal lymphoma with morphologic features identical to those of lymph nodes involved by extranodal marginal zone lymphoma of MALT but without evidence of extranodal disease. It is extremely rare. Most patients have localized disease, commonly head and neck lymph nodes. Treatment is similar to MALT and transformation to large-cell lymphoma may occur.
3) *Splenic marginal zone lymphoma*
   This is a rare disorder, accounting for only 1–2% of B cell leukemias. Typically the patients present with splenomegaly often accompanied by anemia and thrombocytopenia. Lymphocytosis is common but they usually do not have peripheral lymphadenopathy. In some cases, circulating neoplastic lymphocytes with villous appearance is present in the blood. The course is extremely indolent. Splenectomy is the treatment of choice and prolonged remission may be achieved post splenectomy.

Diffuse large-cell lymphoma

This entity is composed of a heterogeneous group of entities.[6] Histologically, they are composed of large, transformed B cells with prominent nucleoli and basophilic cytoplasm, with a diffuse growth pattern. It is the most common type of non-Hodgkin's lymphoma representing about 30% of all cases. Typically, patients present with a rapidly enlarging, symptomatic mass and B symptoms in one-third of cases. Three distinct subtypes are recognized in the WHO classification with unique clinical behavior: mediastinal (thymic), intravascular, and primary effusion. However, there probably exist other distinct entities within this category of diffuse large B cell lymphomas that remain to be characterized and would warrant different therapies.

Typically this disease is considered an aggressive lymphoma in contrast to the indolent or low grade histologies such as lymphoplasmacytic lymphoma, follicular lymphoma and marginal zone lymphoma. The prognosis is strongly associated with the IPI score (Table 6).[7]

Therapy is indicated on diagnosis and the treatment goal is cure, which can be achieved in approximately half of the cases with combination chemotherapy. Newer modalities of therapy include combination chemotherapy with monoclonal antibody and preliminary studies in advanced stage III, IV elderly patients suggest a survival advantage with this combination treatment.

Burkitt's lymphoma

This tumor was discovered in 1958 by Burkitt who described its occurrence in the jaws of African children.[6] It occurs in certain endemic areas in Africa but is sporadic outside Africa. The morphology of the sporadic,

Table 6 Survival of Patients with Diffuse Large B Cell Lymphoma in Relation to IPI Score

Risk Based on IPI Score	% of Cases	5-Year Survival
Low risk	35	73%
Low-intermediate risk	27	51%
High-intermediate risk	22	43%
High risk	16	26%

endemic and immunodeficiency types are similar with the classic starry-sky pattern secondary to macrophage ingestion of the remnants of tumor cells. In the endemic cases, facial bones, particularly the jaws, are affected, while in the sporadic cases, the majority present with disease in the abdomen, most commonly in the distal ileum, cecum or other abdominal organs. Burkitt's lymphoma is commonly a disease in children while in adults, it occurs predominantly in the AIDS population. Most cases have a translocation of c-myc from chromosome 8 to either the immunoglobulin heavy chain region on chromosome 14 — t(8;14), or light chain loci on chromosome 2 — t(2;8) or t(8;22). This lymphoma is highly aggressive and requires intensive combination chemotherapy.

MULTIPLE MYELOMA

Introduction

Myeloma is a plasma cell malignancy arising from a single cell of the mature B-lymphocyte lineage that undergoes malignant transformation and proliferation. It is thought to be preceded by a more common pre-malignant plasma cell disorder called monoclonal gammopathy of undetermined significance (MGUS), which has an incidence that increases with age and a progression rate to multiple myeloma of 20% over 10 years. Myeloma is a disease affecting mainly the elderly with a median age at diagnosis of 60–65 years.

Classification

Most cases of myeloma produce a unique immunoglobulin type from one of the immunoglobulin classes, with approximately the same prevalence as the normal proportion of these immunoglobulins. Therefore the monoclonal immunoglobulin can be either IgG (52%), IgA (21%), IgM (12%) or rarely IgD (2%) and IgE (0.01%), κ or λ light chain only (11%). Non-secretary myeloma is also reported but is rare.

To distinguish myeloma from MGUS, a set of diagnostic criteria is used, which in essence involves a triad of marrow plasmacytosis, monoclonal protein and skeletal lytic lesions. The details are as listed in Table 7.

The diagnosis of myeloma requires a minimum of 1 Major + 1 Minor criterion, or 3 Minor criteria that include a + b.

Table 7 Criteria for Diagnosis of Multiple Myeloma

Major criteria
a) Plasmacytomas on tissue biopsy;
b) Bone marrow plasmacytosis with $>30\%$ plasma cells; or
c) Monoclonal globulin spike on serum electrophoresis exceeding 3.5 g% for G peaks or 2.0 g% for A peaks. ≥ 1.0 g/24 hours of kappa or lambda light chain excretion on urine electrophoresis in the absence of amyloidosis.
Minor criteria
a) Bone marrow plasmacytosis 10–30%;
b) Monoclonal globulin spike present, but less than the levels defined above;
c) Lytic bone lesions; or
d) Normal IgM less than 50 mg%, IgA less than 100 mg% or IgG less than 600 mg%.

The presence of certain non-specific disease features will support the diagnosis, particularly if of recent onset:

1) anemia;
2) hypercalcemia;
3) azotemia;
4) demineralization and compression fractures; or
5) hypoalbuminemia.

Staging and Prognostication

The universally adopted staging system for myeloma is that introduced by Durie and Salmon in 1975,[9] which classifies myeloma into stages I, II or III, each reflecting low, intermediate to high tumor load, and with subclassification into A or B based on creatinine level. Table 8 shows the staging system correlating with tumor load. Median survival based on this staging system is 61 months for Stage I, 55 months for Stage II, 30 months for Stage IIIA and 15 months for Stage IIIB myeloma. Other important prognostic parameters were discovered in subsequent studies, including β2-microglobulin, which is an indicator of tumor load; plasma cell labeling index, which reflects myeloma growth rate; C-reactive protein which corresponds to IL-6 activity (a growth factor for myeloma cells); and age.

Table 8 Durie and Salmon Staging System

Stage	Tumor Load
Stage I: All of the following: Hemoglobin value > 100 g/L Serum calcium value normal or < 2.60 mmol/L Bone X-ray, normal bone structure or solitary bone plasmacytoma only Low M-component: IgG value < 50 g/L IgA value < 30 g/L Urine light chain M-component on electrophoresis < 4 g/24 hr	Low
Stage II: Fitting neither stage I nor stage III	Intermediate
Stage III: One or more of the following: Hemoglobin value < 85 g/L Serum calcium value > 3.00 mmol/L Advanced lytic bone lesions High M-component: IgG value > 70 g/L or IgA value > 50 g/L Urine light chain M-component on electrophoresis > 12 g/24 hr	High
Subclassification (A or B) A: Relatively normal renal function (serum creatinine value < 170 μmol/l) B: Abnormal renal function (serum creatinine value ⩾ 170 μmol/l)	

Laboratory Investigations

For diagnosis, complete staging and a baseline for monitoring response after treatment, the necessary tests include:

1) A marrow study for confirmation of plasmacytosis. Cytogenetic studies are also useful as presence of abnormal karyotypes, in particular 13q deletion and 11q abnormalities carry a poor prognosis. Plasma cell labeling index is useful to risk stratify — an elevated index is associated with worse prognosis.
2) Serum and urine protein electrophoresis, including immunofixation using monoclonal antibodies against IgG, IgA, IgM, κ and λ light chains, to identify the subclass of the paraprotein.
3) Skeletal survey: X-rays may show punched out lytic lesions or diffuse osteopenia and is more sensitive than bone scan in myeloma.
4) Other blood tests: full blood count, serum creatinine, calcium, uric acid to detect any metabolic disturbance, β2-microglobulin and C-reactive protein.

Clinical Manifestation

Myeloma presents with a wide variety of symptoms as a result of its direct infiltration and immunoglobulin secreted, as summarized below:

1) Skeletal involvement: in up to 70% of cases, causing bone pain especially lower back. Sudden severe pain may indicate pathological collapse of vertebra or fracture of long bones.
2) Renal impairment: in up to 50% of cases and is multifactorial. Filtered light chain results in tubular casts, causes tubular dysfunction and decreases glomerular filtration. Amyloid deposition causes proteinuria and nephrotic syndrome. Other reversible causes are dehydration, hypercalcemia, hyperuricemia, hyperviscosity, pyelonephritis, analgesic nephropathy, etc.
3) Hematological: marrow failure as a result of plasma cell infiltrate or transient suppression by chemotherapy.
4) Bleeding disorder: either due to thrombocytopenia, platelet dysfunction or monoclonal immunoglobulin, which interferes with function of clotting factors.
5) Neurological: spinal cord or nerve root compression, which could result in paraplegia or radicular pain; or peripheral neuropathy as a paraneoplastic syndrome.
6) Metabolic: hypercalcemia with symptoms ranging from anorexia, nausea, constipation and somnolence to overt stupor and coma. Hyperuricemia could cause gouty arthritis.
7) Infection: as a result of neutropenia and hypogammaglobulinemia.

Treatment

The mainstay of treatment is chemotherapy, sometimes followed by an autologous stem cell transplant. However supportive treatment plays an equally important role. This includes treatment of hypercalcemia, hyperuricemia, correction of dehydration, treatment of concomitant infection, correction of anemia and pain control. Of note, bisphosphonate has proven its efficacy in preventing skeletal events, and treatment of anemia with erythropoietin has been proven to improve quality of life. Radiation therapy is important for local pain management.

Chemotherapy is indicated for symptomatic patients, while those with asymptomatic Stage I disease may run an indolent course before

treatment is necessary without any benefit in early intervention, as shown in randomized controlled trials.[10]

The choice of definitive treatment depends on the patient's age, other comorbidities and objectives. Pulse oral melphalan and prednisolone for 4–7 days every 4–6 weeks has been used for 4 decades and is still a good regimen for those patients deemed unfit for more aggressive treatment. It improves survival from 7 months in untreated cases to 3 years, but does not result in a cure. Combination chemotherapy regimens are used with the hope to improve outcome. These regimens involve lymphoid active drugs, e.g. cyclophosphamide, vincristine, adriamycin, BCNU, melphalan and prednisolone, in various combinations. Some of these have reported superior disease-free survival. However a meta-analysis done on 27 such randomized trials found no survival benefit from combination chemotherapy as compared to melphalan and prednisolone for any prognostic subgroups.[11]

Over the past few years VAD (vincristine, adriamycin as continuous infusion over 96 hours, and dexamethasone for 4 days) has become the preferred treatment, as high-dose dexamethasone achieves more rapid control of disease as well as avoid potential damage to stem cell in contrast to other alkylating agent-containing regimen. Response rate of 60–80% with CR rate of 10–25% is achievable. There is, however, no long-term survival advantage of this over combination chemotherapy or melphalan/prednisolone, thus VAD is used for those with a plan to proceed to autologous transplant.[10]

After achieving a plateau phase, i.e. stable levels of marrow plasmacytosis and serum monoclonal protein, there is no benefit from continuing with chemotherapy. Alpha interferon as maintenance has been used in various studies. A meta-analysis showed a delay to progression by 6 months but with a negligible survival benefit.[12]

Autologous stem cell transplant has emerged as the consolidative treatment of choice for patients with responsive disease after initial response to chemotherapy. It gives a superior 7-year overall survival of 43% and disease-free survival of 16% as compared to 25% and 8% respectively with chemotherapy alone, as shown in randomized controlled trial.[13] Transplant-related mortality in these generally elderly patients (up to 60–65 years old) is acceptable at $< 5\%$ in most studies. The main problem remains the continuous risk of relapse, underscoring the fact that this procedure only delays relapse and does not effect in a cure. Tandem

autologous transplant has produced higher CR rate in single arm studies but result of randomized controlled trials comparing single vs tandem transplant is still immature. Alpha interferon maintenance has also produced delay in relapse in the post autologous transplant setting.

Allogenic transplant from matched sibling donors has limited application in myeloma as few are young enough with suitable HLA-matched siblings. It has the advantage over autologous transplant of infusing a disease-free marrow and being a proven graft vs the myeloma effect.[14] In registry data comparing allogenic vs case-matched historical autologous cases,[15] relapse rate is lower but the high transplant-related mortality resulted in inferior eventual survival. Reduction in transplant-related mortality with use of non-myeloablative regimen may improve the outcome, but no concrete recommendation can be made as yet.

Over the past few years, a new discovery that makes an impact in myeloma treatment is the use of thalidomide. Even in patients with refractory disease failing several lines of prior treatment, a response rate of 20–30% is still possible with this oral treatment.[16]

Other Plasma Cell Disorders

There are other plasma cell dyscrasias besides multiple myeloma, including MGUS, smouldering myeloma and solitary plasmacytoma. Table 9 summarizes the distinguishing features of each.

Table 9 Comparison between the Various Plasma Cell Disorders

	MGUS	Smouldering Myeloma	Solitary Plasmacytoma	Multiple Myeloma
Marrow plasma cell	$<10\%$	$>10\%$	$<5\%$	$>10\%$
"M" band	<3 g/dL, stable	>3 g/dL, stable	nil or low	>3 g/dL, rising
Labeling index	$<0.8\%$	$<0.8\%$	$<0.8\%$	often $>0.8\%$
Symptoms	absent	absent	absent	often present
Lytic lesions	absent	absent	solitary	often present
Progression to myeloma	20% in 10 years	stable for many years	1/3 in 3 years, $>1/2$ in 10 years	
Treatment	watch	watch	excision or irradiation	chemotherapy

REFERENCES

1. Devita VT Jr, Hellman S, Rosenberg SA, *Cancer, Principles and practice of Oncology*, 5th ed., New Jersey, Lippincott Williams and Wilkins Healthcare, 1999.
2. Harris NL, Jaffee ES, Diebold J, Flandrin G, Muller–Hermelink HK, Vardiman J, *et al*., World Health Organization Classification of neoplastic disease of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee Meeting, *J Clin Oncol* **17**:3835–3849, 1999.
3. International Workshop on Chronic Lymphocytic Leukemia: Chronic lymphocytic leukemia: Recommendations for diagnosis, staging and response criteria, *Ann Intern Med* **110**:236–238, 1989.
4. Rai KR, Han T, Prognostic factors and clinical staging in chronic lymphocytic leukemia, *Hematol Oncol Clin North Am* **4**:447–456, 1990.
5. Binet J-L, Anquier A, Digihiero G, Chastang C, Piguet H, Goasguen J, *et al*., A new prognostic classification of CLL derived from multi-variate survival analysis, *Cancer* **48**:198–207, 1981.
6. Beutler E, Coller BS, Lichtman MA, Kipps TJ, Uri Seligsohn, *Williams Hematology*, 6th ed., international ed., McGraw–Hill Company, 2001.
7. International Non-Hodgkin's Lymphoma Prognostic Factors Project, A predictive model for aggressive Non-Hodgkin's lymphoma, *N Engl J Med* **329**:987–994, 1993.
8. Isaacson PG, The current status of lymphoma classification, *Br J Haematology* **109**(2): 258–266, 2000.
9. Durie BGM, Salmon SE, A clinical staging system for multiple myeloma: correlation of measured myeloma cell mass with presenting clinical features, response to treatment and survival, **36**:842–854, 1975.
10. The UK myeloma Forum guidelines Working Group, Diagnosis and management of multiple myeloma, *Br J Haematology* **115**:522–540, 2001.
11. The Myeloma Trialists' Collaborative Group, Combination chemotherapy vs melphalan plus prednisone as treatment for multiple myeloma: an overview of 6633 patients from 27 randomized trials, *J Clin Oncol* **16**:3832–3842, 1998.
12. The Myeloma Trialists' Collaborative Group, Interferon as therapy for multiple myeloma: and individual patient data overview of 24 randomized trials and 4012 patients, *Br J Haematology* **113**(4-II):1020–1034, 2001.
13. Attal M *et al*., A prospective, randomized trial of autologous bone marrow transplantation and chemotherapy in multiple myeloma, *N Engl J Med* **335**:91–97, 1996.
14. Singhal S *et al*., Anti-tumour activity of thalidomide in refractory multiple myeloma, *N Engl J Med* **341**:1565–1571, 1999.
15. Bjorkstrand B *et al*., Allogeneic bone marrow transplantation versus autologous stem cell transplantation in multiple myeloma: A retrospective case-matched study from the European Group for blood and marrow transplantation, *Blood* **88**:4711–1748, 1996.

16. Mehta J and Singhal S, Graft-versus-myeloma, Bone marrow transplantation, **22**:835–843, 1998.

17. Dimpoulos MA, Panayiotidis P, Moulopoulos LA, Stikakis P and Dalakas M, Waldenstrom's macroglobulinemia: clinical features, complications and management, *J Clin Oncol* **18**:214–226, 2000.

37

Bleeding Disorders

Lee Lai Heng

NORMAL HEMOSTASIS

Hemostasis evolved such that bleeding can be arrested at vascular injury and the clotting process regulated to maintain blood circulation within the vascular tree. Vascular damage causes vasoconstriction and also exposes subendothelial structures to flowing blood. Platelets are activated upon exposure to collagen to adhere and aggregate on the injured site. Simultaneously, coagulation proteins are activated to form a fibrin mesh over the aggregated platelets. Platelet contractile activity then draws the attached fibrin polymers more tightly over the injured vascular surface and away from the luminal blood flow.

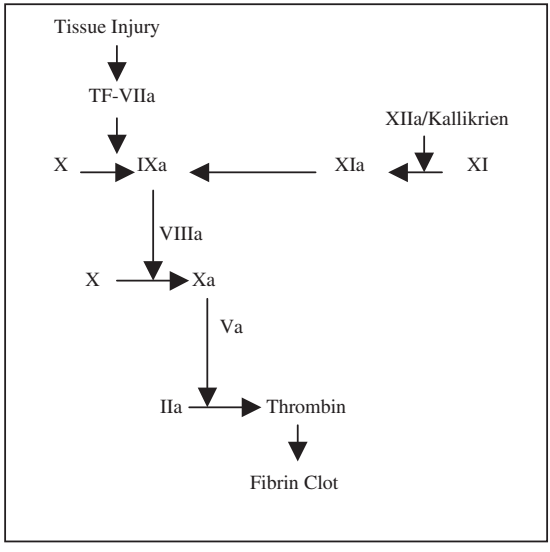
The coagulation and fibrinolytic systems are biological amplification systems that work in unison to maintain the integrity of the vascular tree. While thrombosis is promoted by the coagulation system, spontaneous coagulation is prevented by the naturally occurring anticoagulants and the fibrinolytic system removes any fibrin that is formed within the circulation.

The coagulation system consists of a number of serine protease enzymes. A classic model of the blood clotting cascade gave equal importance to an

"intrinsic" pathway triggered by exposing "contact factors" (XII and XI) to negatively charged subendothelial surfaces and an "extrinsic" pathway initiated by tissue factor, where each converged on factor X. The presently favored model emphasizes cells and cell surface-associated processes, and the dominant influence of the tissue factor. These enzymes may be sequentially activated, in the presence of an adequate supply of co-factors and platelets, to generate thrombin. Thrombin in turn converts fibrinogen to fibrin monomer, which polymerizes to form a fibrin mesh. Clot stability is achieved through covalent cross-linking of fibrin strands and incorporation of platelets, white cells and red cells in the fibrin meshwork form a hemostatic plug that seals the area of blood vessel injury.

Any fibrin formed within the vasculature can be removed by plasmin, which is a protein that degrades fibrinogen and fibrin molecules. Plasmin is generated on the surface of fibrin clots from the circulating precursor molecule plasminogen and thus fibrinolytic activity is relatively confined to the clot and with minimal degradation of fibrinogen within the circulation. The most important physiological activator of plasminogen is endothelial-derived tissue plasminogen activator (tPA). Impairment of regulation of the components of the fibrinolytic system can lead to circulating to active tPA and plasmin with breakdown of both fibrin clots and circulating fibrinogen, a state referred to as "hyperfibrinolysis".

Table 1 Schematic Representation of Blood Coagulation Reactions



CLINICAL APPROACH TO BLEEDING DISORDERS

Inappropriate bleeding may occur as a result of an abnormality of blood vessels, coagulation factors or platelets. Almost all bleeding disorders are caused by deficiency or dysfunction, inherited or acquired, of blood clotting factors or platelets. Inherited disorders are uncommon but require lifelong care. Their study has led to detailed understanding of the genetics, molecular structures, and functions of the plasma blood clotting factors. Acquired disorders are far more frequent than inherited ones and are therefore more important to the generalist. Platelet dysfunction as well as vascular disorders must be excluded before attributing bleeding to coagulation disorders. Bleeding symptoms *per se* are not indicative of bleeding disorders. Underlying bleeding disorder is suspected only if the bleeding is recurrent or inappropriately excessive and prolonged for the situation. It is important to exclude anatomical causes of bleeding such as post-surgical vascular bleed, concealed lacerations or retained placenta before attributing the cause of bleeding to coagulation disorders.

Clinical Presentations

Symptoms of bleeding disorders vary from mild, easy bruising to catastrophic post-traumatic bleeding. Mucocutaneous bleeds such as petechiae, bruising, epistaxis, gum bleeding and menorrhagia suggest underlying platelet defects while hemarthrosis, muscle hematoma and internal bleeding suggest coagulation factor defects. Previous history of spontaneous bleeding, post-dental treatment bleeding, post-surgery bleeding as well as obstetrics and menstrual history is important. Bleeding episodes since childhood implies an inherited disorder while one of recent onset is indicative of acquired disorders. However mild hereditary bleeding disorders may be asymptomatic and bleeding

Table 2 Causes of Coagulopathy

Congenital	Acquired
Hemophilia A and B	Vitamin K deficiency
Other coagulation defects	Liver disease
von Willebrand's Disease	Anticoagulant overdose
Bernard–Soulier Disease	Disseminated intravascular coagulation
Glanzmann's Disease	Acquired inhibitors of coagulation factors
	Massive blood transfusion

tendency only unmasked by major surgery and trauma. A positive family history for bleeding diseases indicates hereditary causes.

Many systemic diseases are associated with acquired bleeding disorders. Autoimmune diseases are often complicated by vasculitis, thrombocytopenias and occasionally give rise to acquired inhibitors to coagulation factors. Systemic infections, bone marrow failure, liver and renal diseases are well known to cause thrombocytopenias as well as acquired coagulopathies. A systemic enquiry of present and past medical history of systemic diseases is essential.

Bleeding tendency is associated with many drugs. Warfarin causes deficiencies in factors II, VII, IX and X. Aspirin, non-steroidal anti-inflammatory drugs; ticlopidine and other anti-platelet drugs inhibit platelet function. Certain antibiotics, particularly the extended spectrum beta lactamases, can cause thrombocytopenias. In the local setting, it is important to exclude traditional and herbal medicines as they can cause a variety of bleeding problems ranging from immune thrombocytopenias, platelet dysfunctions and bone marrow hypoplasias.

A detailed physical examination is required to access the sites and severity of bleeding. Ecchymosis, petechiae and hemarthrosis, if present, are usually apparent. Fundal hemorrhages, signs of intracranial bleed, psoas hematoma and internal bleeding may be less obvious and may be missed unless the index of clinical suspicion is high. Signs of systemic disorders, such as anemia, lymphadenopathy, hepatosplenomegaly, vasculitis, septicemia and arthritis, must be diligently sought.

Laboratory Investigations

1) *Full Blood Count (FBC)*. This indicates the severity of blood loss with hemoglobin level and excludes thrombocytopenia as the cause of bleeding if platelet count is normal.
2) *Clotting times.* Prothrombin Time (PT) accesses the tissue factor pathway. Elevation of PT in isolation indicates Factor VII deficiency. Partial Thromboplastin Time (PTT) measures the clotting activities of Factors VIII, IX and XI. Significant deficiencies in any of these factors will cause the PTT to become prolonged. Elevated PT and PTT values simultaneously indicate rare Factor X, V, or II deficiency. PT and aPTT are global clotting tests; therefore normal values do not exclude mild factor deficiencies.

3) *Assays for coagulation factors* are performed when a certain factor deficiency is suspected. If the deficiency of a coagulation factor is present, inhibitors to the deficient factor must be performed.
4) *Platelet Aggregomety*. Platelet function tests are performed by exposing a patient's platelets to various agonists such as adenosine diphosphate, collagen, epinephrine, and ristocetin that induce platelet aggregation. The degree of aggregation defects does not correlate to severity of bleeding symptoms.

INHERITED BLEEDING DISORDERS

Hemophilias

Hemophilia A and B are hereditary bleeding disorders due to deficiencies in clotting factors VIII and IX respectively. Both are transmitted as sex-linked recessive disorders. The classification of hemophilia is based on the coagulant activity of the factor VIII and IX and reflects the severity of disease. Mild hemophilia has factor levels above 5%, moderate hemophilias has factor levels between 1 and 5% and severe hemophilia has factor level less than 1%.

The clinical manifestations for hemophilia A and B are similar and indistinguishable. Bleeding and its complications cause symptoms and signs of this disease. The severity of the disease varies greatly. Mild hemophiliacs are often asymptomatic and may only present in later life after trauma and surgery. Moderate hemophiliacs have bleeding of intermediate and variable severity. Severe hemophiliacs present early in first years of life with bleeding symptoms.

Clotting factor deficiencies classically cause bleeding in joints, muscles and brain. The frequencies of joints affected are, in order, knees, elbows, ankles, hips and wrists. Recurrent joint bleeds lead to crippling arthritis and joint deformities. Spontaneous muscle bleeding can affect the arms, legs, ileo-psoas or any site, and can lead to compartmental syndrome, nerve compression and eventual muscle fibrosis. Hematuria is common; retroperitoneal and central nervous system bleeds are rare but life-threatening when they occur.

The diagnostic laboratory test is to demonstrate a deficiency of the affected factor. The PT is normal while aPTT is usually prolonged but a normal aPTT does not exclude mild disease. Factor assays will show a

decrease Factor VIII in hemophilia A and a decrease Factor IX is hemophilia B. In cases with Factor VIII deficiency, tests must be done to exclude von Willebrand's disease before the diagnosis of hemophilia A can be established. In all cases of factor deficiencies, assays for inhibitors to the deficient factor must be performed.

Management of hemophilias

Lifelong support and supervision are required in management of hemophiliacs and involves the collaboration between the hemotologists and specialists from other disciplines. It is important to check for hepatitis viruses and HIV at presentation. Patients should receive vaccinations against hepatitis A and B if not immune. Patient education is essential; they must be taught on care of joints, the management of minor bleeds and to avoid medications such as aspirin.

Replacement of the deficient factor remains the cornerstone in the treatment of severe hemophiliacs and life-threatening bleeding. This is the only logical way to restore coagulation in these situations. Pharmacological products including antifibrinolytics (e.g. tranexamic acid) and desmopressin can be used alone in mild bleeding episodes. They are also useful adjuncts to factor replacements in treating severe bleeding episodes. However one-third of the hemophilia A patients do not respond to desmopressin and tachyphylaxis occurs in responders with repeated use.

Cryoprecipitate and fresh frozen plasma preceded factor concentrates as sources of Factor VIII and IX respectively. These products are freshly frozen from donated blood and are not virally inactivated, hence they are no longer recommended. Presently, there is a variety of factor VIII and IX concentrate products available commercially, and all are processed to inactivate hepatitis and HIV viruses. Factor replacements are adjusted to achieve the desired factor levels in the circulation according to the nature of bleed.

Factor level VIII and IX levels that are 10–20% of normal are usually adequate to sustain hemostasis after minor trauma. For major injuries, surgery and bleeding into dangerous sites (e.g. central nervous system, lungs or retropharyngeal space), factor levels must be maintained at 50–100% for several days. The half-life of factor VIII is 8–12 hours, so bolus infusions should be repeated every 8–12 hours to maintain hemostasis in cases of

severe bleeding and major surgery. Treatment with factor IX is more unpredictable than with factor VIII due to the variable half-life of infused factor IX of 18–24 hours. Factor IX infusions should be repeated every 12–24 hours to achieve the desired level.

Continuous factor infusion therapy was shown to be superior to bolus injection in achieving a stable hemostatic effect with a saving of 20–50% in the required factor. Continuous infusion of Factor VIII at a rate of 2 u/kg/h gave a plasma level of 50 u/dL and infusion of Factor IX at a rate of 7.5 u/kg/h gave a plasma level of 1 u/dL. This mode of therapy is useful for patients who require prolonged factor replacement.

Home treatment with self-injections of factor concentrates is encouraged, as early treatment of bleeding episodes is essential. It reduces the severity of bleed and the long-term morbidity and joint deformities. Parents of young patients and older patients from the age of 7 onwards can be trained and taught self-injections. Central lines for venous access in the very young patients may be considered.

Gene therapy is aimed at treating the root of the problem since hemophilia is basically a genetic disorder. After many years of research and experiments, Clinical trials for this modality of treatment has finally begun.

Complications of disease

All complications of the disease are sequelae of previous bleeds. Repeated hemarthrosis lead to chronic arthropathy and joint deformities that are crippling. Intramuscular hematomas can cause compartmental syndrome, muscle atrophy, fibrosis and eventual contractures. The functions of affected joints should be assessed regularly to see if joint surgery or replacements are required. Permanent neurological deficits result from previous intracranial bleeds.

Complications of therapy

Development of inhibitors to clotting factors is an immunological phenomenon. After repeated transfusions of blood products, some patients develop antibodies against the factor they lack. About 20% of patients with hemophilia A and less than 5% of patients with hemophilia B develop inhibitors to factors VIII and IX respectively. Inhibitor levels are

measured in Bethesda units (BU). Therapy for patients with inhibitors poses a great challenge in the management of hemophiliacs. For low titer factor VIII inhibitors of less than 5 BU, a large bolus dose of factor VIII can be used to "swarm" the inhibitors, followed by factor VIII infusion to arrest the bleed. For high titer inhibitors, porcine factor VIII can be used for patients with hemophilia A while activated prothrombin complex (containing activated factors II, VIII, IX and X) can be used for both hemophilia A and B. The recent development of recombinant VIIa paves the way for a new and attractive approach in the treatment of hemophiliacs with inhibitors. VIIa binds to any exposed tissue factor and directly activates factor X, bypassing the factor IX–VIII step. Treatment to eradicate the inhibitor is difficult. Steroids, cytotoxic agents and immune globulin are not effective against hemophilic allo-antibody inhibitors. Plasmapheresis and affinity chromatography that removes immunoglobulins using protein A sepharose can transiently lower the inhibitor level to allow factor VIII infusion therapy. This approach is useful in preparation for elective surgery but the logistics in organizing such treatment limits its use in acute bleeding episodes.

Allergic reactions are usually due to the proteins present in the blood product. It is more commonly seen with use of intermediate purity products and less with the use of highly purified factor concentrates.

Transfusion-related infectious diseases pose a very real problem in the treatment of hemophilias. Previously when cryoprecipitate and factor concentrates, which were not treated for viral inactivation, were used, HIV and hepatitis viruses were transmitted to patients. Currently, all factor concentrate products approved and registered for use are virally inactivated.

Other Congenital Coagulation Deficiencies

Deficiencies of all the coagulation factors have been described and the commoner ones are factors IX, II, V, VII, X and XIII. Some families may have combined deficiencies known as Familial Multiple Coagulation Factor Deficiencies. Their prevalence of 1–2 per million is rare compared to that of hemophilia A or B. They are mostly of autosomal recessive inheritance. Deficiencies in multiple factors due to a single gene defect are seen in consanguineous families. Most of these patients with inherited coagulation factor deficiencies do not bleed unless exposed to trauma or

surgery. When uncontrolled bleeding arises, the aim is to restore plasma concentration of the required factor up to the minimal hemostatic value. Factor concentrate products are not available for these other coagulation proteins, so fresh frozen plasma remains as the main source of coagulation factors. Fresh frozen plasma is not virally inactivated and the risk of acquiring transfusion-related infections remains high. Management of patients with inherited coagulation factors deficiencies should reflect knowledge of specific disorder being treated and clinical circumstances in which treatment is ordered.

von Willebrand's Disease

von Willebrand's Disease (vWD) is the commonest form of autosomally inherited bleeding disorder that occurs worldwide. It has a high prevalence of the disease that affects up to 1% of the population and an estimated incidence of 1 in 5000. The disease occurs as a result of either a reduction in number or function of von Willebrand's Factor (vWF). All vWD is caused by mutations at the vWF locus on chromosome 12.

vWF functions as the mediator for platelet adhesion to subendothelium and platelet aggregation. It also serves as a transport and carrier protein for factor VIII. Factor VIII not attached to vWF and is labile in the plasma. vWF is synthesized and stored in the endothelial cells and megakaryocytes. Secreted vWF protein monomers form dimers that subsequently polymerize to form high molecular weight multimers. Hence, the vWF has molecular weight ranging from 500 000 to 2 000 000 depending on the degree of polymerization. The multimers of higher molecular weight are hemostatically more competent than those with lower molecular weight.

Classification of the disease is necessary as the complex functions of the vWF gives rise to the wide variations in presentation and severity of von Willebrand's Disease. More than 20 subtypes have been described. For simplicity, the disease is classified into 3 main types with type II disease further subtyped into IIa and IIb. Other entities worth mentioning are the von Willebrand-like disorders. These are not due to genetic mutations at vWF locus but its clinical presentations may be similar.

The clinical features vary markedly. Bleeding may be intermittent. Most times bleeding is related to platelet dysfunction giving rise to mucocutaneuos bleeds such as epistaxis, menorrhagia, easy bruising and

Table 3 Classification of vWD

Disease, Inheritence, Pathogenesis	Clinical Features	Laboratory Features
Type I vWD — autosomal dominant. Due to decreased production of normal vWF.	70% of vWD. Bleeding usually mild.	Reduction in vWF antigen, RiCof activity and factor VIII. Multimeric pattern is normal.
Type IIa — autosomal dominant. Caused by a structural abnormality of vWF resulting in loss of high molecular weight multimers.	20% of vWD.	Low level of RiCof activity compared to vWF antigen. Multimeric pattern shows absence of largest vWF multimers.
Type IIb — autosomal dominant. Increased affinity of vWF to GpIb leading to absence of high molecular weight multimers in the plasma.	5% of vWD. Desmopressin (DDAVP) may worsen thrombocytopenia.	Increased RIPA. Multimeric pattern shows absence of largest vWF multimers. Thrombocytopenia.
Type III — autosomal recessive. Almost no production of vWF.	Uncommon. Severe bleeding disorder with mucocutaneous bleed, bleeding at other sites and hemarthrosis.	Complete absence of vWF antigen, RiCof activity and multimers from the plasma.
Platelet type vWD (Pseudo vWD) — autosomal dominant. Mutation in platelet GpIb gene causing increased binding of normal vWF to GpIb.	Cryoprecipitate can precipitate *in vivo* thrombocytopenia.	Decreased amount of high molecular weight-multimers, thrombocytopenia and increased RIPA. These abnormalities not corrected by the addition of normal vWF as the defects lie within the platelets.
Normandy variant von Willebrand's disease — autosomal recessive. Due to point mutations in the amino portion of vWF which is the binding site for FVIII:C	Often misdiagnosed as hemophilia A, but the response to purified factor VIII is not optimal. Appropriate treatment is infusion of vWF concentrates or cryoprecipitate.	

prolonged bleeding after cuts, dental extractions and surgery. Patients with type III vWD and Normandy variant vWD have severe bleeding like that of hemophilia A with hemarthrosis, muscle hematomas and spontaneous bleeding. Type IIb may cause thrombocytopenia, often noted in full blood counts done for unrelated reasons.

Laboratory investigations are required for diagnosis and classification of the disease. vWF protein is measured by its antigenic levels; it is an acute phase protein and its level may be elevated by stress and in circumstance such as pregnancy, estrogen, neoplasm, infections and thyrotoxicosis. This test result must be interpreted in context of clinical settings. Ristocetin co-factor assay (RiCof) is the functional assay for vWF; it measures the activity of vWF in plasma by promoting the aggregation of fixed or washed platelets in response to the antibiotic ristocetin, which induces aggregation of platelets through GpIb. Ristocetin Induced Platelet Aggregation (RIPA) is useful in screening for variants of vWF that cause increased platelet aggregation, such as type IIb and platelet type vWD. This test measures platelet aggregation in fresh platelet-rich patient plasma induced by high and low doses of ristocetin. Factor VIII levels can be reduced in some patients. Multimeric analysis is essential for classification of vWD; it allows the determination of the amount and size distribution of the vWF multimers through gel electrophoresis and immuno-staining using anti-vWF antibodies.

Treatment of von Willebrand's disease

While the general management is similar to that of hemophilias, there are specific treatment and special caution pertaining to management of vWD. Pharmacological agents play a greater role in the treatment of vWD. Desmopressin or DDAVP given intravenously increase factor VIII and RiCof 3 to 5 times the baseline levels in 80% of patients with vWD. It stimulates the release of vWF from the endothelial stores and repeated doses cause tachyphylaxis. It is recommended for patients with mild to moderate disease undergoing moderate hemostatic stress. However, it is contraindicated in type IIb vWD as it induces *in vivo* thrombocytopenia and it is ineffective in type III vWD. Patients who are diagnosed to have vWD and who are not type IIb should have their response to DDAVP documented; it is generally reproducible in a single patient and in other members of that kindred. A concentrated nasal spray is available for home treatment.

vWF with high molecular multimers are present in intermediate purity factor VIII concentrates and cryoprecipitate. Recently, chromatography-purified concentrates particularly rich in vWF have become available. These are used as vWF replacements to be given to patients not responding to DDAVP and in major bleeding episodes or surgery. The factor concentrates are virally inactivated and are much preferred over cryoprecipitate. Cryoprecipitate is not virally attenuated and carry the risk of transfusion-transmitted infectious diseases.

Pregnant patients with vWD will require special care during pregnancy and puerperium. vWF and factor VIII rise during the third trimester of pregnancy and many patients do not bleed during labor. In patients with severe disease with factor VIII levels less than 30%, prophylactic factor replacement will be required for delivery. In the postpartum period, as the vWF levels fall, patients will be at risk of hemorrhage for up to one-month postpartum. They may require factor replacement if bleeding is severe.

ACQUIRED BLEEDING DISORDERS

Vitamin K Deficiency

The essential role of vitamin K in coagulation is as a co-factor for the gamma carboxylation of the precursor proteins for factors II, VII, IX and X which are produced in the liver. Vitamin K is a fat-soluble vitamin obtained either by dietary intake from vegetables and liver and absorbed in the small gut or produced by bacterial synthesis in the gut and absorbed by the colon. Causes of vitamin K deficiency are multiple and include malnutrition, malabsorption, obstruction of bile flow, prolonged use of antibiotics and rat poisons.

Prolongation of PT occurs before PTT because of the short half-life of factor VII. Eventually both PT and PTT are being prolonged if the condition is left untreated. Both PT and PTT prolongations are correctable with addition of normal plasma. Fibrinogen level and platelet count will be normal in the absence of liver disease. A therapeutic trial of vitamin K will confirm the diagnosis with rapid correction of PT after 24 hours.

Treatment is aimed at the underlying diseases. Prophylactic vitamin K should be given to patients at risk. In asymptomatic vitamin K-deficient patients, it is sufficient to give vitamin K replacement at 10 mg per day for

a few days. For patients who are bleeding, fresh frozen plasma can be given for immediate replacement of clotting factors in addition to vitamin K. In life-threatening conditions, prothrombin complex concentrates (PCC) or Factor VIIa can be considered. Brodifacoum, a commonly used rodenticide, has a long half-life and irreversibly inhibits vitamin K action. Brodifacoum poisoning will require high doses of vitamin K at 25–50 mg three times a day for several months till the poison is confirmed to be excreted by drug levels.

Liver Disease

Decreased clotting factors, hyperfibrinolysis and thombocytopenia all contribute towards the bleeding tendencies in liver disease. Most coagulation factors are made exclusively in the liver; diseases affecting the synthetic function cause rapid reduction of coagulation factors concentrations. The liver also functions as a reticuloendothelial organ; clearing activated clotting factors from the circulation. Impairment of this function together with the release of procoagulant substances from the diseased liver can cause disseminated intravascular coagulation and hyperfibrinolysis. Thrombocytopenia is invariably present and is due to decreased production, increased consumption and splenic pooling.

In acute liver disease caused by viruses, toxins or drugs, there is impaired carboxylation of vitamin K-dependent clotting factors, giving rise to prolongation of PT; moderate thrombocytopenia and mild or absent hypofibrinogenemia. However bleeding rarely occurs and complete normalization of hemostasis occurs with the recovery of liver. In contrast, acute liver failure with encephalopathy is always associated with prolongation of PT, aPTT, thrombocytopenia and hypofibrinogenemia. Defective synthesis of clotting factors, DIC triggered by procoagulant substances from the injured liver and hyperfibrinolysis all contribute to the badly deranged hemostatic functions. In chronic hepatocellular disease, there is progressive loss of liver synthetic functions. Due to the short half-life of factor VII, PT is prolonged before PTT becomes prolonged. Normal or elevated levels of fibrinogen are present in stable chronic liver disease but abnormal sialic acid content causes dysfibrinogenemia. Hypofibrinogenaemis eventually occurs in the terminal stages of the disease. Associated thrombocytopenia further aggravates bleeding diathesis.

Management of hemostatic failure in liver disease involves the use of blood component therapy and pharmacological manipulation of coagulation and fibrinolytic systems. Vitamin K should be given as a prophylaxis. Administration of fresh frozen plasma will partially correct the clotting times and reduces the consumption of coagulation factors as it contains all the clotting factors as well as natural anticoagulants and inhibitors of fibrinolysis. Platelet transfusions are required in bleeding patients with thrombocytopenia. Infusion of desmopressin shortens the bleeding time in patients with cirrhosis. Aprotinin via continuous infusion is associated with marked suppression of fibrinolysis. These drugs may play an important role as adjuncts in addition to blood component therapy in patients with uncontrollable hemorrhage. Prothrombin complexes (PCCs) and antifibrinolytic drugs such as tranexamic acid are contraindicated because of their potential in inducing DIC.

Acquired Anticoagulants

An acquired anticoagulant or inhibitor is usually an immunoglobulin that interferes with normal coagulation by inhibiting the functions of clotting factors. Inhibitor levels are measured in Bethesda units (BU). Inhibitors to many clotting factors have been described, of which factor VIII inhibitor is most commonly seen. Factor VIII inhibitors occur spontaneously in non-hemophilia patients who are never exposed to exogenous factor VIII; especially in elderly patients, pregnant and postpartum women and patients with immunological disorders such as systemic lupus erythematosus and rheumatoid arthritis. Clinical manifestations are similar to those with hemophilia, such as ecchymosis, muscle hematomas and hemarthrosis. Transfusion therapy to achieve hemostasis is identical to the treatment for hemophilia patients with inhibitors. In contrast to hemophiliacs, most patients with acquired inhibitors respond to treatment to eradicate the inhibitor. Prednisolone and oral cyclophosphamide daily have been used separately or in combination with high response rates.

Lupus-type anticoagulants are immunoglobulins with antiphospholipid activities, usually of IgG or IgM class. They can occur with or without association with other autoimmune disorders. It causes an *in vitro* prolongation of clotting time in phospholipid-dependent coagulation assays. Such abnormal blood investigation results often herald a host of investigations to exclude coagulopathy. In reality, antiphospholipid antibodies do

not cause abnormal bleeding symptoms. Paradoxically they are associated with hypercoagulable states predisposing to thromboembolism.

Disseminated Intravascular Coagulation (DIC)

Disseminated Intravascular Coagulation (DIC) results from inappropriate activation of blood coagulation and fibrinolysis by pathogenic stimuli. It is caused by procoagulants that are introduced into or produced in the blood which activate the clotting system so much that the natural anticoagulation systems are overwhelmed. This causes the formation of microthrombi in the circulation, hence the term "disseminated intravascular coagulation". In normal response to tissue injury, thrombin generation is contained to the site of injury which leads to clot formation to stop blood loss. In DIC, the pathogenic procoagulant stimulus causes an unregulated thrombin explosion, with the release of excessive free thrombin into the circulation and resulting in the activation of platelets and formation of fibrin clots. Widespread microvascular thrombosis causes depletion through consumption of the clotting factors, platelets and the natural anticoagulants proteins such as anti-thrombin III, Protein C and S. In an attempt to maintain vascular patency, excess plasmin is generated so that systemic as well as local fibrinolysis occurs. It is the generation of free thrombin that causes the thrombotic manifestations of tissue ischemia and organ damage while plasmin generation together with depleted clotting factors and thrombocytopenias are responsible for the hemorrhagic manifestations. The derangements of coagulation and fibrinolysis seen in DIC are mediated by cytokines such as interleukin-6 and tumor necrosis factor that are released in response to systemic infections and inflammation.

There are multiple causes of DIC. The endothelium may be disrupted so that tissue factor is released from tissue damaged by trauma, ischemia, excessive metabolic stress, toxins, infective organisms, or activation of complement system. White blood cells release tissue factor in response to tissue insults, leukemic cells may spontaneously release tissue factor and cancer cells can directly activate factor X. Snake venoms are capable of activating many components of the clotting system. In essence, anything that leads to an overproduction of thrombin will cause DIC and this results from an immense number of clinical situations.

The main clinical features of DIC are thrombosis and bleeding. Patients can be asymptomatic with laboratory evidence of DIC. This is often seen in

Table 4 Causes of DIC

Acute DIC	Chronic DIC
Infections	Malignancy
Obstetric complications	Retained dead fetus
Liver disease	Liver disease
Trauma, toxins	Severe localized intravascular coagulation
Massive hemolysis	

early sepsis, malignancies and liver disease; with progression of disease, these patients can deteriorate rapidly. In acute DIC, massive activation of the coagulation does not allow for any compensatory effort, resulting in depletion of coagulation and anticoagulation proteins, and thrombocytopenia. Bleeding symptoms occur due to clotting factor depletion, platelet dysfunction, thrombocytopenia and excessive fibrinolysis. Such bleeding is often generalized with multiple ecchymosis, bleeding from intravenous access sites and surgical wounds, and endotracheal and gastric aspirates are heavily blood stained. Thrombosis in DIC occurs at the microvascular level giving rise to tissue ischemia and multiple organ failure that are extremely life-threatening. In chronic DIC, steady low-level or intermittent activation is variably compensated by increased production of coagulation and anticoagulation factors as well as platelets. Venous thrombosis of big vessels is rare in the setting of acute DIC but often seen mainly in patients with malignancies and chronic DIC.

DIC is primarily a clinical diagnosis. Laboratory tests are used to confirm the diagnosis and monitor replacement of blood components. PT and PTT are both prolonged in severe and fulminant cases but may be normal in chronic cases. In the very early phases of acute, fulminant DIC, the PTT may be shortened because of the activated state of the coagulation system. The platelet count, coagulation factor concentrations, in particular fibrinogen are all decreased in the consumptive phase of severe DIC but may be in the normal range in chronic DIC. Fibrin monomers are formed when thrombin cleaves fibrinogen to form fibrin; this reflects the excessive thrombin generation present in DIC. Fibrin Degradation Products and D-Dimers are formed when there is plasmin-induced enzymatic degradation of fibrin clot. Hence they are measures of plasmin generation brought about by thrombin production.

It is most important to treat the underlying cause. Maintenance of blood volume and tissue perfusion is essential, and replacement of blood

components is indicated if the patient is bleeding or an invasive procedure is required and must be performed. The use of heparin and anticoagulants is controversial as there is no firm evidence for its benefit, however, the risk of drug-induced bleeding in a hemostatically compromised patient is real. Infusions of Antithrombin and Protein C concentrates have been reported to be beneficial. The use of monoclonal antibodies to abolish the derangements of coagulation and fibrinolysis in DIC is still at the experimental phases.

Massive Blood Transfusion

Massive blood transfusion is defined as replacement of more than 1 blood volume (5 L) within 24 hours. Hemostatic failure results from several mechanisms. There is loss of clotting factors and platelets from bleeding. Large volume replacements with fluids and packed red blood cells lacking in clotting factors and platelets leads to dilution and further lowers their concentrations within the circulation. DIC triggered by sepsis, tissue trauma, acidosis and hemolytic transfusion reaction lead to consumption of clotting factors and platelets, further aggravating the hemostatic insult.

The main laboratory tests to access blood volume and hemostatsis are FBC, PT, PTT, fibrinogen levels. The management is aimed at keeping the hematocrit above 0.3 by red cell transfusions, platelet transfusions to keep the platelet count above 50×10^9/L and fresh frozen plasma infusions to maintain normal clotting times. Cryoprecipitate infusions can be considered in patients whose fibrinogen level is less than 125 mg/dL. All transfusions should be monitored by repeated basic laboratory tests. For patients with very severe bleeding, platelets and fresh frozen plasma should be given empirically even before the laboratory results are available.

The metabolic complications seen with massive transfusions such as hyperkalemia, hypocalcemia, hypothermia and acidosis should be carefully monitored and corrected.

Anticoagulant Overdose

Bleeding is a complication intrinsic to any anticoagulation therapy. In many occasions, bleeding is a result of existing pathologies unmasked by anticoagulation, e.g. fibroids, peptic ulcers and hemarhoids. Management

of bleeding has to be directed at the underlying cause in addition to reversal of anticoagulation.

Standard heparin therapy is monitored by the aPTT, the therapeutic range is kept at 1.5–2.5 times the average normal control value. In heparin over dosage, the aPTT is markedly prolonged. In overdosed patients who are not bleeding, heparin infusion is discontinued and the aPTT should become normal within 6 hours, as the half-life of infused heparin is only 1–2 hours. For patients with life-threatening bleeding, the effects of heparin are neutralized by intravenous protamine sulphate at a dose of 1 mg per 100 iu of heparin up to a maximum dose of 40 mg.

Low molecular weight heparins (LMWH) can be partially reversed with intravenous protamine sulphate at a dose of 1 mg per 100 iu of LMWH. The pharmacokinetics of presently available LMWHs (enoxaparin, nadroparin and deltaparin) is such that the anti-Xa effects peaks at 3–4 hours after subcutaneous injection and is markedly reduced thereafter. In the event of serious bleeding, protamine sulphate is given; otherwise discontinuing the injections will correct the anticoagulation effects within 12 hours.

Warfarin has a long half-life, so discontinuation will take at least 48 hours and vitamin K will take at least 24 hours to reverse the anticoagulation effects. For asymptomatic patients with an INR between 5–8, warfarin can be ommited for 1–3 days. However, vitamin K should be given to these patients if they have other risks for bleeding, as well as those with INR above 8. Small doses of vitamin K at 0.5 mg and 1 mg are effective in reducing the INR. Patients with serious bleeding require immediate reversal of anticoagulation; fresh frozen plasma in addition to intravenous vitamin K should be given. Factor concentrates containing II, IX and X (PCCs) in addition to specific factor VII have been recommended as a rapid way of reversing severe warfarin overdose in patients with life-threatening bleeding.

PLATELET DISORDERS

Thrombocytopenia can occur as a result of numerous conditions. Decreased platelet production or failure of platelet production occurs in aplastic anemia, hematological malignancies, myelodysplasia, myelofibrosis, bone marrow infiltration by other malignancies and vitamin B12 and folate deficiencies. Increased platelet destruction occurs in DIC where there is increased consumption. More commonly, it is immune-mediated

which can occur on its own as in idiopathic thrombocytopenia or as part of other diseases such as collagen vascular disorders, lymphoproliferative disorders, viral infections or drug-induced. Splenic pooling of platelets occurs in liver cirrhosis, myelofibrosis and thalassemia. Dilutional thrombocytopenia occurs in massive transfusions.

Platelet dysfunctions may give rise to bleeding symptoms. Many congenital disorders with platelet dysfunctions are well-described, e.g. Bernard–Soulier syndrome, Glanzmann's thrombasthenia, Storage pool disease, Wiscott–Aldrich syndrome and many more. In clinical practice, it is the acquired causes that are more important as they are far more commonly seen. Common causes for acquired platelet dysfunction are uremia, liver disease, myeloproliferative disease, dysproteinemia and medications such as aspirins and non-steroidal anti-inflammatory drugs.

Except in the treatment of hematological malignancies where prophylactic platelet transfusions are given, platelet transfusions are indicated only if there is significant bleeding. Pharmacological agents such as tranexamic acid may be useful.

Idiopathic Thrombocytopenic Purpura (ITP)

Idiopathic thrombocytopenic purpura (ITP) is an autoimmune disorder in which anti-platelet autoantibodies bind to the antigens on platelet surfaces, resulting in their destruction by the reticuloendothelial system. Hence, it is also known as primary autoimmune thrombocytopenic purpura. The spleen plays an important role as the major site of platelet destruction and anti-platelet antibody production. There are two forms of ITP — acute and chronic. The latter is defined as persistence of thrombocytopenia for more than 6 months. Almost all adult patients with ITP have the chronic form. Like most other autoimmune disorders, there is a female preponderance in adults; and although it can present itself in any age, tend to occur in young to mid-life. It is a hematological disorder for which diagnostic and treatment strategies are not well-defined. In the past decade, there have been attempts for consensus in the management of patients with ITP resulting in practice guidelines being published.

Patients may be asymptomatic and thrombocytopenia diagnosed on routine full blood counts. Less commonly, patients may present with bleeding symptoms characteristic of thrombocytopenia such as petechiae, mucosal bleeding and easy bruising. If the thrombocytopenia is chronic,

recurrent menorrhagia prior to pregnancy and bleeding GIT may lead to iron deficiency anemia. The most serious and potentially fatal complication is bleeding into the central nervous system; this is fortunately very rare.

ITP is still a clinical diagnosis based on exclusion of other diseases that can cause thrombocytopenia. There is no reliable confirmatory laboratory test. Laboratory tests are done to confirm presence of thrombocytopenia, access and to exclude other diseases causing thrombocytopenia. A full blood count is mandatory to confirm thrombocytopenia and the hemoglobin level indicates severity of bleeding. The stained peripheral blood film is useful to exclude pseudo-thrombocytopenia from ethylenediamine tetra-acetic acid (EDTA) induced clumping of platelets. It is also useful in diagnosing thrombocytopenia due to thrombotic thrombocytopenic purpura and pre-eclampsia. Thyroid function test is usually done as a small proportion of patients with ITP have concomitant thyroid dysfunction. If there are no symptoms to suggest an autoimmune disorder, an antinuclear antibody as a screening test is sufficient. If there are suspicions of other autoimmune disorders, serology tests are required to exclude diseases such as systemic lupus erythematosis, rheumatoid arthritis and antiphospholipid syndrome. Other tests such as liver and renal function tests should be ordered based on the clinical picture. Measurement of platelet-associated IgG (PAIgG) as a diagnostic tool remains unsatisfactory despite improved serological techniques. They offer little diagnostic and prognostic information about patients with suspected ITP. The clinical impact of these antibody tests is controversial. Bone marrow biopsy is an invaluable test for exclusion of other hematological disorders. It is generally ordered for older patients and for patients with a clinical diagnosis of ITP but remained non-responsive to treatment.

The treatment of ITP is not curative and is directed towards the inactivation of anti-platelet antibody production and platelet destruction or removal of a major site of platelet destruction. The degree of thrombocytopenia required for hemorrhage varies between $10–30 \times 10^9$/L. No specific treatment is required if the platelet count is above 30×10^9/L as there is very little risk in bleeding. However, these patients will require follow-up visits for purpose of monitoring the platelet counts. If progressive decrease in platelet counts is observed, specific treatment will be recommended if the platelet count is below 30×10^9/L.

Corticosteroids are the preferred first-line treatment of moderate to severe thrombocytopenia. Corticosteroids work by inhibiting the

production of autoantibodies as well as inhibiting the phagocytosis of the sensitized platelets. Corticosteroids usually take effect within 48 hours with adequate dosing and about 40% of patients achieve an effective response with the platelet count rising to above 100×10^9/L. Dosage of steroids is adjusted by titration down to the lowest efficacious dose to maintain a platelet count above 30×10^9/L. It is inexpensive and efficacious. However, there are many potential adverse effects, namely diabetes, hypertension, excessive weight gain, osteoporosis, acne and psychological disturbances.

Splenectomy removes a major site of destruction for antibody-coated platelets as well as a significant source of antibody production. Splenectomy is indicated if there is no response after 4 weeks of corticosteroids at adequate dose in severely thrombocytopenic patients, for patients who had responded to high-dose corticosteroids but recurs during tapering dosage of steroids, and for patients who are unable to tolerate the side-effects of steroids. The response to splenectomy varies from 50% to 92% in reported series. The rise in platelet count is immediate with a peak in 1–2 weeks. Ten percent of patients will relapse after an initial successful splenectomy, usually within the first year. Splenectomized patients have increased risks of severe pneumococcal infections; therefore all patients should receive pneumococcal vaccine 10–14 days before splenectomy.

Patients with ITP require emergency treatment if the platelet count is less than 10×10^9/L in the presence of bleeding or if there is extensive bleeding or neurological bleed regardless of platelet count. Such emergency treatment involves the simultaneous use of high-dose corticosteroids, platelet transfusion and high-dose intravenous immunoglobulins (IVIG). IVIG is very efficacious in raising the platelet counts rapidly. Its mechanism of action is mainly at reticulo-endothelial cell blockage and acting as anti-idiotype antibodies. It is given by intravenous infusion at the dose of 1 g/kg body weight over 6–8 hours for 2 consecutive days. There are very few adverse effects with IVIG. Headaches, backache, nausea, aseptic meningitis and fever occur in less than 5% of patients. There are rare but serious complications such as anaphylaxis, alloimmune hemolysis, and hepatitis C infection. It is important that only virally inactivated products are used. The main disadvantages are the high cost and the transient nature of responses in raising the platelet counts, requiring repeated infusions that will further escalate the costs.

About a quarter of patients are refractory to corticosteroids and/or splenectomy. Refractoriness is defined as non-responsiveness to steroids or

requiring more than 15 mg of prednisolone per day to maintain a platelet count of above 30×10^9/L. Azathiaprine is used in conjunction with steroids in such cases. Other alternatives include cyclophosphamide, danazol and vincristine. There have been reports of successful use of alpha interferon, cyclosporin A and mycophenolate in patients with refractory ITP.

The management of ITP patients involves long-term follow up to prevent bleeding as well as to manage the many side-effects and complications of the drugs used to treat this illness.

FURTHER READING

1. Mann KG, Biochemistry and Physiology of Blood Coagulation, *Thromb Haemost* **82**(2):165–174, 1999.
2. Rapaport SI, Pao VM, The Tissue Factor Pathway: How It Has Become a "Prima Ballerina", *Thromb Haemost* **74**(1):7–17, 1995.
3. Beutler E, Lichtman MA, Coller BS, Kipps TJ, *William's Hematology*, 5th ed., pp. 1413–1516, 1995.
4. Haemophilia: State of the Art. Guest Editors — Varon D, Martinowitz U, Heim M, *Haemophilia* **4**:331–696, 1998.
5. Lusher JM, Gene Therapy for Haemophilia A and B: Patient Selection and Follow-Up, Requirements for a Cure, *Thromb Haemost* **82**(2):572–575, 1999.
6. Lee LH, Baglin T, Haemostasis Associated with Liver Disease and Liver Transplantation, *Ann Acad Med Singapore* **23**(s):77s–83s, 1994.
7. Baglin T, Disseminated intravascular coagulation; diagnosis and treatment, *BMJ* **312**:683–687, 1996.
8. Levi M, deJonge E, van de Poll T, ten Cate H, Disseminated intravascular coagulation, *Thromb Haemost* **82**(2):695–705, 1999.
9. ten Cate H, Timberman JJ, Levi M, The pathophysiology of disseminated intravascular coagulation, *Thromb Haemost* **82**(2):713–717, 1999.
10. Karpatkin S, Autoimmune (idiopathic) thrombocytopenic purpura, *Lancet* **349**: 1531–1536, 1997.
11. George JN, Woolf SH, Raskob GE, Wasser JS, Aledort LM, Ballem PJ *et al.*, Idiopathic thrombocytopenic purpura: A practice guideline developed by explicit methods for the American Society of Hematology, *Blood* **88**:3–40, 1996.

38

Thrombotic Disorders

Lee Hai Heng

Thrombotic disorders occur when a solid clot forms within the circulation during life. Thrombotic events in both arterial and venous systems are exceedingly common, associated with significant morbidity, and frequently fatal. Pathogenesis and background disorders giving rise to thrombus formation in arteries and veins are different. Myocardial infarctions, occlusive strokes and peripheral vascular diseases occur against the background of atherosclerotic disease, hyperlipidemia, hypertension and diabetes mellitus. Venous thrombosis in the form of deep vein thrombosis and pulmonary embolism occur in subjects where immobility and underlying prothrombotic states are contributory causes.

PATHOGENESIS OF THROMBOSIS

Hemostatic mechanisms and formation of thrombus has been described in Chapter 38. Thrombin plays a central role in hemostasis and thrombosis but unopposed thrombin generation is limited by many physiological factors that include natural anticoagulants (protein C, protein S and antithrombin) protective elements of the endothelial cell surface and the

fibrinolytic system. More than a hundred years ago, Virchow observed that abnormalities in the vessel wall, blood components and the dynamics of blood flow all contributed towards the pathogenesis of thrombosis. This observation still holds true today. Deficiency or dysfunction of control mechanisms may predispose to thrombosis (thrombophilia).

The Role of Blood Components

The fibrinolytic system limits formation of thrombus to the sites of vascular damage in order to maintain vessel patency. Plasminogen incorporated within the thrombus is converted to plasmin resulting in clot lysis while plasminogen bound to endothelium adjacent to a thrombus can be activated to prevent clot propagation. Other blood components such as anti-thrombin, protein C and protein S act in association with the vascular endothelium to form negative feedback mechanisms to halt thrombin formation, thereby preventing thrombus formation. Cellular components, the platelets in particular, play an important role in thrombosis.

The Role of the Endothelium

The most important negative feedback mechanism in preventing thrombosis involves thrombomodulin, an endothelial cell surface receptor that can bind with thrombin. Complex formation between thrombin and thrombomodulin renders thrombin inactive towards platelets or fibrinogen. Thrombomodulin also mediate the negative feedback pathway of protein C and S. Thrombin-thrombomodulin complex rapidly activates protein C to generate an enzyme (Activated Protein C) that requires protein S as its co-factor to cleave susceptible sites on factors VIIIa and Va, and curtail further generation of factors Xa and IIa, thereby preventing further thrombin generation.

Antithrombin (previously called antithrombin III), a circulating general purpose serine protease inhibitor with a molecular weight of 58 kilodaltons and plasma concentration of about 140 μg/ml, forms covalent complexes with clotting factors XIa, IXa, Xa and IIa (all are enzymes from the serine protease family) that are inactive and cleared by the liver. Complex formation is greatly accelerated by heparan sulphates on the endothelial cell surface. These heavily sulphated heparin-like polysaccharide chains induce a conformational change in antithrombin that greatly increases its affinity for serine proteases.

Endothelial cells have at least two other defenses to restrict extension of hemostatic plugs onto undamaged areas of vessel wall. Stimuli that cause platelets to generate and release thromboxane A_2 (a powerful vasoconstrictor and platelet proaggregant) also cause endothelial cells to generate prostacyclin (another prostaglandin derivative that is a powerful vasodilator and anti-aggregant). Local hypoxia from overlying thrombus deposition causes endothelial cells to release tissue plasminogen activator and initiate physiological fibrinolysis.

The Role of Disturbed Flow

Stasis of blood lead to accumulation of platelets and clotting factors, especially thrombin, promote thrombus formation and hypoxia in areas of static flow impairs the antithrombotic abilities of the vascular endothelium further encourage thrombosis. This situation is made worse in diseased vessels in which the endothelium has impaired protective antithrombotic mechanisms. Mechanisms of thrombosis differ according to the part of the vascular system involved, as the dynamics of blood flow are different. Arterial thrombosis usually occurs as a result of atheroma causing vascular endothelial damage, thereby exposing subendothelial tissue to the components of blood. This leads to adherence and aggregation of platelets around the damaged vascular endothelium. Platelet aggregation is more pronounced in the arterial system, forming the "white clot" that is characteristic of arterial thrombosis. Venous thrombosis usually occurs in normal blood vessels. In most cases it originates from venous saccules of calf veins or in valve pockets of veins in proximal lower limbs. Venous thrombi consist mainly of red blood cells and fibrin, also known as "red thrombi". Important causes for venous thrombosis are stasis and hypercoagulable states.

DIAGNOSIS OF THROMBOSIS

Arterial Thrombosis

Arterial thrombosis causes a large variety of diseases including coronary arterial diseases, strokes and peripheral vascular diseases. A full discussion of the diagnosis and treatment of all the arterial thrombotic diseases is beyond the scope of this chapter. One fundamental point is the lack of evidence of a dominant thrombotic component in many occlusive arterial events.

Venous Thrombosis

The correct diagnosis of venous thromboembolism (VTE) is necessary for the appropriate management as most methods of treatment carry major risks, particularly of bleeding. Incorrect diagnosis of VTE, if not present, will therefore expose the patient to these risks unnecessarily. Conversely, an undiagnosed DVT left untreated can result in death. Deep vein thrombosis (DVT) and pulmonary embolism (PE) are different spectrums of the same disease caused by venous thombosis. It is well-established that this diagnosis based on clinical criteria alone is notoriously inaccurate. None of the clinical features of DVT and VTE are unique to the disease; the clinical features of venous thromboembolism are elusive and the disease often unsuspected. Objective testing is a necessity as clinical diagnosis alone is inaccurate and fatality can occur in untreated patients.

Up to two-thirds of venous thrombi are asymptomatic and thrombi are present only half the time when signs and symptoms suggest their presence. The classical clinical features of lower limb DVT are pain, warmth, tenderness and swelling of the lower limb. Homan's sign may be positive. However, identical features can be present in other conditions such as ruptured Baker's cyst and cellulitis. PE usually originates from clots in larger veins of thighs and pelvis. Smaller thrombi causing submassive PE result in transient symptoms. Similar to DVT, the clinical diagnosis is unreliable as the clinical features such as chest pain, cough, hemoptysis, tachycardia and breathlessness are non-specific and can be present in many chest disorders. Bigger thrombi may lead to cardio-vascular decompensation. Unfortunately, the first presentation of a massive PE may be sudden death.

While venography is the "gold standard" in the diagnosis of DVT, it is invasive and associated with significant morbidities such as pain, allergic reaction to the contrast material, infections and thrombosis induced by the procedure itself. The duplex ultrasound has more than 90% sensitivity and specificity for patients with proximal, symptomatic DVT and about 70% for calf DVT. In cases of negative duplex studies where the clinical suspicion of venous thrombosis is high, a repeat duplex scan or venography should be done.

Pulmonary angiography remains the gold standard for establishing the diagnosis of pulmonary embolism, but it is invasive and not suitable for ill patients. The V/Q lung scan is the most frequently requested test for PE, the results are reported terms of probabilities of the patient having

PE and there are considerable problems in interpreting intermediate probability lung scans. Since its introduction in the 1990s, the spiral computed tomographic angiography (CTA) has emerged as a useful tool for the diagnosis of PE as it is not invasive and accurate. Indeed, there are suggestions that the CTA should replace the V/Q lung scan as the first-line investigation for PE.

D-DIMER

The use of D-Dimer as an aid in excluding DVT or PE.

When fibrin monomers bond to form a thrombus, a factor XIII acts to bind their "D" domains. This bond is resistant to plasmin and thus this degradation fragment is the "D-Dimer". Elevated levels of D-Dimer indicate the presence of fibrin clots being lyzed by plasmin.

Studies have shown that a D-Dimer assay below 500 ng/ml had virtually no chance of having pulmonary embolism or clot. Rapid accurate facts have been developed. In situations where there is only a low or moderate suspicion of DVT/PE, a normal D-Dimer renal with rapid accurate tests would essentially rule out thrombosis.

The drawback is that D-Dimer is sensitive but not specific. Patients after surgery, with disseminated intravascular coagulation and trauma will have elevated D-Dimers.

IDENTIFICATION OF PATIENTS AT RISK FROM THROMBOSIS

Thrombophilia is a term describing inherited or acquired defects leading to a predisposition to venous or arterial thrmbosis. Other synonymous terms are hypercoagulable state or prothrombotic state. They imply changes in the blood that are conducive to thrombosis and thrombus formation occurs if there is an additional trigger factor such as stasis or oral contraceptives.

Clinical Risk Factors for Arterial Thrombosis

Risk factors for arterial thrombosis are essentially those related to the development of atherosclerosis associated with coronary arterial disease, strokes and peripheral vascular disease. Epidemiological studies such as the Framingham studies have identified many risk factors for coronary

Table 1 Common Primary Hypercoagulable States

Antithrombin III deficiency
Protein C deficiency
Protein S deficiency
Factor V Leidon
Homocysteinemia
Prothrombin G20210A

Table 2 Secondary Hypercoagulable States

Increasing age
Immobilization
Obesity
Previous history of deep vein thrombosis
Family history of deep vein thrombosis
Inherited thrombophilia
Cancer/Anti-cancer drugs
Major abdominal, pelvic and hip surgery
Trauma to lower limb, especially fractured neck of femur
Pregnancy
Oral contraceptives and hormone replacement therapy
Hematological diseases, e.g. myeloproliferative disorders, paroxysmal nocturnal hemoglobinuria, heparin-induced thrombopathy
Medical disorders, e.g. antiphospholipid syndrome, vasculitis, hyperlipidemias nephritic syndrome, obesity

arterial disease. Hypertension, hyperlipidemia, diabetes mellitus, polycythemia, cigarette smoking and genetic influences are the commonly present predisposing factors for arterial thrombosis. Hyperhomocysteinemia and antiphospholid syndrome are other established risk factors. In general, the more risk factors present and the greater the degree of abnormality of any factor, the greater the risk for arterial thrombosis.

Clinical Risk Factors for Venous Thrombosis

There are many clinical risk factors identified for the development of venous thrombosis. Thrombosis usually occurs if there is a combination of risk factors. An asymptomatic person with protein C deficiency may not develop thrombosis normally, but may develop a DVT after surgery. Inherited conditions predisposing to thrombosis are termed primary

thrombophilia while acquired causes are called secondary thrombophilia. While acquired risk factors are much more common in clinical practice, primary thrombophilia should be investigated in younger patients and in those with recurrent VTEs as well as those with a family history of VTE.

Inherited Abnormalities of Control Mechanisms

Deficiency or dysfunction of antithrombin, protein C, and protein S are important inherited thrombophilic abnormalities. These 3 conditions are inherited in an autosomal recessive manner; they have a worldwide distribution and were well-studied. These deficiencies are associated with familial venous thrombosis and pulmonary embolism, idiopathic or recurrent venous thromboembolism at an early age, and venous thrombosis in unusual locations (mesenteric, axillary, cerebral). Generally, they do not predispose to artery occlusion although there are reported exceptions.

Two thrombogenic genetic polymorphisms that affect factors V and II are common in European and related populations but are rare or absent elsewhere. In factor V Leiden, a point mutation (G1691A) found in 4–7% of Europeans substitutes Arg506 for Gln at the cleavage site for activated protein C. Thrombin generation is greater and lasts longer than usual because the abnormal factor V resists degradation by activated protein C (APC resistance). A prothrombin mutation (G20210A) affects 1–3% of European populations and causes a roughly 20% increase in the circulating level of plasma prothrombin. These polymorphisms are weak predispositions for venous thromboembolism when they act alone but gain importance when other risk factors are coinherited or superimposed (including treatment with estrogen-containing oral contraceptives). Their prevalence can reach 30–60% in Europeans with otherwise idiopathic or recurrent venous thrombosis.

Other Thrombophilias

Hyperhomocysteinemia — or elevated levels of homocysteine, is increasingly recognized as a risk factor for both venous and artery thrombosis. It can result from various genetic disorders that affects the transsulphuration or remethylation pathways of homocysteine metabolism, e.g. cystathione beta synthase deficiency, methylene-tetra-hydrofolate

reductase (MTHFR) deficiency and a thermolabile variant of MTHFR. Acquired causes of hyperhomocysteinemia include folate or vitamin B12 deficiency, renal failure, hypothyroidism and drugs such as methotrexate and phenytoin.

Antiphospholipid antibodies (anticardiolipin antibody and the "lupus anticoagulant") are associated with an increased risk of venous or artery occlusion in systemic lupus erythematosus, and the antiphospholipid antibody syndrome.

Case control studies suggest that several other blood coagulation changes are also associated with an increased risk of developing venous thromboembolism. These include sustained elevation of factors VIII or IX coagulant activity, factor XI antigen and the Thrombin Activable Fibrinolysis Inhibitor (TAFI). The clinical utility of these observations remains uncertain.

Thrombophilias and Adverse Pregnancy Outcomes

The association between adverse pregnancy outcomes and the presence of lupus and the lupus anticoagulant is well-recognized. There has also been much interest of late in observations that pregnancy loss (especially repeated pregnancy loss) and other adverse pregnancy outcomes may be associated with a high prevalence of inherited thrombophilia (including hyperhomocysteinemia, activated protein C resistance, factor V or prothrombin mutation, protein C or S, and antithrombin deficiency) and may be preventable with an anticoagulant prophylaxis. Studies to date have been relatively small and highly selected patient populations. Systematic reviews indicate it is too early to judge the contribution of thrombophilia to adverse outcomes, and that routine screening for thrombophilias in pregnancy is not indicated. Large recent case control studies from Scotland and Canada have found no evidence of an excess pregnancy risk attached to factor V, prothrombin or methyltetrahydrofolate reductase polymorphisms in mothers or their newborn.

When and What to Measure for Suspected Thrombophilia

The present recommendation is to restrict screening for thrombophilia to young people with unprovoked thrombosis or when there is vein thrombosis in unusual territories (mesenteric or cerebral). Patients with recurrent venous thromboembolism and those with a family history should

also be screened for thrombophilia. It is reasonable to measure antithrombin, protein C and protein S in familial thrombosis (and also test for factor V Leiden and prothrombin polymorphism in Caucasians). Testing for lupus anticoagulant and anticardiolipin antibody is useful when suspecting acquired thrombophilia. Measuring the homocysteine level is relevant to both. Other tests should be restricted to clinical research.

THROMBOSIS IN ASIA

The true incidence of VTE in the Asian populations is unknown. In Caucasian populations, venous thromboembolism is an established entity with an annual community prevalence of 1–2 per thousand in Europe and North America. By contrast, it is widely thought that venous thromboembolism (VTE) is rare in Asia and that VTE is mostly a "Caucasian" problem. Differences may be due to lifestyle and genetic predisposition. This view is bolstered by the rarity of some prothrombotic clotting factor polymorphisms (factor V Leiden and prothrombin G20210A) in Chinese and other Asian ethnic populations. As a result, the study of VTE in Asian populations has attracted limited attention. There is, however, a growing body of evidence to show that VTE is not uncommon in this region after all. These evidences are from autopsy studies, audits of admission diagnoses to major referral hospitals in large cities, and prospective screening for subclinical deep vein thrombosis in "high risk" settings like major joint surgery or stroke.

Autopsy Studies

There is a wide discrepancy in the incidence of fatal PEs based on autopsy findings. The rate of fatal PE in studies involving Caucasian populations reported in-hospital deaths due to PE ranged widely from 4–13%. Pulmonary embolism was unusual and fatal embolism was rare among adult coroner's autopsies done in Singapore between 1952 and 1966 (1.6 per 1000 necropsies), but prevalence had increased to 1.05% by 1989–93. Autopsy reports from Queen Mary Hospital in Hong Kong that span 30 years showed a rising prevalence of embolism from 0.75% to 4.7%, thereby approaching the levels reported in Europe and North America. The difference in the type of patients subjected to autopsies may account for the different rates of fatal PE. Autopsy studies in hospital patients in Caucasian populations are much commoner than in Asian populations

due to cultural and religious differences; autopsies performed in this region are mainly for coroner's cases and those with unnatural causes of death.

Clinically Diagnosed Venous Thromboembolism in Reports from Asia

The true incidence of deep vein thrombosis in Asian populations remains uncertain, but data from recent reviews of hospital admissions in various countries around the region is suggestive evidence of a rising prevalence. A 1988 paper from Hong Kong reported 2.7 cases per 10 000 admissions to hospital that is similar to the rate of 2.8 per 10 000 hospital admissions cited in a 1990 paper from Kuala Lumpur. A 1992 report from Singapore indicated a prevalence of 7.9 per 10 000 admissions to the National University Hospital and a recent report from Singapore General Hospital observed a prevalence of 15.8 per 10 000 admissions for acute deep vein thrombosis between 1996 and 1997. This is much higher than what was previously reported. Almost 10% also had symptoms and confirmation of pulmonary embolism. This represents a minimum since it records only patients with objectively confirmed diagnoses. Even so, it remains considerably lower than the 0.9% of hospital admissions reported from Worcester, Massachussets. The causes of a rising prevalence venous thromboembolism in Asia may include an ageing population, changing lifestyles, increasing surgical and medical interventions for the diseases of older age and more intensive treatment for malignancy.

The clinical presentation of deep leg vein thrombosis in Asia appears to be similar to that described elsewhere — with leg pain and swelling the predominant complaints. The same is true for pulmonary embolism. A significant minority of patients with thrombosis (roughly 25%) has no apparent predisposition. The clinical features of venous thromboembolism are elusive as 75% of patients who present with VTE are asymptomatic and the disease often unsuspected. Objective testing is a necessity as clinical diagnosis alone is inaccurate and fatality can occur in untreated patients. In postmortem studies, 60–70% of patients who died from fatal PE were misdiagnosed and not suspected to have PE while they were still alive. Its first presentation can be fatal in individuals who are otherwise healthy. There is no data to show any racial differences in clinical features and presentations of VTE. The major risk factors for developing DVT in our patients are that of increasing age, immobilization, surgery and

malignancy, which are also the major risk factors cited in Western reports for Caucasian populations. There is a suspicion that genetic predisposing factors may differ in Asia and the West. Recent studies have shown that factor V Leidon mutation and mutation at position 20210 in the prothrombin gene are extremely rare in Chinese although common in the Caucasian patients with thrombosis. There also appears to be a higher prevalence of inherited antithrombin, protein C and protein S deficiency among Chinese patients presenting in Hong Kong or Taiwan with venous thromboembolism.

There are some data to address the issue of post-surgical VTE and the impression from clinical surveys indicates that thrombosis is not truly infrequent in Asia. Several surveys of symptomatic and confirmed venous thromboembolism suggest that patients who would be considered to be at "high risk" in the West are also at risk in Asia. Almost 2:1000 pregnancies and deliveries in Chinese women from Hong Kong were complicated by venous thrombosis or pulmonary embolism, mostly postpartum and after cesarean section. Studies done in this region between 1974 and 1993 on VTE after general surgery reported VTE rates in the absence of prophylaxis ranging from 3% to 19.1% that overlapped with the range of 6–35% reported for Caucasian populations. In Asian patients undergoing orthopedic surgery, there are huge variations in the reported DVT rates between different Asian centers ranging from 0% in Thailand, to 64.3% in Malaysia for DVT post total hip replacement. Hip replacement in Korea and Japan was followed by thrombosis rates of 15–25% that are high but somewhat lower than the rates of 50–60% reported from Europe and North America. Regarding DVT developing post total knee replacement, the reported rates ranged from 11.3% in Korea, 14% in Singapore and 76.5% in Malaysia. Thrombosis rates observed in Asia after abdominal or gynecological surgery, and after ischemic stroke, tend to be lower than those observed in the West. Reports vary and comparisons with Western studies are only approximate, as the study design and interpretation of diagnostic screening test results affect the outcome of these comparisons.

There is a long-standing belief that venous thromboembolism is extremely rare in Asia and clinically insignificant. However, recent surveys suggest that this may now not be true and that in Asia the incidence of venous thrombosis and pulmonary embolism in some "high risk" clinical settings is increasing and may approach that seen in the West. If this is so, then there is a need in Asia for an increased diagnostic

sensitivity to venous thrombosis and pulmonary embolism and to consider formal studies of risk factors and prevention in Asian populations.

ANTITHROMBOTIC TREATMENT

Treatment and Prevention of Arterial Thrombosis

The mainstay in prevention of arterial disease is to reduce the number of predisposing risk factors. Treatment of established arterial thrombosis includes the use of antiplatelet drugs, thrombolytic therapy and bypass surgeries. These would have been discussed in the appropriate sections of this book.

Primary Prophylaxis of Venous Thrombosis

Venous thrombosis with its potentially fatal complication pulmonary embolism is preventable. In Caucasian populations, venous thromboembolism is well-studied resulting in many established guidelines for its prevention and treatment. The role of primary prophylaxis in preventing its occurrence of venous thromboembolism (VTE) is well-established, particularly in the post-surgical setting. Patients are stratified into different risk levels of developing VTE. Early mobilization and mechanical preventive measures such as the compression stockings are recommended. For

Table 3 Risk Stratification in Caucasian Patients Undergoing Surgery

Low Risk Group: Less than 10% risk of DVT
• Absence of thrombophilia and diseases associated with hypercoagulable states
• Absence of other risk factors
• Minor surgery, trauma or medical illness at any age
• Major surgery age under 40 years
Moderate Risk Group: 10–40% risk of DVT
• Absence of thrombophilia and diseases associated with hypercoagulable states
• All major surgery with presence of other risk factors
• Minor surgery in patients with previous history of DVT
High Risk Group: 40–80% risk of PE/DVT
• Major surgery in patients with single or recurrent previous DVT
• Family history of DVT/PE
• Presence of thrombophilia and diseases associated with hypercoagulable states
• Fractures involving pelvis, hip, lower limbs
• Major surgery involving pelvis, hip, lower limbs
• Major cancer surgery

patients with moderate and high risks of developing VTE, pharmacological agents such as standard heparin and low molecular weight heparin (LMWH) are used in addition to mechanical measures like the compression stockings and pneumatic leg pumps. It is suggested that longer duration of prophylaxis that extends beyond hospital stay further reduce thrombotic rates for very high-risk patients. It is also recommended that non-surgical patients with medical conditions prone to VTE on prolonged bedrest be given VTE prophylaxis.

In Asia, we do not have enough data for such risk stratification in our patients and thrombo-prophylaxis is not a routine practice. Nonetheless, recent studies on VTE prophylaxis done in Asia have shown that DVT can be easily preventable by prophylactic LMWH without increase in bleeding complications. In Singapore the use of prophylactic LMWH completely abolishes the occurrence of DVT for patients undergoing total knee replacements and colorectal surgery. Surveys from Asia therefore suggest that routine thrombosis prophylaxis is probably indicated after major joint surgery and certainly required for patients with underlying thrombophilia in the perioperative periods. Further studies in Asia are needed to define the need for prophylaxis in other medical and surgical settings. In addition, since average body weight tends to differ between Western and many Asian populations, there is a need for Asian studies about the safety of anticoagulant dosing regimens.

Treatment of Established Venous Thrombosis

Unless contraindicated by bleeding complications, anticoagulation is indicated in established venous thrombosis. The aim of anticoagulant treatment is to prevent further thrombosis and pulmonary embolism while resolution of the thrombi occurs by the natural fibrinolytic activity. Immediate anticoagulation is indisputable for proximal lower limb DVT with or without PE, and for DVT in unusual sites, as the risk of recurrence up to 25% is seen in these patients with inadequate anticoagulation. For lower limb DVTs confined to the calf, propagation and recurrence occur in about 30% of patients if left untreated, and pulmonary embolism can then occur. There is also a 12% recurrence rate within a year for calf DVTs that are inadequately treated. Generally, anticoagulation is recommended for calf DVTs. Untreated patients should be closely monitored and treatment given if there is clot propagation or recurrence.

Anticoagulation is initiated by either unfractionated heparin or low molecular weight heparin (LMWH) as these drugs produce an immediate anticoagulation effect. The dosage of unfractionated heparin is titrated to maintain a partial thromboplastin time twice normal. Low molecular weight heparin has been shown to be as efficacious and safe as unfractionated heparin in the treatment of venous thromboembolism. Its longer half-life and predictable bioavailability makes it easier and more convenient to use. The dosage is calculated according to body weight and no monitoring is required in most instances. It can be self-administered by subcutaneous injections once to twice daily, making home treatment possible. Compliant patients without coexisting hemorrhagic risk factors need not be hospitalized for the treatment of venous thrombosis. Warfarin therapy can be initiated at the same time of heparin. Usually, it takes 3–4 days for warfarin to produce the desired anticoagulant effect that is measured by the International Normalized Ratio (INR). Heparin injections are discontinued once the INR reaches therapeutic level.

The role of thrombolytic therapy in deep vein thrombosis is not established although it can be used in very extensive venous thrombosis compromising circulation. It is used more often for patients with massive pulmonary embolism. Patients need to be monitored closely for bleeding complications. Thrombolytic therapy should be followed by anticoagulation with heparin and subsequently warfarin.

The role of surgical embolectomy is reserved mainly for massive pulmonary embolism. Very rarely, the procedure is performed for patients with extensive venous thrombosis that threatens the arterial blood supply to the limb.

Some patients with extensive venous thrombosis and high chances of pulmonary embolism are unable to receive anticoagulation treatment because of high risk of bleeding complications. There are many such examples such as bleeding peptic ulcer, bleeding from active ulcerative colitis, postpartum hemorrhage, recent neurosurgery and recent extensive surgery. Inferior vena cava filter can be considered for such patients; its main role is to prevent pulmonary embolism but these patients should receive anticoagulation once their bleeding risks are resolved.

Secondary Prophylaxis of Venous Thromboembolism

Initial treatment of VTE with either unfractionated heparin or LMWH to be followed by oral anticoagulation is well-defined. After the initial

episode of venous thrombosis, the annual cumulative incidence of recurrence is 5–8%, with an overall recurrence rate up to 30% of patients within 8 years of the first acute event. Secondary prophylaxis is required to prevent its recurrence.

Warfarin is the anticoagulant most frequently used in secondary prophylaxis as it is inexpensive and taken orally. For secondary prevention of VTE, the INR is maintained at 2–3 in most cases and frequent blood tests and monitoring are necessary. Unfractionated heparin is an alternative agent but it requires frequent injections and its long-term use is associated with significant osteoporosis. LMWH can be used as an alternative drug in secondary prophylaxis. It is self-administered, has a constant and adequate anticoagulation profile, does not cross the placenta (with no adverse report of its use in pregnancy) and not associated with significant osteoporosis in long-term use. There are some data to support the use of LMWH in secondary prophylaxis of VTE in situations where warfarin is not suitable. LMWH can be considered during pregnancy, for patients with recurrent venous thrombosis despite therapeutic INR, for the short duration of secondary prophylaxis and for patients unable to attend outpatient clinic regularly for monitoring of INR. However the routine use of prolonged therapy with LMWH is expensive.

Table 4 Recommendations of Duration of Anticoagulation after the First Thromboembolic Event based on Risk Stratification

Low Risk Group
- One event with a known temporary risk factor
- Calf DVT — 4 to 6 weeks anticoagulation after risk factors cease to be present
- Proximal DVT — 12 weeks anticoagulation after risk factors cease to be present

Moderate Risk
- One event
- Provocative stimulus — medical rather than surgical risk factors
- No known provocative risk factor
- Calf DVT — anticoagulation for 12 weeks after acute thrombotic event
- Proximal DVT — anticoagulation for 24 weeks after acute thrombotic event

High Risk — Indefinite Anticoagulation
- Two or more spontaneous events
- One spontaneous life-threatening event, e.g. near-fatal PE cerebral, mesenteric, portal venous thrombosis
- Antithrombin, Protein C, Protein S deficiencies
- Combined genetic abnormalities
- Antiphospholipid syndrome
- Uncontrolled Cancer

It is important to balance the benefit of anticoagulation in preventing recurrences against the risks of significant bleeding. In long-term anticoagulation, reported annual rates of hemorrhagic complications range from 7.6–16.5% (0.25–0.64% fatal, 1.1–2.7% major and 6.2–13.8% minor). The duration of anticoagulation should be limited to shortest period necessary to obtain the desired effects, as the risk of VTE must be balanced against the risk of bleeding. The optimum duration of secondary prophylaxis is still a matter of debate. Current recommendations based on available data are made according to the thrombotic risk stratification of patients. More comprehensive studies on VTE risk assessment and stratification are required to further define the optimal intensity and duration of anticoagulation therapy following an initial episode of VTE.

CONTRAINDICATIONS TO ANTICOAGULATION

Contraindications to the use of anticoagulation are seldom absolute and in most times they are relative. The bleeding risks should not be greater than the benefits derived from anticoagulation in thrombotic disorders. Anticoagulation should be avoided for patients who are actively bleeding or had recent life-threatening bleeds, patients with recent neurosurgery and eye surgery, those with severe uncontrolled hypertension and those with severe underlying hemostatic defects.

NEW ANTITHROMBOTIC DRUGS

The observations on thrombosis rates make it relevant to consider the mechanisms of action and clinical role of old and newer antithrombotic drugs used for thrombosis prevention and treatment. Two major groups of anticoagulants are now in general use: One includes heparin and the low molecular weight heparins — both given by injection; the other consists of the oral vitamin K antagonists (warfarin and its analogs). Newer anticoagulants have earned a limited clinical niche but may become more widely accepted. These include a synthetic heparin pentasaccharide with anti-factor Xa but no antithrombin activity (fondaparinux), several direct and specific antithrombins (hirudin and its analogs, melagatran, ximelagatran, argatroban), and a factor VIIa inhibitor (NAPc-2). In addition, there is a range of novel antiplatelet drugs. Newer antiplatelet drugs that have undergone extensive clinical evaluation include inhibitors of the

platelet surface ADP receptor (ticlopidine and clopidogrel), and a series of glycoprotein IIb/IIIa inhibitors — starting with the humanized chimeric monoclonal antibody Absciximab and now includes eptifibatide and tirofiban. These have value in peripheral vascular disease (clopidogrel), in acute coronary syndromes, and after coronary artery interventions (glycoprotein IIb/IIIa inhibitors).

FURTHER READING

1. Marder VJ, Matei DE, Hereditary and acquired thrombophilic syndromes, in: Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN (eds.), *Hemastasis and Thrombosis. Basic Principles and Clinical Practice*, Philadelphia, Lippincott Williams & Wilkins, pp. 1243–1275, 2001.
2. Hirsh J, Colman RW, Marder VJ, George JN, Clowes AW, Overview of thrombosis and its treatment, in: Colman RW, Hirsh J, Marder VJ, Clowes AW, George JN (eds.), *Hemostasis and Thrombosis. Basic Principles and Clinical Practice*, Philadelphia, Lippincott Williams & Wilkins, pp. 1071–1084, 2001.
3. Reiner AP, Siscovick DS, Rosendaal FR, Hemostatic risk factors and arterial thrombotic disease, *Thrombos Haemostas* **85**:584–595, 2001.
4. Geerts WH, Heit JA, Clagett GP, *et al*., Sixth ACCP Consensus Conference on Antithrombotic Therapy. Prevention of venous thromboembolism, *Chest* **119** (Suppl): 132S–175S, 2001.
5. Hyers TM, Hull RD, Morris TA, Tapson V, Antithrombotic Therapy for Venous Thromboembolic Disease, *Chest* **119** (Suppl):176S–193S, 2001.
6. Lee LH, Clinical Update on Deep Vein Thrombosis in Singapore, *Ann Acad Med Singapore* **31**:248–252, 2002.
7. Gallus AS, Lee LH, Coghlan DW, New Aspects of the Blood Coagulation Cascade, Anticoagulants and Vein Thrombosis in Asia, *Ann Acad Med Singapore*, 2002 (In Press).

This page intentionally left blank

39

Immunotherapy in Hematology and Oncology

William Y.K. Hwang and Gerrard K.H. Teoh

Immunotherapy is the modulation, augmentation or supplementation of the immune system for the treatment of various diseases. It is a Science that has its roots and fruits in the treatment of cancer and hematological maladies.

Many ailments are the result of inefficiency or aberrancy in the mounting of a concerted immune response against abnormal cells or proteins that arise as the body tries to heal itself in its most natural way. In the attempt to generate useful immunological responses against disease, various strategies have been devised. Over a century ago, Dr William Coley gave injections of crude bacterial extracts to patients with inoperable sarcomas, in the hope that natural immune responses will be augmented and directed towards eradication of the cancer. While most patients succumbed to their disease, some patients showed remarkable responses.[1] Since then, a plethora of strategies have been devised to stimulate the immune responses of patients against tumor cells. Hence, in a broad sense, immunotherapy can be thought of as "natural assisted self-healing".

From the ocean of immunological maneuvers to combat disease, this chapter will distil 3 main approaches in hematology and

oncology — namely cytokines, monoclonal antibodies and cell-based anti-cancer vaccines.

CYTOKINES

Cytokines are biological proteins that act as messengers. They operate in an intricate interacting meshwork to cause activation, apoptosis (programmed cell death), or proliferation of the target cells. Some cytokines have an anti-tumor effect; these include the interferons (IFNs), tumor necrosis factor (TNF), and the interleukins (ILs). Other cytokines, including granulocyte macrophage-colony stimulating factor (GM-CSF) and IL-3 may help activate dendritic cells (DC) for cancer vaccination. Yet other cells play a supporting role — for example granulocyte-colony stimulating factor (G-CSF), which has been used to accelerate the recovery of leukocyte counts following chemotherapy.

Interferons

In 1957, IFNs were identified and found to be produced by cells infected with virus to prevent other cells from getting infected.[2] There are 5 main varieties of IFNs: alpha (α), beta (β), gamma (γ), omega (ω), and tau (τ). In general, the IFNs have anti-proliferative and immunomodulatory effects on cells. Interferon-β was found to be useful in multiple sclerosis and IFN-γ in chronic granulomatous disease. Interferons- ω and τ are still being studied for their effect on cells. Among all IFNs, IFN-α has demonstrated the greatest utility in hematology, and is particularly effective in inducing an anti-proliferative effect on tumor cells of hematopoietic origin. In fact, IFN-α has been approved in the treatment of a wide variety of ailments including hairy cell leukemia, chronic myeloid leukemia (CML), non-Hodgkin's lymphoma, multiple myeloma (MM), AIDS-related Kaposi's sarcoma, high-risk melanoma, chronic hepatitis B/C and condylomata acuminata.

Interferon-α is effective, when given alone or in combination with other agents, in controlling the progression of these diseases. Its utility in hematological malignancies has been discussed in other chapters in this book and we will not be redundant except to emphasize that fever, chills, fatigue, depression and insomnia are common toxicities of IFN, and these usually improve with time. It is important to prepare the patient for these side-effects and to pre-medicate (e.g. with paracetamol) where necessary.

Moreover, current opinion suggests that it is better to start at lower doses (e.g. 3 million units thrice weekly) of IFNs before subsequently escalating to higher doses.

Interleukins

The ILs are natural proteins in the human immune system, which are produced by hematopoietic cells as well as endothelial cells, fibroblasts, astrocytes and thymocytes. Their primary function is the immunomodulation and immunoregulation (including growth and cytokine secretion) of leukocytes (Table 1).

Table 1 Functions and Clinical Uses of ILs[3]

Interleukin	Function	Clinical Utility
IL-1	Immunomodulation Promotion of hematopoiesis Mediation of inflammation Mediation of disease	Accelerating hematopoietic recovery Tumor responses uncommon
IL-2	Activation of cytotoxic T cells, NK cells, and monocytes Production of lymphocyte-activated killer (LAK) cells B-cell growth and antibody production	Treatment of renal cell cancer, melanoma, relapsed/refractory acute myeloid leukemia, non-Hodgkin's lymphoma, etc.
IL-4	B-cell growth and differentiation factor	Various cancers (though effect so far is less marked than with IL-2)
IL-6	Megakaryocyte maturation, B-cell differentiation, T-cell growth and maturation, osteoclast development, etc.	Modest effects on stimulating platelet recovery and on cancer
IL-10	Inhibition of cytokine synthesis Proliferation of activated T cells	Effect on down-regulating cytokines being investigated for autoimmune and infectious disease
IL-12	Activates NK and T cell cytotoxicity Enhances survival and proliferation of stem cells Enhances IFN-γ production Promotes Th1 production (at expense of Th2)	Experimental role in renal cell cancer and melanoma

When administering the ILs (most commonly IL-2) it is important to be on the lookout for complications of therapy. Interleukin-2, for example, may cause skin desquamation (sometimes precipitated by sunlight exposure), sepsis secondary to gut bacterial translocation, and a fatal capillary leak syndrome.

Tumor Necrosis Factor

In the 1970s it was discovered that the active ingredients in Dr Coley's extracts were bacterial endotoxins,[4] and that leukocytes produce a substance called TNF in response to the endotoxins — substances which seemed to be responsible for the clinical effects observed by Dr Coley.[5] The TNFs are thus named because of their ability to cause necrosis of tumor tissue. Of the two forms, TNF-α and TNF-β, TNF-α has been more extensively studied. Activated macrophages, monocytes and lymphocytes produce TNF-α, which is an essential participant in endotoxin-induced tumor regression through the initiation of cell-mediated anti-tumor immunity. In laboratory studies, intratumoral injection of TNF has resulted in endotoxin-induced-hemorrhagic necrosis and regression of established sarcoma.[6] However, the intravenous administration of this drug could be associated with substantial toxicity and the most promising use of this substance is when given in combination with chemotherapy directly by isolated perfusion into the limbs or organs involved by tumor, in which case the major effect of TNF appears to be an increased permeability of the tumor vascular bed resulting in augmented accumulation of co-administered drug in the tumor.[7]

MONOCLONAL ANTIBODIES

History and Early Development

In 1894, Paul Ehrlich proposed that cells have receptors on their surface, which bind to foreign toxins. When these cells were repeatedly exposed to these toxins, their "receptors" would be shed into the circulation and bind with further toxins in the blood.[8] Though this concept was viewed with great scepticism initially, it later proved to be remarkably close to the body's humoral response to foreign Ags. The cells shedding "receptors" in response to foreign toxins are now known as B-lymphocytes, the toxins

now known as antigens (Ags), and the anti-toxin receptors now known as antibodies (Abs).

Consequently, there was great interest in the mid-20th century to determine the properties and functions of Abs. However, despite the resolution of the basic structure of the immunoglobulins (Igs) in 1962, we were still unable to produce antibodies of predetermined specificity. Jens Kohler and Cesar Milstein made an enormous breakthrough in 1975 by fusing spleen cells of mice immunized against specific antigens, with immortal murine plasmacytoma cells. In this way they were able to produce specific Abs (from the mouse spleen cells) in a continuous manner.[9] They were later awarded the Nobel Prize for this work. Since then thousands of Abs have been produced by this rapid and reliable method. These Abs initially found their utility in diagnostics. Tests like the enzyme-linked immunosorbent assays (ELISAs), immunoflourescence studies, and indirect immunofluorescence flow cytometric analysis, are dependent on antibodies for the detection of specific antigens or epitopes.

Antibodies in Therapy

Murine monoclonal antibodies (mAbs)

Monoclonal antibodies (mAbs) later found their way into clinical use. OKT3, a murine mAb widely used in organ transplantation to treat acute rejection,[10] has also been used in hematopoietic stem cell transplantation (HSCT) to treat graft-versus-host disease (GVHD). In fact, 58–90% of patients who had predominantly steroid refractory acute GVHD responded to the mitogenic anti-CD3 mAb, OKT3. The non-mitogenic anti-CD3 mAb, BC3, which induced apoptosis of activated T cells without stimulating resting T cells to proliferate, is possibly even more effective.

Anti-idiotype Abs: an initial foray into cancer therapy

As the dream of these scientists had always been to be able to cure cancer, Abs were developed that could specifically target tumor cells. Initially, anti-idiotype Abs were made against the unique antigen-binding sites of Igs on the surface of lymphoma cells.[11] These Abs proved reasonably effective, but the effect was not always sustained as the recipient often developed human-anti-mouse-Abs (HAMA) leading to rapid clearance of the mouse protein.

Chimeric antibodies: fusion for a new age

The HAMA response remained a problem until it became possible to make chimeric* Abs by substituting the constant (Fc) regions of the Abs with human sequences. This way, mouse Abs could be dressed up to look more human and evade clearance by the host immune system. In addition, the Fc regions could also be chosen to engineer the best-desired effect for the Ab. For example, the IgG1 Fc region may be chosen when Ab dependent cellular cytotoxicity (ADCC) is desired and IgG2 may be chosen when this effect is not desired. Where the Fc regions are not wanted, Fab fragments (Abs without the Ig "tails") may be produced. Using this technique, Abciximab (ReoPro), a chimeric Fab which inhibits platelet aggregation by binding to the glycoprotein receptor of human platelets, became the first chimeric Ab to win FDA approval for use in preventing cardiac ischemic complications post-percutaneous coronary angioplasty.

But could chimeric Abs be useful for patients with cancer? As the cluster differentiation Ag 20 (CD20) is found on the surface of most B-lymphoma cells, an anti-CD20 murine variable region was fused to a human IgG1 Fc region to direct ADCC against the lymphoma cells. This chimeric antibody, Rituximab, obtained FDA approval in 1997 (see Table 2).

Table 2 Results of Clinical Trials of Rtiuximab for Non-Hodgkin's Lymphoma

Relapsed follicular NHL (Pivotal trial)	42% CR	6% PR
Initial therapy (Solal-Celigny 99)	41% CR	69% RR
(Hainsworth 00)	5% CR	54% RR
R-CHOP in indolent NHL (Czucman, Blood 99)	58% CR	100% RR
R-CHOP in aggressive NHL (Vose, Blood supp 99)	61% CR	97% RR
Elderly NHL		
CHOP	63% CR	57% 2 yr surv
R-CHOP	76% CR	70% 2 yr surv

CR = complete remission; PR = partial remission; surv = survival; R-CHOP = Rituximab + CHOP chemotherapy (comprising cyclophosphamide, adriamycin, vincristine and prednisolone).

*Chimera was a mythological creature with the head of a lion, the body of a goat and the tail of a snake. The term is currently used to describe a living being or a substance where there are separate entities forming the whole.

Humanized antibodies: improving efficacy

Despite this success, human anti-chimeric Ab (HACA) responses could still be generated, resulting in diminished clinical efficacy as the recipient's immune system could mop up even the chimeric Abs. As such, Greg Winter and his group produced the first humanized Ab where only the complementarity determining regions (CDRs) of the immunoglobulins still retain their murine sequences (Fig. 1).[12] They later also produced the first such Ab for clinical use, Campath-1H, which targets the CD52 Ag (expressed abundantly on all human lymphocytes).[13] This antibody proved useful in preventing GVHD as well as in treating lymphoproliferative disorders like lymphoma and chronic lymphocytic leukemia.[14] Another example of a humanized Ab is Daclizumab, which targets the IL-2 receptor. When used for the treatment of acute GVHD, up to 47% of patients achieved complete responses.[15]

"Antibody-plus": tagged antibodies for increased efficacy

Abs have also been used to target chemotherapeutic or radioactive agents to tumor cells. An example of a chemotherapeutic agent linked to an antibody is Mylotarg (Gemtuzumab ozagomicin), which is an anti-CD33 antibody linked to calicheamicin. An example of an Ab in which a radioactive agent is linked to an antibody is Bexxar, which is iodine-131 labeled

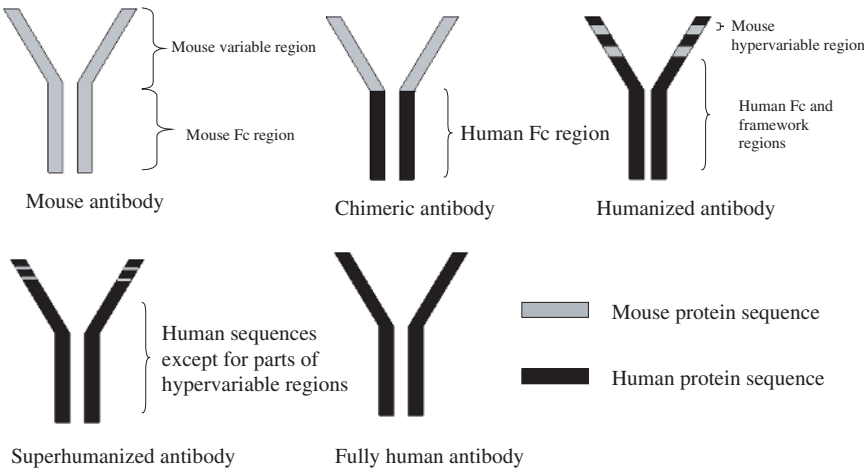


Fig. 1 Of mice and men: An evolution of monoclonal antibodies.

Tositumomab. Besides using the iodine-131 radioisotope (which requires treatment in radiation isolation because of the high gamma energy component) to radiolabel Abs, yttrium-90 (which has no gamma component and hence no need for radiation isolation may also be used. Abs have also been labeled with high energy, short path length alpha emitters like bismuth-213.

Table 3 List of some Major FDA Approved Monoclonal Antibodies

Mab (year approved)/ Description	Indication	Clinical Results
Abciximab (1994)/ chimeric anti-GpIIb/ IIIa Fab	Adjunct to percutaneous transluminal coronary angioplasty or atherectomy (PCTA) for the prevention of acute cardiac ischemic complications in high-risk patients	Mortality, myocardial infarction and urgent revascularization were reduced from 12.8% to 8.3% at 30 days post PTCA
Alemtuzumab (2001)/ humanized anti-CD52	Treatment of patients with B-cell chronic lymphocytic leukemia (B-CLL) who have been treated with alkylating agents and who failed fludarabine therapy	21% to 33% overall responses in B-CLL. Reduction in post-HSCT GVHD and mortality
Basiliximab (1998)/ chimeric anti-CD25	Prophylaxis of acute organ rejection in renal transplantation when used with immunosuppression with cyclosporine and corticosteroids	Significant reduction in graft loss, acute rejection and death (38–42% versus 55–57% in controls)
Daclizumab (1997)/ humanized anti-CD25	Prophylaxis of acute organ rejection in patients receiving renal transplants, to be used as part of an immunosuppressive regimen that includes cyclosporine and corticosteroids	22% of patients who received daclizumab for renal transplants developed rejection compared to 35% of those who did not
Denileukin difitox (1999)/ anti-CD25 linked to diptheria toxin	Anti-CD25 antibody linked to a diptheria toxin for treatment of patients with persistent or recurrent	For patients with advanced CTCL 30% good partial responses lasting an average of

(*Continued*)

Table 3 Continued

Mab (year approved)/ Description	Indication	Clinical Results
	cutaneous T-cell lymphoma (CTCL) whose malignant cells express the CD25 component of the IL-2 receptor	4 months; and 10% complete responses lasting an average 9 months
Etanercept (1997)/ 2 soluble TNF receptors fused to human Fc	Reduction in signs and symptoms of moderately to severely active rheumatoid arthritis in patients who have had an inadequate response to one or more disease-modifying anti-rheumatic drugs (DMARDS)	With etanercept treatment 59% of patients versus 11% controls had a significant reduction in symptoms after 6 months of treatment
Ibritumomab Tiuxetan (2002)/ I-131 labeled mouse anti-CD20	Treatment of patients with relapsed or refractory low-grade, follicular, or transformed B-cell non-Hodgkin's lymphoma, including patients with Rituximab (Rituxan) refractory follicular non-Hodgkin's lymphoma	Patients no longer responding to chemotherapy or Rituximab had a 74% overall response to the drug
Infliximab (1998)/ chimeric anti-TNF-α	Treatment of moderately to severely active Crohn's in patients who have an inadequate response to conventional therapies; or enterocutaneous fistula(s). Expanded to include rheumatoid arthritis in 1999	Treatment reduced the number of draining fistulas that occur in some cases of Crohn's disease — a benefit that lasted five months at most
Palivizumab (1998)/ humanized anti-RSV	Prophylaxis of serious lower respiratory tract disease, caused by respiratory syncytial virus, in pediatric patients at high risk of RSV disease	Palivizumab is the second product licensed for RSV disease after RSV human immunoglobulin
Rituximab (1997) (see text)	Treatment of patients with relapsed or refractory	See text

(*Continued*)

Table 3 Continued

Mab (year approved)/ Description	Indication	Clinical Results
	low-grade or follicular, B-cell non-Hodgkin's lymphoma	
Trastuzumab (1998)/ humanized anti-Her2	For treatment of patients with chemoresistant metastatic breast cancer which overexpress the HER2. Trastuzumab in combination with paclitaxel is indicated for treatment of patients with metastatic breast cancer whose tumors overexpress HER2 protein.	When used alone tumor response rate was 14% with 3% complete responses. When added to chemotherapy it improved survival from 68% to 79%
Gemtuzumab Ozogamycin (2001)/ humanized anti-CD33 linked to calicheamycin	Treatment of patients 60 years or older with CD33 positive acute myeloid leukemia who are not considered candidates for cytotoxic chemotherapy	20% good remissions in this refractory group

Beyond humanized Abs

Despite their success, humanized Abs can still induce human-anti-humanized Ab responses (HAHA).[16,17] As such, techniques to develop antibodies of lower immunogeneicity have been developed. One is the use of transgenic mice that are able to produce human Igs when immunized with Ag; the other is the use of human germline sequences for constructing Abs ("superhumanized" Abs).[18] Both are beyond the scope of this chapter.

CELL-BASED ANTI-CANCER VACCINES

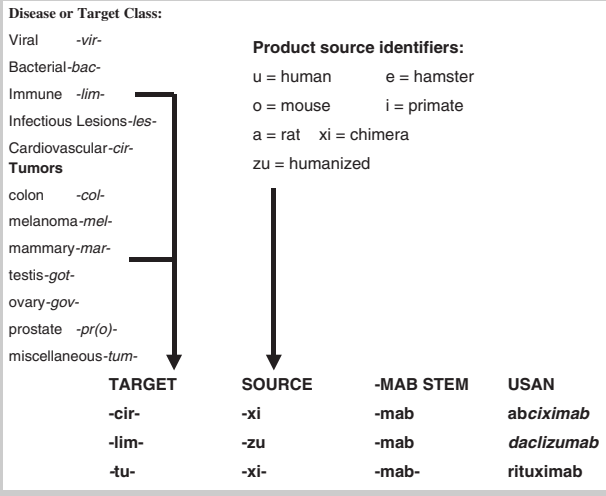
Overview of Cell-based Anti-cancer Vaccines

Cells in the immune system help to regulate health and disease. When an individual is immunosuppressed, his risk of developing lymphoma is increased 90-fold[19] and his risk of developing skin cancer increased 30-fold. Conversely, in tumors that demonstrate spontaneous regression, lymphocytes have frequently been found to infiltrate deep inside tumor tissue, suggesting that these lymphocytes could mediate tumor rejection.[20]

Naming antibodies

With the myriad forms of antibodies available and their increasing usage, it is important to understand their current nomenclature.

Essentially, the first two letters are free for the developers/owners of the antibody to designate. The next two to three letters define the target and the two letters after that define the type of monoclonal used (see figure).



When the antibody is radiolabeled or conjugated to another chemical such as a toxin a separate, second word or other acceptable chemical designation must be used. For example, if the antibody is conjugated to a toxin: "*-tox*" stem must be included (e.g. *zolimomab aritox … aritox* selected to identify ricin A-chain).

For *Gemtuzumab ozagomycin … mycin* is selected to denote conjugation to a mycin antibiotic.

For radiolabeled products, the word order is as follows: name of the isotope, element symbol, isotope number, and name of the monoclonal antibody. For example, *technetium Tc 99 m* biciromab and *indium In 111 altumomab pentetate*. A separate, distinct name assigned to any linker/chelator used to conjugate, or pegylate monoclonal antibodies, e.g. *indium In 111 satumomab pendetide* and *enlimomab pegol*.

A current "blunt" form of cellular therapy is HSCT, which has resulted in permanent remissions for many patients with cancer and which is discussed elsewhere in this book. In this section we will discuss cellular immunotherapy administered as cell-based anti-cancer vaccines (Fig. 2), which is not aimed at replacement of a defective immunity, but rather at repair and/or augmentation (i.e. reconstitution) of an appropriate host-resident anti-tumor immune response. This could be performed via manipulation of autologous tumor cell lysates, T cells, natural killer (NK) cells and/or DCs — leading hopefully to control of disease. The importance of using cells (rather than cytokines, Abs or other humoral elements) in this form of immunotherapy, is the potential of increasing precision and durability of the anti-cancer immune response. Therefore, the primary goal of cell-based anti-cancer vaccines is to reconstitute host tumor immunosurveillance with a precise, long-lived and natural anti-cancer immunity.

Tumor Ag based Cellular Immunotherapy

Tumor Ags are central elements in cellular immunity. Early, less successful attempts at using tumor Ags involved the reinfusion of autologous tumor cell preparations. Another strategy involved the infusion of tumor lysates with immunogeneic adjuvants like BCG and DETOX®. Despite occasional successes, these approaches were met with less than resounding victories.

The ability, in more recent years, to identify and target true tumor-specific Ags (TSAs) may result in effective immunotherapeutic regimens. The process of tumor recognition by $CD8^+$ T lymphocytes is complex and involves the presentation of tumor-associated Ags (TAA) as well as TSA by antigen presenting cells (APCs). Presentation of both TAAs and TSAs requires the loading of antigenic peptides onto the human lymphocyte Ag (HLA) molecules on APCs. Subsequently, engagement of TAAs and/or TSAs expressed on HLA molecules by $CD8^+$ T cells induces tumor-specific cytolytic function in responding cytolytic T lymphocytes (CTLs).

Tumor-associated Ags

Numerous TAAs have previously been described. These include carcinoembryonic Ag (CEA) in colon cancer, alpha-fetoprotein (AFP) in liver cancer, prostate-specific Ag (PSA) in prostate cancer, and melanoma-associated Ag (Melan-A) in malignant melanoma. Moreover, in malignant melanoma, there are at least 14 groups of melanoma-related TAAs that have been identified.[21]

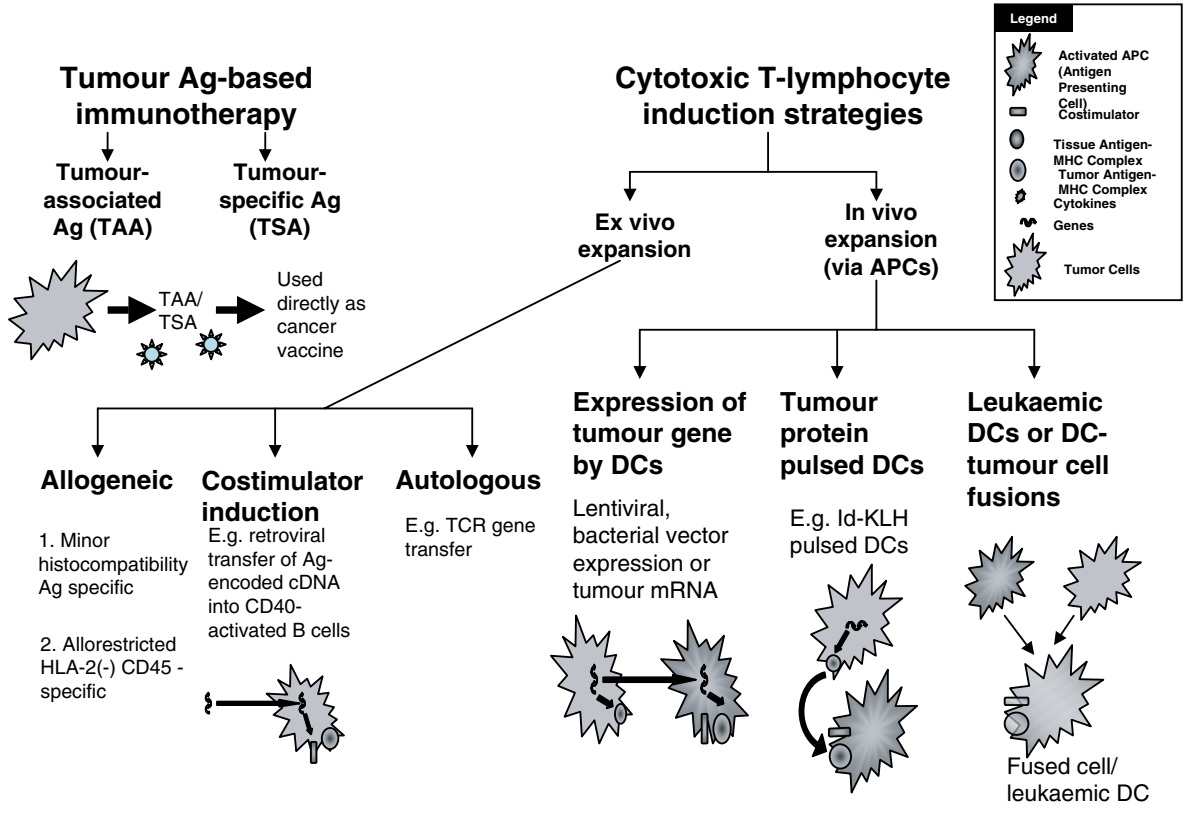


Fig. 2 Strategies in APC augmentation for cancer vaccination.

Tumor-specific Ags

In contrast to TAAs, TSAs should only be expressed in malignant cells and not in non-malignant cells. However, this does not occur in practice, except for anti-Id Abs in chronic lymphocytic leukemia,[22] and monoclonal Igs (M-protein), in MM;[23] which are both highly specific proteins secreted only by the malignant cells. When the definition of TSAs is more loosely applied, several other examples of TSAs can be considered. These include the bcr/Abl protein in Philadelphia chromosome-positive chronic myeloid leukemia (CML)[24] and the pml/RARa protein in t(15:17)-positive acute promyelocytic leukemia (APL).[25] The processes by which TSAs are produced can result in Ags that are either cancer-specific (i.e. present only in all cancers), or cancer-type-specific (i.e. present in only a certain type of cancer), or cancer-patient-specific (i.e. present in a particular patient with that particular cancer).

Importance of DCs

Dendritic cells are ideal anchor-points for the manipulation of the immune system as they are able to initiate and direct immune responses of T cells against tumor cells and various TAAs. Although B cells and macrophages can present Ags to T cells, the most potent APCs are DCs. Thus far, it has been found to be possible to grow large numbers of DCs *in vitro* from hematopoietic cells of the blood, marrow or cord blood. The cytokines GM-CSF, stem cell factor (SCF), flt-3 ligand, TNF-α and IL-4 have all been found to be useful cytokines for growth of these cultures. It is also possible to get these DCs to present TAAs (Table 4). DCs may be loaded with Ags *in vitro* by incubating them with the tumor cells themselves or with cell lysates/eluates. Such an approach has been employed in lymphoma or MM (against the tumor Id protein), as well as advanced melanoma and prostate cancer (against PSA and prostatic acid phosphatase).

A more complex approach involves the introduction of DNA or RNA vectors to DCs. These will encode tumor Ags for presentation on the surface of the DCs. Yet another approach involves the fusion of the DCs to tumor cells. When this technique was applied to patients with metastatic renal cell cancer, 4 out of 12 patients achieved a stable disease state.[26] A recent intriguing approach involves the growth and differentiation of leukemia cells into DCs. These leukemic DCs have been shown to be

Table 4 Tumor-Associated Antigens

Solid Tumor Antigens	Viral Antigens	Oncogenic Proteins and Carbohydrates
MAGE 1,2,3	Associated with:	P53
MART	Hepatitis B virus	Ras
gp100	Epstein Barr virus	Her2
CEA	Human papilloma virus	
Her2		
Mucin		
PSA		
PAP		

MAGE = melanoma antigen gene; MART = melanoma antigen recognized by T cells; CEA = carcinoembryonic antigen; PSA = prostate specific antigen; PAP = prostatic acid phosphatase.

effective in presenting Ags to T cells; thus eliciting leukemia-specific responses, especially in acute and chronic myeloid leukemia.[27] Interestingly, immune cells may also be used to induce tolerance. NK cells and IL-10 driven CD4 T cells have been shown to dampen auto and alloreactivity.[28] This has implications and potential applications in the treatment of autoimmune disease and the establishment of graft tolerance.

Strategies for CTL Expansion

The generation of CTLs involves a multistage sequential process called lymphocyte priming. Primed CTLs produce granzyme and perforin, which mediate cytolysis of target cells, including tumor cells. Expansion of CTLs can be achieved either *ex vivo*, i.e. by removing peripheral blood mononuclear cells and then exposing them to Ag; or *in vivo*, by introducing suitably expressed Ags (via APCs) directly into the host. *Ex vivo* expansion of CTLs involves the priming of lymphocytes against tumor cells or tumor antigens *ex vivo* before infusion of the CTLs into the patient (Table 5). Similarly, *in vivo* expansion of CTLs can be performed by the infusion of APCs (mainly mature DCs bearing tumor Ags) into the patient, instead of infusion of CTLs. These APCs stimulate host anti-tumor immunity, leading to the *in vivo* expansion of host CTLs (Table 6). Due to the complexity and rapidly evolving nature of this subject, only a selection of strategies in both *in vivo* and *ex vivo* expansion of CTLs will be covered in this review.

Table 5 Strategies for *ex vivo* Expansion of CTLs for Adoptive Immunotherapy

Strategy	Target	Key Features
Allogenic CTL expansion:		
mHags-specific CTLs	HA-1 or HA-2	spares non-hematopoietic tissue
allorestricted HLA-2(−) CD45-specific CTLs	CD45-derived peptides	GVL > GVH
Allogenic and autologous CTL expansion:		
retroviral transfer of Ag-encoded cDNA into CD40-activated B cells	transduced antigenic peptides	any HLA haplotype
Autologous CTL expansion:		
TCR gene transfer	endogenous tumor peptides	breaks tumor tolerance preserves self-tolerance

Table 6 Selected Strategies for *in vivo* Expansion of CTLs for Adoptive Immunotherapy (via APCs)

Strategy	Target	Key Features
lentiviral or bacterial vector expression systems in DCs	tumor-derived RNA	simple and safe presentation
tumor mRNA-transduced DCs	tumor-derived RNA	polyclonal CTL response
anti-NHL Id-protein/KLH-immunity induced pulsed DCs	NHL Id protein	anti-Id
leukemic-DCs	leukemic Ags	leukemia cells function as APCs
DC-tumor cell fusions with or without IL-12	all tumor cell Ags	most effective vaccine

Ex vivo Expansion of CTLs for Adoptive Immunotherapy[29]

Method 1 — Minor histocompatibility antigen (mHAG)-specific CTLs[30]

The mHags HA-1 and HA-2 are exclusively expressed on hematopoietic cells, including leukemic cells and leukemic precursors, and induce HLA-A*0201-restricted CTLs *in vivo*. These *ex vivo* generated HA-1- and HA-2-specific CTLs lyse leukemia cells but not non-hematopoietic cells, and it is

possible to generate them in sufficiently large numbers for the purpose of adoptive immunotherapy. This strategy thus provides a possible novel therapy for patients with relapsed leukemia after HSCT.

Method 2 — Allorestricted CD45-specific CTL[31]

In an attempt to improve the GVL effect (but not worsen GVH effect), it is possible to generate allorestricted CTLs directed against human CD45 Ag. Such CTLs recognize patient hematopoietic cells including leukemic cells, enhanced donor cell engraftment and improved GVL effect. In contrast, these CTLs do not recognize host non-hematopoietic tissues or donor cells from the graft. Since CD45-derived peptides bind to HLA-A2 molecules, these peptides could be used to generate CTL lines from HLA-A2(−) donors by sequential stimulation with peptide-pulsed HLA-A2(+) stimulators, and screened for peptide-specific cytotoxicity. Specific allorestricted CTLs having potent activity against leukemic progenitors can thus be produced and adoptively transferred. Moreover, such a strategy may be useful in restoring the GVL effect after HLA-A2-mismatched haploidentical transplantation.

Method 3 — Retrovirally transduced CD40-stimulated B cells[32]

Expression of Ags by APCs can be enhanced via CD40 activation. Hence, CD40-activated B (CD40-B) cells transduced retrovirally with Ag-encoding complementary DNA (cDNA), can be made into potent APCs for the generation of Ag-specific CTLs *in vitro*. Moreover, the antigenic epitopes can be presented by any of the HLA Ags on the APC; i.e. non-HLA restriction. The advantages of using CD40-B cells are that these cells can be maintained in culture without loss of APC function; and be made to express co-stimulatory molecules at levels comparable to that of DCs. Hence, prior knowledge of peptide epitopes is not required, and CTL induction can be performed in patients with any HLA haplotype.

Method 4 — T cell receptor (TCR) gene transfer[33]

Autologous T cells are tolerant to both self and tumor Ags. The ability to break tolerance to tumor Ags whilst preserving self-tolerance potentially leads to tumor rejection. In this method, the specificity of T cells from a

existing donor CTL line is transferred to the patient by performing a TCR gene transfer. This approach obviates the need to generate allorestricted CTLs for each patient (see above), and circumvents self-tolerance of autologous T lymphocytes to tumor Ags. Moreover, this method would make it possible to generate CTLs for immunotherapy for a broad spectrum of malignant diseases.

Strategies for *in vivo* Expansion of CTLs for Adoptive Immunotherapy (via APCs)

Method 1 — Lentiviral or bacterial vector expression systems[34]

Transduction of DCs with an Aag-bearing recombinant viral vector has been shown to improve Ag processing and presentation. This can be performed using lentiviral vectors (lentivectors), which is safe, since no viral integration occurs. Prior studies have reported significant T-cell responses against tumors, or even cause tumor regression using the lentiviral system.

Method 2 — Tumor antigen RNA-transduced DCs[35]

DCs transfected with tumor antigen mRNA induce specific T cell responses in patients with cancer. For example, DCs transduced with prostate antigen (e.g. PSA) mRNA were consistently found to be more effective than single Ag-pulsed DCs in inducing tumor-specific (and polyclonal) CTL response against prostate cancer. The polyclonal CTLs response has been postulated to reduce clonal tumor escape, especially when tumor bulk is low.

Method 3 — Tumor-protein/keyhole limpet hemocyanin (KLH)-pulsed DCs[36]

In an example of this method, Id proteins from B-cell NHL, conjugated to a highly-immunogenic carrier protein, KLH, have been used to pulse DCs. The pulsed DCs, which are infused into patients, present the tumor Id protein to the patient's immune system. Patients receiving this vaccine demonstrated induction of anti-Id mAbs which bound to autologous tumor cells, and tumor regression.

Method 4 — Leukemic DCs

As previously discussed, leukemic cells may be converted by cytokines into leukemic DCs. It has been shown that leukemic DCs can be generated *ex vivo* from myelomonocytic precursors in CML and acute myeloid leukemia (AML) cells. These DCs have potent ability to stimulate leukemia-specific cytolytic activity in autologous lymphocytes and may be useful in immunotherapy of leukemia patients.[27]

Method 5 — DC-tumor cell fusion[37,38]

The fusion of DCs with tumor cells, followed by vaccination of fused cells has been shown to increase the repertoire of anti-tumor polyclonal CTLs in mouse models of MM, melanoma and lymphoma. Since IL-12 upregulates the expression of co-stimulatory molecules in DCs, stimulates $CD4^+$ Th1 reactivity, augments immunologic responses to tumor Ags, and expands Ag-specific $CD8^+$ T cells; addition of IL-12 to this vaccination strategy was associated with increased CTL activity, T-cell proliferation, eradication of established disease, and potentiation of anti-tumor activity. This is a particularly effective method of anti-cancer vaccination.

CONCLUSION

The development of immunological therapies in hematology and oncology have revolutionized the practice of medicine and given many patients hope in the face of despair. No longer can the student or practitioner of medicine afford to be ignorant of the latest innovations and nuances in the field of cytokines, mAbs and cell-based anti-cancer vaccines. While each approach has its advantages and disadvantages, it is likely that a synergy of the techniques will yield the best possible results. As we learn how to harness the power of specific therapy like this, we will be able to bring greater hope and better outcomes for our patients with hematological disorders. We hope this chapter has served to give a succinct overview of this complex subject.

ACKNOWLEDGMENTS

Special thanks to my mentors: Dr Jefferson Foote, Dr Brenda Sandmaier and Dr Philip Tan of the Fred Hutchinson Cancer Research Centre. The

contribution of Dr Ho Liam Pock to the initial representation of figure 2 is gratefully acknowledged.

REFERENCES

1. Coley W, The treatment of malignant tumors by repeated inoculations of erysipelas, with a report of ten original cases, *Am J Med Sci* **105**:487–511, 1893.
2. Isaacs A, Lindemann J, Virus Interference I. The Interferon, *Proc R Soc Ser* **B147**:258–267, 1957.
3. Gale DM, Sarokin P, Biotherapy: a comprehensive overview, in *The Interleukins*, Chapter 6, 2nd ed., pp. 195–244.
4. Hartwell J, Shear M, Adams J, Chemical treatment of tumors: Nature of the haemorrhage producing fraction from derratia marcescens (B prodiziosus) culture infiltrate, *J Natl Cancer I*, 1943.
5. Carswell E, Old L, Kassel R, *et al.,* An endotoxin-induced serum factor which causes necrosis of tumors, *Proc Natl Acad Sci USA* **72**:3666–3670, 1975.
6. North RJ, Havell EA, The antitumor function of tumor necrosis factor (TNF) II. Analysis of the role of endogenous TNF in endotoxin-induced hemorrhagic necrosis and regression of an established sarcoma, *J Exp Med* **167**(3):1086–1099, 1988.
7. Hagen TL, Eggermont AM, Manipulation of the tumor-associated vasculature to improve tumor therapy, *J Liposome Res* **12**(1–2):149–154, 2002.
8. Erlich P, On immunity with special reference to the cell of life, *Proc Roy Soc London* **66**:424–448, 1890.
9. Kohler M, Continuous cultures of fused cells secreting antibodies of predefined specificity, *Nature* **256**(5517):495–497, 1975.
10. Goldstein G, Monoclonal antibody specificity: Orthoclone OKT3 T-cell blocker, *Nephron* **46** (1):5–11, 1987.
11. Miller R, Maloney D, Warnke R, *et al.,* Treatment of B cell lymphoma with monoclonal anti-idotype antibody, *N Engl J Med* **306**:517–522, 1982.
12. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G, Replacing the complementarity-determining regions in a human antibody with those from a mouse, *Nature* **321**:522–525, 1986.
13. Riechmann L, Clark M, Waldmann H, Winter G, Reshaping human antibodies for therapy, *Nature* **332**:323–327, 1988.
14. Hale G, Dyer MJS, Clark MR, Phillips JM, Marcus R, Riechmann L, Winter G, Waldmann H, Remission induction in non-Hodgkin lymphoma with reshaped human monoclonal antibody CAMPATH-1H, *Lancet* **17**(2):1394–1399, 1988.
15. Przepiorka D, Kernan NA, Ippoliti C, *et al.,* Daclizumab, a humanized anti-interleukin-2 receptor alpha chain antibody, for treatment of acute graft-versus-host disease, *Blood*, **95**(1):83–89, 2000.

16. Isaacs JD, Watts RA, Hazleman BL, Hale G, Keogan MT, Cobbold SP, Waldmann H, Humanised monoclonal antibody therapy for rheumatoid arthritis, *Lancet* **340**: 748–752, 1992.

17. Hwang WYK, Teoh G, Foote J, An Analysis of Reported Human Anti-Humanized Antibody (HAHA), Human Anti-Chimeric Antibody (HACA) and Human Anti-Human Antibody (HAHA') Responses, *IBC's 14th Annual International Antibody Engineering Conference* (abstract), San Diego, 2003.

18. Tan P, Mitchell DA, Buss TN, Holmes MA, Anasetti C, Foote J, "Superhumanized" antibodies: reduction of immunogenic potential by complementarity-determining region grafting with human germline sequences: application to an anti-CD28, *J Immunol* **169**(2):1119–1125, 2002.

19. Challis GB, Stam HJ, The spontaneous regression of cancer. A review of cases from 1900 to 1987, *Acta Oncologica* **29**:545–550, 1990.

20. Ward RL, Hawkins NJ, Coomber D, *et al*., Antibody immunity to the HER-2/neu oncogenic protein in patients with colorectal cancer, *Human Immunology* **60**:510–515, 1999.

21. Lindauer M, Stanislawski T, Haussler A, Antunes E, Cellary A, Huber C, *et al*., The molecular basis of cancer immunotherapy by cytotoxic T lymphocytes, *J Mol Med* **76**:32–47, 1998.

22. Bogdan CA, Alexander AA, Gorny MK, Matute R, Marjanovic N, Zolla–Pazner S, *et al*., Chronic lymphocytic leukemia with prostate infiltration mediated by specific clonal membrane-bound IgM, *Cancer Res* **63**:2067–2071, 2003.

23. Kunkel HG, The "abnormality" of myeloma proteins, *Cancer Res* **28**:1351–1353, 1968.

24. Daley GQ, Van Etten RA, Baltimore D, Blast crisis in a murine model of chronic myelogenous leukemia, *Proc Natl Acad Sci USA* **88**:11335–11338, 1991.

25. Redner RL, Rush EA, Faas S, Rudert WA, Corey SJ, The t(5;17) variant of acute promyelocytic leukemia expresses a nucleophosmin-retinoic acid receptor fusion, *Blood* **87**:882–886, 1996.

26. Marten A, Renoth S, Heinicke T, Albers P, Pauli A, Mey U, Caspari R, Flieger D, Hanfland P, Von Ruecker A, Eis–Hubinger AM, Muller S, Schwaner I, Lohmann U, Heylmann G, Sauerbruch T, Schmidt–Wolf IG, Allogeneic dendritic cells fused with tumor cells: preclinical results and outcome of a clinical phase I/II trial in patients with metastatic renal cell carcinoma, *Hum Gene Ther* **14**(5):483–494, 2003.

27. D Choudhury BA, Liang JC, Thomas EK, Flores–Romo L, Xie QS, Agusala K, Sutaria S, Sinha I, Champlin RE, Claxton DF, Dendritic cells derived *in vitro* from acute myelogenous leukemia cells stimulate autologous, antileukemic T-cell responses, *Blood* **93**(3):780–786, 1999.

28. Takayama T, Tahara H, Thomson AW, Differential effects of myeloid dendritic cells retrovirally transduced to express mammalian or viral interleukin-10 on cytotoxic T lymphocyte and natural killer cell functions and resistance to tumor growth, *Transplantation* **71**(9):1334–1340, 2001.

29. LimTY, Gullo C, Hwang W, Teoh G, The Role of Cytolytic T Lymphocytes in the Eradication of Multiple Myeloma, SGH Proceedings, in press.

30. Warren EH, Greenberg PD, Riddell SR, Cytotoxic T-lymphocyte-defined human minor histocompatibility antigens with a restricted tissue distribution, *Blood* **91**: 2197–2207, 1998.

31. Amrolia PJ, Reid SD, Gao L, Schultheis B, Dotti G, Brenner MK, *et al*., Allorestricted cytotoxic T cells specific for human CD45 show potent antileukemic activity, *Blood* **101**:1007–1014, 2003.

32. Kondo E, Topp MS, Kiem HP, Obata Y, Morishima Y, Kuzushima K, *et al*., Efficient generation of antigen-specific cytotoxic T cells using retrovirally transduced CD40-activated B cells, *J Immunol* **169**:2164–2171, 2002.

33. Giralt S, Bensinger W, Goodman M, Podoloff D, Eary J, Wendt R, *et al*., 166Ho-DOTMP plus melphalan followed by peripheral blood stem cell transplantation in patients with multiple myeloma: results of two phase 1/2 trials, *Blood* **102**:2684–2691, 2003.

34. Esslinger C, Chapatte L, Finke D, Miconnet I, Guillaume P, Levy F, *et al*., *In vivo* administration of a lentiviral vaccine targets DCs and induces efficient CD8(+) T cell responses, *J Clin Invest* **111**:1673–1681, 2003.

35. Heiser A, Dahm P, Yancey DR, Maurice MA, Boczkowski D, Nair SK, *et al*., Human dendritic cells transfected with RNA encoding prostate-specific antigen stimulate prostate-specific CTL responses *in vitro*, *J Immunol* **164**:5508–5514, 2000.

36. Timmerman JM, Czerwinski DK, Davis TA, Hsu FJ, Benike C, Hao ZM, et al., Idiotype-pulsed dendritic cell vaccination for B-cell lymphoma: clinical and immune responses in 35 patients, *Blood* **99**:1517–1526, 2002.

37. Wang J, Saffold S, Cao X, Krauss J, Chen W, Eliciting T cell immunity against poorly immunogenic tumors by immunization with dendritic cell-tumor fusion vaccines, *J Immunol* **161**:5516–5524, 1998.

38. Gong J, Koido S, Chen D, Tanaka Y, Huang L, Avigan D, *et al*., Immunization against murine multiple myeloma with fusions of dendritic and plasmacytoma cells is potentiated by interleukin 12, *Blood* **99**:2512–2517, 2002.

40

Hematopoietic Stem Cell Transplantation

Patrick Tan

INTRODUCTION

In recent years, the term "hematopoietic stem cell transplantation" has been used to replace "bone marrow transplantation". It is a more precise term and emphasizes "hematopoietic stem cell" as the key element. The hematopoietic stem cells are the earliest, most immature group of cells that possess the greatest potential for self-renewal and long-term marrow repopulating capacity. These cells can be harvested from the bone marrow, peripheral blood, umbilical cord and even fetal liver. The sources of these stem cells may be a compatible family member or an unrelated donor (allogeneic transplantation), an identical twin (syngeneic transplantation) or the patient's own cells previously collected and suitably stored (autologous transplantation).

Hematopoietic stem cell transplantation sprang to life and became a reality in the late 1960s when 3 patients with congenital immunodeficiency disorders were successfully transplanted with bone marrow cells from their siblings.[1–3]

Since then, stem cell transplantation from a healthy donor has been increasingly used with a high level of success in many patients with malignant and non-malignant life-threatening disorders. Currently, more than 20 000 autologous and allogeneic marrow and peripheral blood and cord blood stem cell transplants are being carried out worldwide every year.

THE TYPES OF TRANSPLANTS AND INDICATIONS

The Classical Allogeneic Transplant

In the classical hematopoietic stem cell transplant, high-dose chemotherapy +/− radiotherapy (conditioning regimen) is always given before stem cells were infused. The belief that "marrow space" has to be created before infused cells can grow has led to this practice. However, it is now known that the conditioning regimen prevents donor cell rejection by virtue of their strong immunosuppressive effects rather than physically creating space in the bone marrow. Nevertheless, high-dose chemotherapy +/− radiotherapy acts not only as a strong immunosuppressant to prevent graft rejection, but it also exerts significant anti-tumor activities when the transplant is being done for malignancy.

Restoration

Restoration of marrow and immunological functions are the basis for allogeneic marrow transplant in severe aplastic anemias[4] and immunodeficiency diseases. In the classical allogeneic stem cell transplant, donor cells are introduced to repopulate the recipient's failing marrow but only after high-dose chemotherapy, which by destroying host immune cells, provides the needed immunosuppression to prevent rejection. Sustainable long-term engraftment of the donor hemopoietic stem cell is the key factor in marrow and immune functions restoration.

Conditions that may benefit from the restoration brought about doing hematopoietic stem cell transplant (HSCT) include: 1) stem cell disorders — severe aplastic anemia, Fanconi anemia, paroxysmal nocturnal hemoglobinuria; 2) congenital immunodeficiency disease — Ataxia telengiectasia, adenine deaminase, Wiskott–Aldrich syndrome and many others; 3) liposmal storage disorders; 4) inherited red cell

abnormality — thalassemia major, pure red cell aplasia, sickle cell anemia; and 5) other inherited disorders — osteopetrosis, Lesh–Nyhan syndrome, chronic granulomatous disease, Chediak–Higashi syndrome.

Graft-versus-tumor effect

Hematopoietic stem cells can generate lymphocytes that can react against targets in the recipient. This lymphocytes generated from normal donor stem cells can exert an anti-tumor effect. In leukemics, it has been shown that the anti-leukemic effect by the donor T-lymphocytes play an important role in preventing relapse. It is well known that the chance of acute leukemia relapse is high (60%) after a syngeneic (twin) marrow transplant. The absence of HLA disparity, which is essential for T cell recognition and reaction in the twins, is associated with the little or no graft-versus-leukemia effect. Such transplant is usually uncomplicated but relapse rate is high. Consistent with this is the high relapse rate seen when the donor marrow underwent rigorous T-depletion processing in an attempt to remove all T cells.[5–7] In contrast, much lower relapse rates (10–20%) are being reported with matched (related or unrelated) allogeneic marrow transplant despite the use of similar high-dose myeloablative regimen. Furthermore, relapse rate is significantly lower in patients who had developed some form of graft-versus-host disease than those who had developed none.

Such observations are consistently seen in many large studies and have led us to conclude that the graft-versus-leukemia effect of the accompanying T-lymphocytes plays a vital role in the prevention of relapse. Nevertheless, the initial donor stem cell engraftment is a prerequisite for the long-term survival of the donor T cells and graft-versus-leukemia effect. Donor T cells will be rapidly rejected by the host without an initial donor stem cell engraftment.

Conditions that may benefit from the graft-versus-tumor effects as a result of doing an allogeneic stem cell transplant are the: 1) acute leukemias — acute lymphoblastic leukemia, acute myeloid leukemia and the acute biphenotypic leukemia; 2) chronic leukemias — chronic myeloid leukemia, chronic lymphatic leukemia; 3) myelodysplastic syndrome; and 4) plasma cell disorders — multiple myeloma, plasma cell leukemia, Waldenstom's macroglobulinemia.

The Autologous Rescue

High-dose chemotherapy made possible[8–12]

The use of autologous blood stem cells as a form of rescue after high-dose chemotherapy +/− radiotherapy came into fashion in the late 1980s. The discovery of the surface molecule CD 34, which is the surrogate marker for the hematopoietic stem cell, in the early 1990s allowed easy and rapid evaluation of the stem cell dose in a given collected sample. This led to rapid increases in the number of autologous stem cell rescue procedures done for many conditions worldwide. The CD 34 molecule does not mark the stem cell but rather the population of cells that contain the stem cells.

Hematopoietic stem cell rescue has allowed us to use high doses of chemo-radiotherapy to treat resistant tumors. The use of high dose of chemotherapy +/− radiotherapy followed by autologous peripheral stem cell rescue is currently being widely explored in the treatment of some solid tumors like lymphoma, breast cancer, as well as certain cases of leukemias. Beneficial outcomes have been reported in selected groups of patients.

Autologous stem cell transplant has been shown to be useful in acute myeloid leukemia (AML), relapsed lymphoma and multiple myeloma. Patients with AML with no matched donor can achieve a 40–50% cure rate with autologous peripheral blood stem cell (PBSC) transplant as compared to a cure rate of 20% from conventional chemotherapy. Those patients with high-grade lymphoma (Hodgkin's and non-Hodgkin's) whose disease has relapsed after primary therapy and yet still chemosensitive, may benefit from autologous PBSC transplant. A cure rate of greater than 50% can be achieved in contrast to a cure rate of only 10% achieved with conventional salvage therapy.[8] In multiple myeloma, a median survival greater than 5 years with autologous PBSC transplant can be achieved as compared to a median survival of 2 years achieved with conventional therapy. Solid malignancies like neuroblastoma, Ewing's sarcoma, and renal cell carcinoma may also benefit from autologous hematopoietic stem cell transplant.[4]

Immune reconstitution and immune tolerance[13–16]

Immune reconstitution resulting in immune tolerance can happen following myeloablative doses of chemotherapy made possible by transplanting autologous stem cells. Hematopoietic stem cells (HSC) are progenitors of B- and T-lymphocytes, macrophages, and dendritic cells. These cells

mediate cellular and humoral immunity. Self-tolerance and immune recognition of foreign antigens is not innate but learned during ontogeny. Ablation or near ablation of a pathologically self-reactive immune system followed by reconstitution of the immune ontogeny from autologous HSC may reintroduce self-tolerance.

There are many anecdotal reports of patients treated with hematopoietic stem cell transplant (HSCT) for hematologic or neoplastic diseases that have experienced subsequent remission of their coincidental autoimmune disorders. In general, most allogeneic and autologous HSCT have resulted in at least short-term remissions. Because allogeneic HSCT has a higher rate of treatment-related mortality, consensus conferences have recommended autologous HSCT as the preferred form of stem cell transplant for severe autoimmune disorders.

Currently, autologous peripheral blood stem cell transplant has been found to be beneficial for the following conditions: Resistant rheumatoid arthritis, systemic vasculitis, severe systemic lupus erythematosus, polymyositis, scleroderma, relapsing post-splenectomy autoimmune hemolytic anemia and multiple sclerosis.

SOURCES OF HEMATOPOIETIC STEM CELLS

The Bone Marrow

The bone marrow is a well-established source of hematopoietic stem cells. The bone marrow is not only rich in stem cells but it contains many other hematopoietic cells, including key immune cells, which may be involved in graft-vs-host disease as well as graft-vs-tumor effects.

Bone marrow harvesting procedure has not changed significantly since it was first described in the late 1960s. Bone marrow harvesting is a safe procedure with minimal risk of complications. The overall incidence of life-threatening complications associated with the harvest procedure was 0.27%.[1,2] The collection of bone marrow can only be done from the posterior iliac crests where there is no vital organ beneath it. The amount of hematopoietic stem cells that can be collected is limited.

Peripheral Blood Stem Cells

Mobilized peripheral blood as the source of stem cells is currently widely used in autologous transplants. Currently, marrow is still the main source of stem cell in allogeneic transplants, although the use of mobilized peripheral blood stem cell is becoming increasingly popular. The concern

raised over the use of pharmacological doses of cytokines (such as the granulocyte-stimulating factor) to mobilize normal donors faded with the lack of acute, serious side effects and lack of evidence of any long-term hematopoietic disturbance.

Advantages of PBPC include rapid and durable trilineage hematologic engraftment, improved tolerance of the harvesting procedure (without general anesthesia) and possibly reduced tumor contamination in the autologous setting. Although there is a ten-fold increase in the number of donor T-lymphocytes in the peripheral blood harvest as compared to marrow harvest, the actual incidence and severity of acute graft-versus-host disease appears to be similar in both situations. Nevertheless, there is evidence that the incidence of chronic graft-versus-host disease is increased with mobilized peripheral blood.

1) *Allogeneic PBPC donations*. In the allogeneic setting, peripheral blood stem cells can be harvested from a normal donor through repetitive apheresis procedures after the donor has received colony-stimulating factors that augment the number of circulating stem cells. The procedure is usually performed on an outpatient basis and most donors will be able to continue their daily routine afterwards.
2) *Autologous PBPC collections.* In autologous PBPC transplant, stem cells can be collected following an early recovery from chemotherapy, which functions both as a mobilization agent and an *in vivo* purging agent. Colony-stimulating factors are often added to improve the yield.
3) *Cord Blood Stem Cells.* Human umbilical cord blood contains pluripotent HSC and successful hematopoietic reconstitution has been accomplished using cord blood stem cells. Initial success using cord blood from donors with two or three HLAs mismatched with the recipients, the apparent ease of engraftment and the lack of GVHD suggest that cord blood stem cells might have advantages over bone marrow in a wide variety of situations, especially for transplants between unrelated persons.

CONDITIONING REGIMEN

Allogeneic Stem Cell Transplant

The primary purpose of the conditioning regimen in the allogeneic transplant is to promote engraftment of the donor cells. The agents used for an

allogeneic stem cell transplant must not only be cytotoxic, but they also need to have immunosuppressive properties to help facilitate donor engraftment. High-dose chemotherapy +/− radiotherapy have been widely used to induce adequate immunosuppression to ensure donor cell engraftment. When used in malignancies, the high doses of chemotherapy provide a strong anti-tumor effect as well. The doses given are often myeloablative; that is, they will completely ablate the hematopoietic system of the host such that recovery will not be possible without any external stem cell rescue.

Autologous Stem Cell Transplant

The agents used in the conditioning regimen are generally cytotoxics and their primary goal is to eradicate malignant or unwanted cells. The stem cells are used as a rescue, to support ablative doses of chemotherapy +/− radiotherapy so that maximum anti-tumor effect can be achieved. There is no necessity to ensure strong immunosuppressive effect since the graft is from the patient and will not initiate any immune reaction.

THE NEW METHOD

The Non-myeloablative Allogeneic Hematopoietic Stem Cell Transplant[15–16]

The classical method also called "the myeloablative allogeneic stem cell transplant" is associated with significant toxicities, which are responsible for considerable transplant-related mortality. In many series, transplant-related mortality in the classical allogeneic stem cell transplant might reach up to between 20% and 30%. These toxicities are largely due to the use of high-dose chemo-radiotherapy prior to transplantation which was thought to be the main "engine" that led to the cure of malignancy, and the transplant itself was merely a supportive measure designed to allow the patient to receive myeloablative treatment without experiencing permanent aplasia.

However, it is now clear that the results seen with allogeneic transplantation are partly attributable to an immune effect mediated by donor lymphocytes, recognized as the graft-versus-tumor effect. The success of donor lymphocyte infusions in inducing remissions in patients relapsing

following allogeneic transplantation suggests that long-term disease control may be feasible without high-dose conditioning regimen. The emphasis has now switched to attempting to achieve a reduction in transplant-related morbidity and mortality with subsequent improvements in the quality of life.

Cellular therapy is the main element in the non-myeloablative allogeneic stem cell transplant (NASCT). Hematopoietic stem cells as well as immune cells like the T-lymphocytes are the main elements used to mediate the cure. Immunosuppressive agents with little cytotoxic effects are used as conditioning regimens to suppress host immune cells in order to prevent graft rejection; thus allowing donor stem cells and other progenitor cells to proliferate and occupy the marrow.

Restoration of hematopoiesis in patients with severe aplastic anemias and correction of immunodeficiency states can be achieved by allowing long-term engraftment of donor stem cells. Non-myeloablative regimens followed by transfusion of allogeneic progenitor cells have been investigated in a number of centers as a way of harnessing the graft-versus-tumor effect. One of the most well-studied NASCT protocol is the use of fludarabine, and low-dose total body irradiation as conditioning regimen and cyclosporine A and mycophenolate mofetil as the post-transplant immunosuppressants by Storb *et al.*, Fred Hutchinson Cancer Research Center in Seattle. Many patients with life-threatening malignant and non-malignant diseases have been entered into this protocol and significant numbers are enjoying disease-free survival of up to 3 years.

NASCT could be used: 1) to replace host with donor immuno-hematopoietic cells in malignant or genetically abnormal diseases characterized by deficiency of such cells; 2) as a replacement therapy for missing stem cell products (e.g. lysosomal enzymes); 3) as a platform for allogeneic immunotherapy with naïve or immune donor lymphocytes; 4) for therapy in autoimmune diseases; and 5) for induction of tolerance to donor alloantigens for cellular allografts and xenografts.

CELL DOSE, ENGRAFTMENT, GVHD PROPHYLAXIS, POST-TRANSPLANT SUPPORTIVE CARE AND COMPLICATIONS

These are as per bone marrow transplant. Please refer to Chapter 24, *Bone Marrow Transplant*.

PRACTICAL CONSIDERATIONS IN SELECTION OF HEMATOPOIETIC STEM CELL TRANSPLANT

In general, HSCT will be considered for following group of patients with 1) refractory disease — those who have failed conventional therapy; 2) recurrent disease — those who have relapsed after an initial success; and 3) poor risk disease — those with conditions known to do poorly with conventional therapy.

Allogeneic hematopoietic stem cell transplant (HSCT) has been found to be beneficial in certain cases of malignant (acute and chronic leukemias, myelomas, lymphomas) and life-threatening but non-malignant (congenital immunodeficiency disorders, severe aplastic anemia, thalassemia major) hematological disorders.

Autologous hematopoietic stem cell transplant is largely being used as a form of rescue therapy for certain solid tumor in particular the high-grade lymphomas. It has also being shown to be useful in older patients (50–70 years old) with acute myeloid leukemia (AML) and multiple myelomas. The use of autologous HSCT is, however, limited to these conditions and generally contraindicated in refractory disease since harvested autologous stem cells may be contaminated with tumor cells when initial disease load is high.

THE FIRST-LINE TRANSPLANTS

Allogeneic Matched Sibling Transplant

The matched sibling transplant is always the first choice transplant to be considered whenever a patient needs an allogeneic transplant. Outcomes with matched sibling transplant have been found to be encouraging for patients up to 45 years old. Long-term disease-free survival of 70% can be achieved when matched sibling transplants are done for the early acute and chronic leukemias and the severe aplastic anemias. These outcomes are much better than those seen with conventional therapies where disease-free survival for the acute leukemias, chronic leukemias and severe aplastic anemias are generally about 10–30%. For older patients, severe graft-vs-host disease becomes a serious problem and classical matched sibling transplant should be best avoided. Mortality from severe graft-vs-host disease rises to above 30% after 45 years of age and 40–50% above 50 years old.

The use of the non-myeloablative conditioning regimen has led to less severe graft-vs-host disease and may allow matched sibling transplant up to 60 years old. Nevertheless, more studies will need to be done to confirm this. Only about 30% of patients will have matched sibling; the others who need hematopoietic stem cell transplants will have to receive the stem cells from alternative sources (volunteer donors or cord blood).

Autologous Hematopoietic Stem Cell Transplant

The high-grade lymphomas

Autologous peripheral blood stem cell (PBSC) transplant has been found to be beneficial and superior to salvage chemotherapy in relapsed high-grade lymphomas and should be the first choice transplant in this condition. This finding has been borne out in several large randomized studies. Autologous PBSC transplant is now the treatment of choice for high-grade lymphoma patients whom relapsed from disease. Long-term disease survival outcome of 50–60% can be achieved compared to conventional salvage chemotherapy where the long-term disease survival outcome for this category of patients are generally below 10%.

The elderly AML and myelomas

In these elderly patients (> 60 years), allogeneic transplants carry a high chance of transplant-related mortality and is generally not recommended. Autologous transplant is generally beneficial in older patients with AML in remission and multiple myelomas.

Autologous PBST is superior to conventional chemotherapy in older AML patients, giving a long-term disease-free survival of 45%. These patients are deemed unfit to receive allogeneic transplants. Young patients with AML may do much better with matched sibling transplant (70% long-term disease-free survival) and should not be recommended for autologous HSCT if they have a matched sibling donor.

Autologous peripheral blood stem cell (PBSC) transplant has been shown to be beneficial for older myeloma patients. The median survival of myeloma patients treated with conventional chemotherapy alone is about 2 years. With autologous PBSC transplant, median survival beyond 5 years can be achieved in myelomas.

SECOND-LINE TRANSPLANTS[17] (FOR THE LEUKEMIAS, SEVERE APSLATIC ANEMIAS, CONGENITAL IMMUNODEFICIENCY DISEASE)

The Unrelated Volunteer Donor Transplants

The unrelated donor transplant, whether a volunteer donor or cord blood sample, should always be the second line and be only considered if there is no matched sibling. The unrelated donors provide an alternative source of hematopoietic stem cells, however, the problem of graft-vs-host disease is much more severe with matched unrelated volunteer donor transplant than with matched sibling transplants. HLA matching has to be very stringent and only full matched transplants should be considered in most patients except for the very refractory leukemias where survival with conventional treatment is 0%.

Despite full matching, severe graft-vs-host disease appears to be about 35–40% in many large series. The problem of severe graft-vs-host disease worsens with older patients and the eventual outcomes for most categories of diseases is unfavourable for patients above 35–40 years old.

The Unrelated Umbilical Cord Blood Stem Cell Transplant

The umbilical cord blood has been found to be rich in hematopoietic stem cell and could be used as an alternative source of hematopoietic stem cells to bone marrow and blood stem cells. Unrelated cord blood stem cell transplants using the classical high-dose chemotherapy are very well-established in children and the outcomes are comparable to adult stem cell transplants and may even be better.

The limitation of cell dose has been the most important barrier in the application of UCB transplant to adults. Nevertheless, with current good processing technology, a large proportion of samples in many banks could be used for a 60 kg male. The first major adult UCB transplant series was recently published by Laughlin *et al*.[18] It is clear that in this study that severe GVHD is much less than with adult volunteer donor transplant.

Volunteer Donor Marrow or Cord Blood Stem Cell Transplant?

Patients below 35 years old

For those patients below 35 years old and do not have matched sibling, a simultaneous search for both bone marrow and UCB donors should be

performed. The selection of unrelated donor UCB versus bone marrow should be based on the urgency of the transplant, the availability of 6/6 matched bone marrow donors, the cell dose of the UCB units and the degree of HLA disparity of the UCB units.

The median time to obtain a full matched volunteer donor up to the point of actual donation is about 3 months in Singapore. It is about 3–6 months in many other places including the US and Europe. Matched volunteer donor stem cell transplant is therefore not suitable for urgent life-threatening situations. Patients with severe pancytopenias from marrow failures or refractory acute leukemias run a high risk of getting life-threatening infections and will not be able to wait for three months.

The median time to obtain a suitable unrelated donor UCB for transplantation is about 3 weeks. Therefore, if urgent transplantation is required, UCB has a major advantage.

For pediatric and adult recipients <35 years old, a full matched volunteer donor transplant is generally preferred to a 2 antigen mismatched UCB transplant. However, 2 antigen HLA mismatched unrelated donor UCB unit is preferred to a 1 HLA-antigen mismatched unrelated bone marrow donor, provided the cell dose of the UCB unit is greater than 15 million nucleated cells per kg body weight of the recipient. When selecting between 1 and 2 antigen mismatched UCB units, the choice of an UCB graft should be based on cell dose.

Older patients who have no matched sibling

Patients who are 35 years or older will not be considered for the volunteer donor transplant because of its inherent high complication rates. Those with hematologic malignancies, other than AML and myelomas or severe aplastic anemias will not benefit from autologous PBSC transplant. These patients will have to be content with conventional salvage therapy, which might only be temporarily effective.

Lately, we have shown that cord blood stem cell transplant using the non-myeloablative approach is promising. Graft failure, which was associated with cord blood transplant using the myeloablative approach, became a less major issue. Autologous recovery will occur with the non-myeloablative approach, if the cord blood stem cell is unable to grow. Graft-vs-host disease appears mild. More cases will need to be done to assess the toxicity of the approach and its efficacy. Will there be higher relapse as compared to matched sibling transplant?

Tandem transplants using autologous PBSC transplant and then followed by cord blood stem cell transplant using the non-myeloablative regimen have been suggested by some experts in the field. Although autologous PBSC transplant may not be beneficial as a stand-alone procedure, significant tumor reduction could be achieved and this will help to prevent subsequent relapse after the cord blood stem cell transplants. Both types of transplants are known to be associated to carry low complications. Graft failure will not pose a major problem and graft-vs-host disease will be tamed. Yet, tumor reduction will be maximal, with both the high-dose chemotherapy exhibited by the autologous PBSC transplant and graft-vs-tumor effect as exhibited by the cord blood immune cells.

The non-myeloablative approach has allowed cord blood to be used as a good alternative to matched sibling marrow for this group of patients. Further studies will need to be done to evaluate this treatment approach.

FUTURE DIRECTIONS

Hematopoietic stem cell transplant is now a well-recognized treatment modality for many serious and life-threatening conditions — the severe aplastic anemia, the congenital immunodeficiency disorders, and the acute and chronic leukemias. Currently, it is being widely explored in the treatment of solid malignancies and the severe autoimmune diseases.

The stem cell transplant is a form of cell therapy. The knowledge and experience gained from the stem cell transplant have allowed us to develop other form of cell therapies. The last decade has produced remarkable advances in our ability to dissect the various cell lineages that comprise the bone marrow. This has allowed us to use various cell types along the principles adopted as for the stem cell transplant to treat various disease conditions. The use of donor T-lymphocytes for post-transplant leukemic relapse and the use of dendritic cells to activate graft-versus-tumor effects are good examples of the other cell therapies. The future will see the use of specific subpopulation of cells tailored and targeted against specific disease entities. There are major scientific, clinical and regulatory hurdles that still need to be overcome to bring the full potential clinical benefits of the stem cell and other form of cell therapies to patients. Nevertheless, it seems clear that problems can be solved. The next decade should be a very exciting period in the development of this field.

REFERENCES

1. Good RA, Meuwissen HJ, Gatti RA, Allen HD, Successful marrow transplant for correction of immumological deficit in lymphopenic aggammaglobulinemiaand treatment of immunologically induced pancytopenia, *Exp Hematol* **19**:4–10, 1969.

2. Gatti RA, Meuwissen HJ, Allen HD, Hong R, Good RA, Immunological reconstitution of sex-linked lymphopenic immunological deficiency, *Lancet* **2**:1366–1369, 1968.

3. Bach FH, Albertini RJ, Anderson JL, Bone marrow transplantation in a patient with Wiskott–Aldrich syndrome, *Lancet* **2**:1364–1366, 1968.

4. Anasetti C, Doney KC, Storb R, Marrow transplantation for severe aplastic anaemia. Long term outcome in fifty untransfused patients, *Ann Intern Med* **104**:461, 1986.

5. Antin JH, Graft versus leukemia: no longer an epiphenomenona, *Blood* **82**:2273–2277, 1993.

6. Thomas ED, Storb R, Clift RA, Bone marrow transplantation, *N Engl J Med* **292**:832–842, 895–902, 1975.

7. Thomas ED, Buckner CD, Banaji M, One hundred patients with acute leukemia treated by chemotherapy, total body irradiation, and allogeneic marrow transplantation, *Blood* **49**:511–533, 1977.

8. Tan P, Chapter 24 Bone marrow transplant, in *A Clinical Approach to Medicine*, pp. 487–513, 2001.

9. Statistical Center of the International Bone Marrow Transplant Registry and Autologous Blood and Marrow Transplant Registry, USA, 1995.

10. Herzig GP, Herzig RH, Current concepts in dose intensity and marrow transplantation, in Gale RP (ed.), *Acute myelogenous leukemia: Progress and controversies*, New York, Wiley-liss, pp. 333, 1990.

11. Appelbaum FR, Deisseroth AB, Graw RG, Prolonged complete remission following high dose chemotherapy of Burkitt's lymphoma in relapse, *Cancer* **41**:1059–1063, 1978.

12. Armitage JO, Jagannath S, Spitzer G, High dose therapy and autologous marrow transplantation as salvage treatment for patients with diffuse large cell lymphoma, *Eur J Cancer Clin Oncol* **22**:871–877, 1986.

13. Skoda RC, Tichelli A, Tyndall A, Hoffman T, Gillessen A, Gratwohl A, Autologous peripheral blood stem cell transplantation in patient with chronic thrombocytopenia, *Br J Haematol* **99**:56–57, 1997.

14. Joske DJL, Autologous bone marrow transplantation for rheumatoid arthritis, *Lancet* **6**:545–548, 1997.

15. Storb, Yu C, Wagner JL, Stable mixed hematopoietic chimerism in DLA identical littermate dogs given sublethal total body irradiation before and pharmacological immunosuppression after marrow transplantaton, *Blood* **89**:3048–3054, 1997.

16. Yu C, Storb R, Deeg HJ, Schuening FG, Nash RA, Graham T, Synergism between mycophenolate mofetil and cyclosporine in preventing graft versus host disease in

lethally irradiated dogs given DLA-nonidentical unrelated marrow grafts, (Abstract) *Blood* **86**(Suppl 1):577a, 1995.

17. Barker JN, Wagner JE, Umbilical cord blood transplantation: current state of the art, *Curr Opin Oncol* **14**:160–164, 2002.

18. Laughlin MJ, Barker J, Bambach B, Koc ON, Rizzieri DA, Wagner JE, Gerson SL, Lazarus HM, Cairo M, Stevens CE, Rubinstein P, Kurtzberg J, Hematopoietic engraftment and survival in adult recipients of umbilical-cord blood from unrelated donors, *N Engl J Med* **14**(344):1815, 2001.

This page intentionally left blank

Immunology

This page intentionally left blank

41

Drug Allergy

Bernard Thong Yu Hor and Chng Hiok Hee

INTRODUCTION

Adverse drug reactions (ADR) are common iatrogenic medical problems that may result in significant morbidity and mortality. It is important for the practicing physician to be familiar with the diagnosis and management of ADR. As defined by the World Health Organization, ADR include all non-therapeutic effects of the drug, with the exception of treatment failure, intentional or accidental poisoning and drug abuse. ADR include: 1) expected pharmacologic side-effects; 2) consequences of abnormal pharmacogenetic drug metabolism; 3) drug-drug interactions; 4) alterations of tissue ecology (e.g. intestinal bacterial overgrowth after extensive antibiotic usage); and 5) drug/disease specific events (e.g. ampicillin-induced rash in acute Epstein–Barr virus infection), and untoward reactions including idiosyncratic, intolerant, pseudoallergic and allergic reactions. Drug allergies are immune-mediated ADR in sensitized individuals. Although allergic reactions to drugs are unpredictable, early diagnosis and appropriate management reduce morbidity and mortality associated with drug allergy.

Epidemiology

ADR have been reported to account for 3–6% of all hospital admissions and to occur in 10–15% of hospitalized patients. In Singapore, the Pharmacovigilance Unit of the Health Sciences Authority (HSA) monitors ADR based on voluntary reports from healthcare professionals. Drug allergy has been estimated to account for 5–10% of all ADR. Data from an electronic network-based, allergist-immunologist case-verified, inpatient drug allergy reporting system found an incidence of drug allergy in hospitalized patients in Singapore of 4.20 per 1000 hospitalizations, incidence of drug allergy developing during the course of inpatient treatment of 2.07 per 1000 hospitalizations and mortality from drug allergy of 0.09 per 1000 hospitalizations.

Although any drug or biologic agent can potentially cause an allergic reaction, the chemical properties and molecular weights of some drugs result in more reactions than other drugs. In the Singapore series, antimicrobials and anti-epileptic drugs together accounted for 75% of all drug allergies reported in a large acute general hospital (Table 1). Allergic drug reactions may manifest in many ways and any tissue or organ may be affected. Cutaneous drug eruptions are the most common clinical manifestations with maculopapular rash being the most common morphology in reported series. Systemic manifestations occurred in 30% in the Singapore series, of which hepatitis was the most common. In this registry, serious allergic reactions like Stevens–Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and generalized exfoliative dermatitis (GED),

Table 1 Types of Drugs Commonly Implicated in Allergic Reactions in a General Hospital in Singapore

Drugs Causing Allergy	Number	Percentage
Penicillins	53	25.2
Cephalosporins	33	15.7
Co-trimoxazole	19	9.0
Phenytoin	17	8.1
Carbamazepine	13	6.2
Allopurinol	12	5.7
Ticlopidine	10	4.8
Quinolones	10	4.8
Others	43	20.5

occurred in 5.2% of the cases, and half of these cases were due to allopurinol, phenytoin, carbamazepine or penicillins.

PATHOPHYSIOLOGY

The allergenic potential of drugs depends in part on their chemical properties and size. Drugs which exist in the form of macromolecules (e.g. insulin and recombinant proteins) or functionally multivalent compounds (e.g. neuromuscular blocking agents), can directly stimulate the immune system and are more likely to produce allergic reactions. Low molecular weight drugs are usually haptens and must bind with macromolecular "carrier proteins" to form immunogenic complexes or undergo metabolism to produce protein-reactive intermediates which then haptenate macromolecules.

Drug allergy can be broadly classified into IgE and non-IgE mediated reactions. IgE-mediated reactions are immediate hypersensitivity reactions, requiring the presence of drug-specific IgE antibodies. Drug allergens binding specific IgE antibodies on the surface of mast cells and basophils cross-link the antibodies and result in cell activation. Mediators released from these cells are responsible for pruritus, urticaria, angioedema, nasal congestion, rhinorrhea, bronchospasm, abdominal pain, diarrhea and hypotension that may occur in any combination in the patient. The β-lactam antibiotics, blood products and vaccines are examples of drugs that may cause IgE-mediated reactions.

Non-IgE mediated reactions may be due to complement dependent cytotoxic reactions (type II hypersensitivity, e.g. hemolytic anemia), immune-complex reactions (type III hypersensitivity, e.g. serum sickness) or T-cell mediated delayed hypersensitivity (type IV hypersensitivity, e.g. allergic contact dermatitis). Many non-IgE mediated drug reactions cannot be classified under any of the Gell and Coomb's classification of hypersensitivity as mentioned. These include delayed dermatologic drug eruptions (e.g. morbilliform rashes, eczematous rashes, erythroderma, exfoliative dermatitis), immunobullous eruptions (e.g. erythema multiforme (EM), SJS and TEN), drug-induced fever and organ-specific manifestations (e.g. hepatocellular necrosis or hepatitis, nephropathy, pneumonitis and aseptic meningitis). However, immunopathogenesis is suspected because a number of these reactions are associated with CD4/CD8 positive T cells, drug-specific T cell clones and positive patch tests. Based on recent research

findings it may be possible to classify drug exanthems under 4 proposed subclasses of T-cell mediated type IV reaction.

RISK FACTORS FOR DRUG ALLERGY

These include drug and host-specific risk factors, and are summarized in Table 2. The most important drug-specific risk factors are the chemical properties and molecular weight of the drug. Other drug-specific risk factors are the dose, route of administration, duration of treatment, repetitive exposure to the drug and concurrent illnesses.

Host risk factors include age, gender, atopy (for reactions to high molecular weight drugs) and specific genetic polymorphisms. Drug allergy has been reported to aggregate in certain families although it has not been conclusively shown that a family history of drug allergy is a definite risk factor for developing new drug allergies. Multiple drug allergies have also been reported in non-HIV populations, where wide-ranging immunological cross-reactivities with drugs containing tertiary and quaternary ammonium groups, which are present in many different pharmacologically active agents, were suggested. However, allergy to one drug has not been shown to be a risk factor for developing allergy to other drugs.

PRINCIPLES IN THE DIAGNOSIS OF DRUG ALLERGY

The temporal relationship of onset of reaction and institution of drug is important. Drug allergies occur only after a sufficient period of sensitization

Table 2 Risk Factors for Drug Allergy

Drug-specific	Host-specific
• Chemical properties, e.g. β-lactam ring • Molecular weight (high > low) • Dose (high > low) • Route of administration (parenteral > oral, topical) • Duration of treatment (longer > shorter) • Frequent repetitive courses • Concurrent medical illness — HIV infection	• Age (older) • Gender (females) • Atopy • Specific genetic polymorphisms

(at least 5–7 days of treatment). If a suspected reaction occurs immediately after the first dose of a drug, a history of previous exposure to that drug or a drug of the same class would support a diagnosis of an allergic reaction. Without prior exposure, the onset of drug allergy usually occurs within the first 1–3 weeks of starting the medication but may be delayed for as long as 2–3 months. For example, reactions to anti-tuberculous drugs, allopurinol, anticonvulsants and propylthiouracil may first appear up to 1 month or more after the onset of therapy.

A past history of drug allergy may be relevant as allergic drug reactions may be due to cross-reactivity. Common examples include β-lactam antibiotics, e.g. penicillins and carbepenems (IgE-mediated allergy) and the aromatic anti-convulsants, e.g. phenytoin, phenobarbitone and carbamazepine (non-IgE mediated allergy). However, there has been much misconception about the cross-reactivity between the cephalosporins and penicillins. If the reaction is a maculopapular rash, the rate of cross-reactivity is unknown and is probably very low. It is when the reaction is an urticaria or anaphylaxis that the cross-reaction may be up to 10% for first generation cephalosporins (case reports dealt only with cephalothin and cephaloridine). In the case of newer generation cephalosporins, the risk is much lower. Similarly, with regard to allergies to sulphonamide-containing drugs (e.g. sulphonyureas, cotrimoxazole, celecoxib), although patients should be advised on the potential for cross-reactivity, the use of another sulpha-based drug is not absolutely contraindicated.

Resolution of the reaction upon withdrawal of a drug suggests a cause-effect relationship and aids diagnosis. On the other hand, if a patient has been labeled as allergic to a drug and subsequently receives a course of the same medication without any adverse immunologic reaction, then the person is not allergic.

DIFFERENTIAL DIAGNOSES OF A DRUG ALLERGY

Many cutaneous and systemic reactions may mimic a drug allergy. The empirical use of antimicrobials in viral infections often results in confusion as to whether a drug allergy has occurred when an exanthem due to viral infection develops. Irritant contact dermatitis from topical medicaments may at times mimic drug allergy. Side-effects of drugs (e.g. erythromycin-induced dyspepsia, clindamycin-induced pseudomembranous colitis and salbutamol-induced tachycardia) should be differentiated

from drug allergies as they are known to be non-immunologically mediated effects of therapy.

Some idiosyncratic adverse drug reactions may produce symptoms and signs that mimic a drug allergy. These are sometimes referred to as "pseudoallergic" reactions. In patients with non-steroidal anti-inflammatory drug (NSAID) intolerance also referred to as NSAID sensitivity, cyclooxygenase inhibition by NSAID leads to an overproduction of leukotrienes, resulting in urticaria, angioedema and/or non-IgE mediated anaphylaxis. Opiates and radiocontrast media can induce direct mast cell histamine release and hence urticaria and angioedema. The flushing of "red man syndrome" associated with vancomycin is also the result of direct mast cell histamine release. Angiotensin-converting enzyme inhibitors release kinins causing angioedema, commonly of the tongue. The maculopapular exanthem occurring following ampicillin use during infectious mononucleosis is also an idiosyncratic reaction. Patients may be prescribed ampicillin on another occasion without recurrence of similar lesions.

APPROACH TO THE EVALUATION OF DRUG ALLERGY

The diagnosis of drug allergy depends primarily on the patient's history, clinical findings that are consistent with an allergic reaction and that the onset of reaction is temporally-related to use of the suspected drug. Investigations in the form of skin tests (prick, intradermal and patch), blood tests for drug-specific IgE, lymphocyte transformation test (LTT) and incremental drug challenges, may assist in determining the putative drug causing the reaction.

History

Both the type (cutaneous, systemic or both) and severity (minor or life-threatening) of the allergic reaction should be determined. The onset, distribution and morphology of the drug eruption provide important clues to the diagnosis. Common drug eruptions include maculopapular rashes, urticaria, and fixed drug eruptions. These are usually mild and not life-threatening. Immunobullous eruptions, SJS, TEN, drug hypersensitivity syndromes and anaphylaxis are potentially life-threatening. Systemic symptoms including fever, and lymphadenopathy suggest a

drug hypersensitivity syndrome. Urticaria, angioedema, dysphonia, rhinorrhea, nasal congestion, dyspnea, wheeze, syncope, abdominal pain, diarrhea and vomiting are symptoms of anaphylaxis. A guide to the severity of a previous reaction would include whether the patient needed prolonged hospitalization, intensive care or mechanical ventilation.

A drug history should describe the temporal relationship of the onset of allergy symptoms to the initiation of drug use. The use of medications (previous and current) including over-the-counter preparations, alternative medicine and medications obtained from all medical sources (e.g. other family physicians, hospitals) should be sought. When the patient is unable to name the drugs that he/she is taking, the medical history may provide a clue to the nature of the medication. For example, hyperuricemia recently diagnosed may have been treated with allopurinol, headache with a NSAID or a recent seizure with phenytoin. In addition, the medical history can provide clues to the failure of response to treatment of a drug allergy, e.g. persistent hypotension in the treatment of anaphylaxis in a patient on β-blockers.

The indication for the prescription of the drug should be obtained as the signs and symptoms may be due to the underlying medical condition rather than the drug (e.g. viral exanthem with cervical lymphadenopathy and hepatitis) or may have been precipitated by the underlying condition (e.g. ampicillin-induced maculopapular rash in infectious mononucleosis).

An evaluation of risk factors for drug allergy and knowledge of drugs that commonly result in allergy are useful aids in the diagnosis and management of drug allergy.

Physical Examination

A full physical examination should be carried out. Note the morphology and distribution of the drug eruption and whether there are features of mucocutaneous or systemic involvement (GED, SJS, TEN). Important non-dermatological signs are those of anaphylaxis (tachypnea, stridor, laryngeal edema, rhonchi, hypotension, generalized erythema and impaired consciousness), drug hypersensitivity syndrome (fever, lymphadenopathy, hepatosplenomegaly) and SJS (oral, conjunctival and genital mucosal lesions).

Investigations

Whether further investigations are necessary depends on the probability of a diagnosis of drug allergy after completing the history and physical examination. Using the principles outlined above, if the reaction is consistent with a drug allergy and the putative drug can be identified from history, no further confirmatory tests are needed. However, if drug causality is one of several other possible reasons for the described clinical event or if there are several potential putative drugs, further *in vitro* or *in vivo* tests may be used to come to a more definitive diagnosis of drug allergy and/or to identify the putative drug.

Basic investigations including the full blood count and differential count, platelet, serum creatinine, liver function tests, urine microscopy and chest X-ray should be done to evaluate for other organs involvement in the case of drug allergy with systemic symptoms (drug hypersensitivity syndromes, SJS, TEN). The other diagnostic tests to order depends on whether one is evaluating an IgE or non-IgE mediated hypersensitivity to a drug and the suspected drug. In drug allergy testing the reagents, with few exceptions, are neither standardized for *in vitro* nor *in vivo* tests. Provocation tests may be harmful to the patient and are possibly not sensitive enough since crucial co-factors might be absent during the challenge procedure. These issues of sensitivity and specificity of the tests, safety and risks versus benefits of performing certain tests have to be discussed with the patient. These tests and procedures should be carried out by trained allergists/clinical immunologists.

Skin testing

Skin prick (SPT) and intradermal (ID) tests are used in the diagnosis of IgE-mediated reactions. These are usually done 4–6 weeks after the initial drug reaction as the tests may be falsely negative if done too early. Only if the SPT is negative, an ID test, where a greater volume of the allergen is introduced into the skin, is done. A positive skin test strongly suggests a diagnosis of drug allergy and may predict which persons are at greater risk of future reactions.

The advantages of SPT/ID tests are that these are specific, convenient and relatively inexpensive. They are also relatively safe with systemic reactions occurring in less than 0.02% of cases. Contraindications to skin testing include inability to discontinue medications that interfere with

skin testing, e.g. antihistamine, severe atopic eczema or generalized skin disorders and severe dermatographism. The main limitation is that except for penicillin, insulin, heterologous serum, streptokinase and chymopapain, most of the skin tests for drugs have not been adequately studied, standardized or validated. The positive and negative predictive values of these tests depend on the drug being tested. In penicillin skin testing, both the major and minor determinants are used and a positive skin test has a very high predictive value of an immediate reaction to penicillin.

A late reading intradermal test or a patch test may be used to assist the diagnosis of non-immediate (suspected T-cell mediated reactions) reactions to β-lactams and other drugs manifesting as cutaneous lesions (e.g. maculopapular exanthems, SJS, TEN) appearing several hours after the last drug intake. These tests also need further validation and standardization for such drug reactions. Patch testing is the most reliable test for the diagnosis of allergic contact dermatitis.

In vitro testing

For the diagnosis of IgE-mediated reactions, serum for detection of drug-specific IgE is usually taken about 4–6 weeks after the reaction. The main advantage of using *in vitro* specific IgE testing is when skin testing is contraindicated (see above). An important limitation of the test is that the allergenic drug moiety/metabolite for most drugs are not known and the sensitivity and specificity of many commercial tests have not been well studied. In many instances, this test is far less sensitive than skin tests and generally do not have good negative predictive value. For example, the negative predictive value of this test for penicillin hypersensitivity is poor because they are relatively insensitive and do not test for the minor determinants.

Drug challenge

The main indication for drug challenge (incremental dose challenge) is to exclude rather than to prove an allergic drug reaction. It is carried out when the probability of an allergic reaction to the drug is low. It should not be carried out when the reaction is severe, such as the drug hypersensitivity syndromes, SJS, TEN and anaphylaxis. The procedure involves

giving the patient fractions of the target dose of the drug either orally or parenterally, increasing the dose at regular intervals and observing for any recurrence of the suspected reaction. It is stopped when a positive reaction develops or when the standard dose is reached. The process can be done as an outpatient. If the suspected reaction is urticaria or angioedema, the patient should have an intravenous access that can be used in case a severe reaction is precipitated. Facilities for resuscitation must be available during an incremental drug challenge. Drug challenge should be done under the supervision of an allergist/clinical immunologist experienced with the procedure.

The main limitation of an incremental dose challenge is a potential risk of anaphylaxis or other serious reactions. The safety issues and the risks versus benefits of a drug challenge have to be discussed with the patient and informed consent taken.

Other tests

A clinical diagnosis of anaphylaxis may be confirmed retrospectively by a rise in serum tryptase that peaks at 1–2 hours and remains elevated 2–4 hours after the reaction. Although an increase in the level of 24-hour urinary histamine/N-methylhistamine aids in the diagnosis of anaphylaxis, the test is not readily available. Indirect and direct Coomb's tests are often positive in drug-induced hemolytic anemia. Specific tests for immuno-cytotoxic thrombocytopenia and granulocytopenia are available in some centers. The lymphocyte transformation test (LTT) measures the proliferative response of peripheral blood mononuclear cells (T cells) to the suspected drug under optimal *in vitro* conditions. Some groups have had good experiences with LTT, with sensitivity of 58–78% and specificity of 85–100% but other researchers have found it less useful. This test is not readily available.

TREATMENT OF DRUG ALLERGY

Acute Allergic Reaction

Outpatient treatment of drug allergies may be considered in non-progressive urticaria, maculopapular rash, and fixed drug eruptions in the absence of systemic symptoms and signs. The patient should be admitted for

inpatient treatment in the following instances: life-threatening angioedema, anaphylaxis, rapidly progressive maculopapular rash, erythema multiforme or other immunobullous drug eruptions, drug hypersensitivity syndrome, SJS and TEN or when there is major organ involvement such as severe hepatitis or hemolytic anemia. The immediate management of any patient with an allergic drug reaction would be to stop the culprit drug or all non-essential suspected drugs immediately. Do remember that an allergic reaction may still progress after the putative drug has been stopped although in most instances it will be less severe.

Anaphylactic drug reactions require prompt emergency treatment comprising the "ABCD of resuscitation". This includes securing the airway (endotracheal intubation, oxygenation), breathing (nebulized bronchodilators, mechanical ventilation), circulation (Trendelenberg position, intravenous fluids and vasopressor) and the use of appropriate drugs [IM or SC epinephrine; adults, 0.3–0.5 ml of a 1:1000 (1 mg/ml, wt/vol) dilution every 10–15 minutes up to a maximum dose of 1 ml per dose; children, 0.01 ml (0.01 mg/kg body weight up to a maximum of 0.5 ml per dose of a 1:1000 dilution, repeated every 15 minutes for 2 doses, then every 4 hours as needed), intravenous diphenhydramine (Benadryl®) [1–2 mg/kg body weight or 25–50 mg].

Intravenous hydrocortisone is used in anaphylaxis mainly for late response. Corticosteroid is given to patients with a drug hypersensitivity syndrome especially when there is major organ involvement, e.g. severe skin inflammation, hepatitis and acute renal failure. Patients with immune complex and cytotoxic drug reactions may also benefit from corticosteroids. However, it should be avoided in TEN as there is a high risk of sepsis and its use in severe SJS is controversial. A tapering course of prednisolone starting at 0.5–1 mg/kg body weight per day over 4–8 weeks with calcium and vitamin D supplementation as prophylaxis against steroid-induced osteoporosis is often sufficient.

High-dose intravenous immunoglobulin (IVIG) 1g/kg body weight per day for 2 days was recently found to be beneficial in TEN because of its immunomodulatory effect. When renal insufficiency is present, a lower dose of 400 mg/kg body weight per day for 3 days may be given.

In mild drug reactions, withdrawal of the drug may be all that is necessary. Patients with pruritic skin lesions require symptom relief with antihistamines. Topical emollients (e.g. 10% urea cream, aqueous cream) should routinely be given. Patients with TEN need barrier nursing and

referral to a plastics surgeon. Patients with SJS require oral toilet and referral to an ophthalmologist for ocular lubricants and topical corticosteroids. Ocular involvement and skin pigmentary changes are the most significant long-term sequelae in SJS and TEN.

Opinion from an allergist/clinical immunologist should be sought when there is uncertainty regarding whether the reaction is a drug allergy, which drug is responsible, when diagnostic testing is required or advice on the treatment of a serious drug allergy or desensitization is required.

All suspected or confirmed drug allergies should be reported to the Pharmacovigilance Unit of the Health Sciences Authority (HSA) of Singapore. If in doubt, a referral to an allergist/immunologist should be made rather than labeling the patient as being allergic to multiple medications and thus depriving the patient of potentially useful drugs in the future. A Medik Awas application for a medical alert card or bracelet should be made with the Singapore Medical Association to avoid future exposure to the same or cross-reacting drugs. The patient should be educated on potentially cross-reacting drugs that are best avoided.

Treating through Reactions

"Treating through" refers to the continuation of drug treatment in the presence of a suspected allergic reaction. This may be carried out when the continued use of the drug is essential and the allergic reaction is not life-threatening. It is recommended that an allergist's/clinical immunologist's opinion be consulted. Antihistamines and corticosteroids are often prescribed to suppress the allergic reaction. The patient should be monitored closely for progression of the allergic reaction.

Desensitization

Desensitization is a method of making a patient tolerant to a drug that he/she had previously developed an allergic reaction to. It is carried out when there are no reasonable alternative therapies and treatment is essential. Desensitization is contraindicated if the reaction was a dangerous reaction like SJS or TEN. However, anaphylaxis is not a contraindication. It must be carried out under the supervision of an allergist/clinical

immunologist familiar with the procedure. Patients are given progressively larger doses of the medication according to a schedule until the full therapeutic dose is reached. Two methods of desensitization are available, rapid desensitization for IgE-mediated reactions and slow desensitization for patients with non-IgE mediated maculopapular rashes. Rapid desensitization, which allows the standard dose of the drug to be achieved within about 6 hours, is always done in the intensive care where resuscitation facilities are available. Slow desensitization, which may take up to a month to achieve the required dose, can be carried out in the outpatient setting. Since tolerance can be breached if doses are missed, once desensitization has been achieved the drug must be taken regularly until the full course is completed. Return of clinical sensitivity can occur within 24–48 hours of cessation of the drug. After successful desensitization therapy, the patient must be informed that he/she is still considered to be allergic to the drug. Successful desensitization of patients with IgE-mediated allergy to penicillin, vancomycin and insulin, as well as non-IgE mediated allergy to allopurinol, cotrimoxazole and sulphasalazine have been reported.

Premedication

This is used for patients with a history of anaphylactoid contrast agent mediated reaction. Patients are pre-treated with a combination of corticosteroids and antihistamines. Premedication should not be used for patients with known immune-mediated allergic drug reactions.

PREVENTION OF DRUG ALLERGY

The principles of preventing allergic drug reactions include: 1) a careful history to determine risk factors in the patient; 2) avoidance of drugs that had caused reactions in the past; 3) avoidance of cross-reactive drugs; 4) prudent prescribing of drugs frequently associated with allergic reactions; and 5) use of oral as opposed to parenteral drugs. All patients with a history of drug allergy should be educated on the potential severity of future reactions, measures to avoid future events including knowledge of cross-reacting drugs and the need to carry a Medic Alert (Medik Awas) card or bracelet all the time.

CONCLUSION

Drug allergies are iatrogenic immune-mediated adverse drug reactions. Although allergic drug reactions are generally unpredictable, morbidity and mortality can be reduced with early diagnosis and appropriate treatment. On the other hand, a correct diagnosis and avoidance of mislabeling of drug allergy prevents the patient from being deprived of potentially valuable medications. The most important consideration in evaluating a patient for drug allergy is a suspicion by the physician that an unexplained symptom or sign may be due to a drug being administered. A careful history and clinical examination are the main investigative tools.

SUGGESTED READING

1. Demoly P, Bosquet J, Epidemiology of drug allergy, *Curr Opin Allergy Clin Immunol* **1**:305–310, 2001.
2. Lazarou J, Pomeranz BH, Corey PN, Incidence of adverse drug reactions in hospitalized patients, A meta-analysis of prospective studies, *JAMA* **279**(15):1200–1205, 1998.
3. Thong BY, Leong KP, Tang CY, Chng HH, Drug allergy in a general hospital: results of a novel prospective inpatient reporting system, *Ann Allergy Asthma Immunol* **93**: 342–347, 2003.
4. Gruchalla RS, Understanding drug allergies, *J Allergy Clin Immunol* **105**(6) part 2: S637–644, 2000.
5. Pichler W, Yawalkar N, Schmid S, Helbling A, Pathogenesis of drug induced exanthems, *Allergy* **57**:884–893, 2002.
6. Brockow K, Romano A, Blanca M, Ring J, Pichler W, Demoly P, General considerations for skin test procedures in the diagnosis of drug hypersensitivity, *Allergy* **57**:45–51, 2002.
7. Viard I, Wehrli P, Bullani R, Schneider P, Holler N, Salomon D, Hunziker T, Saurat JH, Tschopp J, French LE, Inhibition of toxic epidermal necrolysis by blockade of CD95 with human immunoglobulin, *Science* **282**:490–493, 1998.
8. Roujeau JC, Treatment of severe drug eruptions, *J Dermatol* **26**(11):718–722, 1999.
9. Bernstein IL, Gruchalla RS, Lee RE, Nicklas RA, Dykewicz MS (eds.), Disease management of drug hypersensitivity: a practice parameter, *Ann Allergy Asthma Immunol* **83**: 665–700, 1999.

42

Food Allergy

Cheng Yew Kuang and Chng Hiok Hee

INTRODUCTION

It is important for the physician to be familiar with the varied presentations of food allergy as well as the principles and approach to its diagnosis and management. This is because the prevalence of perceived food allergy is high amongst the general population and the proliferation of costly, unproven diagnostic and therapeutic techniques add to the mislabeling of food allergy. A misdiagnosis of food allergy, particularly multiple food allergy, and unnecessary food restriction, often lead to malnutrition, psychological distress and impaired quality of life.

CLASSIFICATION AND TERMINOLOGY

An adverse food reaction is any untoward reaction to the ingestion of food or food additive. These reactions can be broadly classified into toxic and non-toxic reactions (Fig. 1). Toxic reactions affect anyone who consumes the food and there is no individual susceptibility, e.g. poison in non-edible mushrooms and staphylococcal toxin in food poisoning.

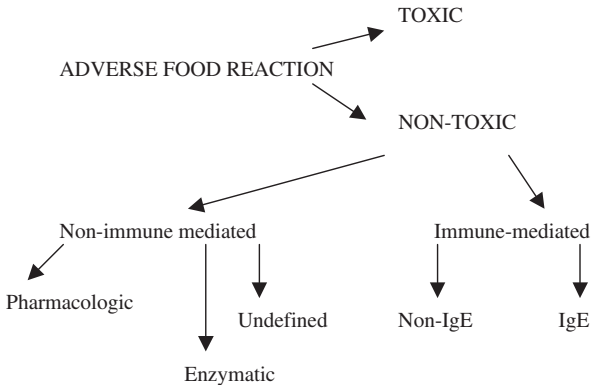


Fig. 1 Classsification of adverse reactions to food.

Non-toxic food reactions depend on individual susceptibility and are classified into immune-mediated and non-immune mediated reactions.

The term food allergy is used when there is an abnormal immunologic response to a specific food resulting in a variety of symptoms. Food allergy can be further subgrouped into IgE and non-IgE mediated reactions. Food intolerance is any adverse food reaction due to a non-immunologic mechanism. This may be the result of the pharmacologic property of the food (e.g. tyramine-induced headache) or metabolic disorders (e.g. milk-induced diarrhea from lactase deficiency). In some instances, history and challenge tests clearly prove the causative role of the food but the mechanism is unknown. Such reactions are classified under food intolerance in the subgroup of undefined reactions.

Psychosomatic food adverse reactions are not included in this classification as the reaction is truly not food-dependent but is primarily a psychological disorder. Doctors should also differentiate the term sensitivity from hypersensitivity as confusion results when these terms are used interchangeably. The term hypersensitivity should be used only when the immune system has a role, i.e. the reaction is allergic while the term sensitivity does not imply a role for the immune system in the adverse reaction.

EPIDEMIOLOGY

Adverse food reactions are not uncommon, perceived food allergy is common and true food allergy is uncommon. The prevalence of food allergy is

greatest in the first few years of life and declines over the first decade. Based on epidemiological studies done in Western populations, approximately 5% of children younger than 3 years old and 1.5% of adult population experience food allergy. Children with atopic diseases tend to have a higher prevalence of food allergy. About one-third of children with moderate to severe atopic dermatitis and 10% of children with asthma have been shown to have food allergy. Data on the prevalence of food allergy in adults with atopic disorders is not available but is believed to be probably also increased. The true prevalence of food allergy in Singapore is not known.

It is fortunate that only a few foods are responsible for most of the allergic reactions. In addition, the majority of patients are allergic to only a few foods. Multiple food allergy is rare. Egg, milk, wheat, soy, peanut, tree nuts, fish, and crustaceans have been shown to be responsible for more than 90% of allergic food reactions in young children in the US. In older children and adults, peanut, crustaceans, tree nuts and fish (in order of frequency) were reported to be responsible for most fatal anaphylactic reactions and account for 90% of food allergy.

In a recent review of food-induced type I hypersensitivity reactions in adult patients treated at our Clinical Immunology/Allergy clinic, crustaceans, molluscs and edible "bird's nest" (made up of saliva of swiftlet, the allergen being a 66 kD glycoprotein) were found to be the main causative foods. More than half of these patients have a history of allergic rhinitis, asthma and/or eczema and a family history of atopic diseases was often present. Similarly, a study by our pediatric counterpart found that "bird's nest" allergy was the single most common cause of food allergy seen at the National University Hospital. These findings suggest that differences in cultural practice determine the types of food that cause allergy.

NATURAL HISTORY OF FOOD ALLERGY

The past 20 years of research has led to a greater understanding of the natural history of food allergy. It has been shown that children tend to lose their allergic reaction to milk, soy, egg and wheat as they grow older. For patients who had severe reactions, this may take a longer time but tolerance is often eventually achieved. However despite clinical tolerance, specific IgE antibodies to the allergenic food as detected by skin prick or *in vitro* test (e.g. RAST) persists. On the other hand, allergy to foods like peanut, tree nuts, fish and shellfish tends to remain lifelong.

PATHOPHYSIOLOGY

The exact mechanism of development of food allergy is unknown and is likely to be multifactorial. Most individuals develop tolerance to ingested food antigens. It is postulated that in genetically predisposed individuals, disruption of the immunologic and non-immunologic gastrointestinal barriers could alter the handling of ingested antigens thus allowing sensitization to occur. Subsequent exposure to food antigens would then result in the clinical manifestations of food allergy. The increased susceptibility of infants to food allergy is believed to be due to immaturity of the immune system and the gut.

Many food allergic reactions are IgE-mediated type I reactions in the Gell and Coomb's classification. These reactions are immediate in their onset; most manifesting within 2 hours and frequently within minutes of food ingestion. The non-IgE mediated reactions are not so well characterized and understood. Such reactions usually take several hours to a few days before becoming clinically evident. Some reactions such as atopic eczema may be both IgE and non-IgE mediated.

Advances in molecular techniques have led to the characterization of a number of food allergens. Examples of the major allergens identified in foods are casein and whey in cow's milk, ovomucoid in egg white and tropomyosin in shellfish. The allergenic components of foods are mainly glycoproteins. They are water-soluble, heat-stable and resistant to acid and proteolytic digestion. Food processing however may alter the antigenicity of food. Fish allergens may be changed with canning process and patients who cannot tolerate fresh fish may tolerate canned tuna and other fish. Another phenomenon of clinical implication is that allergen cross-reactivity is readily demonstrated by skin test, *in vitro* specific IgE test and immunoblotting technique but this often does not reflect clinical cross-reactivity. As an example, an individual may have positive skin tests to peanut and other legumes but only develops allergic reactions to peanut and can consume other legumes without problem.

CLINICAL MANIFESTATIONS

IgE-mediated Reactions

Anaphylaxis

The most serious food allergic reaction is anaphylaxis. Anaphylaxis is a potentially life-threatening clinical syndrome characterized by the

sudden onset of generalized symptoms, affecting multiple organ systems in the body. These symptoms, which typically occur less than 2 hours, often minutes after ingestion of food, include: urticaria and angioedema, pruritus and flushing of skin, tightness in the throat and dysphonia, laryngeal edema, tongue swelling, itch in throat, palate, tongue and lips, dyspnea and wheezing, dizziness, near syncope, syncope and hypotension, nausea, vomiting and abdominal cramps, rhinorrhea and nasal congestion, eye itch and tearing and substernal chest discomfort. Factors reported to contribute to a fatal outcome are a delay in administration of epinephrine, a concomitant diagnosis of asthma, a history of previous severe reactions, failure to recognize biphasic anaphylaxis (recurrence of symptoms several hours after resolution of first episode of anaphylaxis) and the use of beta-blockers.

Cutaneous and other organ system manifestations

Skin lesions are the most common manifestations of IgE-mediated food hypersensitivity. Signs range from acute urticaria and angioedema to a erythematous rash. It is important to note that chronic urticaria is almost never caused by food allergy. In addition in atopic children with food allergy, the predominant cutaneous manifestations are erythema and pruritus rather than urticaria. Gastrointestinal complaints are the second most frequent manifestations. The main clinical features are vomiting, diarrhea and abdominal pain. Airway symptoms like sneezing, rhinorrhea, wheeze and dyspnea when present in isolation without other organ complaints are rarely ever indicative of food allergy. When due to food allergy, these airway symptoms are usually present as part of a generalized anaphylactic reaction. Nevertheless, it has been shown that a very small subset of patients (particularly children) with moderate to severe asthma may have had food allergy triggering the disease.

Oral allergy syndrome

This is a form of contact allergy to food allergens, usually a fresh fruit or vegetable, in contact with the oral mucosa. Symptoms include pruritus with or without angioedema of the lips, tongue, palate and posterior oropharynx. These symptoms resolve rapidly and rarely involve any other organs. Characteristically, the involved food is only an allergen source when fresh as the proteins involved are heat-labile. For example, a patient may not tolerate eating raw apple but could consume apple pie safely. Oral

Table 1 Relationship between Pollen and Food in Oral Allergy Syndrome

Pollen	Food
Ragweed	Melon, banana, cucumber
Birch	Apple, hazelnut, carrot, stone fruits (e.g. apricot, cherry, plum)
Mugwort	Celery, carrot, certain spices

allergy syndrome is associated with specific pollen allergy in individuals with seasonal allergic rhinitis. Studies have shown cross-reactive proteins present in the fruit or vegetable and common pollens (Table 1). It is interesting to note there have been case reports of oral allergy syndrome abating after a course of immunotherapy for coexistent pollen allergy.

Food-related exercise-induced anaphylaxis

This reaction occurs only if the susceptible individual exercises within 2 to 4 hours of food ingestion. With food alone or exercise alone, there is no adverse reaction. It may occur with any type of food in some patients but only specific types of food in others. A skin prick test to the incriminated food is positive but symptoms are only reproduced if food is ingested together with exercise. Such episodes are best prevented with avoidance of food intake 4 hours before or after exercise.

Non-IgE Mediated Reactions

Food-induced enterocolitis

This affects infants between 1 week and 3 months old. The main symptoms are protracted vomiting and diarrhea, which frequently results in dehydration. The infant may appear ill and septic. Cow milk and soy protein are most often implicated. Symptoms usually resolve after 72 hours of allergen avoidance. Rechallenge is characterized by recurrence of symptoms within 1 to 6 hours associated with leukocytosis and heme-positive and leukocytes-positive stools. Jejunal biopsy reveals villous atrophy, edema, and increased lymphocytes, eosinophils and mast cells. There is absence of specific IgE antibodies to the allergenic food. Infants usually outgrow this allergy. Similar but less severe reactions have been reported in adults.

Food-induced proctocolitis

This is similar to enterocolitis but symptoms are milder and involvement is limited to the colon and rectum. The typical clinical picture is that of a well-appearing infant with blood-streaked or bloody stools. Endoscopic findings vary from injected mucosa to aphthous ulcers. Biopsies show eosinophilic infiltrates in the epithelium and lamina propria. Symptoms resolved within 72 hours after allergen avoidance. The infant tends to develop tolerance when they are older. Cow milk and soy are the most commonly implicated foods.

Food-induced enteropathy

The age of onset is in the first few months of life. The clinical features of this malabsorption disorder are anorexia, vomiting, protracted diarrhea and failure to thrive. Cow milk is the most frequent cause but soy, egg, wheat, rice, chicken and fish have been implicated. Endoscopy reveals areas of villous atrophy interspersed with areas of normal mucosa (patchy enteropathy). Jeujunal biopsy shows a predominant mononuclear cell infiltrate in the epithelium and lamina propria. Resolution of symptoms after elimination of allergen is slow and resolution of intestinal lesions may take 6–18 months. Rechallenge brings on symptoms only several days to weeks later. Patients tend to outgrow their food allergy.

Celiac disease

Celiac disease, also known as gluten enteropathy, produces a more extensive enteropathy as compared to food-induced enteropathy. The sensitivity is to gliadin, the alcohol-soluble portion of gluten. Gluten is found in wheat, oat, rye and barley. Symptoms are similar to that of malabsorption. Celiac disease may be associated with dermatitis herpetiformis, abdominal lymphoma, type I diabetes mellitus, selective IgA deficiency, hyposplenism, hypo- and hyperthyroidism, myasthenia, sarcoidosis and rare cerebellar atrophy. Serological markers include IgA antibodies to reticulin, smooth muscle endomysium and circulating IgG and IgA antibodies to gliadin. Titers of these antibodies decrease or disappear after gluten elimination and therefore can be used to follow response to treatment or monitor compliance. Jejunal biopsy classically shows villous atrophy, lymphocytic and plasma cells infiltration of the lamina propria with a

predominance of IgA producing cells. Resolution of villous atrophy on repeat biopsy following gluten-free diet is diagnostic. A lifelong avoidance of gluten containing food is necessary to control symptoms and avoid the increased risk of malignancy.

Allergic eosinophilic gastroenteritis

These patients have multiple food allergies due to IgE or non-IgE mediated mechanisms. Symptoms include postprandial nausea and vomiting, abdominal pain and diarrhea, gastroesophageal reflux, early satiety or refusal to eat, weight loss in adults and growth failure in children. This entity is characterized by eosinophils infiltrating the esophageal, gastric or intestinal walls, thus sporadic and therefore multiple biopsies might be needed as lesions could be missed on single biopsy. There may be presence of peripheral blood eosinophilia, elevated total IgE, and positive SPT to food. These symptoms usually subside within 3 to 6 weeks of allergen elimination although gut histology may not return to normal for months.

Heiner's syndrome

Heiner's syndrome, also called Wilson Heiner–Lahey syndrome, is a rare form of primary pulmonary hemosiderosis associated with cow milk allergy. It presents in infants or toddlers. The syndrome comprises occult rectal blood loss with hypochromic microcytic anemia, hypoproteinemia and recurrent pulmonary infiltrates. Lung biopsy shows hemosiderin-laden macrophages and IgG, IgA and C3 deposit. Skin prick test to cow milk is usually positive and patients may have high titers of precipitins to cow's milk protein.

Diagnosis

The diagnosis of food allergy is based on a thorough history, physical examination, and diagnostic tests including food challenge. The double-blind placebo-controlled food challenge (DBPCFC) is the gold standard for diagnosis, except in the patient with anaphylactic reactions where rechallenge is potentially fatal.

History

History should focus on age of onset, types of symptoms, details of the meal associated with the reaction, quantity of the suspected food

ingested, length of time between food ingestion and development of symptoms, activities at time of or prior to symptoms, e.g. exercise, and length of time since the last reaction. A family and personal history of atopic disorders should be obtained as food allergy often occurs in this group of individuals.

In the clinical setting, many patients are not able to recall the details of the food ingested. Diet diaries are a useful adjunct to the medical history in these circumstances. Patients are advised to record all foods eaten including anything that is just placed in the mouth (e.g. chewing gum, mouthwash, toothpaste), any symptoms experienced and length of time from ingestion to symptoms. Patients who experience problems infrequently, should record these for the period at least 24 hours prior to onset of symptoms immediately after the event. Patients with frequent symptoms should keep the diet diary on a prospective basis until the problem is resolved. As Asian cooking involves the use of many spices and sauces, it is important that these are recorded down as it may be the seasoning that patient is allergic to and not the food itself. Through such diaries, a previously unrecognized association between a food and symptoms may become evident.

Physical examination

Physical examination is often unremarkable in between reactions. However particular attention should be paid to signs of atopic disorders and nutritional status. In children, the height and weight percentiles must be monitored as these may be indicators of failure to thrive. During an allergic reaction, signs and symptoms are dependent on the mechanism of the allergy and may range from localized discomfort to anaphylaxis.

Investigation

The investigation to be ordered depends on the type of hypersensitivity reaction. If an IgE-mediated disorder is likely based on the history, skin prick test (SPT) or *in vitro* measurement of food-specific IgE such as the Pharmacia CAP may be ordered. SPT is a simple and reliable test that provides useful information when done by trained personnel and interpreted appropriately. The incidence of false-negative results is low and the negative predictive value is higher than 95% if reliable extracts are used. The positive predictive value of SPT is, however, significantly lower

and ranges from 60% in individuals in whom the prevalence of food allergy is fairly high to 3% in those in whom the prevalence is low and in whom there is no suggestive food allergy history. Therefore, positive skin tests to food must be interpreted cautiously. In a local study, 4.9% of healthy adults were tested positive to shellfish mix but have no problems when ingesting these foods. Intradermal skin testing (ID test) is not carried out in the evaluation of food allergy because of the higher risk of anaphylaxis. In addition, DBPCFC have shown that when ID test is positive and SPT is negative, the positve ID test has no clinical relevance.

Serum food-specific IgE measurement is ordered if the patient has contraindications to skin testing, such as significant dermographism, severe skin disease, severe anaphylaxis to food or inability to discontinue drugs that interfere with skin testing (e.g. antihistamine). These tests are often less sensitive than SPT, more expensive and results are delayed by several days. The results of such tests should be interpreted in a similar manner as SPT.

Positive results on skin prick or *in vitro* IgE testing should be confirmed with a trial of elimination diet followed by oral challenge except where contraindicated. In patients with a consistent history of anaphylaxis following ingestion of a food, a positive skin test to that food may be considered diagnostic and challenge is never done because of the high risk.

If a non-IgE mediated food allergy is suspected based on the history, the following tests may be useful in the appropriate patient: full blood count and differential count, stools for occult blood, malabsorption studies, anti-gliadin antibodies in suspected celiac disease, or a referral for endoscopy and biopsy.

Elimination diets are therapeutic trials and may be considered in atopic dermatitis (following SPT or *in vitro* test for food-specific IgE) or when history suggests food allergy but a specific food is not yet identified or history suggests food intolerance. These therapeutic trials should be used only for a limited period of time (10–14 days). The outcome should be monitored closely and overzealous elimination must be avoided as these could lead to nutritional deficiencies and affect the quality of life. There are 3 types of elimination diet namely: basic elimination diet, targeted elimination diet, and severe elimination diet. Elimination diets should be followed by a confirmation of allergy to the implicated food using open or preferably double-blind placebo-controlled food challenge and appropriate counseling.

Food challenge may be carried out as an open challenge, single or double-blind with or without a placebo and should only be carried out by trained personnel. DBPCFC is the gold standard for the definitive diagnosis of food allergy. DBPCFC, as with other types of food challenges, is administered in the fasting state. All medications that may interfere with interpretation of symptoms should be discontinued one week prior to the test. In DBPCFC, lyophilized food is often used and it is blinded to the patient and the medical personnel as either a capsule or liquid. If fresh food is used it has to be blended and masked. Since DBPCFC is time-consuming, open or single-blinded oral challenges are sometimes used in the clinical setting when appropriate. Food challenge is contraindicated in those with a history of anaphylaxis following ingestion of food. Food challenges are helpful in conditions such as atopic dermatitis, non-IgE mediated food allergy and food intolerance, in children in whom specific foods may be difficult to implicate, in interpreting positive skin tests or *in vitro* specific IgE tests that appear not to correlate with the patient's history. All blinded negative challenges must be confirmed by an open feeding under observation to rule out a false-negative reaction, which may be secondary to the food processing.

MANAGEMENT

The management of food allergy includes pharmacologic treatment in the event of a reaction, avoidance of the allergenic food to prevent future reactions, education and counseling. At present, immunotherapy for food allergy is still considered experimental.

Pharmacologic treatment

Epinephrine is the treatment of choice for severe type I (IgE-mediated) reactions to food, e.g. airway compromise or anaphylaxis. The adult dose of epinephrine is 0.2–0.5 ml of a 1:1000 dilution (0.2–0.5 mg) intramuscularly or subcutaneously. For serious reactions, intramuscular route is preferred as the drug is more rapidly absorbed. This dose may be repeated every 10–15 minutes for the first hour. If the patient remains hypotensive, plasma volume expanders have to be considered. The dose of epinephrine for a child is 0.01 mg/kg up to a maximum of 0.3 mg or 0.3 ml of 1:1000 dilution given intramuscularly. This dose may be repeated every 15 minutes for up to 3 doses.

Patients on beta-blockers may remain hypotensive because of epinephrine resistance. This may be countered with glucagon injection. Antihistamines and corticosteroids may be given as adjunctive treatment. All patients who have anaphylaxis should be observed for at least 24 hours because delayed, biphasic anaphylaxis may occur.

Food avoidance measures, education, and counseling

Strict avoidance of the food to which a patient is allergic is the only proven therapy for food allergy. Patients, family members and in the case of pediatric patients, their childcare providers and school personnel, should be educated on food avoidance measures, allergen identification including potential hidden food sources, symptom recognition and what to do in the event of accidental ingestion. A treatment plan must be prepared. Providing a list of "allowed foods" may not be beneficial as ingredients frequently change. Instead, patients should be encouraged to learn to read food labels. A child with food allergy must be instructed not to accept food from classmates or friends. The skills of administering epinephrine (prefilled syringes including autoinjectable preparations) should be taught using a dummy, and reinforced regularly to those with type I hypersensitivity reaction. As mentioned, reactions may be biphasic, thus all patients are advised to go to a hospital for extended observation even after they have successfully used the epinephrine injections for an episode of food-induced anaphylaxis. As patients and their families have to make various lifestyle modifications, referral to a support group may be beneficial. In the case of allergies to foods that may be outgrown (e.g. egg, wheat, milk), re-evaluation to confirm if allergy has been "lost" could be done annually.

UNPROVEN METHODS IN DIAGNOSIS AND TREATMENT

Several tests and treatment methods proven to be of no value in food allergy are being used by some practitioners of alternative medicine as well as promoted via lay literature and the internet. Some examples are shown in Table 2. Doctors should be familiar with these techniques so that they can advise their patients appropriately. When in doubt, always refer to a clinical immunologist/allergist.

Table 2 Unproven Techniques of Diagnosis and Therapy of Food Allergy

1) Cytotoxicity testing
2) Provocative testing and neutralization treatment
3) Measurement of immune complexes, serum IgG, and IgG4
4) Electrodermal testing
5) Applied kinesiology
6) Dietary manipulation: Masking and addition, rotary diet, fasting

The production of serum IgG and IgA to food that we eat is normal, thus, the presence of these antibodies to food does not indicate allergy. Normal individuals have also been shown to have circulating immune complexes to commonly ingested foods. These have not been proven to be pathological. Provocation-neutralization test is used by some practitioners to diagnose and treat non-IgE mediated food "allergy". In this procedure, food extracts of various dilutions are either injected intradermally or given sublingually and subjective symptoms over the next 10–20 minutes are recorded. If a symptom has occurred then a different dilution of the food extract is given until the symptom is gone. The "neutralizing dose" is the dose of food extract given that correlated with resolution of symptom. Patients are then instructed to take the neutralizing dose sublingually or by injection prophylactically before or after exposure on a regular basis. A double-blind trial has shown this technique to be ineffective in the diagnosis of food allergy.

SUGGESTED READING

1. Bruijnzeel–Koomen C, Ortolani C, Aas K, *et al*., Adverse reactions to food: position paper, *Allergy*, **50**:623–635, 1995.
2. Sampson HA, Food Allergy. Part I: Immunopathogenesis and clinical disorders, *J Allergy Clin Immunol* **103**:716–728, 1999.
3. Sampson HA, Food Allergy. Part II: Diagnosis and management, *J Allergy Clin Immunol* **103**:981–989, 1999.
4. Metcalfe DD, Sampson HA, Simon RA (eds.), *Food Allergy: Adverse Reactions to Foods and Food Additives*, 2nd edition, Blackwell Science, Cambridge, MA, 1996.
5. Thong BY, Tang CY, Leong KP, Chng HH, Food allergy in adults: a review of 86 local cases with type I hypersensitivity reactions, in *Proceedings of the 1st NHG Scientific Congress*, Singapore, pp. 133, 2002.

6. Shek LP, Lee BW, Food allergy in children — the Singapore story, *Asian Pacific J of Allergy Immunol* **17**:203–206, 1999.

7. Chng HH, Tang CY, Leong KP, Healthy adults demonstrate less skin reactivity to commercial extracts of commonly ingested food than to D. farinae, *Asian Pacific J of Allergy Immunol* **17**:175–178, 1999.

43

Immunodeficiency in the Adult

Leong Khai Pang and Chng Hiok Hee

INTRODUCTION

Immunodeficiency may be primary or secondary to another disorder such as infection with the human immunodeficiency virus (HIV/AIDS). Primary immunodeficiency disorders are rare, with a prevalence ranging from 1:75 000 to 1:500 000 if we exclude selective IgA deficiency. More than half of the primary immunodeficiency disorders are diagnosed in children under 15 years old and many do not survive to adulthood without early diagnosis, prompt treatment of infections and life-saving measures like an allogeneic bone marrow transplant. In this review, we concentrate on the primary immunodeficiency disorders in adults, although we shall also mention immunodeficiencies in childhood and secondary immunodeficiencies.

The conventional method of classifying these conditions is according to the component of the immune system that is affected. Accordingly, antibody deficiencies make up 50–60% of all cases, combined cellular and antibody deficiencies 20–25%, phagocytic deficiencies 15–20%, cellular deficiencies 5–10% and complement deficiencies about 2%. The primary

Table 1 Molecular Defects that Produce Immunodeficiency

DNA repair
Ataxia telangiectasia (ATM mutation)
B-lymphocyte development
XLA (Btk mutation)
Hyper-IgM syndrome (mutations in genes for CD154 (X-linked) or CD40 (autosomal)
T-lymphocyte development
X-linked SCID (mutations of γc gene)
JAK 3 deficiency
Immunoglobulin gene rearrangement (RAG-1 or RAG-2 deficiency)
Adenosine deaminase deficiency
Artemis mutation
T cell signaling
Bare lymphocyte syndrome (defective expression of MHC class I molecules due to TAP1 and TAP2 mutations)
Defective expression of MHC class II molecules (due to mutations at 4 possible genes: RFX5, RFXAP gene mutation, RFXANK and a MHC class II transactivator)
ZAP-70 deficiency
IFN-γ receptor deficiency
Adhesion molecules
Leucocyte adhesion deficiency (CD18 gene mutation)
Phagocytic function
Chronic granulomatous disease (mutations in the genes encoding components of the cytochrome b245 heterodimer)

Btk = Bruton's tyrosine kinase; IFN-γ = interferon-gamma; RAG = recombination activation group; scid = severe combined immunodeficiency; XLA = Bruton's X-linked agammaglobulinemia; ZAP = zeta-associated protein.

immunodeficiency disorders that present in adulthood include the antibody deficiencies (except Bruton's X-linked agammaglobulinemia [XLA]), chronic granulomatous disease (CGD) and complement deficiencies.

With advances in molecular techniques, we now know the precise genetic cause of many primary immunodeficiencies. Not all patients with identical clinical and laboratory findings have mutations in the same gene. Therefore, an alternative classification is based on the molecular defects of immune function (Table 1). This scheme has some limitations because the causes of a few types of primary immunodeficiency disorders have not been elucidated yet. At least 7 genes associated with

immunodeficiency are found on the X chromosome. They are those related to certain forms of CGD, properdin deficiency, Wiskott–Aldrich syndrome (WAS), severe combined immunodeficiency (SCID), XLA, lymphoproliferative syndrome (XLP) and hyper-IgM syndrome (HIM).

PRIMARY ANTIBODY DEFICIENCIES

Antibody-deficient patients are susceptible to encapsulated bacteria such as *Streptococcus pneumoniae* and *Haemophilus influenzae*. They present with recurrent sinopulmonary bacterial infections and sepsis. These patients generally have no problems with fungi or viruses because cell-mediated immunity is often preserved. An exception is chronic encephalomyelitis associated with enterovirus infection in patients with XLA.

Bruton's X-linked Agammaglobulinemia

In 1952, Colonel Ogden Bruton described the first case of primary immunodeficiency. The patient was an 8-year-old boy who had repeated infections since the age of $4\frac{1}{2}$ years. Remarkably, in that paper, Bruton described the clinical manifestations of the disease, illumined the defect (deficiency of the gamma globulin portion of serum protein) and established the treatment (parenteral injection of gamma globulin). The molecular basis for XLA is a mutation in the gene for Bruton's tyrosine kinase (Btk). Btk is involved in cellular signaling. XLA patients have marked deficiency or complete absence of all 5 immunoglobulin classes.

Infants with this disorder become symptomatic from about 7–9 months of age following the disappearance of transplacentally-acquired maternal immunoglobulins. Presenting complaints include recurrent bacterial otitis media, sinusitis, bronchitis and pneumonia, pyoderma and gastroenteritis. While most viral and fungal organisms are not a problem, patients with XLA are susceptible to viral hepatitis, disseminated polio, and chronic enteroviral encephalitis infections. Some patients also develop arthritis, a dermatomyositis-like syndrome or malabsorption. Tonsils and adenoids are typically absent. Despite repeated infections, there is also an absence of lymphadenopathy and splenomegaly. The established treatment is regular intravenous immunoglobulin (IVIG) infusion.

Common Variable Immune Deficiency

The age of presentation of CVID is typically in the second or third decade although children and older adults may be affected. Unlike XLA, it is not sex-linked; patients may have marked lymphadenopathy and splenomegaly, and mature B cells are found in the peripheral blood. There is marked decrease in serum IgG and IgA while the concentration of serum IgM is normal in about half of the patients. CVID patients also have a higher than normal incidence of T-cell abnormality, which often progressively deteriorates with time.

CVID is a heterogeneous group of diseases with a common phenotypic expression, and no single genetic defect has been identified. The majority of patients suffer from chronic pyogenic sinopulmonary infections and the predominant pathogen is *Hemophilus influenzae*. Intestinal pathogens like *Giardia lamblia, Yersinia*, and *Campylobacter jejuni* can also be found. Other unusual infections include *Mycoplasma hominis*, and *Ureaplasma urealyticum* and these are often associated with arthritis. Over time, CVID sufferers are also susceptible to fungal and viral infections. Patients have an increased incidence of atrophic gastritis, autoimmune disease, lymphoma and gastric carcinoma.

Like XLA, CVID is treated with regular IVIG infusions. Timely and adequate immunoglobulin replacement is effective in preventing chronic complications such as bronchiectasis. CVID patients who have absent serum IgA may have anti-IgA antibodies and develop an allergic reaction (including anaphylaxis) when exposed to IVIG.

Selective IgA Deficiency

Selective IgA deficiency is the commonest primary immunodeficiency, with prevalence rates of 1:400 to 1:3000. It is defined as a serum IgA concentration of less than 7 mg/dL with normal serum IgM and IgG levels. Like CVID, the genetic basis for selective IgA deficiency is unknown. Most people with selective IgA deficiency are asymptomatic while others have recurrent infections (sinopulmonary, occasionally meningitis). There is an increased association with allergies, autoimmunity (e.g. systemic lupus erythematosus, rheumatoid arthritis, Sjogren's syndrome, organ-specific autoimmune conditions like insulin-dependent diabetes mellitus, pernicious anemia and chronic active hepatitis) and gastrointestinal tract

diseases such as celiac disease, ulcerative colitis and regional enteritis. An IgG2 subclass deficiency has been described in some individuals with selective IgA deficiency. Cell-mediated immunity is normal. Some patients with selective IgA deficiency develop CVID over time.

Pooled immunoglobulin should not be given to patients with selective IgA deficiency as they are capable of forming normal amounts of other immunoglobulins. IgA-deficient individuals with very low levels of serum IgA may have anti-IgA antibodies and thus may develop an anaphylactic reaction upon infusion of IVIG. Transfusion reactions due to anti-IgA antibodies may be minimized by using washed red cells or prestored autologous transfusions.

Hyper-IgM Syndrome

Patients with HIM have normal or high IgM levels with low or absent IgG, IgA and IgE. It is due to a deficiency of CD154 (previously called the CD40 ligand). The lack of interaction between CD154 on the surface of activated T cells and CD40 on B cells prevents the latter from producing immunoglobulins of classes other than IgM. HIM is largely an X-linked disease but autosomal dominant and recessive forms are recognized. A mutation in the gene that encodes activation-induced cytidine deaminase (AID) is responsible for the autosomal variant. Individuals with HIM present with recurrent pyogenic infections usually from childhood. They are susceptible to *Pneumocystis carinii*, cryptosporidial and fungal infections because of subtle defects in T cell responsiveness.

Selective IgG Subclass Deficiency

When there is recurrent pyogenic sinopulmonary infections in an individual with normal quantitative serum immunoglobulin, low levels of IgG subclasses may be causative. IgG subclass deficiency is defined as a serum IgG subclass level that is more than 2SD below normal for age. The reference range for serum levels of IgG1, IgG2, IgG3 and IgG4 has not been determined for all populations, including that of Singapore. In adults, deficiency in IgG3 subclass is most common while in children, IgG2 is the most prevalent. IgA deficiency and ataxia telangiectasia may accompany IgG subclass deficiencies. It should be noted that healthy individuals without recurrent infections can have low serum IgG subclass levels.

PRIMARY CELLULAR DEFICIENCIES

Primary immunodeficiency disorders associated with isolated defective T cells are rare. In most patients, defective T cell immunity is associated with defective B cell immunity as well. Patients with cellular immune deficiencies often have onset of symptoms in early infancy. Manifestations include recalcitrant thrush, recurrent infections with fungal, viral and protozoal pathogens. Fatal infections occur following live virus and BCG vaccinations. Transfusion of blood products can lead to graft-versus-host disease.

DiGeorge Syndrome

When an embryonic tissue that is supposed to give rise to many specific tissues (a "field") does not develop normally, a polytypic field defect is said to have occurred. DiGeorge syndrome is an example of this, in which the embryonic cephalic neurocrest cells are defective. It is characterized by dysgenesis of the parathyroid glands, heart (right-sided aortic arch), thymus and face, which arise from the third and fourth branchial arches. Deletions of chromosome 22 are found in 90% of cases of DiGeorge syndrome. The commonest presentation is heart failure or persistent hypocalcemic tetany in infancy rather than immunodeficiency. Patients may have incomplete forms of this syndrome without severe immune deficiency. Those with the complete form hardly possess any T cells and do not show spontaneous improvement in cell-mediated immunity. Cultured thymic epithelial explants have been successfully transplanted to infants with complete DiGeorge syndrome.

Chronic Mucocutaneous Candidiasis

This disorder is due to a selective defect in T cell immunity resulting in susceptibility to chronic candidal infection. B cell immunity is intact. Candidal infections involve the mucous membranes, skin, nails and, in older patients, the vagina. Generally there is no susceptibility to systemic candidal infections. Patients may have associated idiopathic endocrinopathy such as hypoparathyroidism, Addison's disease, hypothyroidism, diabetes mellitus or pernicious anemia. Endocrinopathy may be the first presentation. This disorder presents as early as 1 year of age or as late as the second decade.

PRIMARY COMBINED CELLULAR AND ANTIBODY DEFICIENCIES

Combined immunodeficiency disorders have various causes and are of variable severity. The defects in T cell and B cell immunity may be complete as in severe combined immune deficiency (SCID) or partial as in ataxia telangiectasia. Patients are susceptible to a wide spectrum of organisms, viral, bacterial, fungal and protozoal. The onset of symptoms is usually in infancy.

Severe Combined Immunodeficiency

About half of all cases of SCID are X-linked and half are autosomal recessive, producing the observed patient sex ratio of three males to one female. X-linked SCID (XLSCID) is due to mutations in the gene encoding the cytokine common gamma chain (γc), which is a protein that is found in the receptors for IL-2, lL-4, IL-7 and IL-15. The resultant loss of function of these crucial cytokines, particularly IL-2, leads to the severe immunodeficiency.

Patients present in the first 2–7 months of life. They are susceptible to bacteria, viruses, fungi and parasites and most perish from sepsis within the first two years. *Pneumocystis carinii* pneumonia and disseminated BCG infections are common presentations in addition to chronic diarrhea, respiratory symptoms and failure to thrive. The standard treatment for many years has been an extremely restrictive isolation and bone marrow or stem cell transplant, even during the *in utero* stage, but the immune reconstitution is often incomplete. Gene therapy (transduction of marrow stem cells with the γc transgene) has shown excellent results, but this program has been halted because two recipient out of eleven developed leukemia.

About 50% of all cases of autosomal recessive SCID is due to adenosine deaminase (ADA) deficiency. The accumulation of toxic metabolites due to the lack of this enzyme causes apoptosis of developing and circulating lymphoid cells. These patients also manifest cartilage malformation of the ribs and pelvis. Other genetic mutations that produce SCID occur in Artemis and the genes for Janus kinase 3 (JAK3), RAG1 or RAG2, CD45 and IL-7 receptor α chain.

Wiskott–Aldrich Syndrome

Wiskott–Aldrich syndrome (WAS) is a X-linked disease characterized by eczema, recurrent pyogenic infections and severe congenital thrombocytopenia with small platelets. The WAS protein (WASP), involved in cytoskeleton integrity, is defective in this condition. Many patients present with bloody diarrhea in the first month of life. Patients are prone to infections by encapsulated bacteria, *Pneumocystic carinii* and herpes virus. There is a predilection for lymphoid malignancy and leukemia. The prognosis is dismal as 20% develop cancer, 20% die of hemorrhage and 50% die of infections by their teenage years. A bone marrow transplant is the best treatment.

Ataxia-telangiectasia

This autosomal-recessive condition consists of the constellation of progressive cerebellar ataxia, oculocutaneous telangiectasia, predisposition to lymphoid malignancies and combined cellular and antibody deficiencies. Selective IgA deficiency is common. Neurologic manifestations precede the immune deficiency, as ataxia is apparent when the child begins to walk. Endocrine problems such as hypogonadism and glucose intolerance may be found. Patients develop bacterial sinopulmonary infections and occasionally mycobacterial infections. Some patients may not be diagnosed until the second decade of life. The gene of ATM, involved in cell cycle checkpoint pathway, is mutated in patients with this disease and this explains the association with malignancy.

Immunodeficiency with Thymoma

The association of hypogammaglobulinemia with thymoma usually occurs relatively late in adult life. There may be associated T cell defect. Patients have recurrent pyogenic infections, severe diarrhea and infrequently fungal and viral infections.

X-linked Lymphoproliferative Syndrome (XLP)

Male patients are usually asymptomatic until they develop EBV infections. In children, this results in fulminant hepatitis while older patients

may develop Hodgkin's or non-Hodgkin's lymphoma and/or immunodeficiency with low serum IgG and abnormal NK cell function. Less commonly, patients develop EBV-associated hemophagocytic syndrome and vascultis.

PRIMARY PHAGOCYTIC DEFICIENCY

Patients with phagocytic defects are susceptible to gram-negative and catalase-positive bacteria including *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Aspergillus* or other fungal infections. Susceptibility to infections range from mild recurrent skin abscesses to severe, fatal systemic infections.

Chronic Granulomatous Disease

Chronic granulomatous disease (CGD) is a rare and heterogeneous group of diseases distinguished by the inability to generate reactive oxygen radicals, so that neutrophils and macrophages cannot effectively kill the bacteria and fungi they have phagocytosed. These patients have recurrent deep-seated infections (bacterial and fungal) such as abscesses in the lungs, liver and other solid organs that are characterized by exuberant granuloma formation. Granuloma formation may result in symptoms of intestinal and urinary tract obstruction. The main infective organisms are *Staphylococcus aureus*, *Burkholderia cepacia*, *Serratia marcescens*, *Nocardia* species and *Aspergillus* species.

Genes that encode one of the four components of phagocyte NADPH oxidase (gp91^{phox}, gp22^{phox}, gp47^{phox} and gp67^{phox}) are mutated in CGD. Defects of the gp91^{phox} gene are found in two-thirds of CGD patients, while gp22^{phox} mutations occur in most of the rest. Gp91^{phox} is found on the X chromosome while the rest are autosomal. Those with the X-linked form of CGD tend to have more severe disease and present earlier (within first year of life) than the autosomal forms. Rarely, CGD may present first in adulthood.

Prophylactic treatment with cotrimoxazole has reduced the frequency of infections by half and interferon-γ by 70%. Bone marrow transplant and gene therapy are potential cures for this immunodeficiency disorder.

Hyper-IgE Syndrome

The hyper-IgE (HIE), or Job's syndrome, is a rare autosomal dominant disorder characterized by recurrent abscess-forming infections of the skin and lower respiratory tract commencing in infancy or early childhood with persistent eczematoid rashes and markedly elevated serum IgE levels. The main problem lies with the phagocytes, not the IgE antibodies. Other features of HIE are delayed shedding of primary dentition, fractures, hypermobility, scoliosis and coarse facies. The genetic basis is unknown.

Chediak–Higashi Syndrome

This rare condition is distinguished by giant lysosomal inclusions in the neutrophils. Patients have recurrent bacterial infections, hepatosplenomegaly, central nervous system abnormalities and partial albinism.

Leucocyte Adhesion Defect

Leucocytes require adhesion molecules to direct them to sites of inflammation. An important pair consists of lymphocyte function-associated antigen (LFA-1) on the surface of leucocytes and intercellular adhesion molecule 1 (ICAM-1) on endothelial surfaces. LFA-1 is composed of CD18 and CD11. When either of these subunits is defective, the cells lose their homing signal and continue to circulate in high numbers. Patients have marked leucocytosis (wbc $> 25\,000/\text{mm}^3$). There is susceptibility to bacterial (staphylococcus and gram-negative enteric bacteria) and fungal infections. Individuals also demonstrate delayed separation of the umbilical cord, periodontitis, poor wound healing and absence of pus at sites of infection. In severe forms, without BMT, death occurs early. In moderate forms where a small amount of CD18 is expressed, patients may survive to adulthood.

PRIMARY COMPLEMENT DEFICIENCY

The complement system amplifies the power of the antibodies. When the bacterial surface antigens interact with the relevant antibodies in the host (a process called opsonization), the complement system is activated. This

eventually leads to the formation of channels that cause cell lysis and death. Deficiencies of all complement components have been described. Complement C3 deficiency is associated with severe pyogenic infections, glomerulonephritis and SLE. Deficiency of properdin and the complement terminal cascade (C5 to C9, the membrane-attack complex) render the patient susceptible to *Neisseria meningitidis* infection.

Paradoxically, deficiencies of the early components of the complement system (C1q, C1r, C1s, C4, C2 and C3) are associated with the expression of an SLE-like disease. The deficiency of early components of complement (specifically C3b) prevents immune complexes from forming a lattice network and impairs their clearance by the reticule-endothelial system. The excess circulating immune complexes mediate the development of lupus, vasculitis and glomerulonephritis. The features of SLE in complement deficiency tend to be atypical and to be limited to the skin. C1 inhibitor deficiency is associated with hereditary angioedema.

SECONDARY IMMUNODEFICIENCY

Many illnesses result in a defective immune system that can be exploited by microorganisms. These include malnutrition, diabetes mellitus, chronic renal failure, cirrhosis, HIV/AIDS and malignancies. Medications such as corticosteroids and anti-neoplastic agents can cause suppression of the immune system. Before diagnosing primary immunodeficiency in an adult, it is important to exclude these common conditions first.

We should also distinguish those conditions that involve breaches of local defence without an overall loss of immunity. These include bronchiectasis, chronic stasis ulcers, base-of-skull fractures and prosthetic valvular or articular implants. Nasal carriers of *Staphylococcus aureus* who are immunocompetent can develop repeated episodes of skin infections with the same organisms. Many local problems are amenable to treatment.

WHEN TO SUSPECT IMMUNODEFICIENCY

A high index of suspicion is needed to diagnose immunodeficiency, particularly the milder forms that present in adulthood. In children with normal immune system, 6–8 episodes of respiratory tract infection per year is

not unusual if there is close contact with other children such as in a day-care center or older siblings in school. It is also normal for children to have up to 6 episodes of otitis and 2 episodes of gastroenteritis per year for the first 2–3 years of life. Repeated pharyngitis or urinary tract infections in adults do not signify the presence of an immunodeficiency. One group of researchers suggested investigating for immunodeficiency when there are more than 8 episodes of otitis media, 2 of serious sinusitis or 2 of pneumonia per year.

A complete history on the previous episodes of infections is important. This should include the age of onset, frequency and sites of previous infections, the infective organisms, duration of hospitalization, type and duration of therapy and response. A history of recurrent bacteremia or abscesses of the skin or solid organs, infections that do not respond to appropriate antibiotics or isolation of opportunistic organisms (e.g. *Pneumocystis carinii, Serratia marcescen, Aspergillus*) suggest an abnormal host immune system. A significant clue to the seriousness of infections in children is failure to thrive or not keeping up with the expected gains in weight, height and developmental skills. All these information are clues to a defective immune system as well as the specific immune defect (Table 2). A history of delayed umbilical cord detachment and adverse effects to vaccines or transfusions also provide clues to the nature of defect. Family history particularly of affected males is helpful in assessing the pattern of inheritance and may suggest the precise diagnosis.

A complete physical examination is required. During clinical examination, look for evidence of chronic infection and features that distinguish certain types of immunodeficiency. Patients with WAS have petechiae due to the thrombocytopenia and eczema. An immunodeficient

Table 2 When to Suspect Immunodeficiency

Unusually severe infection, e.g. pneumococcal pneumonia with empyema
Infection that persists for uncommonly long duration despite appropriate treatment
Recurrent infections above the norm
Unusual pathogens, e.g. aspergillus osteomyelitis, cryptosporidial infection
Sepsis or meningitis due to gonococcus or meningococcus
Staphylococcal liver abscess or recurrent staphylococcal adenitis
Absent tonsils and lymph nodes in patients with chronic otitis and/or sinusitis
Persistently enlarged lymph nodes and/or spleen

patient with ataxia and telangiectasia on the conjunctivae, the neck and limbs could be suffering from ataxia-telangiectasia. Patients with HIE have coarse facies and eczema. XLA patients do not have lymphoid tissue while patients with CGD and CVID may have generalized lymphadenopathy and splenomegaly. Mucocutaneous candidiasis indicate T cell defect. Chronic periodontitis and dental decay are seen in patients with neutrophil defects. Children should have their growth and developmental parameters charted.

INVESTIGATION OF IMMUNODEFICIENCY

According to the component of the immune system suspected to be defective, an initial screen is done with a full blood count (FBC), platelet count, peripheral blood film (PBF), serum quantitative immunoglobulins (IgG, IgA, IgM and IgE), nitroblue tetrazolium test (NBT) or total hemolytic complement (CH50). If the clinical suspicion of immunodeficiency is low and these screening tests are normal, there is no need for further investigation. If clinical suspicion is high, subsequent laboratory evaluation should be tailored to detect and confirm the suspected specific immune defect. Unfortunately, most of the specialized tests are not readily available.

The FBC must include the peripheral blood film (PBF) with leucocyte morphology, a differential white cell count, platelet count and the mean platelet volume. Look for presence of Howell–Jolly bodies on the PBF, which indicates hyposplenia or asplenia. Remember that children have higher lymphocyte counts than adults and normal adult levels in a child do not exclude immunodeficiency. The lymphocyte count should not be less than $2000/mm^3$ in newborns and $4000/mm^3$ in seven-month-olds.

A patient with recurrent sinusitis or pneumonia may have an antibody deficiency. Quantification of serum immunoglobulins (IgG, IgA and IgM) and isohemagglutinins (anti-A or anti-B antibodies) may indicate the defect. The concentrations of serum immunoglobulins must be interpreted according to the values appropriate to the patient's age. If there is strong suspicion of the presence of an antibody deficiency despite normal serum IgG level, serum IgG subclasses could be ordered. When serum IgG level is low, it is vital to perform a dynamic test to confirm that the rise in specific antibody level post-vaccination is inadequate. This test consists of measuring the antibody levels at baseline and two to three weeks after vaccination with an antigen such as unconjugated pneumococcal capsular

polysaccharide, *Haemophilus influenzae type b* (use only unconjugated vaccine), tetanus and diphtheria. When the antibody increase is four-fold or greater, the patient has normal humoral immune response. B cells can be measured with flow cytometry using the surface receptors CD19, CD20, light chains or surface immunoglobulin.

When cellular immune deficiency is suspected, the simplest test is to obtain the absolute lymphocyte count. In adults, a level below $1500/mm^3$ signifies lymphocytopenia. In a child, a lateral chest X-ray may reveal a small thymus in congenital T cell deficiency. Delayed cutaneous hypersensitivity (the Mantoux test is the prototypic example) requires the presence of functional T cells, and a series of recall antigens are used to assess this. At least 5 antigens (for example, mumps, tuberculin PPD, *Candida*, tetanus toxoid, diphtheria antigens) should be tested on because the patient may not have been sensitized to one or more of them in the past. If these preceding screening tests are abnormal, *in vitro* proliferative tests are performed with T cells cultured with recall antigens (tuberculin, *Candida*, tetanus, streptokinase or diphtheria), non-specific stimulators (phytohemagglutinin, pokeweed mitogen or concanavalin) and allogeneic cells. Enumeration of T cells using surface markers (CD3, CD4, CD8, TCR$\alpha\beta$, TCR$\gamma\delta$) should also be ordered. In selected patients, the ability to secrete cytokines may need to be evaluated. Tests for HIV infection should be considered in the high-risk individual.

Recurrent bacterial or fungal infections, skin or solid organ abscesses with granuloma formation, or periodontal abscesses suggest CGD. The normal neutrophil count is $1500/mm^3$. In CGD, neutrophil count is elevated in the presence of infection but normal when infection is cleared. The NBT is a good screening test for the oxidative burst of neutrophils. The chemiluminescence test is a more sensitive equivalent of the NBT where the oxidative burst is converted to light emission. An elaborate test found only in research facilities is the bactericidal assay, in which the serial percentage killing of bacteria incubated with the patient's leucocytes is measured over 2 hours. When the NBT is normal while the bactericidal test is impaired, we accept that there is a defect in phagocytosis. To obtain accurate results in these functional assays, blood must be sent fresh to the laboratory. In patients with recurrent skin abscesses, an elevated serum IgE is consistent with HIE.

Total hemolytic complement (CH50) is a useful screening test for the majority of complement disorders. A normal CH50 excludes complement

deficiency except for deficiencies of properdin or factor D which are detected by a different hemolytic assay (APH50). Normal C3 and C4 levels in the presence of undetectable CH50 is strong evidence of congenital complement component deficiency. On the other hand, a low C4 and/or low C3 with undetectable CH50 suggests complement consumption. As complement components lose activity very rapidly at room temperature, serum for CH50 assay needs special handling and confirmation of results is required before more elaborate tests are ordered.

Tests to exclude other disorders besides primary immunodeficiency may have to be considered. These include sweat chloride test in patients with recurrent sinopulmonary infections to rule out cystic fibrosis. The criteria and procedures for diagnosing immunodeficiency have been published by the IUIS Scientific Committee (see Further Reading).

TREATMENT OF IMMUNODEFICIENCY

The aims of therapy are to control any current infection, eradicate potential sites that microorganisms may still be harboring, treat subsequent infective episodes in an aggressive and timely manner, and correct the immunologic defect if possible. In addition, live bacterial or viral vaccines (BCG, vaccinia, and vaccines for polio, measles, rubella and mumps) should never be administered to the immunodeficient patient or their relatives. Patients with SCID must avoid blood transfusion as there is a risk of graft-versus-host disease from transfused lymphocytes, which cannot be eliminated when patients lack functional T cells of their own. If transfusion is necessary only irradiated blood products should be given.

For half a century, it is known that regular administration of immunoglobulin to patients with XLA or CVID is successful in preventing infections and their sequelae. Initially, the route of administration of immunoglobulin was intramuscular, but intravenous and subcutaneous deliveries are mainly used today. IVIG should not be given to patients with selective IgA deficiency. The initial dose of IVIG is 0.4 g/kg body weight every 3–4 times weekly, and both the dose and duration may be modified to maintain the IgG trough above 5 g/L. Rapid infusion of IVIG leads to fever, bronchospasm and hypotension. Rare side-effects of IVIG are renal impairment and aseptic meningitis. Modern methods of IVIG production process have reduced the risk of transmitting hepatitis C and other viral diseases. Besides these, the other problems with IVIG are cost and undersupply.

The cellular immunologic defects may be corrected fully only with bone marrow transplant or marrow stem cell gene therapy. Patients with CGD benefit from prophylaxis with cotrimoxazole or subcutaneous interferon gamma (IFNγ). However, the latter is expensive. Granulocyte transfusion and IFNγ may be used when there is severe infection. Bone marrow transplant has been reported to be efficacious in correcting the phagocytic deficiency and gene therapy is likely to play a role in the future in CGD. There is no specific treatment for the congenital deficiency of complement componenets. A purified C1 inhibitor preparation for treatment of HAE attacks is available in some specialized centers.

The management of primary immunodeficiency should include patient/parent education and counseling, the detection of carriers, and prenatal diagnosis when facilities are available.

CONCLUSIONS

The primary immunodeficiencies are uncommon diseases. Recurrent, serious or refractory infections or those associated with opportunistic pathogens are clues that an immunodeficiency is present. Secondary immunodeficiencies or local breaches in bodily defence are more common and should always be considered in the diagnostic process. Simple screening tests available to all clinicians can be use to exclude immunodeficiencies confidently in the majority of cases. Sophisticated tests of the various arms of the immune system are available to confirm a suspected deficiency. Therapy is now available for the majority of immunodeficiency disorders.

FURTHER READING

1. IUIS Scientific Committee, Primary immunodeficiency diseases, *Clin Exp Immunol* **118** (Suppl 1):1–28, 1999.
2. Ballow M, Primary immunodeficiency disorders: antibody deficiency, *J Allergy Clin Immunol* **109**:581–591, 2002.
3. Buckley RH, Primary cellular immunodeficiencies, *J Allergy Clin Immunol* **109**:747–757, 2002.
4. Fischer A, Primary immunodeficiency diseases: an experimental model for molecular medicine, *Lancet* **357**:1863–1869, 2001.

5. Segal BH, Leto TL, Gallin JI, Malech HL, Holland SM, Genetic, biochemical and clinical features of chronic granulomatous disease, *Medicine* **79**:170–200, 2000.

6. Walport MJ, Complement, *N Engl J Med* **344**:1058–1066, 1140–1144, 2002.

7. AAAAI, ACAAI and JCAAI Joint Task Force on Practice Parameters, Practice parameters for the diagnosis and management of immunodeficiency, *Ann Allergy Asthma Immunol* **76**:282–294, 1996.

This page intentionally left blank

Internal Medicine

This page intentionally left blank

44

Severe Acute Respiratory Syndrome: First Plague of the New Millennium

Hsu Li Yang, Tan Ban Hock and Leong Hoe Nam

INTRODUCTION

Severe acute repiratory syndrome (SARS) is an acute infectious respiratory illness caused by the SARS-associated coronavirus (SARS-CoV). It is the first major readily-transmissible disease of the 21st century. In the short 7 months of the epidemic, 8096 cases were reported, with 774 fatalities.[1] These numbers appear unimpressive compared with the millions who perished during the 1918 influenza pandemic, or the huge numbers afflicted by and dying of malaria and tuberculosis. However, when the economic and social toll is factored in, it is apparent that the SARS epidemic had an impact far greater than suggested by the number of infected persons. In some affected countries, public panic was widespread, government officials lost their jobs, and social stability was threatened. Schools, hospitals and markets were closed, and tourist industries practically ground to a halt.

ETIOLOGY

The realization that humankind was being attacked by a novel agent dawned rapidly on clinicians and researchers, and efforts to isolate a causative organism, believed to be a virus, intensified. These efforts were rewarded remarkably quickly when researchers in the virtual network of laboratories formed by the World Health Organization (WHO) isolated a novel coronavirus from specimens collected from SARS patients.[2–4] The complete RNA genome was sequenced by April 2003.[5] This new virus was provisionally named SARS-associated coronavirus (SARS-CoV). Its role as the primary pathogen of SARS was demonstrated by Thijs Kuiken and his team, who fulfilled Koch's postulates when they replicated the disease in four cynomolgus macaques (*Macaca fascicularis*) after infecting them with laboratory vero cell-cultured SARS-CoV.[6]

Coronaviruses are members of a family of large (100–140 nm), enveloped, positive-sense, single-stranded RNA viruses that replicate in the cytoplasm of animal cells. The name derives from their characteristic "sun-burst" morphology on electron microscopy: petal-shaped glycoprotein peplomers projecting from the surface of the spherical viral particles. Previously-identified human coronaviruses (OC43, OC38, 229E) had only been associated with mild respiratory diseases.[7] SARS-CoV is the first coronavirus that causes severe disease in humans.

Coronaviruses are some of the most important veterinary viral pathogens, hence the intense speculation that the SARS-CoV is of animal origin and made the jump from animal to man. Investigators have isolated a remarkably similar virus from both the masked palm civet cat (*Paguma larvata*) and the raccoon dog (*Nyctereutes procyonoides*) from live-animal markets of Guangdong.[8] This animal virus differed genetically from SARS-CoV by only an extra 29-nucleotide sequence. Some humans working at these markets had also tested positive for SARS-CoV antibodies, although none had been symptomatic.[8] It is theorized that the chance deletion of this RNA sequence had resulted in a more virulent form of the coronavirus (viz. SARS-CoV) which had then caused the epidemic. This view was further corroborated by a remarkable study by scientists in China, which traced the molecular evolution of the SARS-CoV from a genotype similar to the animal SARS-like coronaviruses during the initial outbreak in Guangdong to the stable genotype, which was prevalent in the subsequent worldwide epidemic.[9]

EPIDEMIOLOGY (WORLDWIDE)

The first cases of SARS occurred in Foshan city, Guangdong province, China, in mid-November 2002. By 11th February 2003, there were 305 affected persons, with 5 deaths.[10] The pathogen responsible for this "atypical pneumonia" was thought to be a member of the *Chlamydia* family at that time, as it had been isolated in 3 of the cases.

A physician who had been infected by his patients brought the disease out of Guangdong on 21st February 2003. Under circumstances that are still not fully understood, at least 14 guests and visitors to the 9th floor of the Metropole Hotel in Hong Kong — where the physician had stayed — were infected. These people subsequently seeded the outbreaks of SARS in Hanoi, Hong Kong, Singapore and Toronto.

By 15th March, more than 150 cases of this new disease had been reported, and epidemiological analysis indicated that it spread internationally via routes of air travel, and locally via predominantly nosocomial transmission. The WHO coined the term "severe acute respiratory syndrome" to describe this new disease and issued global emergency travel recommendations in an attempt to curb transmission of SARS worldwide. At the same time, affected countries started practicing strict infection control within hospitals, rigorous contact tracing and enforced quarantining of contacts in order to break the local chains of transmission.

From March to June, the epidemic was gradually brought under control. The last case of SARS in this initial outbreak was isolated in Taiwan on 15th June 2003, and the epidemic was declared over on 5th July 2003, when Taiwan was removed from the list of areas with local transmission.[12]

EPIDEMIOLOGY (SINGAPORE)

From 20th to 25th February 2003, three young Singaporean women went on a shopping trip to Hong Kong. They stayed on the 9th floor of the Metropole Hotel during the trip and all of them subsequently fell sick with fever and dry cough. On 1st March 2003, one of the young women (hereafter called IP) was admitted to Tan Tock Seng Hospital (TTSH) and was diagnosed to have an atypical pneumonia. On 2nd March 2003, another of the three was admitted to the Singapore General Hospital (SGH) for a similar illness. The persistent fever, rapidly deteriorating chest radiographs and negative microbiological evaluation puzzled the

doctors managing them. Efforts at obtaining respiratory specimens were enhanced and attempts were made at isolating lesser-known viruses such as the human metapneumovirus.

During her hospitalization in TTSH, IP infected at least 22 other people prior to the institution of infection control measures, and became the index patient of Singapore's SARS epidemic.[13] Her fellow traveling companions did not spread the disease.

On 11th March 2003, a physician who had personally taken respiratory specimens from IP left for a course in New York. However, he felt so unwell after arrival that he decided to come home as soon as possible. Unaware of events back home, he consulted a travel medicine specialist and was given clearance to fly.

By this time, the Ministry of Health (MOH) had become aware that an outbreak of this mysterious respiratory illness was taking place among TTSH healthcare workers (HCW) and other close contacts of IP. A taskforce was formed and coordinating meetings were held. When news reached the taskforce that the physician was planning to return, the decision was taken to stop him at the point where he would transit — in Frankfurt. This made headline news all over the world and focused the world's attention on this mysterious illness.

The virus initially spread among HCW, patients, visitors and their close family contacts. SARS spread to other hospitals when former TTSH inpatients and their contacts were hospitalized and managed without adequate precautions.[14] This occurred partly because ex-TTSH inpatients often had concomitant illnesses that masked the symptoms of SARS. Additionally, the true extent of the outbreak at TTSH was not fully appreciated in the early weeks of the epidemic. As has been the experience elsewhere, the management of patients from known SARS-affected hospitals under less than stringent infection control precautions was related in part to under-appreciation of the extent of the epidemic in the first hospital.[15] Amplification of the outbreak also occurred as a result of several "super-spreaders" — 144 of Singapore's probable cases were linked to contact with only 5 patients.[16]

The identification of a cluster of cases among employees at the Pasir Panjang wholesale market (PPWM) in April 2003 signaled the onset of community transmission of SARS. The market was declared closed by the authorities on 20th April 2003, and more than 400 persons were placed on home quarantine. These sweeping measures limited the "PPWM cluster" to only 15 individuals.

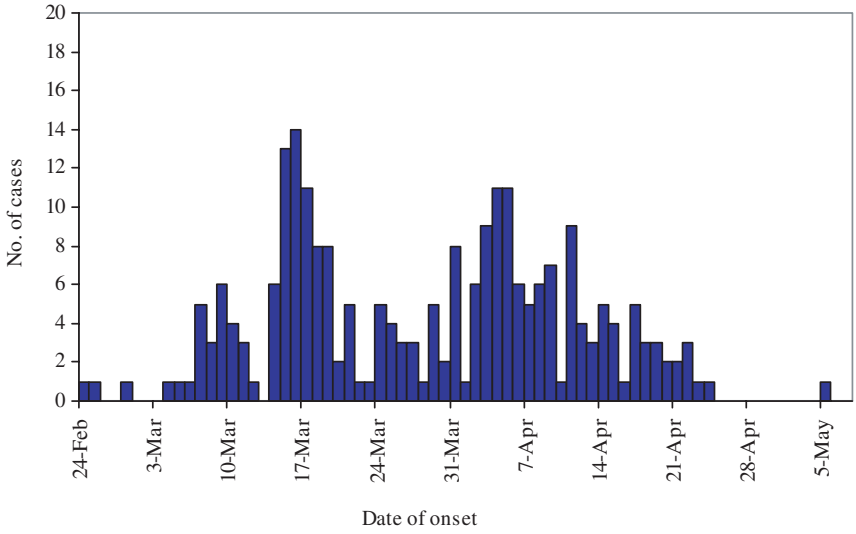


Fig. 1 Outbreak of 206 cases of SARS in Singapore, Mar–Apr 2003.

After a period of intensive infection control, contact tracing, and isolation of contacts, the end came when the last case of SARS in Singapore was isolated on 11th May 2003, and Singapore was removed from the list of areas with local transmission on 31st May 2003.[16] There were 238 cases of SARS diagnosed in Singapore, with 33 deaths.

In Singapore the median age of probable SARS patients was 35 years (range 1 to 90) and 66% were female. The ethnic breakdown was as follows: Chinese (65%), Malay (12%), Indian (11%) and Others (12%). HCW constituted 41% of probable SARS patients. Just over half of these were nurses.[17] The epidemiologic curve of the outbreak is shown in Fig. 1.

TRANSMISSION

Preliminary results on the stability of the virus show that it is stable in urine and stool for 1–2 days, and it may persist much longer in diarrheal stool (with higher pH) or under cold conditions.[18] It must also be mentioned that reverse transcriptase-polymerase chain reaction (RT-PCR) tests for SARS-CoV in convalescent stool specimens may remain positive for more than a month after the onset of illness in some cases. Despite this, however, there is mounting evidence to show that SARS is moderately rather than highly transmissible.[19,20] In Singapore, the majority (81%) of

probable SARS cases had no evidence of transmission of a clinical disease to other persons,[14] although this may have been the result of early detection and isolation of cases. There is also no evidence that asymptomatic cases and those who have recovered from SARS are infectious.

The main modes of transmission of the virus are via droplets and direct/indirect contact. This is borne out by the fact that the majority of new cases occurred in close contacts of patients, and the institution of contact and respiratory precautions was effective in protecting HCW.

The presence of virus in stools suggests that oral-fecal transmission is possible.[2] In the Amoy Gardens outbreak in Hong Kong, it was speculated that the virus may have spread through the sewage system.[21]

Airborne transmission appears to be rare except under circumstances facilitating the aerosolization of the virus, such as endotracheal intubation, bronchoscopy, and nebulization of drugs. At one hospital in Hong Kong, the outbreak was believed to have been magnified as a result of the use of nebulized bronchodilators in a patient with SARS[22] — up to 112 cases were linked to this patient.

The phenomenon of "super-spreading" — situations wherein a single case infected many others — was seen in many countries. The most likely explanation seems to be extensive viral shedding. However, late recognition and hence delayed isolation, as well as circumstances such as the nebulizer therapy case mentioned above, may have contributed to this phenomenon. Unfortunately, there are no clinical or epidemiological features to predict the super-spreader. The resulting epidemiological trail permits but a remorseful analysis.

One trait that led to the spectacular albeit transient success of the SARS-CoV was its ability to be transmitted efficiently within hospitals. HCW, inpatients, and visitors have been identified as the sources of infection in SARS victims all over the world. In Singapore, 75.2% of the patients were infected within a hospital. Hospitals were the sites of disease amplification,[14,19,22,23] and breaking the chains of transmission within hospitals was crucial in bringing the epidemic to an end.

The reasons for the enhanced infectivity within hospitals are manifold. Peiris and co-workers demonstrated (with the use of sequential quantitative RT-PCR analyses) that the viral load in nasopharyngeal secretions and stools peaked at about 10 and 14 days respectively after the onset of illness.[24] This would coincide with the period during which patients were most likely be hospitalized. Close contact between HCW

and patients, and between patients, is unavoidable. In addition, many patient-care procedures carried out in hospitals could result in aerosolization of the virus.

CASE DEFINITIONS

The latest WHO case definitions for SARS[25] is shown in Table 1.

The Center for Disease Control (Atlanta) has a similar set of case definitions, with the additional proviso that cases with negative convalescent-phase (i.e. 4 weeks from onset of illness) serological tests will be excluded from classification.[26]

Table 1 WHO Case Definitions for Surveillance of SARS (1st May 2003)

Suspect Case	1) A person presenting after 1st November 2002[a] with history of: • high fever (> 38°C) AND • cough or breathing difficulty AND one or more of the following exposures during the 10 days prior to onset of symptoms: • **close contact**[b] with a person who is a suspect or probable case of SARS • history of travel, to an area with recent local transmission of SARS • residing in an area with recent local transmission of SARS 2) A person with an unexplained acute respiratory illness resulting in death after 1st November 2002,[a] but on whom no autopsy has been performed AND one or more of the following exposures during to 10 days prior to onset of symptoms: • **close contact**[b] with a person who is a suspect or probable case of SARS • history of travel to an area with recent local transmission of SARS • residing in an area with recent local transmission of SARS
Probable Case	1) A suspect case with radiographic evidence of infiltrates consistent with pneumonia or respiratory distress syndrome (RDS) on chest X-ray (CXR). 2) A suspect case of SARS that is positive for SARS coronavirus by one or more assays. See "Use of laboratory methods for SARS diagnosis", http://www.who.int/csr/sars/labmethods/ 3) A suspect case with autopsy findings consistent with the pathology of RDS without an identifiable cause.

(*Continued*)

Table 1 Continued

Exclusion Criteria	A case should be excluded if an alternative diagnosis can fully explain their illness.
Reclassification of Cases	As SARS is currently a diagnosis of exclusion, the status of a reported case may change over time. A patient should always be managed as clinically appropriate, regardless of their case status. 1) A case initially classified as suspect or probable, for whom an alternative diagnosis can fully explain the illness, should be discarded after carefully considering the possibility of coinfection. 2) A suspect case who, after investigation, fulfils the probable case definition should be reclassified as "probable". 3) A suspect case with a normal CXR should be treated, as deemed appropriate, and monitored for 7 days. Those cases in whom recovery is inadequate should be re-evaluated by CXR. 4) Those suspect cases in whom recovery is adequate but whose illness cannot be fully explained by an alternative diagnosis should remain as "suspect". 5) A suspect case who dies, on whom no autopsy is conducted, should remain classified as "suspect". However, if this case is identified as being part of a chain transmission of SARS, the case should be reclassified as "probable". 6) If an autopsy is conducted and no pathological evidence of RDS is found, the case should be "discarded".

[a] The surveillance period begins on 1st November 2002 to capture cases of atypical pneumonia in China now recognized as SARS. International transmission of SARS was first reported in March 2003 for cases with onset in February 2003.
[b] **Close contact**: Having cared for, lived with, or had direct contact with respiratory secretions or body fluids of a suspect or probable case of SARS.

CLINICAL SYNDROMES

The initial symptoms of SARS are non-specific, and physical examination tends to be remarkable only for the number of negative findings. There is also no single laboratory test sensitive and specific enough to diagnose SARS with any reasonable degree of accuracy. Contact history is crucial. Nevertheless, a combination of the history, physical signs, radiological and laboratory findings may suggest a possible diagnosis of SARS even when the contact history is unreliable.

Symptoms and Signs

The most common symptom of SARS is fever with a body temperature of $>38°C$. The most common fever pattern is one where the baseline temperature is $>37.5°C$, with one or more spikes per day to $>38°C$. Occasionally, a biphasic pattern is seen, where the temperature returns to normal after 48–72 hours, and the patient remains afebrile for another 24–48 hours, before fever mounts again. Whether this represents a prodromal phase of the illness is not known at this time. Fever may be absent (or of minimal duration) in patients with comorbidities resulting in an impairment of the ability to mount a febrile response.[27]

Other symptoms such as myalgia, chills, headache, non-productive cough, vomiting and diarrhea are less common. The frequency of these symptoms within different cohorts is shown in Table 2. In Singapore, the frequency of rhinorrhea is lower than that of most other cohorts.

Patients generally appeared well during the initial stages of the disease. As it progressed, they became more lethargic and apathetic. The presence of auscultatory crackles denoted fairly extensive pulmonary involvement — a finding previously described in atypical (*Mycoplasma*) pneumonia. Interestingly, the lack of a commensurate degree of dyspnea despite the presence of hypoxia was observed in many of the more severe cases, but this is a soft sign and is not pathognomonic of SARS.

Table 2 Symptoms at Presentation (%)

	Leong *et al. Singapore ($n=233$)**	**Lee *et al.*[22] Hong Kong ($n=138$)**	**Booth *et al.*[23] Canada ($n=144$)**	**Donnelly *et al.*[28] Hong Kong ($n=1250$)**
Fever	92.7	100	99.3	94
Cough	38.5	57.3	69.4	50
Myalgia	62.4	60.9	49.3	51
Dyspnea	13.3	N.A.**	41.7	31
Nausea	11.1	19.6	19.4	N.A.**
Sore throat	13.3	23.2	12.5	23
Diarrhea	6.8	19.6	23.6	27
Malaise	10.7	N.A.**	31.2	64
Headache	N.A.**	55.8	35.4	50
Chills/Rigors	N.A.**	73.2	27.8	65
Rhinorrhea	2.1	22.5	2.1	25

*Unpublished data — Singapore.
**N.A. = not applicable.

Hematological Abnormalities

Abnormal hematological results are common during the course of illness. Self-limiting lymphopenia and thrombocytopenia have been described in the majority of cases.[13,22–24] Cui and co-workers demonstrated that T-lymphocytes, especially $CD4^+$T cells, were predominantly affected, although the absolute numbers of B cells and NK cells were also reduced.[29]

Thrombocytopenia, in the authors' experience, was mild, and platelet counts $< 50\,000/\mu l$ was not seen unless it was in the setting of disseminated intravascular coagulation. In most cases, both lymphocyte and platelet counts had recovered by the end of the third week of illness, although reports describing lymphopenia up to the fifth week of illness are known.[30]

Other Laboratory Findings

Common biochemical and electrolyte abnormalities include elevated levels of lactate dehydrogenase (LDH) and creatine kinase (CK), as well as mild hypokalemia and hyponatremia.[13,22–24] Derangement of liver enzymes, in particular aspartate transaminase (AST) and alanine aminotransferase (ALT), were also seen in a minority of cases. These changes are attributed to viral replication in the liver,[31] although a proportion could be due to the effects of treatment. In multivariate analysis, elevated LDH was an independent predictor of adverse outcome in patients with SARS.[22]

Clotting profile (prothrombin time, activated partial thromboplastin time) abnormalities have been described in a subset of patients,[22] but this was not commonly observed among the cases in Singapore.

Chest Imaging

Chest X-ray (CXR) changes remain one of the diagnostic criteria for SARS as well as a marker for progression of disease. However, it is important to point out that a substantial proportion of cases have normal CXRs during the initial phase of the illness.[13,22,23] Initial CXR changes tend to be small, unilateral, peripheral and patchy, often progressing over a few days to become bilateral and generalized, with confluent infiltrates.[22] The initial shadow was more commonly seen in the right lower zone, and is indistinguishable from other causes of bronchopneumonia. CXR findings in patients who had deteriorated were compatible with that of adult respiratory distress syndrome (ARDS).[32] Pleural effusions, cavitations and hilar

lymphadenopathy were not seen.[22,32] Examples of CXR changes are shown in Figs. 2, 3 and 4.

Typical findings on computed tomography (CT) scans of the thorax were ill-defined, ground-glass opacification in the periphery of the

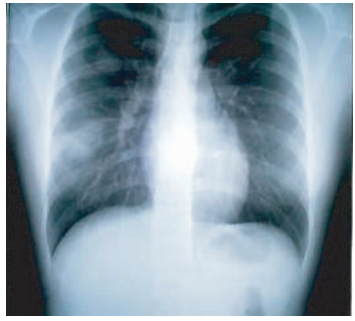


Fig. 2 Chest X-ray of index patient (Day 5 of illness).

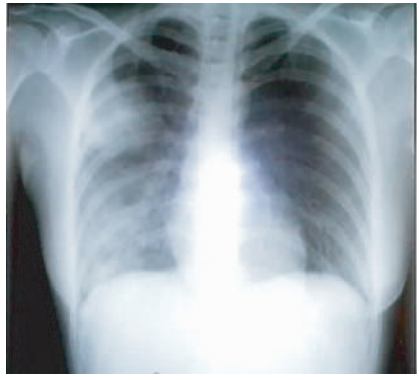


Fig. 3 Chest X-ray of index patient (Day 8 of illness).

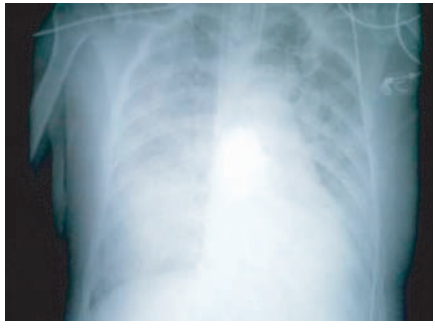


Fig. 4 Chest X-ray of critically ill case (Day 9 of illness), who eventually died.

affected lung parenchyma, usually in a subpleural location.[22] These findings are similar to the radiographic changes seen in acute interstitial pneumonia and bronchiolitis obliterans with organizing pneumonia.

Diagnostic Tests

Tests for the detection of SARS-CoV were developed rapidly due to the efforts of the WHO-led network of collaborating laboratories. Currently, tests are available for the detection of SARS-CoV-specific RNA in clinical specimens, isolation of the virus, and detection of SARS-CoV-specific antibodies.

Unfortunately, both available RT-PCR and viral culture tests are not sensitive enough to exclude SARS if negative results are obtained.[33] This is especially true in the early phase of the disease, when viral loads are low.

Table 3 Current Diagnostic Tests for SARS

Diagnostic Test	Details	Significance of Results
RT-PCR	Both qualitative and quantitative tests are available.	*Positive test*: Presence of viral genetic material; not necessarily infectious virus.
	Different primer sequences and protocols are available from WHO and CDC sites.	*Negative test*: Does not exclude SARS, especially if done within the first week of illness.
Cell Culture	Vero cell-lines. Biosafety Level (BSL) 3 laboratory required.	*Positive test*: Presence of infectious virus.
	Demanding test to perform.	*Negative test*: Does not exclude SARS.
Immunofluorescent Assay (IFA)	Detects IgM and IgG antibodies in serum.	IgM IFA may be positive from Day 10 of illness onwards.
Enzyme-linked immunosorbent assay (ELISA)	Detects a mixture of IgM and IgG antibodies in serum.	Usually yields positive results from Day 21 of illness onwards.
Neutralization Test	Requires BSL 3 laboratory and live virus.	Used at present for confirmation of above antibody tests.

At the same time, contamination of laboratory samples might result in false-positive results. Antibody tests are both more sensitive and specific than currently available RT-PCR tests, but the nature of such tests render them useless as a diagnostic tool in the early stages of the disease, when virus-specific antibodies have not yet been produced. Table 2 lists the available diagnostic tests for the diagnosis of infection with SARS-CoV.

Atypical Presentations

These refer to cases of SARS which do not fit the clinical case definitions (see above). They are fortunately rare, but are of importance because, undiagnosed, they may become new sources for outbreaks. Such cases led to the outbreaks in SGH and the National University Hospital (NUH) here. Several reports have been published,[34–36] and the patients tended to fall into three categories:

- early resolution or absence of fever
- lack of respiratory symptoms with very late development of CXR changes
- clinical syndrome of SARS obscured by comorbidities or synchronous bacterial infections

However, all patients described so far did have lymphopenia and elevated LDH, and such non-specific abnormalities could alert clinicians in affected areas to atypical presentations.[35]

Clinical Course

The incubation period of SARS is short, ranging from 2 to 16 days in 2 large studies,[21,23] with a median incubation period of 6 days. The WHO estimate of the maximum incubation period remains at 10 days.[37]

SARS has a highly variable clinical course. The majority (80%) of local cases developed fever with myalgia initially, followed by respiratory symptoms and CXR infiltrates from the middle to the end of the first week of illness. Dyspnea and lethargy worsened as CXR abnormalities progressed near the end of the second week of illness. This was followed by a period of recovery.

The remainder had a more fulminant course with severe clinical deterioration in the second week necessitating ventilatory support. This group

was characterized initially by the development of single-organ failure (ARDS), although multiorgan failure subsequently developed as a consequence of nosocomial complications.[38] More than half of the patients in this group eventually died.

There were also a few patients at extreme ends of the clinical spectrum, either recovering quickly (this was seen in a small number of young healthy persons) or progressing within the first week to respiratory failure and death.

A comparable clinical course with similar outcomes was seen in Canada[23] and Hong Kong.[22,24] It is thought that the clinical deterioration in the second week of the illness is the result of immunopathological damage to the lungs. This is supported by the observation that this deterioration occurred at a time of decreasing viral shedding from the nasopharynx.[24]

Several investigators have attempted to determine the risk factors for an adverse outcome. In most studies, older age, the presence of comorbidities and elevated LDH were found to be independent predictors of an adverse outcome.[22–24]

Discharge and Follow-up

Varying degrees of pulmonary fibrosis have been reported in a small percentage of patients following recovery. The pathophysiology of this is unknown. The trend has been towards gradual improvement on follow-up CT scans, with eventual resolution.

Many convalescent patients complain of easy fatigability and exertional dyspnea.[20] Still others developed depression and anxiety over various aspects of their disease.[39] These symptoms have resolved with time. We have found that a period of cardiopulmonary rehabilitation helped those in whom dyspnea was a significant complaint. There have been no reports of relapsed or recurrent cases of SARS.

Pediatric Cases

There have been comparatively few cases of children with clinical SARS. Although the presenting symptoms, laboratory results and radiological findings were similar to that of adults, the clinical course appeared to be much milder and shorter in children under 12 years of age.[40,41] The reason for this is unclear at present. All four women who contracted SARS in

Singapore during their pregnancy have since delivered. All the babies were born healthy, and there was no evidence of viral shedding. SARS serology was positive in the babies because of maternal antibodies.

TREATMENT

Multiple therapeutic regimens were attempted during the epidemic, but their true efficacy is unknown, mainly because no randomized, controlled trial was performed. This arose in large part from the pressure and desire to treat all patients in the heat of the epidemic. The results obtained with such regimens were compared with historical controls, and the dangers of placing too much significance on such comparisons are well-known. Some of the better-publicized regimens are described below.

Agents with potent *in vitro* effects against the SARS-CoV have since been discovered, but the question of clinical efficacy can only be answered when another outbreak occurs, and if properly conducted, randomized controlled trials are performed. Some of these agents are briefly described below.

Anti-bacterial Agents

These have little role in patients with probable or confirmed SARS, except in the management of secondary nosocomial infections. Prior to a diagnosis of SARS, however, antibiotics for pneumonia (community- or hospital-acquired) may be appropriate, as differentiating SARS from bacterial pneumonia is difficult.

Anti-viral Agents

Ribavirin

This is a synthetic nucleoside analog with activity against many RNA viruses. It was first used almost simultaneously in Canada, Hong Kong and Singapore, when the etiological agent of SARS was thought to be a paramyxovirus. However, *in vitro* data show that it does not inhibit the replication of SARS-CoV even at highly toxic concentrations, and it was associated with hemolytic anemia in 37.5% of Canadian patients who had been treated with it.[23]

Lopinavir/Ritonavir

This combination protease inhibitor (Kaletra®) is approved for the treatment of HIV. Hong Kong researchers turned to this agent in the final months of the outbreak. In a retrospective analysis of patients from Hong Kong, it was shown that the addition of Kaletra in the initial therapeutic regimen (which included ribavirin and steroids) was associated with a small but significant reduction in the overall death and intubation rate when compared with a matched cohort that received ribavirin and steroids without Kaletra.[42]

Convalescent plasma

The idea of using plasma from patients who had recovered from SARS was conceived in Hong Kong in the early days of the outbreak. It had been shown that serum from convalescent patients could arrest the cytopathic effects of the virus on vero cells, and physicians hoped to duplicate this effect *in vivo*. Presumed lack of efficacy and uncertainty about its risks led to its discontinuation as a form of therapy. Convalescent plasma was transfused into 3 critically ill patients in Singapore, 1 of who ultimately survived.

Interferon

German researcher Jindrich Cinatl and his co-workers tested the activity of recombinant interferons against clinical SARS-CoV isolates from Frankfurt and Hong Kong patients. They found that interferon-β had potent *in vitro* activity in preventing the replication of SARS-CoV.[43] However, this has not been tested clinically.

Immunomodulation/Immunosuppression

Attention shifted to the use of such agents when it became clear that at least part of the lung damage was attributable to immunopathological mechanisms.[24]

Intravenous immunoglobulin

The use of intravenous immunoglobulin (IVIG) as an immunomodulating agent is well-established. At TTSH, A/Prof Chng had designed a

regimen involving the use of IVIG in combination with methylprednisolone, and this was adopted for use in severe cases of SARS at the midpoint of the outbreak. There was a reduction in mortality compared with historical controls, but the numbers treated were too small to achieve significance.[38] IVIG has also been used in other countries,[44] with seemingly favorable results.

Steroids

The best treatment results published to date appear to have been obtained using a protocol combining both ribavirin and steroids.[45] In that study, there was no mortality among 31 cases. However, given the *in vitro* failure of ribavirin, the authors harbor reservations about the actual efficacy of this protocol. The use of steroids is not without the attendant risks of infection,[46] especially in view of T cell lymphopenia seen in most cases of SARS.[29]

Supportive Therapy

Given the lack of definitive therapy, intensive monitoring and supportive care for patients who deteriorate remain of primary importance. Because of the potential for sudden clinical deterioration, local SARS patients had their vital parameters and oxygen saturation monitored at a minimum of 4-hourly intervals.

Patients who required supplementary oxygen of FiO_2 0.5 or higher were transferred to the intensive care unit (ICU) for closer monitoring. Prophylactic endotracheal intubation with mechanical ventilation was initiated when necessary. Aggressive prophylactic anticoagulation with low molecular weight heparin was also carried out for such cases because postmortem findings in the initial fatalities had suggested the presence of a hypercoagulable state. This latter phenomenon was not apparent in milder cases of SARS, but manifested as an increased incidence of thromboembolic complications in the ICU.[38]

Future Directions

A combination of antiviral and immunomodulating drugs might be required for optimal therapy. It has been postulated that decreasing the

viral load may diminish the initial cytolytic lung damage. This may in turn decrease the subsequent immunopathological damage.[24] There are currently few drugs effective against coronaviruses, although some steps unique to coronavirus replication, such as binding of the spike protein on the viral envelope to a specific cell receptor, could be targeted for the development of new antiviral drugs.[47] Other potential therapeutic modalities include RNA interference[48] and protease inhibitors.

Other than drug development, protocols for trials of current "established" regimens should be in place in preparation for the possibility of another outbreak. It would not be ideal if questions about their efficacy remain unanswered following a second SARS epidemic.

INFECTION CONTROL

Unlike the influenza viruses, SARS-CoV is only moderately transmissible, with the number of secondary cases per index case ranging from 2.2 to 3.6 in a mathematical model.[49] A combination of age-old control measures, including strict infection control, shortening the time from symptom onset to isolation of patients, effective contact tracing and quarantine of exposed persons, has resulted in the containment of SARS worldwide.

Protective Measures (Healthcare Setting)

The latest guidelines from the MOH (Singapore) stipulate that HCW looking after SARS patients wear N95 masks, long-sleeved gowns and gloves. It suggests eye protection for anticipated splashes or sprays and also states that, when aerosol-generating procedures are expected, "higher level" respiratory protection through a positive airway purifying respirator is "necessary".[50] Although these recommendations sound intuitive, given that SARS has a mortality risk, data on the exact extent of their usefulness remains sketchy.

Seto and co-workers showed that no infections occurred among HCW from 5 Hong Kong hospitals who reported the use of gloves, gown and handwashing, although on logistic regression analysis, only the use of a mask was significant in preventing infection. Both N95 and surgical masks were equally protective, although a paper mask was not.[51]

Because the viral load in respiratory secretions can be high, and because SARS-CoV can be viable for prolonged periods, the meticulous

avoidance of touch contamination is important. A parallel may be drawn with the respiratory syncytial virus, a well-reported cause of nosocomial respiratory infection in pediatric inpatients, with modes of transmission that are similar to SARS-CoV. Interestingly for a respiratory virus, strict enforcement of gloving proved effective in reducing nosocomial transmission.[52]

In addition to these personal protective measures, administrative measures and engineering controls are probably useful in limiting the spread of SARS. Placing a SARS "suspect" or "probable" in a negative pressure isolation room has received support from local and international experts.[50,53] Limiting patient movement by having radiographs done "portable" was a much-used tactic in Singapore hospitals during the SARS epidemic. Its efficacy cannot be measured but it formed part of the slew of strict measures that worked. It is in line with a general principle of infection control — that the movements of potentially infectious patients should be limited.

Identification and Isolation of Cases

Transmission of SARS-CoV from patients to HCW has been reported despite the use of personal protective measures described above.[27,54] This underscores the importance of reducing the level of exposure in addition to the use of protective equipment. Early identification and isolation of SARS cases is crucial in this respect. While it may be practically and logistically impossible to treat every patient during an outbreak as potentially having SARS, heightened surveillance and rapid contact tracing should be the minimum standard of infection control in such circumstances.

After the First Epidemic

The example of Toronto has been overused, but the lessons to be drawn remain valid. A second outbreak of SARS occurred in Toronto in mid-May probably as a result of infection control measures being lifted too early.[55] With the worldwide epidemic having ended in July 2003, it is unlikely that the first case of a new outbreak of SARS will be diagnosed upon first contact. Maintaining a high degree of suspicion for SARS on the part of HCW is critical. The focus of SARS surveillance activities should now be on the early identification and isolation of suspect and probable cases.

Recent events have again highlighted the importance of the above points. Outbreaks of SARS occurred in Singapore,[56] Taiwan[57] and China[58] as a result of breaches in laboratory safety regulations. Fortunately, the toll was small in each country as a consequence of heightened awareness, resulting in early isolation of cases.

Because of the consequences of the SARS outbreak — disruption of medical services, cost to economies, loss of life — hospitals everywhere have embarked on "SARS preparedness" exercises. A draft CDC (Atlanta) document on preparation for SARS is available online.[59]

Vaccine

Several potential SARS vaccines have been developed, and one is currently undergoing Phase II trials in China. However, it will take at least a couple more years before its efficacy is proven and a viable vaccine becomes available widely.

CONCLUSION

Notwithstanding the recent laboratory cases, the most important question at this point is whether or not the SARS epidemic will occur. Certain viruses, like Ebola or Sin Nombre viruses, cause outbreaks when conditions facilitating their transmission from animal to human hosts (still unknown in the case of Ebola) are met. Others, such as the Nipah virus, have occurred only once, but the potential for recurrence exists due to the presence of animal reservoirs. Even when diseases are apparently eradicated, the possibility of unnatural recurrences due to acts of bioterrorism must be considered. Although smallpox appeared to be the greatest threat at the beginning of 2003, the chaos and economic damage created by SARS may have made this novel coronavirus a particularly attractive agent in this regard.

That an animal reservoir exists for SARS-CoV seems highly likely, and though the precise circumstances that resulted in the virus being transmitted to humans, and subsequently mutating to a more virulent form, are not known, these are unlikely to be so extreme that future replication will be impossible. Subsequent outbreaks are probable, although these will be smaller and more easily contained if the lessons learnt from this epidemic are not forgotten. The heterogeneity of the clinical presentations of SARS is now well known, and there is consequently less chance

of missing atypical cases that might go on to amplify the outbreak. Nosocomial cases constituted the bulk of the epidemics in all countries involved, and the institution of strict infection control measures has been pivotal in ending these outbreaks. All hospitals should now have protocols in place, and the delay between the appearance of a new SARS case and implementation of appropriate measures should be greatly reduced in the future. The lack of an antiviral drug of proven efficacy and of a vaccine underscores the need for continued vigilance and preparedness to ensure that future outbreaks are contained as effectively as the one that has passed.

The SARS-CoV was a timely wake-up call. As man challenges its physical boundaries, there is little doubt that another novel pathogen will emerge once again to cause another outbreak.

REFERENCES

1. WHO, Summary of probable SARS cases with onset of illness from, 1 November 2002 to 31 July 2003, http://www.who.int/csr/sars/country/table 2004_04_21/en/
2. Drosten C, Gunther S, Preiser W, *et al.*, Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *N Engl J Med* **348**(20):1967–1976, 2003.
3. Peiris JS, Lai ST, Poon LL, *et al.*, Coronavirus as a possible cause of severe acute respiratory syndrome, *Lancet* **361**(9366):1319–1325, 2003.
4. Ksiazek TG, Erdman D, Goldsmith CS, *et al.*, A novel coronavirus associated with severe acute respiratory syndrome, *N Engl J Med* **348**(20):1953–1966, 2003.
5. Marra MA, Jones SJ, Astell CR, *et al.*, The genome sequence of the SARS-associated coronavirus, *Science* **300**(5624):1399–1404, 2003.
6. Kuiken T, Fouchier RA, Schutten M, *et al.*, Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome, *Lancet* **362**(9380):263–270, 2003.
7. Hamre D, Connelly AP, Procknow JJ, Virologic studies of acute respiratory disease in young adults. IV. Viral isolations during four years of surveillance, *Am J Epidemiol* **83**(2):238–249, 1966.
8. Guan Y, Zheng BJ, He Ya, *et al.*, Isolation and characterization of viruses related to the SARS coronaviruses from animals in southern China, *Science* **302**(5643):267–268, 2003.
9. Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China, *Science* **303**(5664):1666–1669, 2004.
10. He JF, Xu RH, Yu DW, *et al.*, Severe acute respiratory syndrome in Guangdong province of China: epidemiology and control measures, *Zhonghua Yu Fang Yi Xue Za Zhi* **37**(4):227–232, 2003.

11. WHO, Acute respiratory syndrome in China — Update, http://www.who.int/csr/don/2003_02_12a/en/

12. WHO Update 96 — Taiwan, China: SARS transmission interrupted in last outbreak area, http://www.who.int/csr/don/2003_07_05/en/

13. Hsu LY, Lee CC, Green JA, *et al*., Severe acute respiratory syndrome (SARS) in Singapore: clinical features of index patient and initial contacts, *Emerg Infect Dis* **9**(6):713–717, 2003.

14. CDC, Severe Acute Respiratory Syndrome — Singapore 2003, *MMWR* **52**(18):405–411, 2003.

15. Scales DC, Green K, Chan AK, *et al*., Illness in intensive-care staff after brief exposure to severe acute respiratory syndrome, *Emerg Infect Dis* **9**(10):1205–1210, 2003. http://www.cdc.gov/mmwr/preview/ mmwrhtml/mm5218a1.htm

16. WHO Update 70 — Singapore removed from list of areas with local SARS transmission, http://www.who.int/csr/don/2003_05_30a/en/

17. Singapore Ministry of Health SARS website, http://www.moh. gov.sg/sars/information/healthcare.html#epidemiology/

18. WHO, First data on stability and resistance of SARS coronavirus compiled by members of WHO laboratory network, http:// www.who.int/ csr/sars/survival_2003_05_04/en/index.html

19. Riley S, Fraser C, Donnelly CA, *et al*., Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of Public Health Interventions, *Science* **300**(5627):1961–1966, 2003.

20. Avendano M, Derkach P, Swan S, Clinical course and management of SARS in healthcare workers in Toronto: a case series, *CMAJ* **168**(13):1649–1660, 2003.

21. WHO Update 33 — Hong Kong: Investigation of Amoy Gardens cluster, http://www.who.int/csr/don/2003_04_18/en/

22. Lee N, Hui D, Wu A, *et al*., A major outbreak of severe acute respiratory syndrome in Hong Kong, *N Engl J Med* **348**(20):1986–1994, 2003.

23. Booth CM, Matukas LM, Tomlinson GA, *et al*., Clinical features and short-term outcomes of 144 patients with SARS in the greater Toronto area, *JAMA* **289**(21):2801–2809, 2003.

24. Peiris JS, Chu CM, Cheng VC, *et al*., Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study, *Lancet* **361**(9371):1767–1772, 2003.

25. WHO, Case definitions for surveillance of severe acute respiratory syndrome (SARS). http://www.who.int/csr/sars/casedefinition/en/

26. CDC, Updated interim US case definition for severe acute respiratory syndrome (SARS), http://www.cdc.gov/ncidod/sars/casedefinition.htm

27. Singh K, Hsu LY, Villacian HS, *et al*., Severe acute respiratory syndrome: Lessons from Singapore, *Emerg Infect Dis* **9**(10):1294–1298, 2003.

28. Donnelly CA, Ghani AC, Leung GM, *et al.*, Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong, *Lancet* **361**(9371): 1761–1766, 2003.

29. Cui W, Fan Y, Wu W, *et al.*, Expression of lymphocytes and lymphocyte subsets in patients with severe acute respiratory syndrome, *Clin Infect Dis* **37**(6):857–859, 2003.

30. Wong RS, Wu A, To KF, *et al.*, Hematological manifestations in patients with severe acute respiratory syndrome: retrospective analysis, *BMJ* **326**(7403):1358–1362, 2003.

31. Cui HJ, Tong XL, Li P, *et al.*, Serum hepatic enzyme manifestations in patients with severe acute respiratory syndrome: Retrospective analysis, *World J Gastroenterol* **10**(11):1652–1655, 2004.

32. Wong KT, Antonio GE, Hui DS, *et al.*, Severe acute respiratory syndrome: radiographic appearances and pattern of progression in 138 patients, *Radiology* **228**(2):401–406, 2003.

33. WHO Update 39 — Caution urged when using diagnostic tests, http://www.who.int/csr/sarsarchive/2003_04_25/en/

34. Hon KL, Li AM, Cheng FW, *et al.*, Personal view of SARS: confusing definition, confusing diagnoses, *Lancet* **361**(9373):1984–1985, 2003.

35. Fisher DA, Lim TK, Lim YT, *et al.*, Atypical presentations of SARS, *Lancet* **361**(9370):1740, 2003.

36. Tan TT, Tan BH, Kurup A, *et al.*, An atypical case of SARS masked by E. coli bacteremia resulting in a cluster of infection, *Emerg Infect Dis* — in press.

37. WHO Update 49 — SARS case fatality ratio, incubation period, http://www.who.int/csr/sarsarchive/2003_05_07a/en/

38. Lew TW, Kwek TK, Tai D, *et al.*, Acute respiratory distress syndrome in critically ill patients with severe acute respiratory syndrome, *JAMA* **290**(3):374–380, 2003.

39. Maunder R, Hunter J, Vincent L, *et al.*, The immediate psychological and occupational impact of the 2003 SARS outbreak in a teaching hospital, *CMAJ* **168**(10):1245–1251, 2003.

40. Chiu WK, Cheung PC, Ng CK, *et al.*, Severe acute respiratory syndrome in children: experience in a regional hospital in Hong Kong, *Pediatr Crit Care Med* **4**(3):279–283, 2003.

41. Hon KL, Leung CW, Cheng WT, *et al.*, Clinical presentations and outcome of severe acute respiratory syndrome in children, *Lancet* **361**(9370):1701–1703, 2003.

42. Chu CM, Cheng VC, Hung IF, *et al.*, Role of lopinavir/ritonavir in the treatment of SARS: Initial virological and clinical findings, *Thorax* **59**(3):252–256, 2004.

43. Cinatl J, Morgenstern B, Bauer G, *et al.*, Treatment of SARS with human interferons, *Lancet* **362**(9380):293–294, 2003.

44. Chiang CH, Chen HM, Shih JF, *et al.*, Management of hospital-acquired severe acute respiratory syndrome with different disease spectrum, *J Chin Med Assoc* **66**(6): 328–338, 2003.

45. So LK, Lau AC, Yam LY, *et al.*, Development of a standard treatment protocol for severe acute respiratory syndrome, *Lancet* **361**(9369):1615–1616, 2003.

46. Wang H, Ding Y, Lee X, *et al.*, Fatal aspergillosis in a patient with SARS who was treated with corticosteroids, *N Engl J Med* **349**(5):507–508, 2003.

47. Holmes KV, SARS coronavirus: a new challenge for prevention and therapy, *J Clin Invest* **111**(11):1605–1609, 2003.

48. He ML, Zheng B, Peng Y, *et al.*, Inhibition of SARS-associated coronavirus infection and replication by RNA interference, *JAMA* **290**(20):2665–2666, 2003.

49. Lipsitch M, Cohen T, Cooper B, *et al.*, Transmission dynamics and control of the severe acute respiratory syndrome, *Science* **300**(5627):1966–1970, 2003.

50. MOH, Manual for SARS Infection Control, 13th Oct 2003.

51. Seto WH, Tsang D, Yung R, *et al.*, Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS), *Lancet* **361**(9368):1519–1520, 2003.

52. Leclair JM, Freeman J, Sullivan BF, *et al.*, Prevention of nosocomial respiratory syncytial virus infection through compliance with glove and gown isolation precautions, *NEJM* **317**:329–334, 1987.

53. CDC, Public Health Guidance for Community-Level Preparedness and Response to Severe Acute Respiratory Syndrome (SARS), November 3, 2003, http://www.cdc.gov/ncidod/sars/sarsprepplan.htm

54. Ho AS, Sung JY, Chan-Yeung M, An outbreak of severe acute respiratory syndrome among hospital workers in a community hospital in Hong Kong, *Ann Intern Med* **139**:564–567, 2003.

55. CDC, Update: severe acute respiratory syndrome — Toronto, Canada, 2003, *MMWR* **52**(23):547–550, 2003.

56. Lim PL, Kurup A, Gopalakrishna G, *et al.*, Laboratory-acquired severe acute respiratory syndrome, *N Engl J Med* **350**(17):1740–1745, 2004.

57. ProMED-mail. SARS — Worldwide (185): Taiwan, probable lab exposure. ProMED-mail 2003; 17 Dec 2003; 1217.3081, http://www.promedmail.org/, 31st May 2004.

58. ProMED-mail. SARS — Worldwide (29): China, cases. ProMED-mail 2004; 22 May 20040522.1373, http://www.promedmail.org/, 31st May 2004.

59. CDC. Public Health Guidance for Community Level Preparedness and Response to Severe Acute Respiratory Syndrome (SARS) Version 2, http://www.cdc.gov/ncidod/sars/guidance, 31st May 2004.

45

Infections in the Immunocompromised Host

Tan Ban Hock

INTRODUCTION

The immunocompromised patient is one whose immunity to infection is impaired. The consequence of such compromise in immunity is increased susceptibility to infection with a wide variety of organisms. Infection and immunity are inextricably linked because the principal function of the immune system is defence against microbes.[1]

In recent years, there has been an increase in the number of immunocompromised patients. The main reasons are the success of bone marrow and solid organ transplantation, successes in the use of immunosuppressives for the treatment of connective tissue disorders, and the human immunodeficiency virus (HIV) epidemic. Indeed where the ordinary clinician, could, in the past, regard opportunistic infections in the immunocompromised host as exotica meant for the academic physician and the sub-specialist, he would, in this day, ignore the subject at the peril of his patient's safety.[2]

What Constitutes an Immunocompromised State

As the term "the immunocompromised host" gains currency, it is becoming commonplace to attribute many infectious disease syndromes to immune compromise. Infections with uncommonly encountered pathogens, infections with common organisms in unusual locations, persistent or recurrent bacteremias, for example, are all frequently attributed to immune compromise. A precise understanding of the issues involved in the causation of infection by different pathogens as well as in the causes of recurrence and relapse is needed. This will ensure that care is appropriate. The following example illustrates this important principle.

> TKM, a 70-year-old woman with ischemic heart disease, had a stormy postoperative course after a total gastrectomy for gastric carcinoma. Pulmonary edema and a non-ST elevation myocardial infarct were managed in the intensive care unit. Upon discharge into the high dependency unit, she developed persistent fever and blood cultures repeatedly grew Methicillin-resistant *Staphylococcus aureus*, despite the rapid institution of intravenous vancomycin. Careful clinical examination revealed that the central venous line exit site was erythematous, with purulent drainage. The catheter was removed, vancomycin was continued and blood cultures became sterile.

This short history reveals the importance of a careful physical examination and of not attributing the persistent bacteremia to "immune compromise" from the gastric carcinoma. What then constitutes the immunocompromised state?

Immune compromise is usually the nett result of a combination of factors, each of which in some way affects the optimal functioning of some arm of the immune system. It is this nett state of immunosuppression that is important. This is best exemplified by the solid organ transplant recipient, in whom these principles were first expounded. The nett state of immunosuppression is the extent to which the patient is susceptible to infection, an extent defined by the sum and interaction of underlying disease, immunosuppressive treatment and a multitude of intrinsic and extraneous variables.[3]

Factors that contribute to this nett state of immunosuppression include:

- dose, duration and temporal sequence of immunosuppressive therapy employed,

- defects in host immunity caused by the underlying disease,
- breaches in anatomic integrity of non-specific bodily defences,
- metabolic abnormalities such as those induced by uremia and hyperglycemia, and
- presence of the immunomodulatory viruses (CMV, Epstein–Barr virus (EBV), HIV, Hepatitis B and C viruses).

Certainly this interaction of factors is nowhere more prominent than in the organ transplant recipient. It has been well shown, for example, that the likelihood of getting CMV infection (and therefore the disease) depends on factors such as the donor and recipient CMV status and the type of anti-rejection regimen employed. Eighty-two percent of recipients who are seronegative for CMV (the R^- host) and who receive kidneys from a cadaveric donor seropositive for CMV (the D^+ person) will seroconvert.[4]

The rate of seroconversion for such D^+R^- recipients is, however, 95% if an anti-lymphocyte preparation had been used in addition to conventional immunosuppression. About half of the patients who are at risk in this manner will become clinically ill. These figures continue to hold true to this day, as seen in a more recent publication, where 45% of seronegative recipients in the placebo arm of a CMV prophylaxis trial developed CMV disease.[5] On the other hand, only 20% of renal allograft recipients who were seropositive before transplant fell ill with CMV disease.[6] The incidence of CMV disease (i.e. of reactivation disease) in such patients is 59% if they had received OKT3 as part of the anti-rejection regimen.[6] Therefore, where CMV infection in the solid organ transplant recipient is concerned, the likelihood of infection (and therefore of disease) depends on a combination of factors such as donor and recipient serologic status, immunosuppressive regimen and use of prophylactic agents.[5]

A FEW CONCEPTS

A few truisms about infections in immunocompromised persons need to be emphasized.

The range of diagnostic possibilities is wide

The differential diagnoses to be entertained for any syndrome in the immunocompromised patient includes infective as well as non-infective conditions.

Infections may be those caused by the entire spectrum of pathogens — bacteria, viruses, fungi, parasites and mycobacteria. Non-infectious causes may be related to the underlying disease, or to the treatment or neither. For example, cough, breathlessness and lung infiltrates in a patient with cancer may be caused by recurrent tumor, radiation pneumonitis, pulmonary emboli or infection. A similar syndrome in the bone marrow transplant recipient may be related to a transfusion reaction, or fluid overload, or infection.

In a review of 51 renal transplant recipients with fever and pulmonary infiltrates, Ramsey *et al.* found that 15 patients had causes other than infections —– the non-infectious etiologies included pulmonary edema and pulmonary emboli.[7] In a more recent review of pulmonary infiltrates in the liver transplant patient admitted to the intensive care unit, Singh *et al.* reported the etiologies to be pulmonary edema (40%), pneumonia (38%), atelactasis (10%), ARDS (8%), and contusion (3%).[8] Among the infective causes of pneumonia in that study were: MRSA (27%), *Pseudomonas aeruginosa* (27%), invasive aspergillosis (20%), *Enterobacter*, *Serratia*, and *Pneumocystis carinii*.[8] In 3% of cases, the etiology of the pulmonary infiltrate could not be determined. Cisneros *et al.* reviewed 307 heart transplants in Spain over a two-year period.[9] Sixty-one patients had 65 episodes of pneumonia in that period. The etiologies were varied. Among heart transplant recipients presenting with community-acquired pneumonia, the microbic etiologies included *Haemophilus influenzae, Streptococcus pneumoniae*, and *Mycobacterium tuberculosis*. Of those with nosocomial pneumonia, the microbial culprits included *Acinetobacter baumannii, Pseudomonas aeruginosa, Legionella pneumophila*, and *Staphylococcus aureus*. In addition, the opportunistic pathogens found responsible for pneumonia in their series of heart transplant recipients included CMV, Aspergillus, *Pneumocystis carinii, Nocardia asteroides* and *Rhodococcus equi*.[9]

In spite of the wide range of diagnostic possibilities that have to be considered when evaluating the immunocompromised patient, a systematic approach is still possible. In essence, the approach does not deviate much from the age-old adage of a thorough history and a complete physical examination. However, as these patients tend to have had multiple admissions, the history usually involves more than just the conventional interview. Review of casenotes is just as important, as patients frequently cannot remember all the various problems they have encountered.

A thorough knowledge of the literature is of course important. In transplant recipients, a timetable of infection has been developed. This is a crucial part of the core knowledge of transplant physicians — it will be dealt with separately.

In addition, investigations have to be exhaustive and rapidly undertaken. This is by itself, almost a truism and also will be dealt with separately (*vide infra*).

Apart from the diagnostic approach, however, in some instances an internationally accepted management protocol has evolved that combines therapeutics with diagnostics. One commonly encountered example will be mentioned in some detail: The planned progressive antibiotic care of the cancer patient rendered neutropenic by chemotherapy (*vide infra*).

Certain defects in host immunity render a person more susceptible to infection with certain microorganisms

Asplenic patients, for example, are at increased risk of meningococcal and pneumococcal sepsis.[10] Patients with terminal deficiencies of the complement system are at risk of meningococcal infections.[11] Such correlations are most useful in the pediatric population, where patients with pure, isolated immune defects may be found.

In clinical adult practice, however, the unmodified patient is a rarity. The patient is a complex mix of immune defects and conditions outside the immune system that predispose to infection. These relate to the patient's underlying disease, his lifestyle, the environment, the procedures he undergoes and the therapies he receives. For example, while patients with multiple myeloma are classically said to have an increased susceptibility to bacterial infections,[12] in clinical practice the patient usually also has other conditions (diabetes mellitus, for example) that will affect his or her immune status. This is well illustrated in a recent discussion, in which a patient with myeloma, multiple courses of chemotherapy and diabetes (complicated by acidosis) suffered from mucormycosis of the orbit.[13]

The clinical presentation can be subtle

The inflammatory response to infection is greatly attenuated by immunosuppressive drugs given to the patient. Steroids, in particular, inhibit or

suppress the inflammatory response to a variety of inciting stimuli, be they of infective, chemical or mechanical origin. Steroids, therefore, are both anti-inflammatory and immunosuppressive in effect. Hence, typical signs and symptoms of infection may not be present in the patient who has been (and continues) on chronic steroid therapy.

The consequence is that patients on steroids may present with a relative scarcity of clinical signs. Numerous reports in the literature describe the deceptively benign abdominal findings in such patients despite the presence of a perforated viscus within.[14–17] The presence of active bowel sounds in spite of gross peritoneal soilage is well documented.[16] In fact, renal allograft recipients have even been reported to have asymptomatic pneumoperitoneum from perforated diverticulitis.[17] The dose of steroids seems to correlate inversely with the number of clinical findings. Remine and McIlrath described 79 patients with bowel perforation who had been on steroids, and divided them into three groups based on their steroid dosage. Almost all patients complained of abdominal pain, but clinical findings associated with peritonitis (guarding, rigidity, diminished bowel sounds) were much less prevalent in the group with the highest steroid dose than in the other two groups.[16] They also found a correlation between steroid dosage and mortality: The mortality of patients with bowel perforation was 77% in those receiving more than 20 mg of prednisolone equivalents per day. In addition, they found a significant delay in diagnosis in patients treated with greater than 20 mg of prednisolone equivalents per day. Tyau *et al.* reviewed the records of 209 patients with acute diverticulitis and found a greater risk of free perforation and need for surgery in the immunocompromised.[18] In their series, operations were performed in 58% of immunocompromised patients but only 33% of immunocompetent patients. They believed that the impaired inflammatory response of the immunocompromised host prevented them from walling off the disease process, hence increasing the risk of free perforation and need for surgery.

This altered presentation of serious illness in the immunocompromised means that clinicians managing patients with underlying immunocompromise (especially those on steroid therapy) need to be constantly on the lookout for possible catastrophic conditions (e.g. brain abscess or perforated viscus) when evaluating their patients' complaints.

One may have to resort to the sophisticated test, or even the invasive one, early in order to establish the diagnosis

Because of the immunocompromised patient's blunted immune response and the potentially catastrophic consequences of delayed diagnosis, the early use of a sophisticated or invasive test is sometimes inevitable.

The depressed inflammatory response of the immunocompromised person may modify the appearance of a pulmonary lesion. This is most frequently seen in the neutropenic, especially those with a white cell count below $100/mm^3$. Indeed neutropenia is generally listed as one of the causes of "pneumonia with a normal chest X-ray".[19] It is in this context that the CT scan of the chest has been said to have revolutionized the evaluation of the immunocompromised host with a febrile pneumonitis.[20]

The timetable of infection in transplant recipients

Despite the wide array of factors that influence the likelihood of infection, infections in transplant recipients are more or less predictable and generally occur in accordance with a timetable of infection (see Fig. 2).[21, 22] Such a timetable reflects the net effects of the myriad influences that interact to produce infection in the solid organ transplant recipient. If one goes through the timetable and reflects on the significance of the individual illnesses and why one patient gets one infection and not another, it will be obvious that while immunosuppressive therapy may be the driving force behind the net state of immunosuppression, factors like the patient's geographic origin and the length and duration of hospitalization play a large part in determining the components of the timetable.

In the first month after the transplant, the most common infections are those associated with the surgery and with instrumentation. Renal allograft recipients are at risk of hematomas and urinary leaks.[22] Fluid collections can become secondarily infected. In a local series of pediatric renal transplant recipients, for example, urinary tract infections was significantly associated with the presence of ureteric stents.[23] Heart and heart-lung transplant recipients are at risk of aortic suture line infection and mediastinitis, complications that are devastating.[24,25] Liver transplant recipients may suffer from hepatic artery, hepatic vein or portal vein thrombosis. Abdominal abscesses may complicate difficult abdominal

surgery. In a review of infections after liver transplantation, Kusne *et al*. found that most of the abdominal and liver abscesses occurred within 30 days of the transplant.[26] They also found that all patients with liver abscesses had technical problems with the implanted liver. In addition, all transplant recipients, by virtue of their status as post-surgical patients, may suffer from pneumonia, line sepsis and wound infection. Of the 101 liver allograft recipients in the series of Kusne *et al*., there were 15 episodes of bacterial pneumonia, 10 of which were nosocomial.[26] Technically perfect surgery is crucial to the success of transplantation and this is most obvious in the early post-transplant period. Interventions to handle the above-mentioned complications — drainage catheters, repeat visits to the operating theatre and to the intensive care unit — add to the infection rate.

The major viral infection of significance in the early post-transplant period is herpes simplex infection. This has virtually been eliminated with the routine use of acyclovir.[22]

In the period after the first month and until about the sixth month, two groups of infections predominate. The first are the immumodulating viruses, particularly CMV and EBV (and also Hepatitis B and C and HIV), which begin to exert their direct clinical effects. The second group consists of the classical opportunistic infections — Listeriosis, Nocardiosis, *Pneumocystis carinii* pneumonia, etc.

CMV disease (whether primary infection or reactivation) classically manifests at about six to eight weeks after transplant. Common symptoms include fever and malaise and common findings include leucopenia and thrombocytopenia. This is the probably the main infection that will be clinically significant, and occurs in a significant percentage of patients, with a range of manifestations determined by the degree of immunosuppression. In liver transplant recipients, for example, the use of OKT3 was associated with an increased risk of disseminated disease in primary infection.[27] Most of the other opportunistic infections have been successfully prevented by the routine use of trimethorprim-sulfamethoxazole prophylaxis.[28]

One category of non-viral infection that manifests early, sometimes within the first month but usually within the second or third month, is that carried by the patient into the transplant. These are caused by organisms that can remain latent in the host for prolonged periods without causing harm. The classical example is strongyloidiasis. *Strongyloides stercoralis* can remain latent in the host for prolonged periods,[29,30] and can

be reactivated by immunosuppression, especially steroids. Reports of disseminated strongyloidiasis in transplant recipients abound in the literature.[31–36] Disseminated strongyloidiasis may present as gram-negative sepsis, or bronchopneumonia, or gram-negative meningitis. Most occur in the second and third month after transplant.

Beyond the third month, infections are largely determined by the stability of immunosuppression and the general health of the patient.[21] Patients who are on stable, low-dose immunosuppression with rare episodes of hospitalization get very much the same community-acquired infections that immunocompetent people do.[21] On the other hand, those who require repeated increases in immunosuppressive therapy for rejection and who are maintained on high-dose immunosuppression (especially steroids) will develop opportunistic infections.

From the above discussion, it will be obvious that the immunosuppressive therapy is the main component of the total quantum of immunosuppression that transplant recipients labor under, and that the immunosuppressive therapy employed determines the types of infections and their manifestations to a large extent. It must be emphasized that the infections we tend to associate with immunosuppression — the opportunistic infections — are rarely seen in the first month, when immunosuppression is at its highest. On the other hand, from the timetable, we can see that these occur later in the post-transplant course. This observation underlies one important fact about the link between immunosuppression and infection — that it is the *cumulative* immunosuppression that is important, and not just the dose at one particular point in time.[21]

The timetable of infection is therefore extremely useful for helping to generate a differential diagnosis when confronted with a transplant recipient and his symptoms.

Immune reconstitution is crucial for recovery, but sometimes can by itself, produce a disease

The remarkable success of highly-active anti-retroviral therapy (HAART) in the treatment of HIV infection is now very well known, and has brought into focus the importance of "immune reconstitution" in the immunocompromised person's battle against infections. In the patient whose immune compromise is caused by immunosuppressive agents,

reducing immunosuppression seems an intuitive step to take when managing an infection. Before the advent of ganciclovir, transplant physicians ameliorated the symptoms of CMV disease by reducing the doses of cyclosporine and azathioprine.[37]

It has also been increasingly recognized that restoration of immunity can be associated with disease. The phrase "immune reconstitution syndrome" has been coined to describe the paradoxical deterioration of a pre-existing infection temporally related to the recovery of the immune system. In one of the earliest reports of this dramatic occurrence, Jacobson *et al*. described five patients who had very low absolute CD4 counts and then developed CMV retinitis four to seven weeks after initiation of HAART, despite a concomitant large increase in absolute CD4 count.[38] Two patients had vitritis early in the course of their retinitis. It was hypothesized that the patients may have had subclinical CMV retinal infection when HAART was started, and progression to symptomatic disease may have been accelerated by an improvement in CMV-specific immunity resulting from HAART. The presence of vitritis, not previously associated with AIDS-related CMV retinitis, possibly reflected the heightened inflammatory response. This theory is generally supported by other authors.[39,40]

ANTIBIOTIC THERAPY IN THE FEBRILE NEUTROPENIC

By febrile neutropenic it is usually meant the person who has been rendered neutropenic (absolute neutrophil count $\leqslant 1000/mm^3$) by chemotherapy for hematologic or solid-organ malignancies and who experiences a fever (single oral temperature $\geqslant 38.3°C$ or temperature $\geqslant 38°C$ for one hour.) The use of a series of antibiotics applied sequentially is part of an international consensus resulting from numerous clinical trials. This well-studied subject represents one of the better refined approaches to a difficult problem in the subject of infections in immunocompromised hosts. In essence, it is based on a few well-documented observations. These include the following:

> The post-chemotherapy patient who is neutropenic is very susceptible to infection with bacteria and fungi. Because phagocytosis by neutrophils is an important form of natural immunity to these microbes,[1] neutropenic patients can deteriorate swiftly from invasion of the blood

stream by bacteria and fungi. The rapidity of the decline in the white cell count, the depth of the neutropenia, and the duration of the neutropenia affect the degree to which the patient is immunosuppressed.

The likelihood of coming down with an infection increases with the depth and duration of neutropenia.[41] The risk of infection rises when the absolute neutrophil count decreases to levels below $1000/mm^3$. The risk further increases with further decline in the neutrophil count.

Neutropenic patients tend not to demonstrate many of the characteristic signs and symptoms of infection when they are infected. This is because neutrophils are agents of inflammation and many of the symptoms and signs (especially the signs) of an illness result from inflammation. Hence, for example, only a small percentage of neutropenic patients with a chest infection produce purulent sputum. Similarly, only a small fraction of neutropenic patients with a urinary infection will have pyuria. One consequence is that there is a need to be very thorough when assessing the febrile neutropenic, and as in other immunocompromised patients, to be constantly on the alert for catastrophic conditions when evaluating seemingly trivial complaints. The other important consequence is that there may well be a need to resort to sophisticated imaging early (this has already been emphasized).

Clinical Evaluation of the Febrile Neutropenic

A neutropenic patient who spikes a fever should be evaluated thoroughly, and given broad-spectrum antibiotics intravenously at high doses once all the relevant cultures have been obtained. Evaluation should include a thorough physical examination, with particular attention paid to sites such as the periodontium, the pharynx, the chest, the perianal area, the central venous catheter exit site and bone marrow aspiration sites. In addition, an attempt should be made to place the patient and his temperature in the context of the broader clinical picture. The interval between the temperature spike and the day of onset of cytotoxic chemotherapy is significant, as an estimate of the expected duration of neutropenia can then be made. Furthermore, if the temperature elevation occurred during or soon after the infusion of a blood product, and physical examination is unrevealing, then an infective cause for fever is less likely. The occurrence of

shock during an infusion, on the other hand, should alert one to the possibility of an infusate-related bacteremia. A review of the antibiotics the patient is already on is useful, as a superposed infection with multi-resistant organisms may well be the cause of the new fever.

Blood investigations should include, at the very least, two sets of blood cultures (for bacteria and fungi), preferably drawn through peripheral veins, at two different times. Since time is of the essence, one does not need to wait several hours to obtain the two sets of blood cultures.

Other investigations include a chest radiograph, as well as urine culture and stool studies. Stool studies for patients who have been hospitalized for prolonged periods should begin with a test for *Clostridium difficile* cytotoxin.

Antibiotic Therapy

Once the necessary cultures have been taken, broad-spectrum antibiotics should be administered intravenously. Many well-designed, randomized controlled trials have been published and there is good evidence for the three most commonly employed approaches to initial antibiotic therapy in the febrile neutropenic. These are: Monotherapy, duotherapy without a glycopeptide, and a glycopeptide plus one or two other drugs.

Monotherapy

There are no striking differences between monotherapy and multi-drug combinations for the empirical treatment of uncomplicated episodes of febrile neutropenia. There is evidence to support the use of ceftazidime, or cefepime, or imipenem, or meropenem as monotherapy. Extended-spectrum β-lactamases and Type-1 β-lactamases have reduced the utility of ceftazidime.[42] Cefepime, imipenem, and meropenem have the advantage of activity against *pneumococci* and *viridans streptococci*. Although cefepime is a relatively new agent, several published trials exist to support its use in this fashion.[43–45]

Duotherapy without a glycopeptide

The most commonly used combinations are an aminoglycoside and an anti-pseudomonal carboxy penicillin or ureidopenicillin, or an aminoglycoside and a third or fourth generation anti-pseudomonal cephalosporin.

Ticarcillin, mezlocillin and azlocillin are not available in Singapore. Several studies support the use of piperacillin-tazobactam plus amikacin.

A glycopeptide plus one or two other agents

With the rising incidence of vancomycin-resistant *enterococci* and increasing concern about other vancomycin-resistant organisms, the empiric use of vancomycin is coming into question. This holds true even in the febrile neutropenic, especially as several good studies have shown that vancomycin is not a necessary part of the initial antibiotic regimen.[46] In hospitals where there is a high MRSA rate, on the other hand, including vancomycin in the initial regimen may be indicated, especially in patients with exit-site erythema or drainage, or in patients already known to be carriers of penicillin-resistant pneumococci or MRSA. (Although vancomycin is frequently mentioned, teicoplanin may be used instead, as there are also studies supporting its playing a role akin to vancomycin's in the management of the febrile neutropenic.)[47]

Recent Trends in the Antibiotic Management of Febrile Neutropenics

In recent years several studies have evaluated the safety of oral antibiotics in the initial empirical therapy of febrile neutropenia.[48,49] In essence, these should be restricted to low-risk patients; antibiotics that have been used in this fashion include augmentin-clavulanate and ciprofloxacin, as well as cefixime. Fairly strict criteria have been applied in the major trials on this issue — patients eligible for such trials have generally had to have an anticipated duration of neutropenia <10 days, and be free of impairment in all organs (e.g. no pulmonary infiltrate, and no hepatic or renal dysfunction). In addition, they had to have no vomiting or diarrhea and had to be able to swallow their medications. Results comparable to intravenous therapy have been obtained in those trials where the patients were hospitalized for the duration of fever.

In summary, infection is a major cause of morbidity and mortality in the immunocompromised patient. A fascinating and bewildering array of infectious agents cause disease with myriad manifestations in the patient whose immune function has been compromised. Different types of immune compromise cause different types of infections. The approach to

infection in the immunocompromised demands meticulous attention to detail in patient evaluation. Entire books have been dedicated to this topic alone, and many monographs on individual organisms (of untold fascination to the infectious disease physician) exist. One chapter cannot do justice to this exciting subject, which continues to evolve before our eyes. We will always be students of our patients and their illnesses. The reader is invited to avail himself to the richness of the literature on this topic.

REFERENCES

1. Abbas AK, Lichtman AH and Pober JS (eds.), Cellular and Molecular Immunology, Philadelphia, WB Saunders Company, pp. 319–36, 1994.
2. Russell PS, Foreword to the third edition, in: Rubin RH and Young LS (eds.), *Clinical Approach to Infection in the Compromised Host*, 3rd ed, New York, Plenum Medical Book Company, pp. vii, 1994.
3. Rubin RH, Ikonen T, Gummert JF, Morris RE, The therapeutic prescription for the organ transplant recipient: the linkage of immunosuppression and antimicrobial strategies, *Transplant Infect Dis* **1**:29–39, 1999.
4. Rubin RH, Tolkoff–Rubin NE, Oliver D *et al.*, Multicenter seroepidemiologic study of the impact of cytomegalovirus infection on renal transplantation, *Transplantation* **40**:243–249, 1985.
5. Lowance D, Neumayer HH, Legendre CM *et al.*, Valacyclovir for the prevention of cytomegalovirus disease after renal transplantation, *N Engl J Med* **340**:1462–1470, 1999.
6. Hibberd PL, Tolkoff–Rubin NE, Cosimi AB *et al.*, Symptomatic cytomegalovirus disease in the cytomegalovirus antibody seropositive renal transplant recipient treated with OKT3, *Transplantation* **53**:68–72, 1992.
7. Ramesy PG, Rubin RH, Tolkoff–Rubin NE *et al.*, The renal transplant patient with fever and pulmonary infiltrates: etiology, clinical manifestations, and management, *Medicine* (*Baltimore*) **59**:206–222, 1980.
8. Singh N, Gayowski T, Wagener MM, and Marino IR, Pulmonary infiltrates in liver transplant recipients in the intensive care unit, *Transplantation* **67**:1138–1144, 1999.
9. Cisneros JM, Munoz P, Torre–Cisneros J *et al.*, Pneumonia after heart transplantation: a multiinstitutional study, *Clin Infect Dis* **27**:324–331, 1998.
10. Bohnsack JF, Brown EJ, The role of the spleen in resistance to infection, *Ann Rev Med* **37**:49–59, 1986.
11. Figueroa JE, Densen P, Infectious diseases associated with complement deficiencies, *Clin Micro Rev* **4**:359–395, 1991.

12. Longo DL, Plasma cell disorders, in: Isselbacher KJ, Braunwald E, Wilson JD, Martin JB, Fauci AS and Kasper DL (eds.), *Harrison's Principles of Internal Medicine*, 13th ed., New York, McGraw–Hill, 1618–1625, 1994.

13. Kelley MA, Wu CL, Case records of the Massachusetts General Hospital, *N Engl J Med* **341**:265–274, 1999.

14. Warshaw AL, Welch JP, Ottinger LW, Acute perforation of the colon associated with chronic corticosteroid therapy, *Am J Surg* **131**:442–446, 1976.

15. Cosimi AB, Surgical aspects of infection in the compromised host, in: Rubin RH and Young LS, (eds.), *Clinical Approach to Infection in the Compromised Host*, 3rd ed., New York, Plenum Medical Book Company, pp. 707–729, 1994.

16. Remine SG, McIlrath DC, Bowel perforation in steroid-treated patients, *Ann Surg* **192**:581–586, 1980.

17. Lederman ED, Conti DJ, Lempert N, Singh TP, Lee EC, Complicated diverticulitis following renal transplantation, *Dis Colon Rectum* **41**:613–618, 1998.

18. Tyau ES, Prystowsky JB, Joehl RJ, Nahrwold DL, Acute diverticulitis — a complicated problem in the immunocompromised patient, *Arcg Surg* **126**:855–859, 1991.

19. Bartlett JG, *Management of Respiratory Tract Infections*, Baltimore, Williams and Wilkins, pp. 1–117, 1997.

20. Rubin RH, Greene R, Clinical approach to the compromised host with fever and pulmonary infiltrates, in: Rubin RH and Young LS, (eds.), *Clinical Approach to Infection in the Compromised Host*, 3rd ed., New York, Plenum Medical Book Company, pp. 121–161, 1994.

21. Rubin RH, Infection in organ transplant recipients, in: Rubin RH, Young LS, (eds.), *Clinical Approach to Infection in the Compromised Host*, 3rd ed., New York, Plenum Medical Book Company, pp. 629–705, 1994.

22. Snydman DR, Infection in solid organ transplantation, *Transplant Infect Dis* **1**:21–28, 1999.

23. Sim SK, Yap HK, Murugasu B, Prabhakaran K, Ho CL, Infections in paediatric renal transplant recipients, *Ann Acad Med Singapore* **26**:290–293, 1997.

24. Dowling RD, Baladi N, Zenati M *et al.*, Disruption of the aortic anastomosis after heart-lung transplantation, *Ann Thorac Surg* **49**:118–122, 1990.

25. Thomson D, Menkis A, Pflugfelder P *et al.*, Mycotic aneurysm after heart-lung transplantation, *Transplantation* **47**:195–197, 1989.

26. Kusne S, Dummer S, Singh N, Iwatsuki S, Makowka L, Esquivel C, Tzakis AG, Starzl T, Ho M, Infections after liver transplantation — an analysis of 101 consecutive cases, *Medicine* **67**:132–143, 1988.

27. Singh N, Dummer S, Kusne S *et al.*, Infections with cytomegalovirus and other herpesviruses in 121 liver transplant recipients: transmission by donated organ and the effect of OKT3 antibodies, *J Infect Dis* **158**:124–131, 1988.

28. Fox BC, Sollinger HW, Folkert FO *et al*., A prospective, randomized, double-blind study of trimethorprim-sulfamethoxazole for prophylaxis of infection in renal transplantation: clinical efficacy, absorption of trimethorprim-sulfamethoxasole, effects on microflora, and the cost-benefit of prophylaxis, *Am J Med* **89**:255–274, 1990.

29. Gill GV, Bell DR, Reid HA, Strongyloidiasis in ex-Far East prisoners of war, *BMJ* **1**:1007, 1977.

30. Pelletier LL, Chronic strongyloidiasis in World War II Far East ex-prisoners of war, *Am J Trop Med Hyg* **33**:55–61, 1984.

31. Fagundes LA, Busato O, Brentano L, Strongyloidiasis: fatal complication of renal transplantation, *Lancet* **ii**:439, 1971.

32. Meyers AM, Shapiro DJ, Milne FJ *et al*., *Strongyloides stercoralis* hyperinfection in a renal allograft recipient, *S Afr Med J* **50**:1301–1302, 1976.

33. Scoggin CH, Call NB, Acute respiratory failure due to disseminated strongyloidiasis in a renal transplant recipient, *Ann Int Med* **87**:456–458, 1977.

34. Venizelos PC, Lopata M, Bardawil WA *et al*., Respiratory failure due to *Strongyloides stercoralis* in a patient with a renal transplant, *Chest* **78**:104–106, 1980.

35. Weller IVD, Colpand P, Gabriel R, Strongyloides stercoralis infection in renal transplant recipients, *BMJ* **282**:524, 1981.

36. Morgan JS, Schaffner W, Stone WJ, Opportunistic strongyloidiasis in renal transplant recipients, *Transplantation* **42**:518–524, 1986.

37. Tobin J O'H, Cytomegalovirus infection, in: Weatherall DJ, Ledingham JGG and Warrell DA, (eds.), *Oxford Textbook of Medicine*, 2nd ed., Oxford, Oxford University Press, 5.76–5.80, 1987.

38. Jacobson MA, Zegans M, Pavan PR, O'Donnell JJ, Sattler F, Rao N, Owens S, Pollard R, Cytomegalovirus retinitis after initiation of highly active antiretroviral therapy, *Lancet* **349**:1443–1445, 1997.

39. Stein DS, Cytomegalovirus retinitis after initiation of highly active antiretroviral therapy, *Lancet* **350**:589–590, 1997.

40. Koval CE, Gigliotti F, Nevins D, Demeter LM, Immune reconstitution syndrome after successful treatment of *Pneumocystis carinii* pneumonia in a man with human immunodeficiency virus type I infection, *Clin Infect Dis* **35**:491–493, 2002.

41. Bodey GP, Buckley M, Sathe YS, Freireich EJ, Quantitative relationships between circulating leukocytes and infection in patients with acute leukaemia, *Ann Intern Med* **64**:328–340, 1966.

42. Hughes WT, Armstrong D, Bodey GP, Bow EJ, Brown AE, Calandra T, Feld R, Pizzo PA, Rolston KVI, Shenep JL, Young LS, 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer, *Clin Infect Dis* **34**:730, 2002.

43. Biron P, Fuhrmann C, Cure H *et al*., Cefepime versus imipenem-cilastation as empirical monotherapy in 400 febrile patients with short duration neutropaenia, *J Antimicrob Chemother* **42**:511–518, 1998.

44. Engervall P, Kalin M, Dornbusch K, Bjorkholm M, Cefepime as empirical monotherapy in febrile patients with hematological malignancies and neutropenia: a randomized, single-centre phase II trial, *J Chemother* **11**:278–286, 1999.

45. Chandrasekhar P, Arnow PM, Cefepime versus ceftazidime as empiric therapy for fever in neutropenic patients with cancer, *Ann Pharmacother* **34**:989–995, 2000.

46. Anonymous, Vancomycin added to empirical combination antibiotic therapy for fever in granulocytopenic cancer patients. European Organization for Research and Treatment of Cancer (EORTC) International Antimicrobial Therapy Cooperative Group and the National Cancer Institute of Canada-Clinical Trials Group, *J Infect Dis* **163**:951–958, 1991.

47. Menichetti F, The role of teicoplanin in the treatment of febrile neutropenia, *J Chemother* **12** (Suppl 5):34–39, 2000.

48. Freifeld A, Marchigiani D, Walsh T, Chanock S, Lewis L, Hiemenz J, Hiemenz S, Hicks JE, Gill V, Steinberg SM, Pizzo PA, A double-blind comparison of empirical oral and intravenous antibiotic therapy for low-risk febrile patients with neutropenia during cancer chemotherapy, *N Engl J Med* **341**:305–311, 1999.

49. Kern WV, Cometta A, De Bock R, Langenaeken J, Paesmans M, Gaya H, Oral versus intravenous empirical antimicrobial therapy for fever in patients with granulocytopenia who are receiving cancer chemotherapy, International Antimicrobial Therapy Cooperative Group of the European Organization for Research and Treatment of Cancer, *New Engl J Med* **341**:312–318, 1999.

This page intentionally left blank

46

Malaria

Nicholas Paton

INTRODUCTION

Malaria is the most important protozoal infection in man, with more than one-third of the world's population exposed to the risk of infection. There are an estimated 300–500 million clinical cases per year, and between 1–3 million deaths. The greatest burden of disease is in Africa, and African children account for most of the deaths worldwide.

Malaria control programs that were instituted in the 1950s and 1960s initially had dramatic effects on the incidence of malaria in many areas, but lack of funding and waning enthusiasm led to the collapse of most of these initiatives and the resurgence of the disease. To add to this bleak picture, drug resistance continues to spread at an alarming rate. In many areas of South-East Asia, the parasite is now resistant to all of the traditional first-line treatments for the disease. The "Roll Back Malaria" initiative of the World Health Organization seeks to have a major impact on malaria incidence by improving prevention (especially use of impregnated bed-nets) and also increasing access to appropriate medication. The recent sequencing of the genomes of the parasite *P. falciparum* and of the

mosquito vector are important advances that may ultimately lead to new drugs and strategies to control the spread of the disease.

MALARIA EPIDEMIOLOGY IN SINGAPORE

The situation in Singapore is less alarming than many places, although the disease remains a significant problem here. Several countries in South-East Asia have substantial problems with drug resistance, and the potential for importation of drug-resistant malaria is therefore present. It is critical to remain aware of the disease and its complexities in order to recognize and avert a potential outbreak.

The annual incidence of malaria in Singapore is approximately 300 cases, and more than 90% of those are imported. Approximately 80% of cases are vivax and 20% of cases are falciparum. The majority of the imported cases are from the Indian subcontinent (60%) and South-East Asia (35%). Local outbreaks within Singapore are reported from time to time. Examples include the Punggol outbreak (27 cases) in 1993, Tanjung Rhu/East Coast Park outbreak (41 cases) in 1993 and Dairy Farm Road outbreak (17 cases) in 1996. The last two outbreaks were traced to foreign workers with relapsing vivax malaria who were not detected until transmission had already occurred. Malaria is also occasionally acquired from the islands offshore from Singapore.

Thus, although the malaria situation in Singapore is remarkably good in comparison to the majority of countries in the region, the country remains vulnerable to the introduction of the disease due to the large and increasing traffic of both locals and foreigners to and from endemic areas, and the presence of suitable mosquito vectors (*Anopheles maculatus* and *Anopheles sundaicus*) here.

CLINICAL PRESENTATION

Uncomplicated Malaria

Fever is universal and characteristically comprises three stages: A feeling of intense cold with shivering and rigors lasting approximately 1 hour; a febrile stage lasting 2–6 hours; a sweating stage lasting 2–4 hours, during which there is profuse sweating and the temperature falls to normal. When infection is well-established and the parasite life cycle

synchronized, the fever may become periodic, occurring only every few days. In the first few days of symptoms there is usually no discernible pattern. In practice, it is hard to differentiate malaria from other bacterial or viral infectious diseases based on the fever pattern alone.

Other non-specific features of malaria include headache, vomiting and mild diarrhea. Examination may reveal pale conjunctivae and splenomegaly, but the patient is otherwise normal. The blood count usually shows a normal white cell count and mild thrombocytopenia.

Complicated Falciparum Malaria

P. falciparum has the unique pathogenic ability to cause red blood cell sequestration in the deep capillaries. This property results in the potential for patients with falciparum malaria to develop a range of complications and for the infection to take a more severe course than is seen with malaria due to one of the other parasite species. These complications are described below:

Cerebral malaria

Cerebral malaria usually develops after several days of fever and constitutional symptoms. It often starts with a generalized convulsion, which is then followed by persisting unrousable coma. It is important to distinguish minor degrees of unconsciousness, which are commonly associated with any fever, from cerebral malaria: For the latter to be diagnosed the patient should have, at best, a non-localizing response and an inappropriate verbal response to painful stimuli. There may be mild neck stiffness, but frank rigidity is not a feature of cerebral malaria. Retinal hemorrhages may occur and are a poor prognostic sign. Disorders of conjugate gaze are common. Examination of the peripheral nervous system usually reveals upper motor neuron signs, although almost any pattern can be seen. Opisthotonus, decerebrate and decorticate extensor posturing are all well-recognized. The neurological features resolve fully with recovery from malaria, with less than 10% of patients showing any persistent deficits.

Cardiovascular complications

Patients are often hypotensive, but the development of clinical shock (i.e. hypotension with signs of peripheral circulatory shut down) may

indicate the presence of a secondary bacterial infection, pulmonary edema or a large gastrointestinal hemorrhage. Although postmortem studies have shown that parasites are sequestered in the cardiac vessels, myocardial failure and arrhythmias are rare.

Pulmonary edema

This is common in adult patients in South-East Asia and is the most frequent complication that results in a fatal outcome. Pulmonary edema tends to develop late in the course of the illness (after one or two days of treatment) and may appear suddenly in a patient who otherwise seems to be doing well. The pathogenesis is unclear: It may develop in patients with fluid overload but has also been shown to develop spontaneously in patients with normal or reduced pulmonary capillary wedge pressures.

Gastrointestinal

Nausea, vomiting and watery diarrhea are common. Vomiting necessitates administration of anti-malarial medication by the parenteral route.

Renal failure

Acute renal failure is common and associated with a poor outcome. It may be oliguric (usually) or polyuric. It is usually reversible with treatment and recovery from malaria.

Hepatic impairment and jaundice

Jaundice is most commonly due to hemolysis and is frequently seen in patients with severe malaria. Mild disturbances of hepatic function (transaminases a few times the upper limit of normal) can be seen, but clinical signs of liver failure and dramatically elevated transaminases are very rare.

Hematological complications

Anemia is common and sometimes out of proportion to the degree of parasitemia (especially in children). Thrombocytopenia is almost universal

but is not an indicator of severity and does not have much clinical significance. Features of disseminated intravascular coagulation are sometimes seen and this can lead to minor spontaneous bleeding from the gums, epistaxis, subconjunctival hemorrhages and clinically significant bleeding in the gastrointestinal tract. Hemoglobinuria (blackwater fever), which used to be common in patients repeatedly self-medicating with quinine, is now seen rarely. It is probably due to extensive hemolysis in patients with G6PD deficiency and/or the oxidant effects of antimalarial drugs.

Hypoglycemia

This is a frequent complication of falciparum malaria and its treatment. The symptoms and signs of hypoglycemia (anxiety, palpitations, dizziness and breathlessness; ultimately decreased consciousness, convulsions and coma) may be readily attributed to features of malaria *per se* and hence hypoglycemia is easy to miss. The hyperinsulinemic effect of quinine, the disturbance of glucose regulation by hepatic dysfunction and the high metabolic requirements imposed by a large parasite load all contribute to hypoglycemia.

Secondary infections

Septicemia (most often due to Gram-negative organisms) is a common complication of severe malaria that must be considered in any patient whose condition deteriorates during treatment. Bronchopneumonia, urinary tract infections and bacterial meningitis can also occur.

DEFINITION OF SEVERE AND COMPLICATED MALARIA

A definition of severe malaria is useful in the assessment of patients with falciparum malaria and in determining the appropriate treatment. A consensus document was produced by a WHO expert committee in 1990 and updated in 2000. Table 1 is adapted from that document and a useful clinical checklist for evaluating the severity of infection with falciparum.

Table 1 Manifestations of Severe *P. Falciparum* Malaria in Adults

Clinical

1. Cerebral malaria (impaired consciousness or any cerebral dysfunction)
2. Repeated, generalized convulsions (> 2 within 24 hours)
3. Shock (systolic blood pressure less than 70 mmHg with cold, clammy skin)
4. Respiratory distress (acidotic breathing)
5. Pulmonary edema (radiological)
6. Spontaneous bleeding from gums, nose, GI tract
7. Hemoglobinuria (if not due to hemolysis by anti-malarials in G6PD deficiency)
8. Jaundice

Laboratory

9. Hyperparasitemia (> 5% in non-immune)
10. Severe normocytic anemia (HCT < 15% or Hb < 5 g/dL)
11. Substantial laboratory evidence of disseminated intravascular coagulation
12. Hypoglycemia (< 2.2 mmol/L)
13. Acidosis (arterial pH < 7.25 or plasma bicarbonate < 15 mmol/L)
14. Hyperlactatemia (> 6 mmol/L)
15. Creatinine > 265 μmol/L (with oliguria)

DIAGNOSIS

The essential step in diagnosing malaria is to suspect it in the first place. Thus any febrile patient who has been to a malarious area in the previous few months (the usual incubation period is 2 weeks) should be considered to have malaria until proven otherwise. Due to the difficulty in extracting an accurate and comprehensive travel history from patients, it is also wise to routinely consider the possibility of malaria in any patient with fever and thrombocytopenia who lacks additional indicators of a viral etiology (e.g. rash). Diagnosing dengue fever is easy and unlikely to affect the outcome for the patient. To miss or delay the diagnosis of malaria is also easy, but may have potentially fatal consequences for the patient.

The only certain way to confirm a diagnosis of malaria is to perform a blood film. Thick films are performed as a rapid way of detecting parasites, whereas thin films are used to speciate the parasite. The latter is crucial for determining appropriate treatment; if in doubt, the species should be considered to be falciparum. A single negative blood film makes the diagnosis unlikely, but if the diagnosis of malaria is seriously entertained, the film should be repeated after 12–24 hours until three negative films have been obtained.

Non-microscopy tests for the diagnosis of malaria have been developed in recent years that have potential utility in field situations where the facilities and expertise for microscopy may not be available. The tests use a monoclonal antibody coated on a test strip to detect a malaria protein (*P. falciparum* histidine-rich protein 2) in the blood of infected people. The test strip is easy to use, the result can be read with the naked eye and only a small amount (a finger prick) of blood is required. The sensitivity and specificity are good when used by skilled laboratory staff and when there is a high-grade parasitemia, but these decrease markedly when used by lay people or when the parasitemia is low grade.

TREATMENT

Non-falciparum Malaria

Drug therapy

The "benign" malarias (*P. vivax*, *P. malariae* and *P. ovale*) should be treated with chloroquine. With few exceptions (see below) the parasites remain sensitive. Chloroquine is usually well tolerated. In the relatively high dose used to treat acute malaria it may cause nausea, dysphoria, transient neuropsychiatric syndrome or cerebellar dysfunction. Chloroquine is contraindicated in patients with psoriasis as it can lead to a flare of disease. The usual treatment schedule is 600 mg (of base) followed by 300 mg 6 hours later followed by 300 mg on two sucessive days, although this schedule can be compressed into 36 hours for convenience.

A two-week course of primaquine is also required to treat infections with *P. vivax* and *P. ovale* in order to eradicate the dormant liver forms which can lead to relapse at a future time. Primaquine causes hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency and is contraindicated in patients with severe deficiency. In mild deficiency, 45 mg once weekly is given for six weeks to achieve eradication of the liver forms. The other common side-effects of primaquine are nausea and abdominal pain.

A high prevalence (80%) of chloroquine-resistant vivax has recently been reported from Irian Jaya. Sporadic cases of chloroquine-resistant vivax have also been reported from Papua New Guinea, the Solomon Islands, Myanmar and India. The optimal treatment for such cases is unknown. Quinine may be effective in standard doses when used for

more than 3 days. Mefloquine may also be used. Standard doses of chloroquine can also be combined with higher doses of primaquine. Expert advice should be sought in these cases.

Primaquine resistance in the cure of *P. vivax* malaria is well known in South-East Asia. When vivax relapses following primaquine therapy, the dose of primaquine should be increased to 30 mg of primaquine base daily for 14 days.

Other points in treatment

Although vivax and the other non-falciparum malarias most often cause mild infection and rarely lead to significant complications, treatment is best conducted in hospital. Some patients may develop severe constitutional symptoms (high fever, malaise and nausea) which can be treated appropriately to avoid dehydration; the clinical condition and the parasite count can be closely monitored to detect cases where a mixed infection with falciparum has been missed or erroneously reported; in Singapore, where the potential exists for spread of the disease to others, treatment is best supervised in hospital and the clearance of the blood film properly documented.

Uncomplicated Falciparum Malaria

Drug therapy

Although *P. falciparum* still remains sensitive to chloroquine and/or sulfadoxine-pyrimethamine in some parts of the world, for practical purposes all falciparum malaria presenting to doctors in Singapore is best assumed to be resistant to these two drugs and should therefore be treated with quinine. For mild malaria quinine can be given orally, although it is not well-tolerated. The taste is bitter and quinine often induces a symptom complex called cinchonism (nausea, dysphoria, tinnitus and high tone deafness). This is rapidly reversible on stopping therapy. Although cure may be obtained using courses lasting 3–4 days, these have not been well studied and the recommended treatment duration remains 7 days.

It is also recommended that quinine be combined with one other antimalarial drug, to cover for the possibility of drug resistance. Although quinine resistant malaria is uncommon, and mainly limited to border areas of

Thailand and Myanmar, it is preferable to cover for this eventuality. The routine use of combination therapy for malaria may also help to slow the spread of drug resistance and this approach is now being endorsed by many malaria experts. The usual drug to combine with quinine is doxycycline in a dose of 100 mg twice daily for 7 days. An alternative for children is clindamycin 10 mg/kg twice daily for 7 days.

There are several other treatment options if quinine cannot be used. Mefloquine is effective against all malaria species and the treatment schedule is simple, comprising just 2 doses (15 mg/kg followed by 10 mg/kg 12–24 hours later). The commonest side-effects are nausea, vomiting, abdominal discomfort, vertigo and minor psychological disturbances. The efficacy of mefloquine montherapy has waned in recent years in Thailand, and hence it is now usually combined with an artemisinin derivative (see below).

The most promising drugs in the battle against the spread of multi-drug resistant falciparum malaria are the artemisinin derivatives. Artemisinin originates from the sweet wormwood plant (Artemsia annua), which has been used to treat fever in China for centuries. The parent compound is insoluble in lipid or water, but various derivatives have been developed which are easier to handle. Artesunate is water-soluble and can be given by intravenous injection, and artemether is oil soluble and can be administered by intramuscular injection. These drugs are now widely used in China and South-East Asia for the treatment of drug-resistant malaria. Large-scale clinical trials have demonstrated equivalence with quinine in the treatment of falciparum malaria. They act more rapidly to clear parasites than do other anti-malarials and are exceedingly well-tolerated in humans. A recent meta-analysis of artemether versus quinine showed superiority of artemether over quinine in terms of overall serious adverse events in all patients, and in terms of mortality in the subgroup of adults with multisystem failure. When given as monotherapy there is a significant risk of recrudescence, and hence the artemisinin derivatives should be combined with a second drug. In many areas of South-East Asia where drug resistance is a problem, the combination of a five-day course of artesunate or artemether followed by mefloquine is now the standard therapy. Not only is this combination highly effective against multi-drug resistant falciparum, it also decreases the gametocyte carriage rate, which may be an extremely useful feature in the control of the spread of drug-resistant malaria.

Atovaquone is a broad-spectrum anti-parasitic agent that has recently been investigated for use in the treatment of malaria. When used alone it is associated with high rates of recrudescence, but in combination with doxycycline or proguanil the cure rate is close to 100%. The combination of atovaquone and proguanil is well-tolerated and is now commercially available in an oral formulation. Clinical trials have demonstrated equivalent efficacy to quinine and tetracycline therapy in the treatment of mild falciparum malaria.

Other points in treatment

It is important to repeat the blood film regularly during treatment. A rise in parasite count may be seen immediately after initiation of therapy, but it should fall progressively thereafter. If the count has not decreased from baseline by 75% after 48 hours, drug resistance should be suspected and the choice of drug reconsidered. Particular attention should be paid to make sure that vomiting of medication does not go unnoticed: If the patient fails to get therapeutic doses, malaria that started out as uncomplicated may deteriorate rapidly.

Severe and Complicated Falciparum Malaria

Drug therapy

This is a medical emergency and intravenous quinine should be instituted as soon as possible after the diagnosis is made. A loading dose of 20 mg/kg should be given (infused over 4 hours) followed, every 8 hours, by a dose of 10 mg/kg (infused over 4 hours). A loading dose should not be given if the patient has received quinine or mefloquine within the previous 24 hours. The main complication of quinine administration is hypoglycemia due to stimulation of insulin secretion. Routine ECG monitoring is unnecessary, except in patients with myocardial disease or arrhythmias. The dose of quinine should be reduced by 50% after 72 hours of treatment if the patient remains seriously ill to avoid accumulation. The treatment can be converted to the oral route as soon as the patient can swallow.

Doxycycline can be instituted immediately, or delayed until later in therapy if the patient is unable to take orally at presentation.

Other points in treatment

Fluid management is critical. Patients are often dehydrated on presentation, and this may contribute to hypotension and acute renal failure. On the other hand, they are vulnerable to develop pulmonary edema with aggressive rehydration. Thus rehydration should be performed cautiously, ideally with monitoring of central venous pressure (CVP) or pulmonary capillary wedge pressure (PCWP). The CVP should be maintained at less than 5 cm water and the PCWP at less than 15 mmHg. Dialysis should be started early if there is evidence of progressive acute renal failure.

The patient should be monitored closely for hypoglycemia with finger prick testing every 3–4 hours for the first few days. An infusion of 10% dextrose should be given to patients when hydration status has been stabilized. Vigilance for seizures is important, and the signs may be subtle. In patients with marked neck stiffness, secondary bacterial meningitis may be present. A lumbar puncture should be considered for diagnostic purposes, although practical considerations (e.g. severe thrombocytopenia) often prevent this and empirical treatment may be given. Secondary bacterial infection (septicemia, pneumonia, urinary tract infection) is common. Blood and other appropriate cultures should be taken and empirical treatment started in patients where secondary infection is suspected. Steroids are contraindicated in cerebral malaria and osmotic agents may also be harmful. The parasite count should be checked twice daily in the early stages of treatment to monitor progress and detect the possibility of drug resistance (see above).

Although there are many who advocate the use of exchange transfusion for the treatment of severe malaria and there are many small case series published showing positive outcomes, there have been no large-scale randomized controlled clinical trials to address the issue definitively. Hence the decision must remain in the hands of the individual physician. There is also no consensus on when to perform exchange transfusion, although the level of >15% parasitemia in patients with complications is accepted as a reasonable threshold. The volume of blood to be exchanged is also not standardized, but a partial exchange (of say 4 units) is usually preferred in Singapore.

CONCLUSION

Malaria is one of the major infectious diseases of mankind and, in spite of the best efforts of international agencies, is likely to remain so for the

foreseeable future. Indeed, with the imminent failure of standard drug therapies in Africa, and the increasing resistance to many newer drugs in Thailand and other parts of South-East Asia, the situation may worsen. The best hope for the future lies in the artemisinin derivatives which remain potent and are extremely well-tolerated, and in the widespread adoption of combination therapy as the routine practice to slow the spread of drug resistance.

FURTHER READING

1. Committee on Epidemic Diseases, Epidemiological surveillance of malaria in Singapore 1992–1996, *Epidemiol News Bull* **23**:43–45, 1997.
2. Funk M, Schlagenhauf P, Tschopp A, Steffen R, Malaquick versus ParaSight F as a diagnostic aid in travellers' malaria, *Trans Royal Soc Trop Med Hyg* **93**:268–272, 1999.
3. Mordmuller B, Kremsner PG, Hyperparasitaemia and blood exchange transfusion for the treatment of children with falciparum malaria, *Clin Infect Dis* **26**:850–852, 1998.
4. Price RN *et al.*, Artesunate/mefloquine treatment of multi-drug resistant falciparum malaria, *Trans R Soc Trop Med Hyg* **89**:523–527, 1995.
5. White NJ, The treatment of malaria, *N Engl J Med* **335**:800–806, 1996.
6. WHO, Severe falciparum malaraia, *Trans Royal Soc Trop Med Hyg* **94**(S1):1–90, 2000.
7. Wilkinson RJ, Severe falciparum malaria: predicting the effect of exchange transfusion, *Q J Med* **87**:553–557, 1994.

47

Antimicrobial Resistance

Ling Moi Lin and Tan Ban Hock

INTRODUCTION

It has been estimated that infectious and parasitic diseases cause 23.3% of deaths in the world. Infections of the lower respiratory tract are the third most common cause of death worldwide, with diarrhea diseases in fourth position.[1] Ninety-eight percent of deaths in children occur in developing countries, mostly as a result of infections.

Bacterial resistance to multiple antimicrobials characterizes the present decade. Although most of the resistant bacteria are present in hospitals, multi-drug resistant bacteria such as *Streptococcus pneumoniae, Mycobacterium tuberculosisis, Salmonella typhimurium* also cause serious community-acquired infections.

With the changing scene in antimicrobial resistance worldwide, we are facing increasing challenges in the management of infectious diseases. The strong relationship between antimicrobial use and resistance development has been well demonstrated in the epidemiology of methicillin-resistant *Staphylococcus aureus* (MRSA), drug-resistant *Streptococcus pneumoniae* and recently, vancomycin-resistant *Enterococcus sp* (VRE).

Most people in both developed and developing countries accept antimicrobials as their right to obtain a prescription at the first trivial sign of infection or treat themselves with what can be bought over the counter. This unbridled use of the antimicrobials has propelled us in the direction of growing resistance in the bacteria isolated.

RESISTANT BACTERIA

Methicillin-resistant *Staphylococcus Aureus* (MRSA)

In the 1950s, *Staphylococcus aureus* became resistant to penicillin via the production of β-lactamase, penicillinase. Cloxacillin was then introduced as a penicillinase stable penicillin to treat *Staphylococcus aureus* infections. Shortly after the introduction of cloxacillin and also the other new class of antimicrobials, streptomycin, tetracyclines, chloramphenicol and erythromycin, resistance appeared to the respective agents. Methicillin-resistant *Staphylococcus aureus* (MRSA) was first discovered in 1961 with outbreaks reported in UK, USA and Australia in the late 1970s.[2,3] MRSA made its first appearance in Singapore in 1985. Since then it has remained endemic in many of our hospitals with isolation rates varying from 30–60% of all *Staphylococcus aureus* isolates despite good infection control practices.

The resistance is due to the presence of a *mec* gene resulting in the production of a new penicillin binding protein, PBP2.[4] This results in the low affinity binding of the bacteria to β-lactams. The antibiogram of the MRSA strains vary from country to country. The strain circulating in Singapore is classically resistant to β-lactams, aminoglycosides, macrolides, with varying susceptibility to fusidic acid, clindamycin, chloramphenicol, ciprofloxacin and co-trimoxazole. Vancomycin is the drug of choice for the treatment of MRSA infections. The antibiogram of our local strains have seen changes over the years, especially to the antimicrobials, co-trimoxazole, clindamycin and ciprofloxacin. Hence, the choice of oral agents for the management of MRSA infections is far more limited now.

Vancomycin Intermediate *Staphylococcus Aureus* (VISA)

This was first discovered in May 1996 in Japan from a boy who had a prolonged course of vancomycin for the treatment of MRSA infection.[5,6]

The MIC of the isolate to vancomycin was 8 μg/ml. Since then there were 5–6 reports of similar isolates from patients on prolonged course of vancomycin for MRSA infections. Thus far, there is none isolated here in Singapore. The resistance has been attributed to mutation in the MRSA strain after prolonged exposure to vancomycin.[7]

Vancomycin-resistant *Enterooccus* Species (VRE)

This was first reported in UK and France in 1987.[8,9] Since then, high incidence and outbreaks have been reported in USA particularly.[10] The emergence of VRE has been attributed initially to the high usage of avoparcin, a glycopetide, in animal feed in Europe and vancomycin in USA. However, recently, evidence seems to point to the high usage of cephalosporins as the incriminating factor.[11,12]

In Singapore, the isolation of VRE from clinical isolates in inpatients appear to be sporadic and low (< 1%). However, in a study performed by Oon *et al.* in Jan–Mar 1997 (unpublished data), 12.3% stool carriage was reported in patients of a teaching hospital. Further larger scale studies are needed to verify this as the presence of VRE colonization in our population poses serious infection control issues and challenges.

Extended spectrum β-lactamase (ESBL) producing *Enterobacteriaceae*

These are the Ambler class A penicillinases that possess an extended hydrolysis ability on late generation cephalosporins (cefotaxime, ceftriaxone, ceftazidime, cefepime, cefpirome) and aztreonam. ESBL-producing *Enterobacteriaceae* were first reported in 1983 in Germany[13] and later became major nosocomial pathogens in Europe, USA and now worldwide.

They emerged as a result of amino acid modifications of the TEM-1 or TEM-2 and SHV-1 enzymes. The development was associated with the wide use of 3rd generation cephalosporins because mutants have been selected for during their use. Local studies in 1995 showed a prevalence of 11% in our intensive care units (unpublished data). This is an emerging problem in many of our units with high-risk patients, e.g. the Hematology and Burns Unit, and poses challenging issues in the management as the only effective drug available currently is carbapenem.

Multiresistant Gram-negative Bacteria

Our common nosocomial pathogens are *Acinetobacter baumannii, Pseudomonas aeruginosa* and *Klebsiella* species. The non-fermentative bacteria do cause serious infections in our hospitalized patients and occasionally pose challenges in management. Outer membrane impermeability, efflux pumps and metallo enzymes are but some of the mechanisms causing resistance in the non-fermenters.[14] Multiresistance to varying agents — aminoglycosides, β-lactams, quinolones — have been seen sporadically. The wide use of carbapenem has given rise to the natural selection of *Stenotrophomonas maltophilia*. However, of growing concern now is the development of carbapenem resistance in some of these non-fermenters.

Drug-resistant *Streptococcus Pneumoniae* (DRSP)

Streptococcus pneumoniae is one of the top three major community-acquired pathogens causing respiratory tract infections. Penicillin resistance was first reported in 1977 in South Africa.[15] It is now a worldwide problem with countries reporting incidences varying from 10–60% in their isolates. The resistance due to a chromosomal change in the penicillin binding protein has been attributed to the wide use of cephalosporins through the years. Isolates obtained from inpatients at Singapore General Hospital showed an increase of penicillin resistance of 2% in 1991 to 26% in 1998. The antibiogram also reflected increased resistance to ceftriaxone (20%), macrolides (36%) and tetracycline (59%). These pose challenging issues in the management of community-acquired respiratory tract infections as alternative therapy may have to be sought for some of our patients who showed no response to empirical therapy of amoxicillin, amoxicillin-clavulanic acid or macrolides. The newer quinolones, levofloxacin, gatifloxacin and moxifloxacin have good activity against penicillin-resistant *Streptococcus pneumoniae.*

CONTROL OF ANTIMICROBIAL RESISTANCE

We have seen development of new antimicrobials over the years. However, in parallel, we have also seen the gradual emergence of antimicrobial resistance in many bacterial species. We have to admit that antimicrobial resistance is an undeniable fact and will continue to be a problem

as long as we use the currently available antimicrobial agents. Bacteria are remarkably adaptable and will continue to evolve and acquire new mechanisms of resistance to antimicrobial agents. The close link between antimicrobial use and resistance has been established clearly over the years.[16]

Current policies in dealing with antimicrobial resistance have been ineffective and carefully planned strategies have to be drawn if we are to win the war against microbes.

Antimicrobial use by physicians and patients are influenced by various factors, e.g. knowledge, peer influence, advertisement, availability of antimicrobial and cost. In attempts to devise strategies to control the development of antimicrobial resistance, the following will first have to be established:

1) an understanding of the factors that promote overuse and the barriers to change; and
2) the implementation of effective strategies for changing behavior.

For it to succeed, targeted and effective interventions will have to be used.

Factors that contribute to antimicrobial overuse include lack of education, patient's expectations, past experience, and economic factors that influence the degree of availability of antimicrobial. Hence, multifaceted strategies will have to be adopted in the planning as well as implementation of an antimicrobial policy.

The antimicrobial policy

This usually includes the following components:

1) Drug formulary;
2) Drug detailing;
3) Drug use; and
4) Drug utilization evaluation.

1) *Drug formulary*
   This is a list of drugs approved by the Pharmacy and Therapeutics or equivalent body in the hospital for purchase and use at the hospital. It is a powerful method to curb the tendency for newer agents to displace older, still-effective ones. The Pharmacy and Therapeutics

Committee of the hospital is responsible for reviewing the formulary regularly to decide whether the putative advantages of newer drugs, major or minor, warrant increased expenditures.

2) *Drug detailing*
This refers to matters related to visits made by pharmaceutical representatives to doctors, distribution of literature and other media presentation made by the company to hospital staff. It is a known fact that pharmaceutical firms stand to profit most by promoting the newer, more expensive, and still-patented products. The limitation of contact time with these representatives help to exercise some "control" in the information released. Although most perceive the pharmaceutical representative as helpful colleagues who provide quick information, guest speakers and funds for education and research, nevertheless this restriction should still be considered.

3) *Drug use*
These may entail ordering policies such as:

a) clinical Practice Guidelines for prophylaxis, empirical and therapeutic management;
b) automatic stop orders;
c) the requirement for written justification through special order forms; and
d) the requirement for approval by a recognized authority, e.g. infectious disease physician on use of selected antimicrobial.

Clinical Practice Guidelines are more effective if well supported by other educational activities. Guidelines for the management of common infectious diseases, e.g. respiratory tract and urinary tract infections, are best drawn locally as these are usually more rational.[17] These are more likely to be accepted and followed than those developed by others without local input or recognition of local needs, especially in relation to cost and availability of antimicrobial agents.

Automatic stop orders require the doctor to re-order a drug after a certain duration of therapy, e.g. 3 or 5 days therapy. This is to encourage the doctor to review the drug for continuation or a change to an alternative drug.

Special order forms require the doctor to justify use, usually by stating the category of use (empirical, therapeutic or prophylactic) or specific indication. Use of such a form may be required for all antimicrobials or selected agents or situations.

The other approach of requiring approval from a recognized authority, e.g. an infectious disease physician, is effective in controlling the use of certain antimicrobials. These may be selected on the basis of cost, toxicity or the need to control a particular resistant pattern, e.g. the control of vancomycin use, in an effort to control the development of VRE.

4) *Drug utilization review*
This would require Clinical Practice Guidelines to be in place first. Criteria of appropriate prescribing are then drawn up. A study is then conducted and analysis of data made. Feedback to clinicians regarding their own antimicrobial prescribing practices is then given and this had been shown to be a successful technique for achieving behavioral change.[18] The feedback includes not only cost of antimicrobial prescribed but also comparisons with peers or an indicator.

CONCLUSION

We have a task to keep therapeutic options open for our patients in the milieu of evolving bacterial resistance against antimicrobials. Whether we win the battle against the microbes will depend on our unified efforts in implementing the multi-prong strategies to keep the pathogens at bay.

REFERENCES

1. Murray CJL, Lopez AD, Mortality by cause for eight regions of the world: global burden of disease study, *Lancet* **349**:1269–1276, 1997.
2. Kayser FH, Methicillin-resistant staphylococci 1965–75, *Lancet* **2**:650–653, 1975.
3. Shanson DC, Antibiotic resistant *Staphylococcus aureus, J Hosp Infect* **2**:11–36, 1981.
4. Hackbarth CJ, Chambers HF, Methicillin-resistant staphylococci: genetics and mechanisms of resistance, *Antimicrob Agents Chemother* **33**:991–994, 1989.
5. Reduced susceptibility of *Staphylococcus aureues* to vancomycin — Japan, 1996, *MMWR Morb Mortal Wkly Rep* **46**:624–626, 1997.
6. Hiratmatsu K, Hanaki H, Ino T, Yabuta K, Oguri T, Tenover FC, Methicillin-resistant *Staphylococcus aureus* clinical strain with rediced canocmycin susceptibility, *J Antimicob Chemother* **40**:135–136, 1997.
7. Smith TL *et al.*, Emergence of vancomycin resistance in *Staphylococcus aureus, N Engl J Med* **340**:493–501, 1999.
8. Uttley AHC, Collins CH, Naidoo J, George RC, Vancomycin resistant enterococci, *Lancet* **1**:57–58, 1988.

9. Leclerq R, Derlot E, Duval J, Courvalin P, Plasmid-mediated resistance to vancomycin and teicoplanin in *Enterococcus faecium, N Engl J Med* **319**:157–161, 1988.

10. Centers for Disease Control and Prevention. Nosocomial enterococci resistant to vancomycin — United States, 1989–1993, *MMWR Morb Mortal Wkly Rep* **42**:597–599, 1993.

11. Morrison AJ, Wenzel RP, Nosocomial urinary tract infections due to enterococcus: ten years' experience at a university hospital, *Arch Intern Med* **146**:1549–1551, 1986.

12. Edmond MB *et al*., Vancomycin-resistant *Enterooccus faecium* bacteremia: risk factors for infection, *Clin Infect Dis* **20**:1126–1133, 1995.

13. Knothe H, Shah P, Kremery V, Antal M, Mitsubishi S, Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of *Klebsiella pneumoniae* and *Serratia marcescens, Infection* **11**:315–317, 1983.

14. Hancock REW, Resistance mechanisms in *Pseudomonas aeruginosa* and other nonfermentative gram-negative bacteria, *Clin Infect Dis* **27**(Suppl 1):S93–S99, 1988.

15. Jacobs MR *et al*., Emergence of multiply resistant pneumococci, *N Engl J Med* **299**:735–740, 1978.

16. Moellering RC Jr, Interaction between antimicrobial consumption abd selection of resistant bacterial strains, *Scand J Infect Dis* (Suppl 70):18–24, 1990.

17. Grimshaw JM, Russell IT, Effect of clinical guidelines on medical practice: a systemic review of rigorous evaluations, *Lancet* **342**:317–322, 1993.

18. Schoenbaum SC, Feedback of clinical practice information, *HMO Pract* **7**:5–11, 1993.

48

Dengue Fever

Tang Ong Teng, Tan Ban Hock and Tang Julian Wei-Tze

VIROLOGY

The dengue virus is an enveloped, single-stranded, positive-sense RNA virus, approximately 50 nm in diameter. Its 11 kilobase genome codes for three structural proteins (core, membrane-associated and envelope proteins), and seven non-structural proteins. The envelope protein contains the domains for host-cell attachment, as well as viral neutralization. Dengue virus has 4 defined serotypes (DEN 1–4) and shares extensive serological cross-reactivity with other flaviviruses, such as yellow fever, West Nile virus, Japanese encephalitis virus and St Louis encephalitis virus. Infection with one dengue serotype results in lifelong immunity to that serotype, but only partial and temporary protection to the other serotypes, a property generally agreed to give rise to risks for dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Dengue virus replication occurs in its invertebrate hosts, namely mosquitoes belonging to the *Aedes* genus, which remain infected for life. Viral transmission takes place via biting of its human host, where most viral amplification occurs. Although there is evidence of vertical, transovarian

transmission in female mosquitoes, it is not thought that this is of particular significance in human transmission.[1] Other vertebrate hosts include several species of Asian and African primates, but other vertebrate species can only be experimentally infected with difficulty.

EPIDEMIOLOGY

The dengue viruses are responsible for the largest number of infections among the arboviruses. Currently, dengue fever (DF) is endemic in over 100 countries, threatening about 40% (2.5 billion) of the world's population, with an estimated 50–100 million cases occurring each year. These include 250 000–500 000 cases (0.5%) of DHF,[2] which affects mainly infants and children aged 10 years or less. The mortality of untreated DHF can be as high as 12–44%, but with appropriate treatment, it may be as low as 0.2%.[3] A number of factors are responsible for the rapid spread of dengue infections in recent years. These include increasing international travel and trade; unplanned urbanization, which frequently gives rise to slums; and, global warming. At a local level, an important cause of dengue infections is the lack, or unreliable supply, of piped water, which gives rise to the practice of storing water in receptacles kept in or close to houses. These water storage containers form ideal habitats for the aquatic stages of the *Aedes* mosquito life cycle.

Vectors

The dengue viruses are transmitted to the human host exclusively by mosquitoes belonging to the *Aedes* (*Ae.*) genus. The main vector is *Ae. aegypti* but other mosquitoes belonging to the *scutellaris* group may also act as vectors. *Ae. aegypti* has a worldwide distribution, limited only by climatic conditions. Thus, it is generally confined to parts of the world bounded by the latitudes 35°N and 35°S (Fig. 1), which correspond approximately to the winter isotherms of 10°C; it is not usually found above 1000 meters. During warmer periods, *Ae. aegypti* may thrive beyond its traditional boundaries and has been known to cause outbreaks of dengue infections in such places as Mediterranean Europe.

Ae. albopictus belongs to the *scutellaris* group of the *Ae.* genus and has an Asian origin. With increasing global traveling and trade (notably in such items as used tyres) it is being dispersed to a growing number of

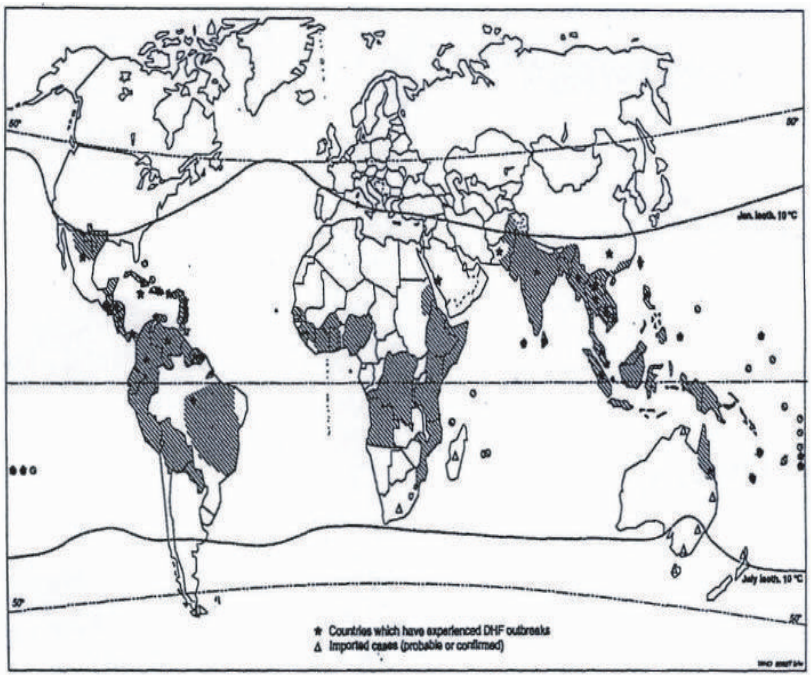


Fig. 1 The general distribution of dengue fever and/or dengue hemorrhagic fever, 1975–1996.
(Reproduced with permission from WHO[1] 1997, Fig 1.1)

countries in the world, thereby assuming increasingly greater significance as a transmitter of dengue infections. *Ae. albopictus* is now not only found in Asia but also North and South Americas, Southern Europe and the Pacific Islands. Other mosquitoes in the *scutellaris* group capable of acting as vectors for the dengue viruses include *Ae. polynesiensis*, *Ae. scutellaris* and, possibly, *Ae. cooki* and *Ae. hebrideus*.

PATHOGENESIS

After the initial bite from the mosquito vector and virus entry into the subcutaneous/dermal space, it is thought that the virus replicates initially in cells of the reticuloendothelial system and/or local fibroblasts. The virus then disseminates and can be detected at the same time as the onset of clinical symptoms in classic DF. However, in DHF/DSS, the viremia is usually over and the clinical manifestations are most likely

due to the immune response to the primary viremia. The virus circulates freely in the plasma, but is also present in circulating peripheral blood mononuclear cells (monocytes/macrophages), and B-lymphocytes. The virus has also been isolated from Kupffer cells and hepatocytes in the liver in several studies, though these have been limited in number, and sometimes contradictory. It has not yet been reliably documented that dengue virus can be isolated from human neurons, though there have been reports of dengue being isolated in CSF samples.[4]

Cytokines

Infection of human host cells leads to various cytokines being produced by the combined immune response of the innate (non-specific, e.g. macrophages, and complement activation) and adaptive (T-cell and specific B-cell responses) immune systems. The cytokines thought to be mainly involved with DHF/DSS are tumor necrosis factor-α (TNF-α), interferon-γ (IFN-γ), and interleukins (IL): IL-1, IL-2 and IL-6. A recent study in Thai children found an IFN-γ response in all subjects, whether or not they subsequently required hospitalization, but TNF-α responses were detected only in four subsequently hospitalized children.[5] Platelet-activating factor (PAF), complement components C3a and C5a, and histamine can induce similar symptoms to those seen in patients with DHF (Gubler and Kuno[6] 1997, Chapter 13.) Dengue-infected endothelial cells may also be destroyed by complement in the presence of cross-reactive antibodies and undergo apoptosis.[7] A recent study on Thai children infected secondarily with DEN 3 has demonstrated that DHF (as measured by degree of plasma leakage and thrombocytopenia) is correlated with early peak plasma IFN-γ levels, and higher plasma viremia levels early in illness.[8]

Antibodies

It is now strongly believed that a combination of B- and T-cell responses underlie the pathology of DHF/DSS. The antibody response to dengue infection will provide lifelong immunity to that particular serotype, by neutralizing that virus. When cross-reacting (heterotypic) or low-levels of neutralizing antibodies bind to cells bearing Fc-γ receptors (i.e. the surface receptor for the Fc portion of IgG), such as macrophages, more dengue virus may be bound. This enhanced binding ("antibody-dependent

enhancement", ADE) enables virus entry and an increased level of replication within these cells leading to an increased viremia and an enhanced specific immune response with increased cytokine production. Studies in infants have shown that maternally-transferred DEN 2 antibodies initially protected infants against DEN 2 infection, but then became enhancing DHF/DSS later on when their levels wanted.[9,10] There is also evidence that the cross-reactive nature of antibodies to other flaviviruses may act also to enhance dengue virus infectivity in patients previously exposed to these other viruses, e.g. yellow fever (Gubler and Kuno[6] 1997, Chapter 11). In contrast, Iquitos, Peru, in 1995, an epidemic of DEN 2 after an epidemic of DEN 1, five years earlier did not lead to any DHF/DSS cases. It was subsequently found that pre-existing sera from these patients were able to neutralize the American DEN 2, but not the South-East Asian DEN 2 strains.[11] This demonstrates that it is the specific antigenic makeup and degree of antibody cross-reactivity that determines the outcome of secondary dengue infections, rather than specific viral virulence. However, the ADE hypothesis does not explain the pathogenesis of DHF/DSS completely, as there are reports of primary dengue infection causing DHF/DSS,[3] and of secondary infections with different dengue serotypes that resulted in no cases of DHF/DSS.[7]

Hemorrhage

The cause of the hemorrhagic manifestations of DHF/DSS is complex and multifactorial. Cytokine and ADE mediated vascular injury, as explained above, may lead to thrombocytopenia (reducing clotting), enhanced permeability (increased bleeding) and there is some evidence of production of procoagulants as well as plasminogen activator inhibitor, from dengue-infected monocytes.[7] A recent study[12] in Vietnamese children with DSS showed significant abnormalities in all major pathways of coagulation cascade, and postulated that both plasma leakage of coagulation proteins, and endothelial, platelet and/or monocyte activation leading to production of procoagulant factors, play a part in the pathogenesis of hemorrhagic diathesis. Disseminated intravascular coagulation (DIC) may be the result of all these factors interacting, as well as possible direct viral liver damage, leading to reduced production of vitamin K-dependent clotting factors, resulting in hemorrhage.[7]

Viral/Host Factors

These factors have been postulated to explain why not all cases of DHF/DSS have been associated with a second infection, and why only a small proportion of second infections proceed to DHF/DSS. Within particular epidemics (e.g. Cuba 1981), the percentage of DHF/DSS cases seemed to increase during the epidemic, which may indicate an increased virulence of the virus with rapid passage through the susceptible human population.[7] This may have been due to the evolution of certain escape mutants, which instead of being neutralized by pre-existing heterotypic antibodies, had evolved sufficiently different epitopes to undergo ADE, giving rise to DHF/DSS cases later on in the epidemic. Host factors that may play a role in the pathogenesis of DHF, include certain HLA types: HLA-B "blank"; HLA-A1, HLA-Cw1, were found more often in the Cuban patients with DHF, and HLA-A29 less frequently;[7] racial differences were also noted in the Cuban epidemic, with fewer blacks developing DHF than whites. Pre-existing medical conditions such as asthma, diabetes mellitus, sickle-cell anemia may increase the risk for severe disease.[7] Other constitutional factors such as age[13] and malnutrition may affect the incidence of DHF occurring in a population, though whether this is mediated through a less robust immune response, is unclear.[3,7]

CLINICAL FEATURES AND DIAGNOSIS

The spectrum of clinical features caused by dengue infections ranges from asymptomatic infection to the potentially fatal DSS. Traditionally, clinical descriptions of dengue infections are separated into DF, and DHF/DSS (see Fig. 2). It must be emphasized that this division is artificial and relies on the presence of significant hemoconcentration, the presence of which puts the disease into the DHF/DSS clinical category. It is virtually impossible to predict in the early stages of dengue infection whether it will progress to DHF/DSS; hence, it is extremely important to monitor the patient closely right from the beginning. In the early stages of both DF and DHF, differential diagnoses to be considered include a wide spectrum of viral (e.g. influenza, chikungunya, measles), bacterial (e.g. typhoid fever, leptospirosis) and other (e.g. malaria) infections. The diagnosis of dengue infection is made on a combination of clinical features and laboratory tests.

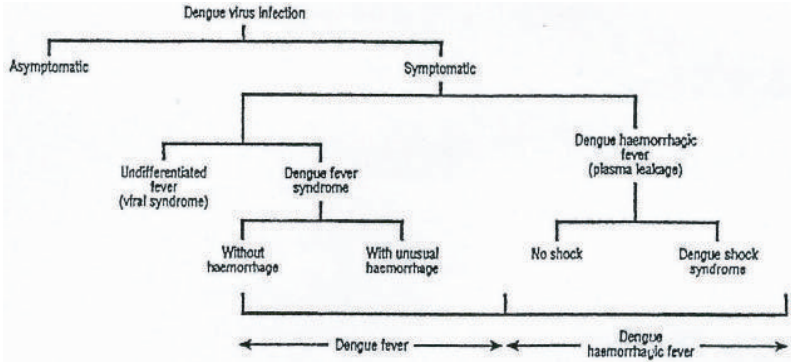


Fig. 2 Manifestations of dengue virus infection.
(Reproduced with permission from WHO[1] 1997, Fig. 2.1)

Classical Dengue Fever (DF)

Classical DF affects particularly older children and young adults. In Singapore the median age has shifted from 14 in 1973 to 28 in 1997, as a result of an effective vector control program put in place in the previous 2 decades.[14] Classical DF is an incapacitating febrile illness with an acute onset. The incubation period is usually 4–7 days but may range from 3–14 days. There may be a prodromal period of some 6–12 hours when the patient may complain of headache, malaise, backache, chills, fatigue, anorexia, and, occasionally, a rash. With the onset of the fever the prodromal symptoms intensify and there is often severe headache with retro-orbital pain, conjunctival congestion and severe myalgia with intense backache ("breakbone fever"). Fever generally lasts from 5–7 days and may reach a maximum of 40°C. Sometimes the fever shows a "saddle-back" diphasic pattern. In the great majority of cases the illness follows a benign course of 5–7 days and symptoms subside rapidly after defervescence. A rash usually occurs, which may be maculopapular, a generalized flushing of the skin; or, more characteristically, it may take the form of a general erythema surrounding areas of normal skin ("islands of sparing"). The tourniquet test is usually positive and petechiae (usually associated with low platelet count) may be noted in the extremities (more often the lower limbs) and on the buccal mucosa and palate. Other non-specific symptoms include mild upper respiratory tract symptoms, epistaxis, anorexia, diarrhea, disturbance of taste, profuse sweating and delirium. There is usually lymphadenopathy often affecting the posterior

auricular nodes and the cervical chains. Hepatomegaly is present in 10–30% of patients. The enlarged liver is soft and tender and may be associated with right hypochondrial pain. Splenomegaly is uncommon. More serious hemorrhages may occur, including menorrhagia. Gastrointestinal hemorrhage is usually occult but, on rare occasions, may be massive, leading to circulatory failure when it should be distinguished from DHF/DSS. In the former situation, there is usually abdominal pain and tenderness and the patient develops clinical feature of hypovolemia despite a fall in hematocrit. Intracranial hemorrhage may occur but is extremely rare. Neurological symptoms are uncommon in DF but have been reported and include headaches, dizziness, restlessness, irritability, insomnia and more seriously, delirium and altered consciousness states, consistent with encephalopathy (or encephalitis[4]). Reye's syndrome and hemophagocytic syndrome are also known to be associated with dengue infections, in rare cases.

The diagnosis of probable DF[1] is based on the presence of an acute febrile illness plus two or more of the following clinical features: Headache, retro-orbital pain, myalgia, arthralgia, rash, hemorrhagic manifestations, or leucopenia. It should also be accompanied by either positive serology on single serum specimen (consisting of one of the following: Titer $\geqslant 1280$ with hemagglutination inhibition test, a comparable IgG titer with enzyme linked immunosorbent assay (ELISA), or a positive titer for IgM antibody test), or occurrence at the same time and location as other confirmed cases of DF.

Dengue Hemorrhagic Fever/Dengue Shock Syndrome (DHF/DSS)

DHF/DSS is distinguished from DF by the presence of features of *plasma leakage* causing hemoconcentration (see Fig. 3). These features (sufficient for making a diagnosis of DHF/DSS) are: The hematocrit must be $\geqslant 20\%$ above the average value for the patient; or must rise by $\geqslant 20\%$ during the course of the illness; or fall by $\geqslant 20\%$ as the patient is given fluid replacement therapy. The initial symptoms of DHF/DSS are very similar to those of DF, but maculopapular rash and arthralgia are observed less frequently. Hemorrhagic manifestations are more severe than in DF and massive gastrointestinal bleeding may occur. The clinical features of DHF tend to manifest themselves at or shortly after defervescence. The patient

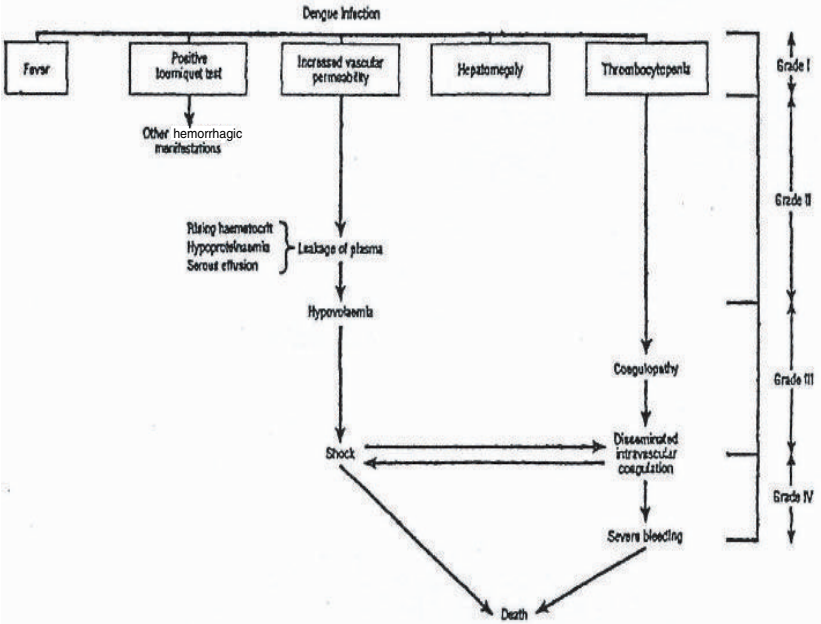


Fig. 3 The spectrum of dengue hemorrhagic fever.
(Reproduced with permission from WHO[1] 1997, Fig. 2.2)

exhibits clinical signs of circulatory disturbances with restlessness, sweating, cold, clammy skin, tachycardia and a fall in blood pressure. Thrombocytopenia is invariably present and may be severe, and there is usually disturbance of blood coagulation with prolonged prothrombin and partial thromboplastin time. With early and appropriate fluid replacement therapy, the patient generally recovers. If treatment is delayed, the disease may progress to the DSS. DSS is characterized by signs of severe circulatory failure with rapid weak pulse; narrowing of the pulse pressure ($\leqslant 20$ mmHg); and *hypotension*: Defined as a systolic pressure of < 90 mmHg for those 5 years or older, and < 80 mmHg for those less than 5 years old. There may be ascites and pleural effusion due to plasma leakage. The duration of shock is short-lived, usually no more than 48 hours. With appropriate fluid replacement, the patient recovers rapidly. If treatment is delayed, profound shock supervenes and this may be complicated by DIC, metabolic acidosis and multi-organ failure including renal and hepatic failure.

DHF may affect other organs. The liver is more frequently involved than in DF. Neurological manifestations include encephalopathy and encephalitis,[4] disturbances of motor and sensory function and, as a late complication, the Guillain–Barre syndrome. The association of severe liver disease and encephalopathy has been documented in both children and adults. If excessive fluid has been administered during correction of plasma leakage, pulmonary edema may occur when plasma leakage ceases and fluid is absorbed. In the heart, rhythm disturbances and ECG changes have been noted. In one report, 44% of Singapore children with DHF had ECG abnormalities.[15] Hematuria may occur but is less common than gastrointestinal hemorrhage. Thickening of the gallbladder wall has been noted in pediatric patients and is thought to be due to edema of serous membranes.

DHF may be diagnosed in the presence of the following features[1]: Fever lasting 2–7 days; hemorrhagic diathesis (as shown by a positive tourniquet test, or petechiae/ecchymoses/purpura, or bleeding from mucosa/gastrointestinal tract/injection sites/other locations, or hematemesis or melena); thrombocytopenia ($\leq 100\,000\,mm^3$); evidence of *plasma leakage* (as defined above). The WHO gives, as an additional feature of plasma leakage, the presence of pleural effusion and/or ascites, which may be used for confirming the diagnosis of DHF in places where hematocrit cannot be easily measured. In these patients, there is usually hypoalbuminemia.

DSS is diagnosed when there is evidence of circulatory failure in addition to the criteria for DHF. The WHO[1] proposes the following features as manifestations of *circulatory failure*: Either tachycardia together with narrowed pulse pressure ($\leqslant 20\,mmHg$), or hypotension for age (as defined in page 833) together with cold, clammy skin and restlessness.

DHF/DSS is graded into four grades of severity by the WHO[1] as follows:

1) *Grade I*: Fever and non-specific symptoms plus minor bleeding diathesis (positive tourniquet test and/or easy bruising);
2) *Grade II*: Spontaneous bleeding (skin or other hemorrhages) in addition to grade I features;
3) *Grade III*: Clinical signs of circulatory failure in addition to above features; and
4) *Grade IV*: Profound shock with undetectable pulse or blood pressure.

LABORATORY INVESTIGATIONS

Non-virological Tests

In DF, the initial white cell count may be normal or slightly increased, but falls in the next 2–3 days with neutropenia. There is monocytosis with an excess of atypical mononuclear cells. Mild to moderate thrombocytopenia (defined as a platelet count of $\leq 100\,000\,mm^3$) is seen in a significant proportion of patients. The hematological abnormalities seen in DHF are similar to those seen in DF, but thrombocytopenia is a constant feature and may be profound. In DHF, prolongation of prothrombin time (PT) and partial thromboplastin time (PTT) is usually present. In severe cases, laboratory evidence of DIC may be present. By definition, DHF is associated with plasma leakage as defined earlier. Hypoproteinemia/hypoalbuminemia is generally present and there may be a reversed albumin to globulin ratio. Elevation of liver transaminases is common and is usually higher than in DF. Characteristically (as also in DF), the level of aspartate aminotransterase (AST) is higher than that of alanine aminotransferase (ALT).

Serology

The initial serological response to primary dengue infection consists of an elevation of dengue-specific IgM 5–7 days after onset of fever, peaking about 2 weeks later, then declining to undetectable levels after 2–3 months.[1] In the absence of dengue-specific IgM testing, the traditional four-fold rise in dengue-specific IgG titers on paired acute and convalescent sera (2–4 weeks later), using a complement fixation test, hemagglutination inhibition test, or a type-specific plaque reduction neutralization assay, will confirm a recent primary dengue infection. Nowadays, commercial testing kits are available for both dengue-specific IgG and IgM testing, generally in an enzyme immunoassay (EIA) format.[16] However, these EIAs are not usually quantitative and cannot be used alone to measure a four-fold rise in dengue-specific IgG titer on paired acute and convalescent sera. It may be possible to characterize severe dengue infections using a combination of antibody profiles, which may allow prediction to DHF/DSS disease in some patients,[17] though this is still experimental.

Virus Isolation and Antigen Detection

Where samples have been collected within 5 days of illness, before substantial amounts of antibody have appeared, viral detection methods can be used for diagnosis. Dengue virus culture gives the best results by inoculation of mosquito tissue, either into mosquitoes themselves, or into mosquito cell lines. The mosquitoes are incubated for 14 days, after which the virus can be detected by immunofluorescence stains applied to a smear of the squashed mosquito head on a glass microscope slide. With high quality specimens, the sensitivity of mosquito culture can be as high as 80% (Gubler and Kuno[6] 1997, Chapter 15). Increased detection sensitivity can be achieved by using up to 5–20 mosquitoes per specimen.[1] Less sensitive, though easier to use, mosquito cell lines are a practical compromise, especially where the laboratory has no insect maintenance service available. Not all dengue viruses produce a visible cytopathic effect in these cell lines and such cell cultures need to be screened by EIA, or fluorescent antibody testing, to demonstrate presence or absence of virus. Incubation periods vary from 5–10 days, depending on the cell line used. Incubation at ambient tropical temperatures allows culture tubes to be brought to the bedside or into the field, though a sterile culture environment is required to avoid contamination.[1] Virus isolation by culture is still relatively insensitive when compared to molecular techniques (see below) — only 36% of confirmed cases have been detected using C6/36 cell lines (Gubler and Kuno[6] 1997, Chapter 15).

Detection of Dengue RNA

The detection of dengue RNA by reverse-transcription polymerase chain reaction (RT-PCR) is becoming more common.[18] One reason is the extreme sensitivity of this approach, which allows amplification of existing dengue RNA by up to a million-fold, allowing an earlier diagnosis, as dengue RNA may be detectable within the first 5 days of infection. Another advantage is that viable virus is not required in the sample, which makes transportation easier, as only the presence of dengue RNA is detected. The RT-PCR also has the ability to simultaneously serotype the virus, by using serotype-specific primers, allowing molecular epidemiological tracking of the dengue virus,[18] especially useful in outbreak situations. Finally, RT-PCR can detect dengue RNA within immune complexes in convalescent samples, as the protein component of the immune complex is removed

during the RNA extraction step, though this is less effective than testing a sample taken in the first 5 days of illness, before an antibody response appears.[19] An interesting application of RT-PCR is the quantitation of dengue viremia in patients. In a recent study of DEN 3 infection in Thai children, a correlation was found between the level of plasma dengue viremia and the severity of clinical diseases.[8] Although studies are lacking, an obvious clinical application would be the rapid inclusion or exclusion of DF early in the course of an undifferentiated viral illness. This has important infection control and public health implications when epidemiological circumstances make SARS a diagnostic possibility.

MANAGEMENT

The management of both adult and pediatric patients with dengue infections follow similar principles, but differ in details. The emphasis of this chapter is on adult patients.

The main pathophysiological abnormality in dengue infection is an increased in vascular permeability leading to plasma leakage and contraction of intravascular volume. This may lead to circulatory failure, ending, in a minority of cases, in DSS. Bleeding diathesis is associated with dengue infections and is due to one or more of the following factors: Capillary fragility, thrombocytopenia and coagulopathy. Currently there is no effective anti-viral agent for treating dengue infections; nor is there a vaccine suitable for clinical use. Treatment of dengue infection is therefore symptomatic and supportive.

Dengue Fever

This is an unpleasant but benign illness from which complete recovery is the rule. Fever, anorexia, nausea and vomiting may lead to dehydration which may be corrected by adequate intake of oral fluid. Analgesics are required for pain control and antipyretics may be given when the temperature rises above 39°C. In patients with a history of febrile convulsions, the threshold for the use of antipyretics may be lowered. Aspirin and non-steroidal anti-inflammatory drugs (NSAIDs) should be avoided because of the risk of a bleeding diathesis. Aspirin may also precipitate Reye's syndrome. Paracetamol is an appropriate analgesic and antipyretic. In the rare cases of unusual hemorrhage (e.g. gastrointestinal), blood transfusion is required.

Dengue Hemorrhagic Fever

Grade I DHF may be treated as an outpatient and the treatment is similar to that for DF. In Grade II DHF, there is not only increased bleeding diathesis but plasma leakage is more severe than in Grade I DHF. In the absence of signs of circulatory disturbance, oral hydration may be given initially. The oral hydration solution may be that used for treating diarrheal diseases, supplemented, if necessary, with fruit juices and water. In both Grades I and II DHF, the patient should be closely monitored for signs of circulatory failure (as defined earlier). If the patient is treated as an outpatient, the physician should ensure that care-givers are adequately trained to recognize such signs. The critical period is around the time of defervescence. A baseline platelet count and hematocrit should be done at presentation, and thereafter daily from the third day of illness.

Patients with signs of significant plasma leakage requiring parenteral fluid therapy should preferably be admitted for inpatient treatment. However, in areas where hospital beds are scarce, this may be undertaken on an outpatient basis, provided there is an adequate setup for this form of treatment. The traditional parenteral replacement fluids recommended by the WHO[1] are 5% dextrose (D) diluted 1:1 or 1:2 in normal saline solution (NSS), or 5% dextrose in Ringer's lactate (RL) or Ringer's acetate (RA) solutions. For a child, the daily fluid requirement is the sum of the daily maintenance fluid (calculated from age and body weight: Formula available from standard pediatric textbooks) plus the estimated amount of replacement fluid. The latter is calculated on the basis of 10 ml/kg for each 1% of body weight loss. Thus for a two-year-old child with an estimated 5% fluid loss from Grade II DHF, and whose normal body weight is 10 kg, the total daily fluid requirement is calculated as follows: Daily maintenance fluid ($10 \times 100 = 1000$ ml), plus estimated fluid loss ($10 \times 10 \times 5 = 500$ ml), giving a total of 1500 ml/day. As fluid replacement proceeds, careful and frequent monitoring of pulse rate, urine output, blood pressure and hematocrit should be made and a close lookout kept for clinical signs of circulatory failure (cold, clammy skin, restlessness and narrowing of pulse pressure). Vital signs should be monitored every 1–2 hours and hematocrit every 2–3 hours. If such monitoring is not possible in an outpatient department, patient should be admitted for inpatient management. The rate of plasma leakage is not constant and may vary over short periods of time. Therefore, the volume of replacement

fluid and its rate of administration should be reviewed regularly throughout the day. For a flow chart of fluid replacement regime, see Fig. 4. Extreme care should be taken not to over-hydrate the patient as this may lead to acute pulmonary edema and other signs of fluid overload when fluid is reabsorbed into the intravascular compartment during recovery.

Inpatient care for DHF is mandatory when the plasma leakage is estimated to be $>10\%$. This is likely when the following signs are present: Tachycardia out of proportion to body temperature; delayed capillary filling (>2s); cold, clammy and mottled skin; continued rise of hematocrit despite intravenous fluid replacement; sudden rise in hematocrit; falling

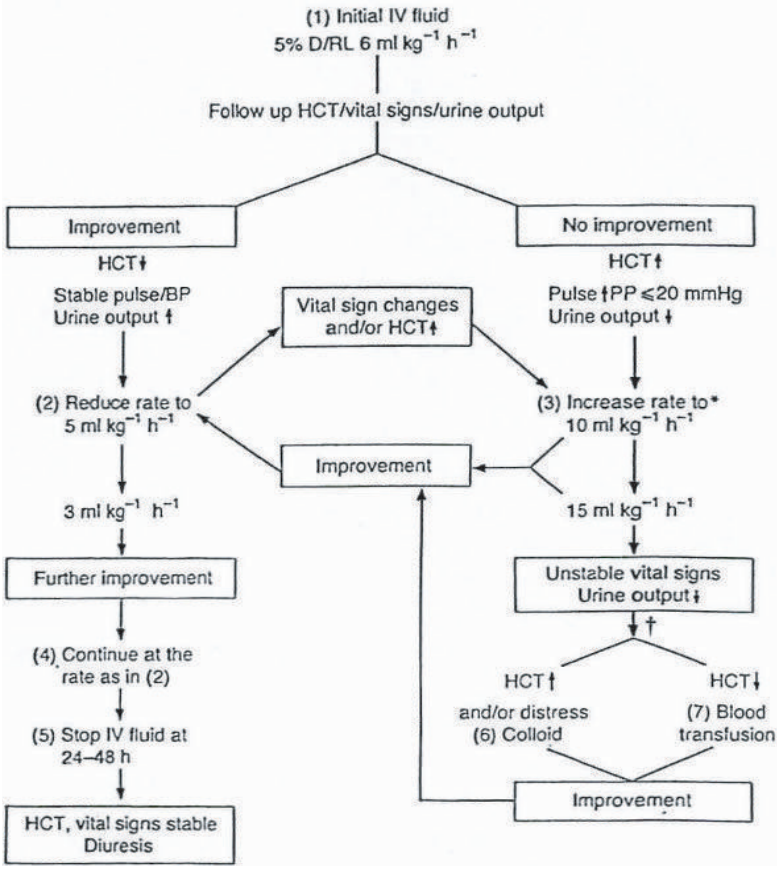


Fig. 4 Volume replacement flow chart in dengue hemorrhagic fever. HCT = hematocrit; PP = pulse pressure; *with signs of shock; † establish CVP catheter and urinary catheter.
(Modified with permission from Gubler and Kuno[6] 1997, Chapter 7, Fig. 7.2)

urine output; and narrowing of pulse pressure (⩽20 mmHg). These features indicate that the patient is progressing into DSS.

Dengue Shock Syndrome

This is a medical emergency and urgent fluid replacement (see Fig. 5) is needed. The initial intravenous fluids to be used are similar to those used in DHF. The fluid should be given as a bolus, in a volume of 10–20 ml/kg, over no more than 20 minutes. If this fails to bring about improvement in the patient's condition, a further bolus should be given bringing the total volume to 20–30 ml/kg. The patient should be given supplementary oxygen. If shock persists and hematocrit continues to rise then a plasma substitute (e.g. Dextran 40) or albumin should be given in a bolus of 10–20 ml/kg, repeated if necessary, to a total of 20–30 ml/kg (see Fig. 5). Internal hemorrhage should be suspected if shock remains but hematocrit declines. When clinical features of shock regresses, further intravenous fluid infusion should be adjusted according to hematocrit level and urine output. DSS may be associated with acid-base and electrolyte disturbances. These should be appropriately treated. Intravenous fluid replacement generally

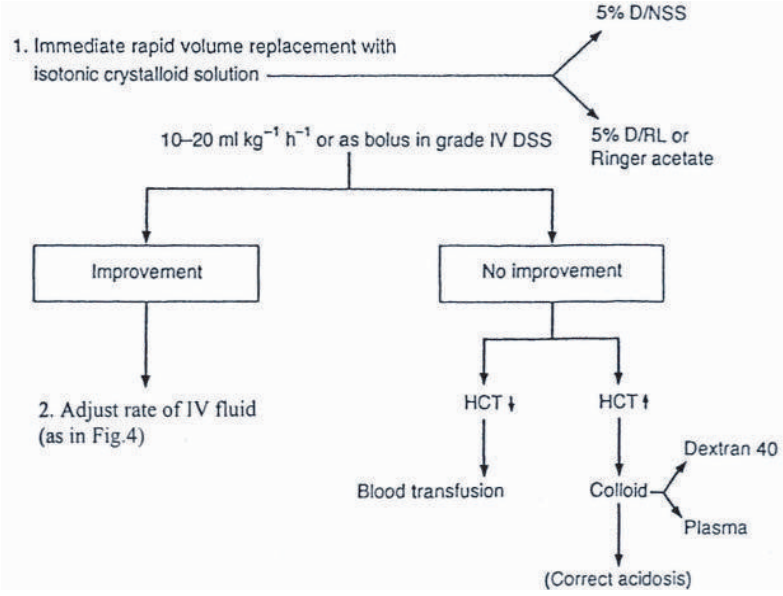


Fig. 5 Diagram for treatment of dengue shock syndrome, grades III and IV. (Reproduced with permission from Gubler and Kuno[6] 1997, Fig. 7.3)

does not need to be given for more than 48 hours following regression of shock. It should be discontinued when hematocrit falls to around 40%. Systemic steroids have not been found to be useful in DSS. If the patient remains well — and in the presence of good blood pressure, pulse volume and urine output — progressive fall of hematocrit towards the normal level is a good sign indicating resorption of extravasated fluid into the intravascular compartment, and should not be confused with internal hemorrhage.

The choice of intravenous fluids as described above follows traditional practices recommended by the WHO.[1] A recent study[20] suggested that normal saline may be used with equal, or possibly better, outcome in place of these traditional crystalloids, and Dextran 70 may replace Dextran 40. Increasingly, in the management of both adult and pediatric patients with DHF, normal saline is being used as the initial parenteral replacement fluid, with addition of other fluids (e.g. 5% dextrose), where needed. Blood transfusion is needed if there is internal hemorrhage. Fresh whole blood is preferable in view of the associated coagulopathy. The latter may need additional fresh frozen plasma for correction. Platelet transfusion may be needed for severe thrombocytopenia — in one of the author's (**TOT**) department, platelet transfusion is routinely given when the count is $< 20\,000/mm^3$. Profound shock in DSS may be complicated by DIC and multi-organ failure, notably the liver and kidney. Such patients should be managed in an intensive care unit and may need renal dialysis, support for the liver and treatment of the DIC. In such cases, management by a multi-disciplinary team offers the best chance of success. Invasive procedures should be kept to a minimum in view of the bleeding diathesis. Plasma leakage, even in DSS, usually lasts for no more than 48 hours. If the patient survives this period, recovery is usually complete.

A patient with DHF/DSS should only be discharged after at least 2 days following recovery from shock, and 24 hours after the fever has settled without antipyretics. The hematocrit, pulse and blood pressure should be stable and within the normal range. Appetite should have returned and there should be good urine output. Platelet count should be $\geqslant 50\,000\,mm^3$. If the patient has ascites and/or pleural effusion, these should have cleared.

PATHOLOGY

The pathology of DHF/DSS has been studied at autopsy and also on tissues obtained from biopsies. Hemorrhage is a constant finding at

postmortem. It is located largely at the skin and mucosa of upper respiratory and the gastrointestinal tracts, but is also found subcapsularly in the liver. Frank hemorrhage into serous cavities is uncommon but free fluid in pleural and abdominal cavities is often seen and has a high protein content ($\geq 4\,g/dL$). Macroscopic examination of skin shows perivascular hemorrhage around small vessels (capillaries and venules) and infiltration by mononuclear cells and lymphocytes. The dermal papillae appear to be the main site of injury. No pathological changes in vessel walls are seen by light microscopy. The liver is enlarged in about 30% of cases at autopsy. Histological changes are similar to those seen in yellow fever, consisting of focal necrosis distributed in the paracentrolobular or midzonel regions. Kupffer cells show swelling and hyaline necrosis. Another feature is the presence of Councilman bodies, which are round or ovoid acidophil bodies with or without nuclear remnants found lying free in sinusoids, space of Disse or liver cell plates. They are believed to represent apoptotic hepatocytes and were first described by Councilman in yellow fever. Renal biopsy shows immune-complex glomerulonephritis, which appears to be transient. Bone marrow examination during the febrile phase of the illness shows general depression of activity affecting most blood elements. In the lymphoid tissues, there is increased activity of the lymphocytes with proliferation of plasma cells and appearance of active germinal centers. In the central nervous system there is perivascular edema and hemorrhage. Traditional descriptions of dengue pathology usually state that encephalitis and meningitis are not a feature; however, a recent publication[4] reported 6 cases of encephalitis in children. Although no pathological examination of brain tissues was made in this series, this traditional view may need to be reviewed if future postmortem examination of similar patients shows features of encephalitis.

PREVENTION

Mosquitoes in the *Aedes* genus share a life cycle in which development of the larvae and pupae takes place in water. Effective vector control should address both the adult population and their aquatic precursors. An effective surveillance program should monitor dengue infections in the human host, and population of adult mosquitoes and their larvae. Surveillance can be of either the disease or the vector.

Disease Surveillance

In general three surveillance methods are used in human disease: Passive, active and sentinel. *Passive surveillance* relies on notification of disease by healthcare workers. Such a disease is usually designated a notifiable disease and notification is required by law. In *active surveillance*, personnel are assigned to retrieve data on the disease. It is more successful than passive surveillance provided sufficient workers are employed to meet the size of the task. In *sentinel surveillance*, certain pre-designated sources are assigned to provide information on key events. Studying patterns of disease behavior longitudinally may give early warning of possible development of an epidemic. Surveillance data are used to formulate effective control and preventive measures.

Vector Surveillance

Surveillance of adult mosquito population uses several indices of which the house index (HI) is one in common use. This is defined as the number of houses positive for containers with *Ae. aegypti* larvae. The HI gives a rough, indirect estimate of the adult mosquito population indoors. In Singapore, when the HI rises above 2%, indoor thermal fogging with insecticides will be carried out.[9] Another method of surveillance consists of collecting adult mosquitoes from houses using a battery-operated backpack aspirator, which sucks mosquitoes living indoors, and provides a direct estimation of the mosquito population. This method may also be used to gauge the effectiveness of control measures. However, it is rather demanding and requires trained personnel. A further indirect method of estimating the mosquito population is to use ovitraps. These are small vessels containing water (which may be enriched with, e.g. hay solution, to attract mosquitoes) laid out in fields. Mosquitoes lay their eggs on the walls of these traps and the egg count provides an estimate of the adult population.

CONTROL

Ae. aegypti lays its eggs on the damp walls of water containers and can resist desiccation for months. They hatch when immersed in water. Their development into larvae and pupae depends on the temperature of the

water; at 25°C the larval and pupal stages last 7–9 days and 2–5 days respectively. An effective program of control must aim at reducing the population of both the adult and the aquatic stages of the mosquito life cycle.

Control of the Larval Population

One of the single most important measures in source reduction is the provision of a reliable piped water supply to homes. This eliminates the need to store water in receptacles, which form ideal larval habitats. Water storage vessels should have tight fitting lids to prevent entry of mosquitoes to lay eggs. Water in flower vases should be changed frequently and water should be regularly emptied from trays placed under flowerpots. Gutters should be cleaned regularly to prevent accumulation of stagnant waters. Outdoor discarded solid wastes (e.g. broken bottles, cans, and plastic containers) which can hold water should be properly disposed of. Used tyres are good larval habitats and should not be left on open ground. Tree holes near houses should be filled. Mechanical devices are available for the control of mosquito larvae and one example is the autocidal trap, developed in Singapore. This is an ovitrap with a floating mesh that prevents eclosion of the pupae that develop from eggs. If a sufficiently large number of these devices are placed in the environment, they will compete with natural larval habitats, thereby reducing the number of larvae that develop into adult mosquitoes. Certain chemicals (e.g. temephos and methoprene) may be added to water at concentrations that are harmless for human consumption, but are, nevertheless, larvicidal. Biological control has also given encouraging results in pilot studies. An example is the copepod Mesocyclops that feeds on newly hatched *Ae. aegypti* larvae. Another is the biocide *Bacillus thuringiensis* H-14 (BTI), a sporing bacterium, which contains a toxin that degranulates solely in the alkaline midgut of mosquito. It kills larval mosquitoes but appears to be entirely safe for humans.[1]

Control of Adult Mosquito Population

This relies on the use of insecticides. Residual insecticides are chemically stable and when applied to surfaces will kill mosquitoes that alight on them. Although undoubtedly effective, their widespread use is labor

intensive. This method is reserved for application around an area where dengue infections have occurred ("perifocal application"). On a smaller scale, insecticides may also be used to impregnate mosquito nets and sprayed from an aerosol gun or pressurized canister in the home. Larger scale indoor spraying of insecticides is used in several countries during dengue outbreaks. Apart from the use of mosquito nets impregnated with insecticides, other methods of personal protection include the fitting of mosquito screens to doors and windows and the use of repellants such as DEET, which can be applied to the skin.

VACCINE DEVELOPMENT

Currently, there is no internationally approved vaccine against DF, though this is one of the WHO's priorities. The main problem arises from the necessity of creating a tetravalent vaccine that will induce protection against all 4 serotypes of dengue virus, to avoid any possibility of ADE, as there is little cross protection between the 4 serotypes. The lack of a suitable, convenient animal model for DHF, and the poor growth of the virus in culture, adds to the difficulty of developing such a vaccine.[2] One of the most promising dengue vaccines is a live-attenuated tetravalent vaccine developed at Mahidol University, Salaya in Thailand, with the support of the WHO.[21] Preliminary trials in adults and children with these tetravalent vaccines induced cross-reacting antibody in 80–90% of subjects and were shown to be safe. These findings were repeated in US volunteers at the Walter Reed Army Institute for Research (WRAIR). On the basis of these successful Phase I clinical trials, Phase II and Phase III clinical trials with the live-attenuated vaccine are now in progress in Thailand and elsewhere.[21] Other approaches to dengue vaccine development include: Infectious clones, inactivated, subunit and nucleic acid vaccines, which are in various stages of development. All have met with some success, but still lag some way behind the tetravalent, live-attenuated approach described above, which (barring any serious early adverse reactions) will probably be the first international vaccine against dengue to be licensed. Most recently, a chimeric dengue vaccine has been created from the existing live-attenuated yellow fever vaccine, by inserting the dengue premembrane and envelope genes into the non-structural portion of the yellow fever 17D vaccine.[21,22] Further developments are eagerly awaited in this field.

ACKNOWLEDGMENTS

The authors would like to thank Professors SK Lam and LC Lum (Department of Medical Microbiology and Paediatrics, respectively, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia), Dr SB Halstead (Uniformed Services University of the Health Sciences, Bethesda, USA), Drs J A Garson and E Aarons (Royal Free and University College Medical School, London, UK) and Ms P Koraka (Erasmus Medical Centre, Rotterdam, The Netherlands) for critically reviewing the manuscript.

REFERENCES

1. World Health Organization, *Dengue Haemorrhagic Fever: Diagnosis, Treatment, Prevention and Control*, 2nd ed., Geneva, WHO, 1997.
2. Gibbons RV, Vaughn DW, Dengue: an escalating problem, *BMJ* **324**: 1563–1566, 2002.
3. Rigau–Perez JG, Clark GG, Gubler DJ, Reiter P, Sanders EJ, Vorndam AV, Dengue and dengue haemorrhagic fever, *Lancet* **352**:971–977, 1998.
4. Lum LC, Lam SK, Choy YS, George R, Harun F, Dengue encephalitis: a true entity? *Am J Trop Med Hyg* **54**:256–259, 1996.
5. Mangada MM, Endy TP, Nisalak A, Chunsuttiwat S, Vaughn DW, Libraty DH, Green S, Ennis FA, Rothman AL, Dengue-specific T cell responses in peripheral blood mononuclear cells obtained prior to secondary dengue infections in Thai school children, *J Infect Dis* **185**:1697–1703, 2002.
6. Gubler DJ, Kuno G (eds.), *Dengue and Dengue Hemorrhagic Fever*, CABI Publishing, CABI International, Wallingford, UK, 1997.
7. Burke DS, Monath TP, Flaviviruses, in *Fields Virololgy*, Vol. 1, Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (eds.), Philadelphia: Lippincott Williams & Wilkins, pp. 1043–1125, 2001.
8. Libraty DH *et al.*, Differing influences of virus burden and immune activation on disease severity in secondary dengue-3 virus infections, *J Infect Dis* **185**:1213–1221, 2002.
9. Kliks SC, Nimmanitya S, Nisalak A, Burke DS, Evidence that maternal dengue antibodies are important in the development of dengue hemorrhagic fever in infants, *Am J Trop Med Hyg* **38**:411–419, 1988.
10. Kliks SC, Nisalak A, Brandt WE, Wahl L, Burke DS, Antibody-enhancement of dengue virus growth in human monocytes as a risk factor for dengue haemorrhagic fever, *Am J Trop Med Hyg* **40**:444–451, 1989.
11. Kochel T *et al.*, Effect of dengue-1 antibodies on American dengue-2 viral infection and dengue haemorrhagic fever, *Lancet* **360**:310–312, 2002.

12. Wills BA *et al.*, Coagulation abnormalities in dengue hemorrhagic fever: serial investigations in 167 Vietnamese children with dengue shock syndrome, *Clin Infect Dis* **35**:277–285, 2002.

13. Guzman MG *et al.*, Effect of age on outcome of secondary dengue 2 infections, *Int J Infect Dis* **6**:118–124, 2002.

14. Goh KT (ed.), *Dengue in Singapore*, Institute of Environmental Epidemiology, Ministry of Environment, Singapore, 1998.

15. Wong HB, Tan G, Cardiac involvement in haemorrhagic fever, *J Sing Paed Soc* **9**:28–35, 1967.

16. Groen J, Koraka P, Velzing J, Copra C, Osterhaus ADME, Evaluation of Six Immunoassays for Detection of Dengue Virus-Specific Immunoglobulin M and G Antibodies, *Clin Diagn Lab Immunol* **7**:867–871, 2000.

17. Koraka P *et al.*, Kinetics of dengue virus-specific serum immunoglobulin classes and subclasses correlate with clinical outcome of infection, *J Clin Microbiol* **39**:4332–4338, 2001.

18. Chow VT, Molecular diagnosis and epidemiology of dengue virus infection, *Ann Acad Med Singapore* **26**:820–826, 1997.

19. Chan SY, Kautner IM, Lam SK, The influence of antibody levels in dengue diagnosis by polymerase chain reaction, *J Virol Methods* **49**:315–322, 1994.

20. Nhan NT *et al.*, 2001 Acute management of dengue shock syndrome: a randomised double-blind comparison of 4 intravenous fluid regimens in the first hour, *Clin Infect Dis* **32**:204–213, 2001.

21. Pugachev KV, Guirakhoo F, Trent DW, Monath TP, Traditional and noval approaches to flavivirus vaccines, *Int J Parasitol* **33**:567–582, 2003.

22. Halstead SB, Deen J, The future of dengue vaccines, *Lancet* **360:** 1243–1245, 2002.

This page intentionally left blank

49

Common Geriatric Problems

Ding Yew Yoong and Philip Choo

The American Geriatrics Society defines geriatrics as the branch of medical science that focuses on health promotion and the treatment of disease and disability in later life. The objective is to restore an ill and disabled person to a level of maximum ability and wherever possible return the person to an independent life at home. Therefore, the main principles and practices in geriatric medicine are:

- *Pre-eminence of function*: Function describes the behaviors necessary to maintain independence in daily life. This encompasses physical, cognitive and psychological domains. Functional status should be assessed routinely in clinical practice as the "sixth vital sign". Although medical diagnoses are important, function is central to the health and well-being of older people. Medical illness can present with functional decline — often as "geriatric syndromes" such as impaired cognition (delirium and dementia), impaired mobility, instability (falls), incontinence and impaired feeding. Functional decline is associated with adverse outcomes such as increased mortality, hospitalization and length of stay, institutionalization and

health expenditure as well as decreased quality of life. Appropriate geriatric intervention can improve some of these outcomes.

- *Multidisciplinary nature of care*: Good care of frail older patients requires good medical management. In addition, nursing, physical therapy, dietetics, pharmacy and social work intervention is often needed to provide comprehensive geriatric assessment (CGA) and management when functional decline is present.
- *Elicitation of preference and goals of care*: The goals of care for older patients are generally more diverse. This is due to heterogeneity in physiological function, health status, beliefs, values, cultural backgrounds and personal preferences. Successful management and satisfying outcomes are better achieved when clinical decision making is based on agreement with patients or their families on the goals of treatment. These goals may change during the same episode of illness through the phases of acute care, rehabilitation and end-of-life care.

SPECIAL CONSIDERATIONS IN OLDER PATIENTS

1) *Individual variability*: Although older patients experience physiological decline in organ function, inter-individual variation is a feature of aging. Knowledge of the baseline (or pre-morbid) function of each patient is important in determining whether a change has occurred.
2) *Multiple pathology*: Older persons often have more than one illness. This can complicate care in the following ways:

    i) Acute illness involving one organ system may put stress on the reserve capacity of another system affected by a pre-existing disease, making interpretation of symptoms and signs less straightforward. (e.g. pneumonia resulting in acute confusion in patients with dementia)
    ii) Clinical symptoms of one disease may be masked by another disease process. (e.g. patients with severe coronary artery disease and significant osteoarthritis may not experience angina because of limited physical activity)
    iii) Treatment of one illness may unmask previously unrecognized pathology. (e.g. drugs with anticholinergic effects may cause urinary retention in older men with asymptomatic bladder outlet obstruction).

3) *Presentation of illness*: There are 2 broad categories.
   i) Usual: the symptoms and signs are similar to those typically seen in younger patients.
   ii) Atypical: symptoms and signs may be different. Disease in one organ system may reflect disease in another system. These presentations of illness represent functional decline and are collectively known as "geriatric syndromes" or the "I"s:
      - intellectual impairment (delirium and dementia);
      - immobility (and deconditioning);
      - instability (and falls);
      - incontinence; and
      - impaired feeding.

   It is important not to dismiss these conditions as part of the aging process but to evaluate them systematically.
4) *Response to illness*: Older persons may have lower expectations of health and function. Many seek medical consultation for only a small proportion of their symptoms. Patients and their carers may mistake impairment as the natural consequences of aging. In view of this, there is a need to seek out specific impairments (especially functional).
5) *Polypharmacy*: The older patient tends to be on more medications at same time. This confers a greater risk of experiencing unwanted effects of medications. Adverse drug reactions should be considered in the differential diagnoses of older patients with recent onset of new symptoms or functional decline. It is essential to ask about and scrutinize all medications taken — including "over-the-counter" and traditional ones.
6) *Narrower limit of diagnostic and therapeutic tolerance*: This arises from losses in homeostatic function and superimposition of disease and disabilities. Diagnostic tests may result in adverse outcomes not usually encountered in younger patients. An example is bowel preparation for colonoscopy causing fluid depletion, resulting in deterioration in renal function, confusion and impaired mobility. Similarly, treatment with benzodiazepines for insomnia may result in falls and confusion. There is less room for misjudgement in clinical care.
7) *Time as an issue of care of older patients*: It cannot be overemphasized that the evaluation and management of frail older patients is more time-consuming. History taking and physical examination takes

longer. There are more comorbid medical conditions to consider. Functional status needs to be assessed and managed. Effective communication with patients and their families also requires more effort and time.

8) *The value of small therapeutic gains*: Cure may not be the aim of management because of chronic disease and disability. Instead, restoration and preservation of function is often the primary objective. Small improvements in function such as ability to transfer independently or improvement of mood with decreased agitation may bring about appreciable improvement in quality of life and satisfaction.

CLINICAL APPROACH IN GERIATRIC MEDICINE

In addition to the usual approach adopted in internal medicine, function is evaluated in the domains of Activities of Daily Living (ADL): Mobility, cognition, mood, continence, swallowing, vision and hearing. Where functional disability is present, a social history that focuses on availability of carers and suitability of the home environment becomes particularly relevant.

Obtaining information from the family or caregiver is important. Information uncovered on function and behavior may otherwise be unavailable.

A brief clinical approach to the major geriatric syndromes follows. Where diagnosis remains unclear or management issues are complex, consultation with geriatricians should be considered.

Impaired Cognition

Impaired cognition is characterized by a decline in mental function. There are 3 main categories of impaired cognition in older people:

1) delirium;
2) dementia; and
3) others: depression and psychosis.

The most useful questions to ask when faced with such a syndrome are:

1) whether the onset is acute (over a few days to weeks) or chronic (over several months to years);

2) whether there are any contributing physical factors: medical illness, drugs, environmental changes; and
3) whether there are any contributing psychological factors: depression or psychosis.

The differentiation between the different categories of impaired cognition is primarily made by history taking. Physical examination helps to confirm the diagnosis and point towards possible underlying causes.

Delirium

This condition is characterized by an acute decline in cognition. It is a common but under-recognized condition in older hospitalized patients. Early identification of delirium is crucial because it is a medical emergency often caused by potentially reversible factors. Prognosis often worsens with delay in treatment.

Evaluation: The Confusion Assessment Method (CAM) is a useful clinical strategy for identification of delirium (see Table 1). It is important to differentiate delirium from dementia because the latter does not require urgent management. Dementia should not be diagnosed in the setting of acute decline or previously unknown cognitive function. When uncertain, delirium should be assumed in the first instance. The causes of the delirium (see Table 2) should be carefully searched for in the history and physical examination. Investigations are selected on the basis of clinical findings. Most patients will require FBC, serum electrolytes and glucose, renal and liver function tests, chest X-ray, ECG and urinalysis. Brain imaging, EEG, lumbar puncture and arterial blood gases are performed where indicated.

Treatment: Hospitalization is usually necessary for diagnosis and treatment of the precipitating condition as well as provision of

Table 1 Confusion Assessment Method (CAM)

The diagnosis of delirium requires the presence of features 1 and 2 and either 3 or 4.
1) Acute onset and fluctuating course
2) Inattention
3) Disorganized thinking
4) Altered level of consciousness

Table 2 Causes of Delirium in Older Patients

Infections
Drugs: Including hypnotics, antidepressants, antipsychotics, antiparkinsonism drugs, antihistamines, analgesics, antimicrobials, H2 antagonists, digoxin
Metabolic disorders: Hypoxia, hypercarbia, hypo- or hyperglycemia, renal or hepatic failure
Fluid and electrolyte disorders: dehydration, acute blood loss, hyponatremia, hypercalcemia
Neurological: Cerebrovascular event, intracranial hemorrhage
Cardiovascular: Acute myocardial infarction, cardiac failure
Withdrawal states: Alcohol, drug (e.g. benzodiazepine)
Acute psychosis
Others: Urinary Retention, fecal impaction, unfamiliar environment

supporting care. A multimodal approach is useful as several factors usually contribute to delirium. Seemingly minor interventions can add up to produce appreciable clinical improvement. The recommended steps in management are:

- Treat offending causes and contributing factors adequately;
- Manage behavioral disturbance: allow familiar persons by the bedside, minimize physical restraints, judicious use of antipsychotic drugs (e.g. haloperidol);
- Optimize the environment: quiet and adequately lit room, clocks, familiar objects; and
- Support other functions: adequate fluid intake and nutrition, early mobilization.

Dementia

This condition is characterized by gradually progressive or sustained cognitive decline. It is common in older people but may be mistaken as part of the aging process by caregivers and even health professionals. The consequences for sufferers and their caregivers may be immense with physical, emotional and financial impact.

Evaluation: The features of dementia are elicited in the history (from caregivers). The DSM-IV diagnostic criteria for dementia can be applied to identify this syndrome (see Table 3). Once dementia is diagnosed, the severity and underlying causes (see Table 4) are searched for in the clinical evaluation. Mental status testing (e.g. Abbreviated Mental Test) is

Table 3 Diagnostic Criteria for Dementia (Adapted from DSM-IV)

A) The development of multiple cognitive deficits manifested by both
   1) memory impairment
   2) one (or more) of the following cognitive disturbances:
      a) aphasia (language disturbance)
      b) apraxia (impaired ability to carry out motor activities despite intact motor function)
      c) agnosia (failure to recognize or identify objects despite intact sensory function)

B) The cognitive deficits in criteria A1 and A2 each cause significant impairment in social or occupational functioning and represent significant decline from a previous level of functioning.

Table 4 Commoner Causes of Dementia in Older Patients

- Non-reversible:
  Alzheimer's disease
  Vascular dementia
  Diffuse Lewy Body Disease
  Parkinson's Disease
- Reversible/arrestable:
  Depression ("pseudodementia")
  Drug effects (including alcohol, hypnotics)
  Metabolic disorders: Hypothyroidism, vitamin B12 deficiency, hypercalcemia
  Neurological: Space-occupying lesions, normal pressure hydrocephalus, neurosyphilis

performed to estimate the level of cognitive functioning. Selected investigations can be useful in ruling out reversible causes. These include:

- Blood investigations: FBC, electrolytes, glucose, calcium, liver function tests, thyroid function, vitamin B12 and folate;
- Imaging: brain; and
- Others: neuropsychological testing (for selected patients only).

Behavioural assessment involves searching for depression, agitation, aggressiveness, hallucinations, delusions and insomnia. Functional assessment identifies associated impairments in mobility, self-care, swallowing and bladder and bowel function. Social assessment uncovers caregivers' difficulties, stress and coping problems.

Treatment: The management of dementia is a multidimensional task that considers medical, behavioral, functional and social aspects of the patient.

- Medical: treatment of potentially reversible causes, cognition-enhancing medication (usually choline esterase inhibitors)
- Behavioral: behavior-optimizing drugs, appropriate handling of patients
- Functional: physical therapy
- Social: caregiver education (explaining the nature of the patient's illness and natural history, teaching coping strategies, advising on legal matters), referral to community-based services

Impaired Mobility

Background: Mobility is essential for activities of daily living. Impaired mobility refers to difficulty in walking with loss of ability to move around in one's environment. It is a common pathway through which many medical illnesses in older patients present. Many causes of impaired mobility are treatable. Even if not, most of its complications such as deconditioning, pressure ulcers, pneumonia, deep vein thrombosis, constipation and urinary retention may be prevented. The causes of impaired mobility span many organ systems (see Table 5). More than one underlying medical condition often contributes in the same patient. For instance, older persons having pre-existing gait problems due to cerebrovascular disease and osteoarthritis of both knees may present with total inability to walk when an acute illness such as cardiac failure occurs. Environment and psychosocial circumstances interact in the relationship between the presence of diseases and immobility.

Evaluation: Optimal management of older patients with impaired mobility requires a thorough clinical evaluation and selected investigations.

- Patients may report walking difficulty or be noticed to walk slower or more unsteadily. Any restriction of mobility or need for walking aids should be asked for. How acutely the patient becomes immobile and presence of other symptoms can provide useful diagnostic clues. History and physical examination should indicate whether neurological illness is present. Concurrent non-neurological causes should be systematically searched for. Complications of immobility should also be evaluated.
- Functional assessment of mobility should also be carried out. One simple method is the "get up and go" test. The patient is asked to sit on a chair, stand up, walk 6 meters, turn around, walk back to the

Table 5 Important Causes of Impaired Mobility

Neurological Causes	Non-neurological Causes
Cerebrovascular disease	Acute medical illness
Parkinsonism	Sepsis
Cervical myelopathy	Cardiac failure/coronary artery disease
Others	Chronic obstructive airways disease (severe)
Cerebellar dysfunction	Dehydration/acute blood loss
Normal pressure hydrocephalus	Electrolyte abnormalities
Peripheral neuropathy	Drug effects
	Antipsychotics
	Sedatives
	Antihypertensives
	Musculoskeletal causes
	Arthritis (inflammatory, degenerative)
	Fractures (especially hip)
	Foot problems
	Sensory causes
	Visual impairment
	Others
	Postural hypotension
	Fear of falling
	Deconditioning (after prolonged bed rest)
	Pain
	Depression
	Environmental (restraints, institutional restrictions)

chair and finally, sit down. The quality of gait and time needed to complete this task is assessed. Taking more than 15 seconds is considered abnormal.

- Depending on the clinical findings, investigations may include blood tests (e.g. FBC, electrolytes, glucose), ECG and imaging (e.g. joints, brain).

Treatment:

- Underlying medical causes should be treated on their own merit.
- A multidisciplinary team approach to functional rehabilitation is needed for most patients. Early assessment and training by physiotherapists and occupational therapists is important if clinically stable. For hospitalized patients, a gradual program of functional

retraining should be started when the clinical condition allows. Movement in bed, range of movement exercise, assisted transfers and limited ambulation is started as soon as possible. Walking aids (sticks or frames) should be provided early if indicated. Attempting to regain the previous level of mobility, preventing complications of immobility, and adapting the home environment to cater to any residual degree of immobility is important in facilitating a smooth transition back to the family environment. Education and training of caregivers in the supervision or assistance of mobility should be provided. Home evaluation may be needed.

Instability (Falls)

Background: A fall is a sudden unintentional change in position causing a person to land at a lower level on an object or the ground. Falls are an important marker of frailty and mortality. The potential consequences of falls are injuries (including fractures), hospitalization, fear of falling, disability and risk of institutionalization. Falls and their complications can often be reduced by identification of the contributing causes through clinical evaluation followed by appropriate intervention.

Changes with aging in postural control and gait play a role in predisposition to falls in older people — although they do not cause falls by themselves. Disease states further contribute to the risk of falling and often precipitate the event. The causes of falls are myriad and they virtually span the breadth of medicine (see Table 6). More than one cause often contributes to gait instability and falls in the same individual. As such, evaluation should not stop with the identification of one cause. Other common causes need to be ruled out as well. In addition, environmental factors like poor lighting, slippery floors and cluttered corridors in the house can further increase the risk of falls.

Although much can be done to reduce falls, it is often impossible to prevent all falls in those at risk. Unrealistic attempts to do so are often accompanied by restriction of mobility and excessive use of physical restraints.

Evaluation: The key is to ask why did the older person fall at the particular time at that particular place.

- Assessment begins with obtaining history on the fall itself — circumstances, and any preceding symptoms such as giddiness, palpitations

Table 6 Causes of Falls

Syncope
Drop attacks
Acute illness (e.g. sepsis, cardiac failure, cardiac arrhythmia, dehydration/acute blood loss, electrolyte abnormalities)
Vestibular disease
Neurological disease (e.g. cereborvascular disease, parkinsonism, cerebellar dysfunction, cervical myelopathy, normal pressure hydrocephalus, peripheral neuropathy, epilepsy)
Musculoskeletal causes (e.g. arthritis, foot problems)
Drugs effects (e.g. sedatives, antipsychotics, antihypertensives, diuretics)
Postural hypotension
Visual impairment
Accidents
True accidents
Environmental hazards (e.g. slippery floors, inadequate lighting, unstable furniture)

or focal neurological symptoms. Establishing whether there was loss of consciousness is crucial as this indicates the possibility of seizures or syncope. However, details may not be able to be obtained from fallers. Accounts of witnesses should be sought.

- A full medical history and drug history should be obtained. The history can help in identifying injuries. Physical examination should focus on likely areas of injuries, vital signs (including postural BP), cardiovascular, neurological, musculoskeletal and visual systems. Cognition is also evaluated. Gait and balance can be assessed by the "get up and go" test (see above). Investigations are guided by clinical findings. They normally include blood investigations (e.g. FBC, electrolytes) and ECG. Brain imaging may be needed where there is suspicion of a neurological cause (e.g. stroke disease) or complication (e.g. subdural hemotoma).

Treatment:

- Appropriate medical treatment of identified causes is the first step.
- Adequate physical retraining and suitable adaptations to the home environment can also reduce risk of further falls. Walking aids (such as sticks and frames) may facilitate safer ambulation. Fallers can be taught the technique of falling in a way that minimizes physical injury and getting up after the fall. Home adaptations may include

rearrangement of furniture and installation of grab bars in strategic places (e.g. toilet) where more delicate maneuvering is necessary.
- Many older persons who have fallen before experience "fear of further falls" (3 F's). This translates to a reluctance to walk again. Gradual physical retraining with constant encouragement and support is crucial.

Incontinence

Background: There are two forms of incontinence — urinary and fecal. This section only examines the first. Urinary incontinence is defined as the involuntary loss of urine severe enough to be a social or health problem. It is more common with increasing age, female gender and physical or cognitive function impairment. However, it is never a normal feature of aging but rather the result of definable medical or functional causes. It is associated with medical (perineal skin rashes), psychosocial (embarrassment, rejection by caregivers, social isolation, depression) and economic (cost of diapers, laundry) consequences. With appropriate treatment, cure is achieved in some, improvement in many, and ability to cope better in all.

Evaluation and management: In any older patient with urinary incontinence, a targeted evaluation should be initiated if it had not been performed recently.

- The first step is to rule out urinary retention by abdominal examination searching for a distended bladder. Portable ultrasound measurement of post-void residual volume (PVR) is a feasible alternative to urethral catheterization. When urinary retention is present (PVR persistently above 100–200 ml), intermittent or indwelling catheterization is usually indicated while undergoing further evaluation.
- Reversible or "transient" causes of incontinence should be ruled out. These mostly originate outside the lower urinary tract and are remembered by the mnemonic, "DIAPPERS" (diapers mis-spelled with an additional 'P') (see Table 7). Targeted history and physical examination with simple investigations such as urinalysis and blood glucose are carried out. When identified, these should be adequately treated while the continence status is monitored.
- If the incontinence does not resolve or if these causes were absent, then empirical treatment according to the type of persistent incontinence (stress, urge, overflow or functional) can be considered (see Table 8). If there are clinical features suggesting the need for

Table 7 "Transient" Urinary Incontinence

Causes	Treatment
Delirium	As for underlying cause
Infection (UTI)	Course of antibiotics
Atrophic vaginitis/urethritis	Topical estrogen
Pharmaceuticals (diuretics, anticholinergics, psychotropics)	Review medications
Psychological factors (depression)	Anti-depressant drugs
Excessive urine output (hyperglycemia, hypercalcemia)	As for underlying cause
Restricted mobility	As for underlying cause
Stool impaction	Disimpaction and laxatives

further evaluation or no improvement with empirical treatment, then specialized evaluation (including urodynamic studies) should be considered to obtain a precise diagnosis of the voiding dysfunction (overactive detrusor, underactive detrusor, outlet obstruction and outlet incompetence). Based on the diagnosis, treatment options including behavioral therapy, drugs, intermittent catheterization and surgery are offered. Long-term use of indwelling catheters and diapers should only be a last resort.

Impaired Feeding

Background: Impaired feeding in older people is an important clinical problem. As it may be due to treatable causes and have preventable consequences, it is important that this condition is recognized in older people. Medical causes of impaired feeding are numerous (see Table 9). In addition, functional status and environmental factors contribute to its occurrence.

Evaluation:

- Recognition of impaired feeding is often achieved from the history given by the patient or caregiver. When not the presenting problem, the presence of any impaired swallowing (e.g. coughing after swallowing, choking on food or drink) should be asked for. Observation of amount of food and drink taken is useful.
- In any older patient with impaired feeding, a search for underlying causes should be made in most patients.

Table 8 Types of Persistent Urinary Incontinence

Type	Defining Features	Common Causes	Empirical Treatment*
Stress	Leakage of urine (usually small amounts) with coughing, sneezing or exercise	Weakness or laxity of pelvic floor muscles Urethral sphincter weakness	Pelvic floor exercises
Urge	Leakage of urine (usually larger but variable amounts) associated with intense desire to void	Overactive bladder associated with neurological disease (e.g. stroke, parkinsonism, suprasacral spinal cord disease) or bladder outlet obstruction Bladder lesions: Cystitis, stones, tumor	Behavioral therapy (e.g. bladder retraining, prompted voiding) Bladder relaxant drug therapy** (e.g. oxybutynin, tolterodine) Specific treatment for bladder outlet obstruction (e.g. alpha blockade drugs) or bladder lesions
Overflow	Leakage of urine (usually small amounts) associated with an over-distended bladder	Bladder outlet obstruction by enlarged prostate, urethral stricture, cystocoele Detrusor weakness associated with DM autonomic neuropathy, cauda equina lesion	Intermittent catheterization Indwelling catheterization

(Continued)

Table 8 Continued

Type	Defining Features	Common Causes	Empirical Treatment*
		Detrusor sphincter-dyssynergia associated with suprasacral spinal cord lesions	
Functional	Leakage of urine associated with inability to toilet because of impairment of physical and/or cognitive functioning or environmental barriers	Neurological and musculoskeletal disease Severe dementia	Behavioral therapy (e.g. prompted voiding) Environmental modifications Toileting aids (e.g. urinal, commode) Pads External collection devices (e.g. penile sheaths)

*Indications for further evaluation (rather than empirical treatment): PVR persistently > 100 ml, symptoms of voiding difficulty, pain, hematuria, recurrent UTI, pelvic mass, previous pelvic surgery or irradiation.
**Caution in males: Consider urodynamic evaluation to rule out bladder outlet obstruction first.

Table 9 Causes of impaired feeding ("D"s)

Disease (any acute medical illness, chronic infections, malignancy, endocrine diseases, upper gastrointestinal disease)
Drugs (especially those that cause decreased appetite, vomiting, sedation)
Dementia (usually in the very advanced stages)
Depression
Dysphagia (oropharyngeal & oesophageal)
Dentition (poor dentition)
Disability (functional impairments such as poor mobility, poor dexterity and cognitive problems that interfere with feeding)

- The evaluation for possible dysphagia (swallowing impairment) by bedside assessment can be performed by asking the patient to drink 30 ml of water while seated upright. If delayed swallowing (exceeding 2 seconds), cough during or within 1 minute of swallowing, dysphonia or drooling is observed, then swallowing impairment is present.
- Referral to a speech therapist for a more detailed swallowing assessment should be made where there is still doubt about swallowing ability. Functional endoscopic examination of swallowing (FEES) or videofluoroscopic study of swallowing is useful in instances when there is uncertainty of diagnosis or when severity of impairment needs to be assessed. Where there is swallowing impairment, the underlying cause should be identified if possible. For oropharyngeal dysphagia, neurological causes (such as cerebrovascular disease and parkinsonism) should be sought as they may merit specific treatment. For oesophageal dysphagia, lesions of the oesophagus should be ruled out.
- Treatable complications of impaired feeding should also be sought. These include malnutrition, dehydration and aspiration pneumonia (for swallowing impairment). Where appropriate, nutritional assessment should also be performed.

Treatment:

- The underlying cause(s) should be treated or removed where possible.
- In the case of impaired swallowing, the application of compensatory measures should be considered. These include modification of feeding technique and foods or liquids taken. Postural strategies and sensory enhancement techniques may be considered in selected patients. The expertise of a speech therapist should be utilized for more complicated cases. Where swallowing impairment is severe, alternative non-oral forms of feeding [such as nasogastric tube and percutaneous endoscopic gastrostomy (PEG) feeding] should be considered if assessed to be helpful in overall management of the patient.
- Consultation with a dietician should be sought to optimize nutritional intake — particularly where there already is malnutrition.
- Complications of impaired feeding should be treated on their own merits.
- For patients with functional impairment contributing to the impaired feeding, rehabilitation and environmental manipulation to address these deficits should be addressed.

Iatrogenesis

Iatrogenesis refers to the problems associated with treatment. It is more common with older people because of:

- Increased susceptibility to adverse effects of treatment: Due to altered pharmacokinetics and pharmacodynamics, polypharmacy and decrease in organ reserve with advancing age and disease
- Increased use of diagnostic procedures and therapeutic measures due to higher disease incidence.

There are 2 main forms of iatrogenesis in older patients:

1) Hazards of hospitalization: Hospitals pose certain risks to older patients (see Table 10) that may be unrelated to the admitting illness. The consideration of risks and benefits of hospitalization need to be carefully considered.
2) Adverse drug reactions: This is the most common form of iatrogenic disease in older people. Although virtually any medication may be the cause, cardiovascular and psychotropic drugs are the most commonly implicated. In addition to specific adverse effects of different drugs, this can present as any geriatric syndrome. In fact, adverse drug reactions should be considered in the list of differential diagnoses of geriatric syndromes and any new symptom or problem encountered by older people. Risk of adverse drug reactions can be reduced with prudent prescribing (see Table 11).

Table 10 Hazards of Hospitalization for Older Patients

Diagnostic procedures: endoscopy, imaging with contrast agents (intravenous or oral)
Therapeutic procedures: intravenous lines, urinary catheters, nasogastric tubes, blood transfusion
Drugs: error, drug-drug interaction, drug-disease interaction, adverse effects
Surgery: anesthesia, hypovolemia, infection, stroke, cardiac events
Bedrest: venous thrombosis, constipation, urinary incontinence and retention, deconditioning with impaired mobility and functional decline, postural hypotension, decubitus ulcers
Infections: nosocomial
Delirium

Table 11 Recommendations for Geriatric Prescribing

Try to make a diagnosis before prescribing medications.
Consider non-drug therapy.
Know the pharmacology of drugs prescribed.
"Start low" (initial dosage) and "go slow" (increase in dosage).
Consider drug-drug and drug-disease interactions.
Check on compliance regularly: Particularly of patients with impaired cognition, poor vision.
Monitor and maintain a high index of suspicion for potential adverse reactions.

REFERENCES

1. Gill TM, Geriatric Medicine: It's more than caring for older people, *Am J Med* **113**:85–90, 2002.
2. Kane RL, Ouslander JG, Abrass B (eds.), *Essentials of Clinical Geriatrics*, 4th edition, McGraw–Hill, New York, 1999.
3. Cobbs EL, Duthie EH, Murphy JB (eds.), *Geriatric Review Syllabus: A Core Curriculum in Geriatric Medicine*, 5th ed., Blackwell Publishing for the American Geriatrics Society, 2002.
4. Burton JR, Approach to the diagnosis and management of the elderly patient, in: *Textbook of Internal Medicine*, 3rd ed., Lippincott–Raven Publishers, Philadelphia, pp. 2478–2482, 1997.
5. Tinetti ME, Inouye SK, Gill TM, Thomas M, Doucette JT, Shared risk factors for falls, incontinence and functional dependence: unifying the approach to geriatric syndromes, *JAMA* **273**:1348–1353, 1995.
6. *Standards of Care in Geriatric Medicine*, Society of Geriatric Medicine, Singapore, 2001.
7. Ding YY, Sahadevan S, Geriatric syndromes and assessment, in: *Medicine and Surgery in the Older Person*, Armour Publishing, Singapore, pp. 43–61, 2000.

Medical Oncology

This page intentionally left blank

50

Pathogenesis of Cancer

Richard Epstein

GENERAL CONCEPTS

Is cancer most accurately regarded as a disease, a group of different diseases, or a "normal" degenerative condition similar to aging? To gain insight into this key question, we need to consider what is already well-established about the natural history of human cancer.

First, malignancy is remarkably uncommon in individuals younger than 40 years old, suggesting a Darwinian predilection for the post-reproductive age group. Consistent with this, human cells are more difficult to transform *in vitro* than are cells of more short-lived mammalian species. Hence, genetic stability is an intrinsic property of cells which appears to vary inversely with cancer susceptibility *in vivo*.

Second, tumor types which can be cured in their advanced stages tend to be rare, whereas common cancers in their advanced stages are usually incurable. This is not simply a matter of bad luck for the human race. Rather, it reflects the fact that rare neoplasms arise due to rare genetic events (which in turn are often flagged by characteristic chromosomal translocations), whereas common cancers arise via the progressive

accumulation of common genetic mutations. This latter process results not only in loss of growth control, but also in profound genetic instability similar to that which normally confers an adaptive survival advantage on bacterial cells. Conversely, the sustained clinical remissions that famously accompany the treatment of certain advanced malignancies are likely to be related to the relatively normal composition (and hence stability) of the genome in those tumor cells. The main problem that most cancer patients face today is not rapid tumor cell growth as such, but rather the presence of genetic instability that transforms the disease into a therapeutic "moving target".

Not surprisingly, then, the distinction between rare and common cancers has implications for pathogenesis at the molecular level. For a single chromosomal translocation event to trigger cell transformation, a critical mutation (including gain-of-function mutations) must be induced to drive or permit the necessary growth advantage. In contrast, the pathogenesis of common cancers depends mainly on the incremental accumulation of loss-of-function mutations. This dichotomy gives rise to the contemporary model of two broad groups of "cancer genes": oncogenes and tumor suppressor genes.

Oncogenes (or proto-oncogenes, which term designates the normal cellular homologue) are genes involved in cell growth. A gain-of-function mutation affecting one of these genes can have profound consequences on cell growth, and as few as two oncogene mutations can fully transform an otherwise normal cell line *in vitro*. On the other hand, the sudden acquisition of an activated oncogene by a completely normal cell can result not in cell growth but in cell death, indicating the existence of intact molecular "checkpoints" to cell growth.

Tumor suppressor genes are most commonly cell cycle control genes, though not invariably. A single cell line or tumor can accumulate numerous mutations of tumor suppressor genes, each one of which may contribute independently to the transformed phenotype. The acquisition of these control defects plays a permissive role in cell growth by conferring resistance to normal cell death (apoptosis). Tumor suppressor gene defects can also permit the viable retention of oncogene mutations, further driving uncontrolled growth. It is this combined genetic picture of suppressor gene loss and oncogene activation that characterizes most common advanced human cancers. Examples of tumor suppressor genes which often incur disabling mutations in human tumors include p53,

p21^{Cip1}, p16^{Ink4}, and pRb (the retinoblastoma susceptibility gene product), whereas dominant oncogenes may undergo constitutive activation by mutations (e.g. the codon 12 mutation implicated in K-*Ras* activation in colon carcinogenesis).

PROCESSES AND PATHWAYS

How and why do mutations occur in the genes of normal cells? Normal DNA is subject to numerous damaging events, with many thousands of such events occurring in each cell every day. That these damaging lesions are so efficiently removed is evidence of a highly efficient repair enzyme network within cell nuclei. This process of DNA damage and repair is an entirely normal one in all living organisms and environments.

There are several different possible outcomes from this interaction between DNA damage and repair. First, the cell can repair the damage and continue to grow. This is by far the most common outcome, and the timing of the cell cycle makes allowance for such repair prior to both DNA synthesis and mitotic cell division (i.e. during the gap — G1 and G2 — phases of the cell cycle). Second, if the amount of damage induced is sufficiently high, extra time will be required for these cells to repair; this is made evident by the occurrence of cell cycle delay. A third possible outcome is that the cell is so overwhelmed by damage that it is unable to repair, resulting in cell death.

With respect to cancer pathogenesis, DNA damage plays a critical role insofar as it increases the probability of gene mutations, some of which may not be repaired. The latter will be clonally preserved if they confer a short-term growth advantage. Such growth-promoting mutations can accumulate progressively within the clonal cell outgrowth, leading to progressive dominance of the abnormal cell clone that may form a tumor. Such tumors may be benign or malignant, depending upon whether the uncontrolled growth is also associated with tissue invasion or distant metastasis. The pathogenesis of these manifestations of tumor progression may well involve positive-feedback growth loops between normal stromal cells and tumor cells mediated via a complex cross-talk between secreted proteases and cytokines.

It is important to note that not all steps in cancer progression are mediated via genetic mutations. So-called epigenetic events may also contribute to tumorigenesis: an instructive example is that of methylase-dependent

cytosine methylation, whereby methylcytosine-binding proteins recruit histone deacetylases to induce heritable transcriptional repression via chromatin condensation. Non-mutated tumor suppressor genes are frequently inactivated via promoter methylation occurring within tumors, just as wild-type proto-oncogenes may become hyperactivated via regional DNA hypomethylation.

ENVIRONMENT VERSUS GENES

To most people cancer is a bewildering phenomenon; understandably, then, many attempts have been made to explain its occurrence. One of the most popular models for experimental human cancer has involved the application of chemical carcinogens to animals. This early work led to the distinction between two qualitatively different cancer-causing chemical categories: genotoxic carcinogens, which directly damage the genetic material, and non-genotoxic carcinogens, which promote the growth of established tumors but do not interact directly with DNA. The distinction between these two classes of cancer-causing chemicals led to the modeling of tumor growth into two phases: initiation and promotion.

The relevance of this model to human cancer growth is debatable. It is certainly possible that tumors can be initiated by genotoxic stimuli, as was evident following the Hiroshima atomic bomb. If the background level of tumor incidence in the human population is related to genotoxic stimuli, however, then for many cancers the risk appears to be distributed so evenly as to suggest that such exposure is ubiquitous. Of course, there are several important exceptions to this. The most obvious is cigarette smoking, an exposure now notorious for the high concentration of carcinogens associated with it. Similarly, ultraviolet irradiation of pale skin is associated with the efficient accumulation of DNA-damaging events, some of which are capable of causing cell transformation. Prolonged viral infections, such as those causing chronic hepatitis or papillomavirus infection of the cervix, are also associated with cancer induction, albeit via uncertain pathways.

An intriguing aspect of human carcinogenesis involves the rising incidence of certain tumors following the putative "Westernization" of lifestyle. This was illustrated by migration studies and longitudinal studies in developing countries, particularly with respect to cancers of the breast, colon and prostate. The most obvious lifestyle variable implicated

in this process is diet, though whether the culprit will turn out to be total calories or a qualitative component (e.g. saturated fat) remains controversial. Certainly, it appears unlikely that any directly genotoxic ingredient will be centrally implicated.

In recent years a number of environmental etiologies have been identified in relation to certain malignancies. Chronic infection with the acid-resistant bacterium, *Helicobacter pylori*, has been linked to both gastric cancer and to the rare gastric MALToma lymphoid neoplasm. A role for chronic injury in predisposing to cancer is supported by the finding of a 45-fold increase in the incidence of esophageal adenocarcinoma in individuals with prolonged severe gastroesophageal influx. Yet another association of interest has been that between bladder cancer and ingestion of a diet low in water content.

IMPLICATIONS FOR CANCER PREVENTION

Approximately 50% of human cancers can be prevented — at least in theory. Such preventive measures include the cessation of cigarette smoking; prevention of cervical papillomavirus infection via the routine use of barrier prophylaxis; vaccination against infections such as hepatitis B; detection of pre-invasive neoplastic change in accessible organs such as the breast (via radiology) and cervix (via cytology); and avoidance of over-exposure to sunlight. In practice, however, it has not proven easy to upgrade the efficacy of existing preventive efforts.

For individuals known to be at high risk of tumor development, recent advances have created new opportunities for prevention. An example is the recognition of breast and ovarian cancer predisposition in those individuals whose families are characterized by BRCA1 mutations. Such individuals can be offered early intensive mammographic and clinical screening; the use of prophylactic tamoxifen to prevent breast cancer in other "high-risk" individuals is also being evaluated. The ultimate status of these interventions, with respect to both high-risk and standard-risk patients, is still being evaluated. Conversely, the popularity of hormone replacement therapy (HRT) for postmenopausal women has recently been dented by its association with increased breast cancer risk.

The relative contribution of environment and genes to cancer pathogenesis seems likely to remain a subject of intense debate. Relatively few cancer patients come from families with a strong history of cancer,

suggesting that inherited genetic predispositions are not responsible for the majority of cancers. On the other hand, striking geographical clustering of cancer has been noted in the context of many occupational and environmental conditions. A complex probabilistic (stochastic) interaction between normal endogeneous and microenvironmental processes thus seems likely to be implicated in the development of most human tumors. Newer high-throughput diagnostic technologies such as DNA microarray and phosphoproteomics seem likely to help accelerate this exciting process of discovery.

51

Population Cancer Screening

Soh Lay Tin

Cancer is currently the top-killer in Singapore, claiming 3881.6 lives annually during the period, 1993–1997.[1] Generally, for most cancers, detection of earlier stage disease leads to improved prognosis. Hence, it is the hope of many people that screening will contribute to early detection of cancerous conditions and improved outcome.

Screening refers to tests that are performed to identify asymptomatic individuals in the general population who are likely to have a particular type of cancer. Like all tests, screening has both advantages and disadvantages. One of the advantages is improved outcome if the disease is detected at an early stage. Detection of earlier stage disease will lead to less radical treatment. There will be reassurance for the ones with negative results. There is also the possibility of lower treatment cost for earlier stage disease as the treatment is less radical.

However, screening has a longer list of disadvantages. It can lead to over-treatment of borderline abnormalities that might never have any clinical significance. Those with false-negative results will have a false sense of reassurance and might ignore subsequent development of symptoms resulting in delayed diagnosis and poorer prognosis. For those with

false-positive results, unnecessary anxiety might result. In addition, false-positive result might lead to additional diagnostic tests, which together with the over-treatment of borderline abnormalities, will lead to additional cost. There is also the potential hazard of the test itself if the screening test is invasive.

Hence, for a screening test to be advocated for a cancerous condition, it needs to have adequate sensitivity (i.e. proportion of those with the disease who are tested positive) and adequate specificity (i.e. proportion of those without the disease who are tested negative). In addition, it should preferably be non-invasive, easily carried out and acceptable to those who need to be screened.

Cancers likely to benefit from screening, usually have the following characteristics. Firstly, there should be a definable high-risk population. Secondly, the condition should have a long natural history with preferably, a pre-malignant phase. Thirdly, the pre-malignant condition should be treatable.

With the current data, screening of the general population can be advocated for only 3 cancerous conditions. This article aims to discuss the role of screening of the average risk population.

CERVICAL CANCER

Cervical cancer is an ideal disease for screening because it has a long treatable pre-malignant phase. Furthermore, the screening test is simple and cheap. It is the Papanicolaou test or PAP smear. In this test, exfoliated or desquamated cells are brushed from the cervical surface for cytologic examination. This test enables the detection of the invasive and pre-invasive cervical cancer.

There is no randomized trial on the benefit of screening in cervical cancer. However, there are evidence to support the use of screening in cervical cancer to improve the incidence and mortality. In the Nordic countries, marked reduction in the incidence and mortality from cervical cancer followed the introduction of screening programs.[2] Organized screening started in most of the Nordic countries soon after 1960. Up to 1985, the screening was most intense in Iceland followed by Finland, Sweden and Denmark. In Norway, screening was only spontaneous in late 1994. During the period 1986–1995, the reduction in mortality and incidence rate was highest in Iceland (mortality: 76% and incidence: 67%)

and Finland (73% and 75% respectively), intermediate in Sweden (60% and 55% respectively) and Denmark (55% and 54% respectively) and lowest in Norway (43% and 34% respectively). Similarly, in Canada[3] and in Scotland,[4] intense screening was also associated with reduction in mortality from the disease. Case-control studies[5,6] also showed that screening improves prognosis in cervical cancer.

In contrast, in North America, much of Europe and in Singapore,[1] where organized screening is absent and screening is mainly opportunistic, reductions in incidence and mortality from cervical cancer have not been dramatic.

In Singapore, PAP smear is only carried out for pre- and post-natal patients and those who turn up at the gynecological clinic with a gynecological complaint. There is no figure for the PAP smear rate for asymptomatic women. The estimates are that 60% of all Singapore women undergo regular PAP smears. The target of the Gynecology Cancer Workgroup is for 80% of all women to have regular PAP smear screening and achieve a crude incidence rate of 13 and a crude death rate of 6.16, which represents a reduction of 15%. It is recommended that all women who ever had sex are advised to have their first PAP smear by the age of 25 years. The second smear is taken one year after the first smear and subsequent smears are taken at 3 yearly interval. A woman can be discharged from screening at age 65 years if the smear taken at 65 years is negative and there was a previous negative smear within the last 3 years. Women who have never had sexual intercourse need not have PAP smear screening unless they have symptoms.

As the sensitivity of the PAP smear is dependent on sampling technique and the interpretation of the findings, several new techniques were promoted to improve the sensitivity. One of these commonly used innovations in cervical cancer screening is liquid-based cytologic collection and analysis. However, reviews[7,8] of the studies using this technique have concluded that the accuracy of the analysis is uncertain. Other ways to improve the sensitivity include reevaluation of conventional smears initially interpreted as negative; either manually or with the assistance of a computerized technique. These methods are only more cost effective when used in the laboratories where the PAP smear has relatively poor sensitivity. In laboratories that can interpret the PAP smear with a high degree of accuracy, these additional tests merely increase the cost of screening.[9] As human papillomavirus (HPV) is a predisposing factor for

the development of cervical cancer, other newer methods of screening tried to incorporate HPV testing in the screening. However, the role of HPV testing either as an adjunct to or substitute for cytologic screening needs to be evaluated.

BREAST CANCER

The second cancer for which screening has an impact is breast cancer. In contrast to the situation in cervical cancer, there are 8 reported randomized trials[10–18] on breast cancer screening. Most trials used mammography either alone or in combination with physical examination. These trials showed that in the 50- to 69-year-old woman, early detection of breast cancer by regular mammography, with or without breast examination, reduces breast cancer mortality by one-third.

However, for women aged 40–49, the result is less clear. There is no reduction in mortality from breast cancer within 7 years after the initiation of screening. There is a trend toward reduced breast cancer mortality only after a follow-up of 10 or more years and the decrease is estimated to be 18%.

The first randomized breast cancer screening trial was conducted by the Health Insurance Plan (HIP)[10] of greater New York. About 62 000 women between the age of 40 and 64 years were randomized in the early 1960s to

Table 1 Characteristics of the Randomized Trials on Breast Cancer Screening

Study	Year	Age at Entry	Screen Round	Screening Interval (Yr)	Modality	Study (n)	Control (n)
HIP	1963	40–64	4	1	2-MM + PE	30 239	30 756
Two-county	1977	40–74	2	2 (< 50) 2.8 (> 50)	1-MM	78 085	56 782
Malmo	1976	45–69	2	1.8	2-MM	21 088	21 195
Gothenburg	1983	39–49	5	1.5	2-MM	11 724	14 217
Stockholm	1981	40–64	2	2.3	1-MM	39 164	19 943
Edinburgh	1976	45–64	6	1	PE	23 226	21 904
			3	2	2-MM		
Canada 1	1980	40–49	4	1	2-MM + PE	25 214	25 216
Canada 2	1980	50–59	4	1	2-MM + PE	19 711	19 694

From Screening Sensitivity and Sojourn Time From Breast Cancer, Early Detection Clinical Trials: Mammograms and Physical Exainations. Yu Shen and Marvin Zelen, *J Clin Oncol* **19**(15):3490–3499, 2001.

screening using annual 2-view mammography and breast clinical examination to a total of 4 screens or to a control group. Seventy-five percent of the women were older than 50 years old. As the trial was conducted in the early years of mammography, only 33% of breast cancers were detected by mammography alone. The majority of the breast cancers were detected by physical examination. In the trial, the reduction in breast cancer mortality was 30% and was restricted to women between the ages of 50 and 64 years.

The Edinburgh randomized breast cancer screening trial[11] recruited women aged 45–64 years from 1978–1981 (cohort 1) and those aged 45–49 years during 1982–1985 (cohorts 2 and 3) and they were cluster-randomized to either screening or to a control group. The women were screened by mammography every 2 years and annual clinical examination. Results were based on 14 years of follow-up and 270 000 woman-years of observation. 28 628 women were offered screening and 26 026 women were not screened. After adjustment for socioeconomic status and censoring for deaths due to diagnosis more than 3 years after the end of the study, the breast cancer mortality rate in the screened group was 0.71 of the unscreened group. However, no breast cancer mortality benefit was observed for women whose breast cancer was diagnosed when they were younger than 50 years. As in the HIP trial, the benefit in women who entered the trial before the age of 50 years was limited to diagnosis of breast cancer after the age of 50 years.

In the Swedish Two-county trial,[12] the women in Kopparberg and Ostergotland were also cluster-randomized to either single-view mammogram at 24-month interval for women aged 40–49 and 33-month interval for women aged 50–70. In this trial, 133 000 women were randomized between 1977 and 1979 to regular screening or to control. The initial results,[12] published in 1985, showed a significant 30% reduction in breast cancer mortality in the screened group and was primarily for women aged 50 years or older. Subsequently, the control group was also invited for screening. After 15 years of follow-up,[13] the 32% reduction in breast cancer mortality was maintained [relative risk (RR) = 0.68, 95% confidence interval (CI) = 0.59–0.80, $p < 0.001$]. The largest effect on mortality (39% reduction) can be seen at ages 50 to 69. The reduction in mortality (5%) for women aged 40 to 49 years are significant only for the Kopparberg county but not the Ostergotland county. In this study, the mean sojourn time (MST), sensitivity and the positive predictive value (PPV) were most favorable for women aged 50 to 69 years.

The Stockholm trial[14] was initiated in Mar 1981 and 40 318 women aged 40 to 64 years were clustered-randomized to single-view mammography screening alone while 20 000 women were randomized to no intervention. Two rounds of screening were done at 28 months interval. In 1986, the control group was invited once to screening. After a mean follow-up of 11.4 years, a non-significant mortality reduction of 26% was observed for the whole study group, with a RR of death in breast cancer of 0.74 (95% CI = 0.5–1.1). For women aged 50–64 years, a significant 38% mortality reduction was observed with a RR of 0.62 (CI = 0.38–1.0). For women aged 40–49 years, no effect on mortality was found, with a RR of death in breast cancer of 1.08 (CI = 0.54–2.17) after 11.4 years of follow-up.

In the Malmo mammographic[15] screening trial, women aged 45–79 years were randomized to 5 rounds of 2-view mammographic screening at 18–24 months intervals. In women < 55 years, more women died of breast cancer [RR = 1.29 (CI 0.74–2.25)] in the first 7 years but the trend reversed after that. As for the women aged > 55 years, there is a 20% reduction in mortality from breast cancer [RR = 0.79 (CI 0.51–1.24)].

The Gothenburg breast screening trial[16] studied only women aged 39 to 49 years. Between September 1983 and April 1984, 11 724 women aged 39–49 years were cluster-randomized to two-view mammographic screening at 18 months interval and 14 217 women were not invited to undergo screening until the 5th screen of the study group (6 to 7 years after randomization). A 45% reduction in mortality from breast cancer was observed in the study group compared with the control (RR = 0.55, p = 0.035, 95% CI, 0.31–0.99). However, the mortality of the two groups only began to separate 6 to 8 years after randomization and the gap continued to widen thereafter.

In the Canadian Study 2,[17] women aged 50–59 years were randomized to annual screening consisting of two-view mammography and physical examination or a control where they were screened with annual physical examination alone. This study involved 39 405 women recruited between 1980 through 1985. Although yearly mammography in addition to physical examination detected considerably more lymph node-negative and small breast cancers than screening with physical examination alone, it had no impact on mortality. It suggests that screening by mammography does not further reduce breast cancer mortality above and beyond the benefit of screening by clinical breast examination alone.

In the Canadian Study I (CNBSS I)[18] involving women aged 40–49 years, 25 214 women were randomized to 4 to 5 rounds of screening at 12 months interval. These women were screened with 2-view mammography and clinical examination. Another 25 216 women were randomized to a control group who had a clinical examination at the initiation of the trial followed by the usual health care. At a follow-up of 8.5 to 13 years, the RR in the screened group appears to be worse at 1.14 compared to the control group. Other authors[19,20] have criticized this trial in view of the excess advanced cancer in the screened arm and questioned whether there was non-random allocation of the women. Analysis revealed that a quarter of the women in the control received one or more mammograms in the course of the trial as in the "usual medical care". Of note is the Malmo trial that also showed an excess mortality in the <50 years old women in the first 7 years and it reversed with subsequent follow-up. At 11.4 years of follow-up, the Stockholm trial also showed a RR of breast cancer deaths of 1.08 in the younger women. Whether this excessive breast cancer death in the screened group in the CNBSS I will reverse with longer follow-up remains to be seen.

In the Swedish overview including all randomized trials in Sweden, with a follow-up of 5 to 14 years, the mortality reduction in the group aged 50 to 69 was 29%. However, in the group aged 40 through 49 years, it was a non-significant 13%.[21] An updated meta-analysis (after an additional 3 to 4 years of follow-up), presented at the Falun meeting in 1996,[22] showed a relative mortality of 0.77 (95% CI 0.59–1.01) in the age group 40–49 years. Combining all trials gave a figure of 0.85 (CI 0.71–1.01). The conclusion was that there was a mortality benefit, albeit a smaller and more variable one than that observed in the older women and the mortality takes longer to appear in the younger women.

In a meta-analysis by Kerlikowske,[23] for women between ages 50 to 74 at entry into these studies, breast cancer mortality in the screened group was significantly less than in the control group after 7 to 9 years of follow-up with a relative risk of 0.74 (95% CI 0.66–0.83). The magnitude of the benefit is not affected by further follow-up. However, for women aged 40 to 49 at entry, the duration of follow-up did affect the risk of death. The relative risk of death in the screened group was 1.02 (95% CI 0.73–1.27) after 7 to 9 years of follow-up but 0.83 (95% CI 0.65–1.06) after 10 to 12 years of follow-up. The same result was also reported by Hendrick *et al*.[24]

However, the design of these trials do not allow an adequate estimation of the extra benefit obtained by starting screening at age 40 instead of at age 50. A possible explanation for the delayed benefit in the younger women is that screening mammogram is probably not effective in the younger women. The delayed effect of screening women below the age of 50 at entry into the clinical trials may be the results of these women undergoing screening beyond the age 50. This "age creep" effect was studied and suggested by De Konig *et al*.[25]

Why should age 50 affect the effectiveness of mammographic screening? This is because age 50 corresponds approximately to the time of the menopause. It is a well-known fact that menopausal status[26] has an impact on the biology of breast cancer and premenopausal breast cancer grows faster than postmenopausal breast cancer. This explains the higher incidence of interval cancer (diagnosed between screening) in premenopausal than in postmenopausal women.[27] Therefore, theoretically, reducing the screening interval to less than 18 months in premenopausal women may reduce breast cancer mortality.

Furthermore, the sensitivity of mammography is lower in the premenopausal women.[28] Rosenberg reported that in women younger than 40 years, the sensitivity is 54%. In the 40- to 49-year-olds, up to one-fourth of all invasive breast cancers are not detected by mammography compared with one-tenth of cancers in the 50- to 69-year-olds. Kerlikowske *et al*.[29] showed that the sensitivity of first screening mammogram increases with age: 77.3% for age 30 to 39 years, 86.7% for age 40 to 49 years, 93.6% for age 50 to 59 and 94.1% for age 60 to 69 years.

Despite these controversies, the proponents of screening mammogram in women aged 40 to 49 feels that it should be advocated. However, one has to consider the risk of radiation-induced breast cancer. Radiation has been shown to cause breast cancer. The risk is related to the dose and the younger the exposure, the greater the life-time risk as it occurs at least 10 years after the exposure. Beemsterboer *et al*.[30] developed a computer stimulation model to calculate the breast cancer deaths induced by exposure to low-dose radiation in mammographic screening and the number of lives saved. With a 2-year screening interval and a mean dose of 4 mGy to each breast from a 2-view mammogram, for women aged 50 to 69, the ratio of breast cancer death prevented to those induced as a result of the screening is 242:1. This ratio decreases to 97:1 in women aged 40 to 49 years. This risk is theoretical as there has been no report of mammogram-induced breast cancer.

Screening appears to have no impact on diagnoses made after the end of the screening period. In the HIP trial,[10] diagnoses of breast cancer made 3.0–3.5 years after the end of the study are not impacted by the screening. In the Edinburgh study,[11] the estimated benefit of screening was slightly larger when deaths from diagnoses more than 3 years after the end of the study was censored. The Swedish study[21] also showed similar findings.

The role of clinical breast examination (CBE) and breast self-examination alone as screening is not as clear as screening mammography. In most trials, CBE is part of the screening carried out together with mammography. In only one trial was CBE carried out in the absence of mammography. It is the Canadian National Breast Screening Studies II (CNBSS II[16]).

CNBSS II recruited women aged 50 to 59. In women aged 50 to 59 years, the two-modalities screening (mammography + CBE) appears to be better as the detection rate for two-modalities was 7.2 per 1000 screening examinations versus 3.45 for CBE alone.[31] This compared with a detection rate of 4.67 and 4.84 in women aged 50 to 59 years in the Swedish Two-county and Stockholm trials that used only mammography alone. The sensitivities of the two modalities versus single modality were 88% versus 63% respectively.[31] Despite the higher detection rate, the difference in mortality rate between single and two-modalities screening in CNBSS II did not appear significantly different at 10 years. However, this does not exclude the possibility of a difference with longer follow-up.

In the CNBSS I that recruited women aged 40 to 49 years, both modalities were used in the screening. The detection rate was 3.89 per 1000 screening examinations[31] that is lower than in the older women in CNBSS II. However, it is still higher than the detection rate of 2.09 and 2.06 in the 40 to 49 years old women in the Two-county and Stockholm trials that used mammography alone.[31] The sensitivity in the CNBSS I was 81% compared to 62% and 53% in the Swedish Two-county and Stockholm studies respectively.

The efficacy of breast self-examination (BSE) is even less clear. Meta-analysis[32] suggests a benefit. Two randomized controlled trials[33,34] on BSE have been completed. Preliminary results are not encouraging although they do not rule out the possibility of benefit with longer follow-up. An important point in BSE is that it has to be performed well and this will require good health education.

In Singapore, a screening project was undertaken in 1994[35] and it spanned over 2 years. In this project, 166 600 women aged 50–64 years were randomized to either 2-view mammography without physical examination (67 656) or observation (97 294, control). 28 231 (41.7%) responded and were screened. Comparing to the general population, the responders were more likely to be married, have more formal education, be working, Chinese and be in a higher socioeconomic group. In the responders, 4.8 cancers were detected per 1000 women screened per year. The incidence of cancer in the control group was 1.3 per 1000 women per year. However, the incidence of cancers in the non-respondents was 1.0 per 1000 women per year and was significantly less than in the control group. This study revealed an interesting social aspect of the population. It suggests that further health education needs to be targeted at the lower socioeconomic group before any organized screening program can succeed. Besides limited knowledge of breast cancer, the poorer women have more urgent competing priorities.

COLORECTAL CANCER

Like cervical cancer, colorectal cancer has a long natural history. Detection of early disease is associated with improved outcome. Hence, screening may reduce the mortality from this disease. In this condition, there are three screening tests, which can be used either singly or in combination.

Among the three tests, guaiac-based fecal occult blood testing (FOBT) is the cheapest and has the most evidence as a screening tool. There are 3 large randomized trials on FOBT, which showed that colorectal cancer could be detected at an earlier stage leading to improved prognosis. The Minnesota Colon Cancer[36] Control Study randomized 46 551 participants' ages 50 to 80 years to screening annually, once every 2 years or to a control. The participants submitted six guaiac-impregnated paper slides with 2 smears from each of three consecutive stools. Participants with positive FOBT were subjected to further evaluation including a colonoscopy. At 13 years of follow-up, the cumulative colorectal cancer mortality ratio were 0.67 (95% CI 0.50–0.87) and 0.94 (95% CI = 0.68–1.31) in the annually and biennially screened participants compared to the control. In an update report at 18 years of follow-up, the cumulative colorectal cancer mortality ratio in the annually and biennially screened participants were 0.67 (95% CI = 0.51–0.83) and 0.79 (95% CI = 0.62–0.97) respectively compared to

the control.[37] Hence, biennial screening also results in a statistically significant reduction in colorectal cancer mortality.

The Danish group[38] randomized 30 967 people aged 45 to 75 years to biennial screening and another 30 966 people as control. There were 3 screening rounds during a 5-year period followed by 5 years of follow-up. The test was carried out with dietary restriction during the 3 days before the stools were collected. People with positive FOBT will undergo further evaluation, including a colonoscopy. Sixty-seven percent of those invited to be screened had completed the first screening round and were invited for further screening. More than 90% accepted repeated screening. During the 10-year period, 481 people in the screened group were diagnosed to have colorectal cancer compared to 483 people in the control group. In the screened group, 205 people died compared to 249 deaths in the control group. Therefore the mortality in the screened group compared to the control was 0.82. In an updated report, after seven rounds of biennial screening, the mortality in the screened group was reduced to < 0.7 compared to the unscreened ones.[39]

The UK[40] group randomized people aged 45 to 74 years who lived in the Nottingham area, to biennial screening (76 466) using FOBT or no screening (76 384). As in the Danish study, the stools are not rehydrated. Positive FOBT is followed-up by a colonoscopy. At a median follow-up of 7.8 years, the cumulative colorectal cancer mortality in the screened group is 15% lower than in the control group.

Hence, FOBT reduces the mortality from colorectal cancer by 15–33%. The magnitude is dependent on the frequency of the FOBT and whether the stool sample is rehydrated.[41] Rehydration increases the sensitivity of the test but also increases the false-positive rate from 2% to 10% in older persons.[36] With positive FOBT, the probability of finding a colorectal cancer or large adenoma in colonoscopy ranges from 17 to 46%.[41]

Sigmoidoscopy as a screening test is supported only by case-controlled studies. Selby *et al*.[42] and Newcomb *et al*.[43] reported reduced recto-sigmoid mortality with screening sigmoidoscopy. Ongoing randomized trials by the National Cancer Institute and United Kingdom on screening sigmoidoscopy will be able to provide further answers in future.

There are no controlled trials of screening double-contrast barium enema and colonoscopy. As colorectal cancer develops in benign adenomatous polyps which are amenable to endoscopic resection, it is reasonable to believe that colonoscopy will complement FOBT in reducing the

incidence and mortality from colorectal cancer. Furthermore, a proportion of early colorectal cancer do not bleed and FOBT is relatively ineffective in detecting polyps. However, colonoscopy has a risk of perforation of about 1/1000 and if colonoscopy is the routine screening test, this risk can be significant.

CONCLUSION

In conclusion, there are evidence to screen the average-risk individuals for the three cancers; cervical, breast and colorectal cancers. Despite the benefits of screening, there is no organized screening program for either of these cancers. Organizing such program is not easy. Ethics and economics are important factors to consider. This will determine the population to be screened, the screening test and the frequency of re-screening. Another obstacle to the success of such program is the poor compliance of those at risk. Health education and overcoming the economic barrier in such individuals are also essential before any screening program can succeed.

REFERENCES

1. Chia KS, Lee HP, Seow A, Shanmugaratnam, *Cancer Incidence in Singapore* 1993–1997.

2. Sigurdsson K, The Icelandic and Nordic cervical screening programs: trends in incidence and mortality rates through 1995, *Acta Obstet Gynecol Scand* **78**(6):478–485, 1999.

3. Miller AB, Lindsay J, Hill GB, Mortality from cancer of the uterus in Canada and its relationship to screening for cancer of the cervix, *Int J Cancer* **17**:602–612, 1976.

4. Walker JJ, Brewster D, Gould A, Raab GM, Trends in incidence of and mortality from invasive cancer of the uterine cervix in Scotland (1975–1994), *Public Health* **112**(6):373–378, 1998.

5. Hakama M, Chamberlain J, Day NE, *et al*., Evaluation of screening programmes for gynecological cancers, *Br J Cancer* **52**:669–673, 1985.

6. Sato S, Makino H, Yajima A, Fukao A, Cervical screening in Japan. A case-control study, *Acta Cytol* **41**(4):1103–1106, 1997.

7. Nanda K, McCrory DC, Myers ER, *et al*., Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review, *Ann Intern Med* **132**:810–819, 2000.

8. Payne N, Chilcott J, McGoogan E, *Liquid-based cytology in cervical screening*. Sheffield, England: University of Sheffield School of Health and Related Research, 2002.

9. Brown AD, Garber AM, Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing, *JAMA* **281**:347–353, 1999.

10. Shapiro S, Venet W, Strax P, Venet L, *Periodic screening for breast cancer. The Health Insurance Plan Project and its sequelae*, 1963–1986, Baltimore: Johns Hopkins University Press, 1988.

11. Alexander FE, Anderson TJ, Brown HK, *et al.*, 14 years of follow-up from the Edinburgh randomized trial of breast-cancer screening, *Lancet* **353**:1903–1908, 1999.

12. Tabar L, Fagerberg CJG, Gad A, *et al.*, Reduction in mortality from breast cancer after mass screening with mammography, *Lancet* I:829–832, 1985.

13. Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, *et al.*, The Swedish Two-County Trial twenty years later — Updated mortality results and new insights from long-term follow-up, *Radiol Clin North Am* **38**(4):625–651, 2000.

14. Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE, Followup after 11 years - update of mortality results in the Stockholm mammographic screening trial, *Breast Cancer Res Treat* **45**:263–270, 1997.

15. Andersson I, Aspergren K, Janzon L, Landberg T, Lindholm K, Linell F, Ljungberg O, *et al.*, Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial, *BMJ* **297**:943–948, 1988.

16. Bjurstam N, Bjorneld L, Duffy SW, Smith TC, Cahlin E, Eriksson O, Hafstrom LO, *et al.*, The Gothenburg Breast Screening Trial — first results on mortality, incidence and mode of detection for women ages 39–49 years at randomization, *Cancer* **80**:2091–2099, 1997.

17. Miller AB, To T, Baines CJ, Wall C, Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years, *J Natl Cancer Inst* **92**(18):1490–1499, 2000.

18. Miller AB, To T, Baines CJ, Wall C, The Canadian National Breast Screening Study: update on breast cancer mortality, *J Natl Cancer Inst* **22**:37–41, 1977.

19. Tarone RE, The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study, *Cancer* **75**:997–1003, 1995.

20. Boyd NF, The review of randomization in the Canadian National Breast Screening Study. Is the debate over? *Can Med Ass J* **156**:207–209, 1997.

21. Nystrom L, Rutqvist LE, Wall S, *et al.*, Breast cancer screening with mammography: overview of Swedish randomized trials, *Lancet* **341**: 973–978, 1993.

22. Breast cancer screening with mammography in women aged 40–49 years. Report of the Organising Committee and Collaborators, Falun Meeting, Falun, Sweden (21 and 22 March 1996), *Int J Cancer* **68**: 693–699, 1996.

23. Kerlikowske K, Grady D, Rubin SM, *et al.*, Efficacy of screening mammography. A meta-analysis, *JAMA* **273**:149–154, 1995.

24. Hendrick RE, Smith RA, Rutledge JH 3rd, Smart CR, Benefit of screening mammography in women aged 40–49: a new meta-anlaysis of randomized controlled trials, *J Natl Cancer Inst Monogr* **22**:87–92, 1997.

25. De Konig HJ, Boer R, Warmerdam PG, *et al.*, Quantitative interpretations of age-specific mortality reductions from the Swedish breast cancer screening trials, *J Natl Cancer Inst* **87**:1217–1223, 1995.

26. Harria JR, Morrow M, Bonadonna, Cancer of the Breast, in *Cancer: Principles & Practice of Oncology*, Vol 1, Devita VT, Hellman S, Rosenberg SA (eds.), JB Lippincott Co, pp. 1264–1332, 1993.

27. Tabar L, Fagerberg G, Day NE, Holmberg L, What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial, *Br J Cancer* **55**:547–551, 1987.

28. Rosenberg RD, Hunt WC, Williamson MR, *et al.*, Effects of age, breast density, ethnicity and estrogen replacement therapy on screening mammographic senstivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico, *Radiology* **209**(2):511–518, 1998.

29. Kerlikowske K, Gady D, Barclay J, Sickles E, Ernster V, Likelihood ratios for moderen screening mammogaphy: risk of breast cancer based on age and mammographic interpretation, *JAMA* **276**:39–43, 1996.

30. Beemsterboer PM, Warmerdam PG, Boer R, de Konig HJ, Radiation risk of mammography related to benefit in screening programmes: a favourable balance? *J Med Screen* **5**:81–87, 1998.

31. Fletcher SW, Black W, Harria R, Rimer BK, Shapiro S, Report of the International Workshop on screening for breast cancer, *J Natl Cancer Inst* **85**:1644–1656, 1993.

32. Hill D, White V, Jolley D, *et al.*, Self-examinations of the breast: is it beneficical? Meta-analysis of studies investigating breast self-examination and extent of disease in patients with breast cancer, *BMJ* **297**:271–275, 1988.

33. Thomas DB, Gao DL, Self SG, *et al.*, A randomized trial of breast self-examination in Shanghai: methodology and preliminary results, *J Natl Cancer Inst* **89**:355–365, 1997.

34. Semiglazov VF, Moiseyenko VM, Breast self-examination for the early detection of breast cancer: a USSR/WHO controlled trial in Leningrad, *Bull WHO* **65**:391–396, 1987.

35. Ng EH, Ng FC, Tan PH, Low SC, Chiang G, Seow A, *et al.*, Results of intermediate measures from a population-based, randomized trial of mammographic screening prevalence and detection of breast carcinoma among Asian women — The Singapore breast screening project, *Cancer* **82**:1521–1528, 1998.

36. Mandel JS, Bond JH, Church T, *et al.*, Reducing mortality from colorectal cancer by screening for rectal cancer for fecal occult blood, *N Engl J Med* **328**:1365–1371, 1993.

37. Mandel JS, Church T, Ederer F, Bond JH, Colorectal Cancer mortality: effectiveness of biennial sreening for fecal occult blood, *J Natl Cancer Inst* **91**(5):434–437, 1999.

38. Kronborg O, Fenger C, Olsen, *et al.*, Randomized study of screening for colorectal cancer with faecal occult blood test, *Lancet* **348**: 1467–1471, 1996.

39. Jorgensen OD, Kronborg O, *et al*., A randomized study of screening for colorectal cancer using faecal occult blood testing: results after 13 years and seven biennial screening rounds, *Gut* **50**:29–32, 2002.

40. Hardcastle JD, Chamberlain JO, Robinson MH, *et al*., Randomized controlled trial of faecal occult blood screening for colorectal cancer, *Lancet* **348**:1472–1477, 1996.

41. Lang CA, Ransohoff DF, What can we conclude from the randomized controlled trials of fecal occult blood test screening? *Eur J Gastroenterol Hepatol* **10**:199–204, 1998.

42. Selby JV, Freedman GD, Quesenberry CP, *et al*., A case-control study of screening sigmoidoscopy and mortality from colorectal cancer, *N Engl J Med* **326**:653–657, 1992.

43. Newcomb PA, Northfleet RG, Storer BE, *et al*., Screening sigmoidoscopy and colorectal cancer mortality, *J Natl Cancer Inst* **84**: 1572–1575, 1992.

This page intentionally left blank

52

Malignant Lymphoma

Lim Soon Thye and Miriam Tao

INTRODUCTION

Malignant lymphomas are fascinating and challenging to diagnose and treat because of the diversity of the clinical presentation and tumor behaviour. This diversity is due to the malignant transformation occuring at different stages of lymphocyte maturation and somatic hypermutation.[1] Lymphomas should always be excluded when a patient presents with metastatic disease, as potentially curative treatment can be given to patients with even very advanced stages of lymphoma. Some of the major advances in the last three decades are: 1) an international consensus on the pathological classification of lymphoid neoplasms; 2) validation of prognostic scores separating patients into different relapse risk categories; 3) identification of specific viral- or bacterial-associated lymphomas; 4) publications of mature multicenterd comparative trials that have established standard therapy that are evidence-based; and 5) development of targeted therapy.

NON-HODGKIN'S LYMPHOMA

Incidence

The overall incidence of NHL in Singapore has increased over the last 30 years, similar to trends in Western countries.[2,3] The ASR (age-standardized rate per 100 000 per year) for males were 7.0 (1993–1997) and 3.2 (1968–1972). This is almost 50% higher than the ASR for females, which were 4.1 (1993–1997) and 1.8 (1968–1972).

Age

The median age at diagnosis is 50+ years for all subtypes except for lymphoblastic lymphomas and Burkitt's lymphoma (which are more commonly seen in children and young adults). Primary mediastinal large B-cell lymphomas and anaplastic large T/null-cell are more commonly seen in 30+ years old.

Clinical Presentation

Nodal

Non-Hodgkin's lymphoma frequently presents with painless generalized or localized peripheral lymph node enlargement. There may be waxing and waning of the lymph nodes for months to years (indolent or low-grade lymphomas) or rapid growth over several weeks (aggressive or high-grade lymphomas). Intra-cavitary lymph node enlargement may be massive before causing symptoms (e.g. mediastinal adenopathy can cause breathlessness, retroperitoneal masses can cause back pain).

Extranodal

Lymphoma can involve any site in the body but the most common sites are the stomach, oropharynx, small intestine and the skin. Patient may present with just fever, weight loss and night sweats, "B symptoms", that can be associated in any stage or subtype of lymphoma.

Evaluation of Suspected Lymphoma:

History

1) Is the swelling/lymph node painful? (Tenderness is usually associated with infection)

2) Is the swelling localized or generalized? Localized swelling may be due to infection whereas disseminated adenopathy are more likely due to malignancy although infections like HIV infection may also cause this.
3) Has there been any weight loss, night sweats and fever (B symptoms)?
4) Any travel history or high risk behavior like multiple sex partners and sharing of needles?
5) Certain situations must always alert the clinician to the possibility of lymphoma: e.g. painless testicular mass in an elderly male.

Examination

1) Examine all lymph node regions, oral cavity for tonsillar involvement and the abdomen for mesenteric or retroperitoneal masses and hepatosplenomegaly. Also look for testicular or ovarian masses, focal neurological signs and skin involvement.
2) Look for infection or malignancy in the areas drained by the enlarged lymph nodes.

Diagnosis

1) *Excision biopsy*: The most representative enlarged lymph node or extranodal site should be biopsied and sufficient material obtained.
2) *Core needle biopsy*: If the only enlarged nodes are retroperitoneal and open biopsy is not possible, CT-guided core needle biopsy is acceptable.
3) *Fine needle biopsy* is *not* adequate for diagnosis of lymphoma.

Careful examination for peripheral lymph node enlargement may save a patient an unnecessary mediastinal biopsy or exploratomy laparatomy. If a high-grade or very aggressive lymphoma is suspected, it is ideal to expedite all investigations as urgent treatment may be needed. Histology should always be reviewed by a hemato-pathologist.

Histopathological Classification

Prior to 1982 there were many different classifications, which caused confusion and difficulty in comparing treatment outcomes at different institutions.

Working formulation[4]

This classification was based on only morphology (cellular feature and tissue architecture) and biological aggressiveness. Only 10 subtypes of NHL were recognized and no distinction was made between B- and T-cell lymphomas. Although this classification is still referred to, it is no longer adequate for current practice.

Revised European-American classification of lymphoid neoplasms (REAL)

This was based on immunological and genetic tests in addition to morphology. It recognized 12 types of B-cell and 11 types of T-cell lymphoid malignancies.

World Health Organization (WHO) classification[5]

This is based on the same principles and similar to the REAL classification with some modifications (Table 1).

Table 1 World Health Organization Classification of Neoplastic Diseases of the Lymphoid Tissues

B-Cell Neoplasms	T-Cell Neoplasms
Precursor B-cell neoplasm:	**Precursor T-cell neoplasm**:
Precursor B-lymphoblastic leukemia/lymphoma	Precursor T-lymphoblastic leukemia/lymphoma
Peripheral B-cell neoplasms	**Peripheral T-cell and NK-cell neoplasms**
B-cell chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma (SLL)	T-cell prolymphocytic leukemia
variant: with monoclonal gammopathy/plasmacytoid differentiation	variants: small cell cerebriform cell
B-cell prolymphocytic leukemia	T-cell large granular lymphocyte leukemia (LGL):
Lymphoplasmacytic lymphoma (immunocytoma)	T-cell type, NK-cell type
Mantle cell lymphoma (MCL)	Aggressive NK leukemia
variant: blastic	NK/T-cell lymphoma, nasal and nasal-type (angiocentric)
Follicular lymphoma (FL)	Sezary syndrome
	Mycosis fungoides

Table 1 Continued

B-Cell Neoplasms	T-Cell Neoplasms
variants: grade I (< 15% centroblasts) grade II (> 15% to 50% centroblasts) grade III (> 50% centroblasts) cutaneous	variants: pagetoid reticulosis MF-associated follicular mucinosis granulomatous slack skin disease
Marginal zone B-cell lymphoma (MZL) of Mucosa-associated lymphoid tissue (MALT-type)	Angioimmunoblastic T-cell lymphoma (AILD)
Nodal MZL +/− monocytoid B cells	Peripheral T-cell lymphomas, unspecified
Splenic MZL +/− villous lymphocytes	variants: lymphoepithelioid (Lennert's) T-zone
Hairy cell leukemia (HCL)	Adult T-cell leukemia/lymphoma HTLV-1 + (ATL/L)
Diffuse large B-cell lymphoma (DLCL)	variants: acute lymphomatous chronic smouldering Hodgkin-like
variants: centroblastic immunoblastic T-cell or histiocyte-rich anaplastic large B cell	Anaplastic large cell lymphoma (ALCL), CD 30+, T- and null-cell types
Diffuse large B-cell lymphoma subtypes	variants: lymphohistiocytic small cell
Mediastinal (thymic) large B-cell lymphoma	Primary cutaneous CD-30 positive T-cell Lymphoproliferative disorders
Intravascular large B-cell lymphoma	variants: lymphomatoid papulosis (type A and B) primary cutaneous anaplastic large cell lymphoma
Primary effusion lymphoma in HIV patients/serous lymphoma	Borderline lesions
Burkitt's lymphoma	Subcutaneous panniculitis-like T-cell lymphoma
variants: endemic sporadic atypical (pleomorphic) atypical, with plasmacytoid differentiation (AIDS-associated)	Enteropathy-type intestinal T-cell lymphoma
	Hepatosplenic + "γ/δ" T-cell lymphoma

Staging

Ann Arbor and Cotswold staging system[6,7]

This system is applicable both NHL and HL.

Table 2 Cotswold Staging Classification for Hodgkin's Disease

Stage	Description
I	Involvement of a single lymph node region or lymphoid structure (spleen, thymus, Waldeyer's ring)
II	Involvement of 2 or more lymph node regions on the same side of the diaphragm
III	Involvement of lymph node regions or structures on both sides of the diaphragm: III1: With involvement of splenic, hilar, celiac or portal nodes III2: With involvement of para-aortic, iliac or mesenteric nodes
IV	Involvement of one or more than extranodal sites in addition to a site for which designation E has been used
Designation Applicable to any Disease Stage	
A	No symptoms
B	Fever (temperature $>38°C$, drenching night sweats, unexplained weight loss $>10\%$ body weight in preceding 6 months
X	Bulky disease (mediastinal mass $>1/3$ the maximum thoracic diameter or nodal mass >10 cm)
X	Involvement of a single extranodal site contiguous or proximal to a known nodal site
CS	Clinical stage
PS	Pathological stage

Mandatory staging investigations:

1) Careful history for "B" symptoms and examination as described above.
2) Bone marrow aspirate and trephine biopsy from posterior iliac crest. It remained controversial whether bilateral bone marrow biopsies is more sensitive in detecting bone marrow involvement. A change in disease stage by performing a second biopsy occurred in only 6% of patients with indolent histology and in 2.5% of patients with aggressive NHL.[8] Flow cytometry and cytogenetics studies of marrow aspirate are indicated for prognostic purposes if marrow involvement is suspected and for selected cases such as lymphoblastic lymphoma, follicular, mantle or Burkitt's lymphoma.

3) CT scan of the thorax, abdomen and pelvis. CT scan of postnasal space and neck if there are symptoms from the upper aerodigestive tract. CT brain scan if there are neurological symptoms or focal neurological signs.
4) CSF examination and cytology is indicated for: (a) lymphoblastic or Burkitt's lymphoma; (b) aggressive lymphoma involving testes, paranasal sinus, extensive marrow or extranodal involvement; (c) HIV positive patients; and (d) primary CNS lymphoma.
5) Upper gastrointestinal and small bowel follow through contrast radiographs only if patient has gastrointestinal primary or Waldeyer's ring involvement.
6) Ultrasound of the contralateral testis in testicular lymphoma.

Pre-treatment investigations:

1) Blood tests — full blood count, erythrocyte sedimentation rate, coagulation profile, electrolytes, renal and liver function tests, lactate dehydrogenase (LDH), serum calcium.
2) Other investigations to assesss patient's ability to undergo therapy: (a) serology for Hepatitis B, Hepatitis C and HIV; and (b) 2D echocardiogram or MUGA scan for cardiac function prior to any anthracycline drugs.

Prognosis

The histologic type and patient characteristics as defined in the International Prognostic Index (IPI) can give an estimate of the prognosis.[9] The 5 prognostic factors are listed below (Table 3). The IPI score is

Table 3 International Prognostic Index and 2- and 5-year Survival for Aggressive Lymphomas

Parameter	Score "1" if Adverse Factor Present	Risk Group	IPI Score	2-year Survival (%)	5-year Survival (%)
Age	>60	Low	0 or 1	84	73
LDH	$>1\times$ normal	Low-intermediate	2	66	51
ECOG status	$\geqslant 2$	High-intermediate	3	54	43
Stage	III, IV	High	4 or 5	34	26
Extranodal involvement	>1 site				

Table 4 Survival by Histologic Type and the International Prognostic Index

Consensus Diagnosis	% 5-yr Overall Survival		% 5-yr Failure-free Survival	
	Index 0/1	Index 4/5	Index 0/1	Index 4/5
Follicular, all grades	84	17	55	6
Mantle cell	57	0	27	0
Marginal zone B-cell MALT	89	40	83	0
Marginal zone B-cell, nodal	76	50	30	0
Small lymphocytic (CLL)	76	38	35	13
Diffuse large B-cell	73	22	63	19
Primary mediastinal large B-cell	77	0	69	0
High grade B-cell, Burkitt-like	71	0	71	0
Precursor T-lymphoblastic	29	40	29	40
Peripheral T-cell all types	36	15	27	10
Anaplastic large T/null-cell	81	83	49	83

defined as the total number of adverse factors present and identifies 4 groups that correlate with survival outcome in aggressive lymphomas. The clinical relevance of REAL/WHO classification and IPI for all NHL histologic types was confirmed by the Non-Hodgkin's Lymphoma Classification Project (Table 4).[10]

Treatment

Treatment recommendations are determined mainly by the subtype of lymphoma, stage and IPI score. Lymphomas may be divided into low-grade and aggressive lymphomas.

Low-grade lymphomas

(E.g. Follicular grades I and II, small lymphocytic lymphoma, lymphoplasmacytic lymphoma)
Only 10–20% have localized disease. The majority present with disseminated disease and about 40% with marrow involvement. Median survival is 9 to 10 years.

- Asymptomatic patients may be observed (watch and wait strategy) without treatment.[11]

- Indications to start therapy are: (a) B symptoms; (b) progression of disease; (c) lymphomatous involvement of vital organs or structures resulting in obstruction or organ dysfunction; (d) marrow involvement; (e) bulky disease; or (f) transformation to a more aggressive lymphoma.
- The therapeutic options are extensive but no treatment has been shown to be curative[12] or clearly superior. The median duration of 1st remission is between 1.5 to 3 years. Recurrence is inevitable but multiple treatment options are available. Median duration of 2nd remission is about 13 months and median survival after recurrence was 4.5 years in one series.

Stage IA and IIA:

Radiotherapy potentially curative with 10-year freedom from relapse of 40–50% with involved field (IF)RT and 85% with total lymphoid irradiation (TLI).

Stage IIIA and IVA:

1) Watch and wait strategy.[13]
2) Alkylating chemotherapy agents, single or combination, e.g. chlorambucil or CVP (cyclophosphamide, vincristine, prednisolone), CHOP (cyclophosphamide, doxorubicin, vincristine, prednisolone).
3) Nucleoside analogues, single or combination, e.g. fludarabine or FMP (fludarabine, mitoxantrone, dexamethasone).
4) Monoclonal antibodies, e.g. Rituximab.
5) Radioimmunotherapy, e.g. Tositumomab (iodine-131 conjugated with anti-B1 antibody).

Marginal zone lymphoma MALT (mucosa-associated lymphoid tissue)

Gastric MALT lymphoma is a low grade lymphoma and frequently presents with non-specific GI symptoms and diagnosis requires endoscopy. *Helicobacter pylori* is present in over 90% of gastric MALT lymphomas. Anti-Helicobacter therapy as sole initial treatment for disease confined to mucosa and submucosa has shown 70% complete histological remission

of gastric MALT. It may take up to 5 months to achieve regression. Relapses may be treated with a second course of antibiotics, single agent chemotherapy, irradiation or surgery.

Aggressive lymphomas

(E.g. Diffuse large B cell lymphoma, follicular lymphoma grade 3, peripheral T-cell lymphoma, anaplastic large T/Null-cell lymphoma, and nasal NK/T-cell lymphomas)

- They are primarily treated with combination chemotherapy.
- CHOP remains the standard chemotherapy and showed complete response rates of 45–55%, with cure rates of 30–35%. A phase III comparison of CHOP with more dose intensive chemotherapy regimens (m-BACOP, ProMACE-CytaBOM, and MACOP-B) for the treatment of patients with aggressive NHL showed that CHOP was just as effective but much less toxic.[14] Patients with stage III and IV disease are usually treated with 6 to 8 cycles of chemotherapy.
- In localized aggressive NHL, a short course of CHOP (3–4 cycles) combined with radiotherapy to regions involved by lymphoma is an option. Although this approach was initially shown to be superior to 8 cycles of CHOP alone, excess late relapses and deaths from lymphoma are now reported with longer follow up. Consequently, patients are now treated with a short course of CHOP followed by radiation only if they have a very low risk of relapse.[15,16]
- Recently, a phase III trial in elderly patients with previously untreated advanced aggressive B-cell NHL compared 8 cycles of standard dose CHOP to the same regimen plus rituximab (CHOP-R). This trial showed a significant advantage with the addition of rituximab, an anti-CD20 antibody, CHOP in terms of complete remission rate (76% vs. 63%) and overall survival 3-year survival (62% vs. 51%).[17]
- Currently, up-front, high dose therapy followed by autologous stem cell transplant cannot be recommended outside of a clinical trial, even for patients at high risk of relapse. Ongoing studies may ultimately define which patients could benefit from these aggressive approaches.[18]

Stage I and II

non-bulky: CHOP × 4 + IFRT or CHOP 6–8 cycles
bulky ⩾ 10cm or large mediastinal mass: CHOP × 6–8 + IFRT

Stage III and IV

CHOP × 6-8 cycles or CHOP × 6-8 + anti-CD20 monoclonal antibody.

Relapse aggressive lymphoma

A number of second-line regimens have been developed using non-cross resistant drugs at higher doses for patients with relapsed aggressive lymphomas. Although response rates of up to 60–70% may be achieved with some of these regimens, they are seldom lasting and long-term disease-free survival is rare. Currently, high dose chemotherapy (HDC) and autologous stem cell transplant (ASCT) is the treatment of choice for these patients. An important factor in determining outcome with HDC followed by ASCT is the demonstration of chemosensitivity at relapse. The advantage of this approach was confirmed in a multicenter trial known as the PARMA trial.[19] In this study, patients with relapsed aggressive NHL first received 2 cycles of a second-line chemotherapy (DHAP) and those who responded were randomly assigned to receive either 4 additional cycles of DHAP or HDC followed by ASCT. With median follow up of more than 5 years, patients assigned to the high dose arm has a better overall survival (53%) compared to patients receiving conventional chemotherapy (32%). In a subsequent analysis of prognostic factors for the patients in this study, superior overall survival was seen in patients with at least on poor prognostic factor. For favorable patients with no adverse prognostic factor (i.e. IPI socre of zero), the outcome between HDC followed by ASCT was similar to DHAP alone.[20]

New treatment for NHL

1) *Monoclonal antibody*
   E.g. Rituximab is a chimeric mouse-human anti-CD20 monoclonal antibody. It attaches to the CD20 receptor, an antigen that is expressed only on B-lineage cells and is important for cell cycle initiation and differentiation. Rituximab is approved for use in relapsed follicular lymphoma.
2) *Radioimmunotherapy*
   In this approach, radioisotopes are chemically attached to antibodies. An example of a radioactive antibody is Bexxar, which is a mouse monoclonal anti-CD20 antibody tagged with the radioisotope I-131. Zevalin is another radioactive anti-CD20 antibody. It is tagged with ytrrium-90 (Y-90). The challenge for the future is to learn how best to incorporate these new options into the care of NHL.

3) *Biological therapy*
   In this approach, biological response modifiers are used to stimulate the patient's own immune system to attack and destroy the lymphoma cells. Examples include interferons and interleukins.
4) *Myeloablative and non-myeloablative transplant*
   Myeloablative allogenic transplant has been tested with some success in relapse and refractory NHL but is associated with a high rate of treatment related mortality and is therefore not routinely recommended. Non-myeloablative transplant is a newer approach and uses much lower doses of chemotherapy than in standard transplant. Immunosuppressive agents are given to allow donor cells to engraft and partly take over the patient's immune system. The donor cells then begin reacting against the lymphoma cells and killing them.
5) *Gene chip*
   Recently, a new molecular technology that has enabled clinicians to genetically distinguish lymphomas. This involves what is called a "lymphochip", which is essentially a small piece of glass that contains thousands of genes expressed by normal B cells in a grid-like pattern. Through this technique, it is discovered that diffuse large B-cell lymphomas (DLBCL) are actually two different diseases. One subgroup is termed germinal center B-like DLBCL and expresses genes that are hallmarks of the germinal center stage of B cell development. The other subgroup is termed activated B-like DLBCL and resembles activated peripheral blood B cells in gene expression. More than 75% of patients with germinal center B-like DLBCL were alive 5 years after treatment compared with fewer than 25% of patients with activated B-like DLBCL. It is hoped that such information would help physicians understand the molecular basis for the differences in treatment outcomes, and also in identifying patients who are not likely to be effectively treated by current treatment so that more specific or experimental treatment may be offered.

HODGKIN'S LYMPHOMA

Incidence

The ratio of NHL to HL is about 10:1. The rate for HL in 1993–1997 was considerably ASR 0.6 and 0.3 for males and females respectively.

Age

Common age range is 15 to 30 years.

Classification

The WHO classification of Hodgkin's lymphoma is as follows:

- Nodular lymphocyte predominance HD
- Classical HD: Nodular sclerosis (Grades I and II)

  Classical HD, lymphocyte-rich

  HD, mixed cellularity

  HD, lymphocyte depletion

Staging

The staging investigations required are essentially similar to that for NHL. In the 1970s, when radiation was primary treatment for HL, staging laparotomy to detect occult intra-abdominal disease was an essential component in staging. Patients who were found to have occult intra-abdominal disease would require more extensive radiation fields that include the para-aortic lymph nodes and spleen. However, the current approach to early stage patients is combination of brief chemotherapy with limited radiation therapy. The inclusion of systemic chemotherapy allows a patient to be treated wiht a smaller radiation field and eliminates the need for staging laparotomy as chemotherapy is able to eradicate occult intra-abdominal disease. Moreover, staging laparotomy was associated with complications such as infections from splenectomy and operative risks. Thus, with combined modality treatment, staging laparotomy and splenectomy are no longer required in the management of HL.

Prognosis

The overall survival for patients with HL is 81% with current treatments. There are prognostic factors that identify those patients who would benefit from less toxic treatment and those in whom standard therapy may fail.

Localized HL (Stage I and II) may be separated into very favorable, favorable and unfavorable prognostic groups based on specific criteria.

Table 5 IPI for Advanced HD and 5-yr Progression-free Survival and Overall Survival

Combined Score Grouped	# Pts (%)	Rate of FFP (%)	Rate of OS (%)
0 or 1	459 (29)	79	90
⩾2	114 (371)	60	74
0–2	939 (58)	74	86
⩾3	679 (42)	55	70
0–3	1317 (81)	70	83
⩾4	301 (19)	47	59

These vary across institutions but include presence or absence of B symptoms, large mediastinal mass, bulky adenopathy, elevated ESR, extranodal involvement and number of nodal sites involved.

A prognostic scoring system is also developed for advanced stage HL. Based on data collected from more than 5000 patients with advanced HL, seven unfavorable factors were identified and were used to create a prognostic index. These seven factors are: serum albumin less than 40 g/L, hemoglobin less than 10.5 g/dL, male gender, age over 45 years, stage IV, while blood cell count ≥ 15000/μL, and lymphocyte count less than 600/μL and/or less than 8 percent of the while blood cell count. Each factor has the same weight in increasing the likelihood of relapse. The prognositc index is thus simply calculated by adding the number of unfavorable factors present. Table 5 showed the spread in the progression free and overall survival at five years associated with the various risk groups.

Treatment

Early stage HL (stage I & II)

Total lymphoid irradiation was shown to be curative in localized HL almost 30 decades ago. However, the complications in long-term survivors and death from causes other than HD became the driving force to redefine the optimal treatment for HD. Randomized studies in early stage Hodgkin's disease showed that combining various short courses of chemotherapy with total nodal irradiation resulted in better freedom from disease progression compared to raidation alone. The high cure rates with combination chemotherapy in advanced stage HD paved the way for combined modality therapy (CMT) with chemotherapy and

smaller volumes and doses of radiotherapy.[21] The goals of current treatment recommendations in early stage HD is to maintain high cure rates while minimizing long-term pulmonary, cardiac complications and second neoplasms.

1) *Very Favorable*
   Criteria: Female with IA or IIA nodular sclerosis HL, $\leqslant 3$ nodal sites and < 26 years.
   Treatment: Extended Field (EF) RT. Alternatively Mantle RT after negative laparotomy or ABVD (adriamycin, bleomycin, vinblastine, tacarbazine) $\times$ 6 cycles. Note: Mantle RT without staging laparotomy has been associated with 6-year relapse-free survival of only 68%.[22]

   Criteria: Male with IA lymphocyte predominant HL unilateral high neck (above the thyroid notch) or epitrochlear.
   Treatment: Involved Field (IF) RT only.
2) *Favorable*
   Criteria: No B symptoms, no bulky adenopathy or mediastinal mass to thoracic ratio (MTR) < 0.35, ESR < 50 mm/hr, Age < 50 years, less than 4 separate sites of nodal involvement.
   Treatment: ABVD $\times$ 4 followed by involved field radiotherapy (IFRT) 36–40 Gy. Alternatively ABVD $\times$ 6 in smokers or females < 27 years.
3) *Unfavorable*
   Criteria — Any of the following: ESR > 50, Age > 50 years, $\geqslant 4$ separate sites of nodal involvement, mediastinal mass to thoracic ratio (MTR) > 0.35, bulky adenopathy $\geqslant 10$cm, B symptoms.
   Treatment: ABVD $\times$ 6 followed by involved field radiotherapy (IFRT) 36–40 Gy.

Advanced stage HL (stage III or IV)

1) *Combination chemotherapy ABVD (adriamycin, bleomycin, vinblastine, dacarbazine) 6 to 8 cycles is currently standard treatment*
   Patients with advanced stage HL are treated with multiagent chemotherapy. The first chemotherapy regimen used was MOPP, which was developed in the 1960s. However, toxicity has been an important limitation to MOPP. It is replaced by ABVD, which has

been shown to be as active as MOPP without the associated toxicities, especially secondary malignancies and sterility. Side-effects with ABVD are mainly transient although irreversible toxicities such as pulmonary or cardiac toxicity may occur in less than 5% of patients. The approximate complete remission rate and 5-year survival rate with ABVD were 70 and 60%, respectively.

2) *Escalated BEACOPP*
   - BEACOPP is a complex chemotherapy regimen containing seven drugs. The German Hodgkin Study Group did a study to compare dose-escalated BEACOPP with standard BEACOPP and a ABVD-like regimen. This study involved 1195 patients and showed that dose-escalated BEACOPP is associated with a higher rate of freedom from treatment failure at five years compared to the other regimens. Overall, five-year survivals were 91% for escalated BEACOPP, 88% for standard BEACOPP and 83% for the ABVD-like regimen. However, escalated-BEACOPP was also associated with greater hematologic toxicity and secondary malignancies. Further follow up will be necessary to determine the safety of this new aggressive approach.

Relapsed HL

1) Patients relapsing after being treated with radiation alone for early stage HL may be treated with systemic chemotherapy such as ABVD for 6–8 recycles.
2) Patients relapsing after having received chemotherapy are still potentially curable and therapy depends on the duration of remission after the initial chemotherapy as well as the presence of other prognostic factors. If the relapse occurs more than 12 months after the first complete remission, the patient may be treated with combination chemotherapy. At five years, about 40–50% of them will remain disease free. Relapses after MOPP may be treated with ABVD. Relapses after ABVD may be treated with MOPP or one of the salvage regimens. Patients relapsing within 12 months, patients relapsing at multiple sites or patients with constitutional symptoms at relapse have much poorer outcomes. They are candidates for high dose chemotherapy followed by stem cell transplant.

Follow-up

Patients with lymphoma are re-evaluated during and after completion of treatment to determine their response and subsequent management. If they are in complete remission, they are monitored every 3 months for the first 2 years as relapses of aggressive lymphomas are most likely to occur during this period. They should remain under lifelong follow-up for late relapses as well as for long-term pulmonary or cardiac complications and secondary malignancies.

REFERENCES

1. Kuppers R, Klein U, Hansmann ML, Rajewsky K, Cellular origin of human B-cell lymphomas, *N Engl J Med* **341**(20):1520–1529, 1999.

2. Chia KS, Seow A, Lee HP, Shanmugaratnam K, *Cancer Incidence in Singapore 1993–1997* Singapore Cancer Registry, Report No. 5, pp. 116–118, 2000.

3. Chia KS, Lee HP, Seow A, Shanmugaratnam K, *Trends in Cancer Incidence in Singapore 1968–1992* Singapore Cancer Registry, Report No. 5, pp. 176, 181, 1996.

4 National Cancer Institute sponsored study of classifications of non-Hodgkin's lymphomas: summary and description of a working formulation for clinical usage. The Non-Hodgkin's Lymphoma Pathologic Classification Project, *Cancer* **49**(10):2112–2135, 1982.

5. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller–Hermelink HK, Vardiman J, Lister TA, Bloomfield CD, World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997, *J Clin Oncol* **17**(12):3835–3849, 1999.

6. Carbone PP, Kaplan HS, Musshoff K, Smithers DW, Tubiana M, Report of the Committee on Hodgkin's Disease Staging Classification, *Cancer Res* **31**(11):1860–1861, 1971.

7. Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, Young RC, Rosenberg SA, Coltman CA, Tubiana M, Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting, *J Clin Oncol* **7**(11): 1630– 1636, 1989.

8. Wang J, Weiss LM, Chang KL, Slovak ML, Gaal K, Forman SJ, Arber DA, Diagnostic utility of bilateral bone marrow examination: significance of morphologic and ancillary technique study in malignant, *Cancer* **94**(5):1522–1531, 2002.

9. The International Non-Hodgkin's Lymphoma Prognostic Factors Project, A Predictive Model for Aggressive Non-Hodgkin's Lymphoma, *N Engl J Med* **329**:987–994, 1993.

10. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project, *Blood* **89**(11):3909–3918, 1997.

11. Horning SJ, Rosenberg SA, The natural history of initially untreated low-grade non-Hodgkin's lymphomas, *N Engl J Med* **311**(23): 1471–1475, 1984.

12. Peterson BA, Current treatment of follicular, low grade lymphomas, *Semin Oncol* **26**(Suppl 14):2–11, 1999.

13. Young RC, Longo DL, Glatstein E, *et al*., The treatment of indolent lymphomas. Watchful waiting Vs aggressive combined modality treatment, *Semin Hematol* **25**:11–26, 1988.

14. Fisher RI, Gaynor ER, Dahlberg S, *et al*., Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin's lymphoma, *N Engl J Med* **328**:1002–1006, 1993.

15. Miller TP, Chemotherapy alone compared with chemotherapy plus radiotherapy for localized intermediate- and high-grade non-Hodgkin's lymphoma, *N Engl J Med* **339**(1):21–26, 1998.

16. Miller TP, Leblanc M, Spier C, Chase E, Fischer RI, CHOP Alone Compared to CHOP Plus Radiotherapy for Early Stage, Aggressive Non-Hodgkin's Lymphomas: Update of the Southwest Oncology Group (SWOG) Randomized Trial, ASH 2001 (abstr 3024).

17. Coiffier B, CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma, *N Engl J Med* **346**(4):235–242, 2002.

18. Shipp MA, Abeloff MD, Antman KH, *et al*., International Consensus Conference on High Dose Therapy with Hematopoietic Stem Cell Transplantation in Aggressive Non-Hodgkin's Lymphoma: report of the jury, *J Clin Oncol* 423–429, 1999.

19. Philip T, Guglielmi C, Hagenbeek A, Somers R, Van der Lelie H, Bron D, Sonneveld P, Gisselbrecht C, Cahn JY, Harousseau JL, *et al*., Autologous bone marrow transplantation as compared with salvage chemotherapy in relapses of chemotherapy-sensitive non-Hodgkin's lymphoma, *N Engl J Med* **333**(23):1540–1545, 1995.

20. Blay J, Gomez F, Sebban C, Bachelot T, Biron P, Guglielmi C, Hagenbeek A, Somers R, Chauvin F, Philip T, The International Prognostic Index correlates to survival in patients with aggressive lymphoma in relapse: analysis of the PARMA trial, Parma Group, *Blood* **92**(10):3562–3568, 1998.

21. Santoro A, Bonfante V, Viviani S, Devizzi L, Zanini M, Soncini F, Gasparini M, Valagussa P, Bonadonna G, Subtotal nodal (STNI) vs involved field (IFRT) irradiation after 4 cycles of ABVD in early stage Hodgkin's disease, *Proc Amer Soc Clin Oncol* **15**:415, 1996 (abstr).

22. Hagenbeek A, Carde P, Noordijk EM, Thomas J, Tirelli U, Monconduit M, Eghbali H, Mandard AM, Henry Amar, Prognostic factor tailored treatment of early-stage Hodgkin's disease. Results of a prospective randomized stage III clinical trial in 762 patients (H7 study), *Blood* **90**:565a, 1997 (abstr, suppl 1).

53

Rational Use of Tumor Markers

Wong Nan Soon and Khoo Kei Siong

INTRODUCTION

Tumor markers are substances that are produced by normal tissues in minute quantities, but are secreted in excess by malignantly transformed tissues and in other pathological states. In recent years, the definition of tumor markers has expanded to include not only circulating products in the blood, urine and cerebral spinal fluid, but also the assay of genes and oncogene products.

The ideal tumor marker is specific to the type of malignancy and can be detected with sufficient sensitivity even in the early stages of disease. To date, no such tumor marker exists.

The use of tumor markers has become increasingly popular among clinicians. Potential applications include screening, diagnosis, prognostication, treatment monitoring, localization and therapy.

The common clinically useful tumor markers in wide usage include carcinoembryonic antigen (CEA), prostatic specific antigen (PSA), alpha fetoprotein (AFP), beta human chorionic gonadotropin (b-hCG), carbohydrate antigen 19-9 (CA19-9), carbohydrate antigen 15-3 (CA15-3) and carbohydrate antigen 125 (CA125).

CEA

Biology

CEA is a glycoprotein with a molecular mass of 180 kDa. It is an oncodevelopmental human tumor marker, found in minute quantities on the luminal aspect of adult colonic epithelial cells but in large amounts on the basolateral membranes of embryonic intestine and colonic tumors. It is postulated to function as an intercellular adhesion molecue.

The CEA gene family has been cloned and is located on chromosome 19.

CEA is elevated in colorectal carcinoma. The incidence of positive CEA assays ranges from 20% in patients with Duke's A to 90% in patients with metastatic disease.[1]

In addition, elevated CEA levels are seen in other malignancies such as carcinomas of the lung, breast, stomach and pancreas. CEA is elevated in 30% of small cell lung cancer and 60–70% of non-small cell lung cancer.

Elevated CEA levels can be seen in a number of benign conditions. These include smoking, bronchitis, emphysema, hepatitis, cirrhosis, cholangitis, pancreatitis, renal disease, inflammatory bowel disease and colonic polyps. Elevated levels are also seen in diabetes and collagen vascular disease.

CEA has a half-life of 1–5 days and clearance is hepatic.

Clinical Utility

Screening

CEA has limited utility in screening for colorectal carcinoma as elevated plasma levels are encountered in a number of benign conditions.

CEA has a sensitivity for the detection of early colorectal carcinoma of 40% and a specificity of 90%. Studies in which asymptomatic individuals with elevated CEA were investigated have shown a low cancer detection rate ranging from 0.1 to 4%.[2,3]

Diagnosis

Due to lack of specificity, elevated CEA cannot replace histologic diagnosis in the diagnosis of colorectal and other malignancies.

In a patient with metastatic carcinoma of unknown origin, CEA has not been found to be useful in identifying the primary site and origin. However, a markedly elevated CEA is highly suggestive of a malignancy of epithelial origin.

Prognosis

Pre-operative CEA is a prognostic marker in colorectal carcinoma independent of stage and tumor grade.

In one series, the risk of recurrence is increased by 1.62-fold if the preoperative CEA was >2.5 ng/mL and by 3.25-fold if the CEA was >10 ng/L. A study has also found a shorter disease free survival of 13 months in patients with pre-operative CEA >5 ng/mL compared to 23 months with CEA <5 ng/mL. Stage specific survival is higher in patients with CEA <5 ng/mL.

In breast cancer, the prognostic significance of elevated CEA is controversial.

Treatment monitoring

After successful resection of colorectal cancer the CEA level should decline to normal levels within 4 to 6 weeks. A persistently elevated CEA indicates incomplete resection or metastatic disease, and is an adverse prognostic indicator.

CEA is currently not recommended as the sole test for monitoring response to chemotherapy in patients with metastatic colorectal carcinoma. However if no other simple test is available to indicate response to treatment CEA monitoring every 2 to 3 months may be used. In this scenario, two values above baseline are adequate to document progressive disease even in the absence of radiological corroboration.[2,4]

Detection of recurrence

In colorectal carcinoma, serial CEA monitoring is presently regarded as the most cost-effective non-invasive method for the detection of resectable recurrences.

Two consecutive elevations in CEA yielded a sensitivity of 84% and a specificity of 100% in the detection of recurrent disease according to one study. However other studies have found false-positive rates of 30%.

CEA monitoring enables detection of recurrences at an earlier stage, which would enable resection of isolated liver metastasis with curative intent. It is thus recommended that CEA is monitored every 2 to 3 months for 2 years after primary treatment for stage II and III colorectal carcinoma in individuals who are medically fit to undergo resection of detected liver metastases.

In breast cancer, CEA monitoring has limited utility in the monitoring of recurrent disease. The sensitivity of CEA for this purpose is 50% with a false-positive rate of 12%. The lead time gained ranges from 2.5 to 8 months before the onset of clinical disease. Due to the lack of curative treatment options early detection of disease recurrence does not translate into improved patient survival or quality of life.

AFP

Biology

Alpha fetoprotein is an oncodevelopmental glycoprotein with a molecular weight of 70 kDa produced in high amounts in fetal life and is homologous to albumin. Levels remain high in infancy until 6 months to 1 year of age.

In adulthood AFP is elevated in 85% of hepatocellular carcinoma and 40–60% of patients with non-seminomatous germ cell tumor depending on stage. Other malignancies associated with elevated levels include gastric, pancreatic and lung carcinoma as well as hepatoblastoma.

Benign conditions such as pregnancy, alcoholic disease, hepatitis, cirrhosis, biliary tract obstruction, ataxia telangiectasia and hereditary tyrosinemia can also cause elevated AFP but levels are generally between 20–400 μg/L.

The half-life of AFP is 5 days.

Clinical Utility

Screening

Screening for hepatocellular carcinoma in the high-risk population of chronic hepatitis B carriers is currently the standard of care. It is pertinent to note however that no study exists to date that documents improved disease specific survival with screening. AFP $>$ 20 μg/L has approximate sensitivity of 60% and specificity of 90% in this setting.[5]

Diagnosis

In the presence of liver lesions on imaging, AFP of > 500 μg/L is diagnostic of hepatocellular carcinoma.

In germ cell tumors, elevated AFP is found in non-seminomatous germ cell tumors but not in pure seminomas, hence enabling the two to be distinguished.

Prognosis

AFP is a prognostic marker in patients with non-seminomatous germ cell tumors. A study by the International Germ Cell Cancer Collaborative Group found that patients with advanced non-seminomatous germ cell tumor with AFP < 1000 μg/L had a 5-year survival of 83%. By comparison the 5-year survival was 57% for patients with AFP $> 10\,000$ μg/L.

Elevated AFP is also associated with poor prognosis in hepatocellular carcinoma. In one study the relative risk of death was 1.8 for AFP positive patients compared to AFP negative ones and median survival was significantly shorter.[6]

Treatment monitoring

In advanced non-seminomatous germ cell tumors the level of AFP should decline at half-life during chemotherapy. Failure of marker decline is indicative of poor response and should prompt a switch in therapy. Persistently elevated AFP after orchidectomy in patients with stage I non-seminomatous germ cell tumor is indicative of disease outside the peritoneum and should prompt systemic chemotherapy.

Detection of recurrent disease

AFP is useful in the early detection of recurrent disease in patients with germ cell tumors. In patients with non-seminomatous germ cell tumor with disease confined to the testis (T1) who opt for observation rather than retroperitoneal lymph node dissection the risk of relapse is 20%. In this setting serial monitoring with AFP and b-HCG is recommended monthly for the first year, 2-monthly for the second year and quarterly for the third year and less frequently thereafter.

Monitoring of AFP is also recommended in patients who have had curative resection for hepatocellular carcinoma.

bHCG

Biology

HCG is a glycoprotein hormone with a molecular weight of 50 kDa composed of an alpha and a beta subunit. The beta subunit is specific to HCG and is detected in current radioimmunoassays (beta-HCG), while the alpha subunit is structurally common to LH, FSH and TSH.

HCG is secreted by placental syncytiotrophoblast cells and functions to facilitate implantation of the embryo and maintenance of the corpus luteum.

Elevated levels of HCG are seen in all patients with trophoblastic disease, in 20–40% of non-seminomatous germ cell tumors depending on stage and 20% of advanced pure seminomas. Elevated levels are also seen in other malignancies including melanomas, breast, gastrointestinal, lung and ovarian cancers.

Benign conditions associated with elevated hCG include pregnancy, cirrhosis, duodenal ulcer and inflammatory bowel disease.

Excretion of hCG is via the renal route and the half-life is approximately 24 hours.

Clinical Utility

Screening

HCG is not indicated as a screening test for germ cell tumors or gestational trophoblastic disease in the asymptomatic population due to low disease prevalence.

Diagnosis

In the setting of gestational trophoblastic disease, persistently elevated HCG after evacuation of molar pregnancy in the presence of metastatic lesions on chest XR or CT scan is sufficient to make the diagnosis of choriocarcinoma and prompt chemotherapy without obtaining histology.

In germ cell tumors elevated HCG is useful to distinguish germ cell tumor from other causes of mediastinal masses.

Prognosis

A HCG level greater than 100 000 mIU/L is associated with a high risk of persistent gestational trophoblastic disease after evacuation of a complete

mole. Although controversial short courses of chemoprophylaxis have been used in this situation.

In patients with choriocarcinomas, the extent of elevation of hCG is used as part of a prognostic scoring system to identify patients who require more intensive chemotherapy.

Similarly hCG is also used to stratify patients with non-seminomatous germ cell tumors into good, intermediate and poor risk groups.

Treatment monitoring

In patients with gestational trophoblastic disease, the level of HCG should become undetectable after 8 to 10 weeks after evacuation of a molar pregnancy. Weekly determinations of HCG until 3 consecutive readings are undetectable followed by monthly tests for 6 months is used in the follow-up of patients post evacuation.

In choriocarcinoma monitoring with HCG to document a 1 log decline in levels within 18 days of starting chemotherapy followed by monthly determinations is recommended.

In advanced germ cell tumors b-HCG should decline at half-life with appropriate chemotherapy. Failure to do so should prompt a switch in therapy. Persistently elevated b-HCG after orchidectomy in patients with stage I non-seminomatous germ cell tumor is indicative of disease outside the peritoneum and should prompt systemic chemotherapy.

Detection of recurrence

HCG is indicated for the detection of recurrent disease in patients with treated germ cell tumors and gestational trophoblastic disease.

CA19-9

Biology

CA19-9 is a glycoprotein mucin first discovered in 1979 using a monoclonal antibody raised against the human colon adenocarcinoma cell line designated SW1116. It has a molecular mass of 36 kDa and is structurally related to the human Lewis, a blood group determinant.

It is postulated to function in cell adhesion.

Elevated levels of CA19-9 are seen in the majority of cases of pancreatic cancer, cholangiocarcinomas, stomach and colorectal carcinomas and

mucinous ovarian carcinomas. Less commonly, elevated levels are also found in lung, breast, uterine and non-mucinous ovarian carcinomas.

Benign conditions associated with elevated CA19-9 include pancreatitis, cirrhosis, cholecystitis, biliary obstruction, gastric ulcer disease and pulmonary diseases.

CA19-9 has a serum half-life of 1–3 days.

Clinical Utility

Screening

In the detection of pancreatic cancer, CA19-9 has a sensitivity of 78–90% and a specificity of 95%. However due to the low prevalence of pancreatic carcinoma in the general population, the positive predictive value of CA19-9 is approximately 2%, making it unsuitable for screening.[7] An exception to this is the use of CA19-9 for early detection of cholangiocarcinoma in the high-risk population of patients with sclerosing cholangitis.

Diagnosis

Numerous conditions mentioned above can cause elevation of CA19-9. Hence marker elevation alone is not sufficient for making the diagnosis of pancreatic carcinoma and tissue diagnosis is required.

Prognosis

Both pre-operative and post-operative elevation of CA19-9 levels have been found to correlate inversely with survival in small retrospective studies.[8] CA19-9 is not routinely used for prognostic purposes in pancreatic carcinoma.

Treatment monitoring

Despite limited data, this is the most common application of CA19-9. A rising value of CA19-9 during palliative chemotherapy should prompt imaging procedures to assess disease response with a view towards changing treatment. A downward trend of CA19-9 has been associated with improved survival in one study.[9]

Detection of recurrence

CA19-9 detects recurrent disease with sensitivity of 88% and a lead time of up to 7 months. However due to the lack of curative treatment modalities for recurrent disease monitoring of CA19-9 after surgical resection has limited impact on patient survival.

PSA

Biology

First identified in 1971, PSA is a glycoprotein enzyme with a molecular weight of 30 kDa. It is produced predominantly by epithelial cells lining the acinii and ducts of the prostate gland.

While initially thought to be tissue specific, PSA has since been found in female periurethral tissue, breast cancer tissue, primary ovarian cancer and in breast milk.

In the serum it exists as both free form as well as a complexed form bound to alpha-1-chymotrypsin and alpha-2-marcoglobulin.

PSA functions as a kallikrein-like serine-protease, and its physiologic role is that of degradation of proteins in semen, leading to semen liquefaction.

To date, prostate cancer is the only malignancy giving rise to elevated serum PSA.

Other benign conditions that may lead to elevated PSA include benign prostatic hypertrophy, prostatitis, urinary retention, rigid cystoscopy, transurethral resection of the prostate, prostate biopsy, prostate massage and ejaculation. Prostatic intraepithelial neoplasia, digital rectal examination and flexible cystoscopy do not appear to cause clinically significant elevations of PSA.

Drugs can cause a decline in PSA levels leading to false-negative studies. These include LHRH analogues such as goserelin, antiandrogens such as flutamide, the 5-alpha reductase inhibitor finasteride and herbal remedies such as PC SPES and serenoa repens.

The PSA gene is located on 13q of chromosome 19. The half-life of PSA is 2 to 3 days after radical prostatectomy. Following radiotherapy decline of PSA to nadir levels is reached within a median of 17 months.

Clinical Utility

Screening

PSA > 4.0 ng/mL has a sensitivity of 75% while a PSA < 4.0 ng/mL has a negative predictive value of 67.5–80%. Hence 20–30% of tumors will be missed when PSA alone is used as a screening test. Digital rectal examination is complementary to PSA and increases the sensitivity of screening.

Various measures have been undertaken to improve the sensitivity of PSA testing. One such method is the age-adjusted PSA. For example a lower cut-off of 2.5 ng/mL is used to define normal PSA levels in younger patients below the age of 40.

Another method to improve sensitivity involves PSA velocity, which follows serum PSA values over time. Some investigators feel that an increase of 0.75 ng/mL per year warrants further work-up.

While both age-adjusted PSA and PSA velocity improve sensitivity the trade-off is an increased number of unnecessary prostate biopsies.

An elevated PSA (> 4.0 ng/mL) has a specificity of 60–70%. The positive predictive value of elevated PSA above 4.0 ng/mL in a patient age over 50 is 20–30%.

Modifications to PSA testing have been attempted to improve specificity but at the expense that some cancers will be missed.

One method is age-adjusted PSA using a higher normal PSA cut-off value for older patients.

Another method involves the free-to-total PSA ratio. For reasons that are unclear a lower free-total ratio is seen in prostate cancer while a higher free-total ratio is seen in benign conditions. A free-total ratio of less than 14–28% is taken by various investigators to warrant a prostate biopsy.

PSA density is defined as the ratio of PSA to gland volume and is another method of improving specificity based on the fact that a larger prostate produces more PSA.

No consensus has been reached to date regarding the optimal use of these maneuvers in view of the various trade-offs in sensitivity and specificity.

The use of PSA in screening for prostate cancer remains controversial at this point in time. Given the long natural history in most cases of prostate cancer and the older patient population for this malignancy, many patients will die from other causes with microscopic clinically insignificant prostate cancer even if early disease is detected through screening.

At present, if early detection is offered, it should be limited to men age over 50 with a life expectancy of more than 10 years.

Two randomized controlled trials are underway in the United States and Europe to address the question of whether PSA screening for prostate cancer will lead to significant reduction in disease morbidity and mortality to justify the cost and significant morbidity associated with treatment.

Diagnosis

PSA has limited utility as a sole diagnostic marker of prostate cancer due to the lack of specificity. However it has a place in the evaluation of metastatic cancer to bone of unknown primary in the male patient to exclude prostatic cancer which is potentially amenable to hormonal therapy.

Pre-treatment staging

Serum PSA is proportional to the extent of prostate cancer. In one series 5% of patients with pre-operative PSA greater between 4.0–10 ng/mL had extraprostatic disease compared with 15% of patients with PSA between 20–30 ng/mL.

Pre-operative PSA is useful for determining the need for staging bone scan and CT scan. In one series the incidence of positive bone scan was 0.8% in patients with PSA < 10 ng/mL and 2.6% in patients with PSA between 15–20 ng/mL. Hence routine bone scan is not indicated in the asymptomatic patient with PSA < 20 ng/mL.

Similarly CT scan to detect pelvic lymphadenopathy is rarely positive in patients with PSA < 25 ng/mL and is not indicated.

Pre-treatment PSA < 10 ng/mL can also help to identify a subset of patients in whom the risk of pelvic lymph node metastasis is very low making lymph node dissection unnecessary.

Pre-treatment PSA is a predictive factor with a level of < 10 ng/mL predicting good response to local therapy.

Treatment monitoring

PSA is used in periodic monitoring after primary local therapy to detect early recurrence. The time to elevation (< 12 months) and doubling time (< 6months) of PSA help to point towards distant rather than local recurrence.

In addition it is used to assess response to treatment in the metastatic setting. The nadir PSA and percent PSA decline while on treatment predict for progression free survival and overall survival.

CA125

Biology

CA125 is a carbohydrate epitope discovered in the 1970s as a result of efforts to develop monoclonal antibodies active against ovarian cancer. While the antibody developed (termed OC125) possessed specificity against ovarian cancer it failed to interact with human effector cells to have value as a therapeutic agent and found application as a diagnostic agent.

CA125 is associated with a family of glycoproteins with variable molecular weights ranging from 200 kDa to 1000 kDa. The precise chemical composition and physiological function of CA125 is as yet undefined, nor has the gene been isolated. The major forms in serum have molecular weights ranging from 200 kDa to 400 kDa.

Elevated levels of CA125 are seen in 80–85% of epithelial ovarian cancers as a group, with a lower proportion of patients with early stage ovarian cancer and mucinous subtypes showing elevated levels.

In addition, levels are elevated in a wide variety of malignancies including colorectal, gastric, pancreatic, liver, endometrial and cervical cancer.

Benign disorders such as endometriosis, menstruation, pelvic inflammatory disease, acute pancreatitis, cirrhosis, peritonitis and ascites also lead to elevated CA125 levels.

The serum half-life of CA125 is 4.5 days.

Clinical Utility

Screening

CA125 determination is not useful in screening for ovarian carcinoma. The sensitivity of a single determination of CA125 is 78% in one study and the positive predictive value is under 10%.

Diagnosis

CA125 measurement is useful in the differential diagnosis of ovarian cysts. In the post-menopausal population, ovarian cysts under 3 cm

which are unilocular with normal doppler flow studies and normal CA125 are benign and can be followed up.

In the pre-menopausal population ovarian cysts associated with rising CA125 during serial follow-up or a single reading greater than 65 to 200 U/mL is indicative of malignancy.

Prognosis

Pre-operative level of CA125 is related to disease bulk and has not been found to have independent prognostic value on multivariate analysis.

Post-operative CA125 however has been found to have independent prognostic value.

Treatment monitoring

CA125 is useful for monitoring response to chemotherapy in ovarian carcinoma. A rising level of CA125 is associated with progression of disease in 90% of cases. Similarly a falling CA125 is associated with disease response to treatment.

CA125 is however less useful in determining the presence of residual disease in that 50% of patients will still have residual disease on second look laparotomy despite return of CA125 to less than 35 U/mL.

Detection of recurrence

CA125 is a useful tool for the detection of recurrent disease in patients who have an intially elevated pre-treatment CA125 level. Coupled with physical examination the sensitivity of this test is 90%. While a rising level of CA125 predicts recurrence a normal CA125 does not exclude the diagnosis.

CA15-3

Biology

MUC1 (CD227) is a transmembrane mucinous glycoprotein expressed by most glandular and ductal epithelial cells. It functions in cell protection, lubrication and reduction of cell-cell and cell-matrix adhesion. Synonyms for this molecule include polymorphic epithelial mucin (PEM), epithelial membrane antigen (EMA) and episalin.

The gene for MUC1 is located on chromosome 1q21.

MUC1 gene is frequently overexpressed in malignant breast tumors.

CA15-3 assay detects an epitope on MUC1 using two monoclonal antibodies DF3 and 115D8 using a double determinant immunoassay. CA27.29 assay using monoclonal antibody BR27.29 detects an epitope on MUC1 which overlaps that recognized by DF3 antibody used in the CA15-3 assay.

CA15-3 is not specific to breast cancer. It is elevated in benign breast diseases in addition to hepatitis and cirrhosis. Elevated levels are also seen in colorectal, lung, ovarian and pancreatic carcinoma.

Clinical Utility

Screening

CA15-3 has a sensitivity of 9% for detection of stage I breast cancer and 19% for detection of stage II disease. The sensitivity increase to 38% for stage III disease and 75% for stage IV disease.

The false-positive rate is 5–6% for healthy subjects, increasing to 30% for patients with benign breast disorders. Hence CA15-3 is not a useful screening test for breast cancer.

Diagnosis

Similarly CA15-3 cannot be relied upon as the sole diagnostic test to confirm the presence of breast cancer due to its low positive predictive value.

Prognosis

Studies have shown that pre-operative CA15-3 does not correlate with prognosis. In addition, a low level of CA15-3 does not exclude the presence of metastatic disease and a given CA15-3 level cannot be used to determine the stage of disease.

Treatment monitoring

CA15-3 cannot be used alone as a definitive assessment of response to treatment of metastatic disease. Studies suggest that up to 34% of patients with disease responding to treatment will show a rising CA15-3 trend while 20% will have stable CA15-3 in the face of progressive disease.[14]

A portion of false elevations may be the result of transient marker rise in response to effective treatment. Hence, levels of CA15-3 should only be drawn several weeks after the initiation of chemotherapy and decision to stop therapy should be made after radiological confirmation of disease progression.

Detection of recurrence

The low sensitivity of CA15-3 for detection of early breast cancer suggests that the marker lacks the necessary sensitivity to detect early recurrence. Several studies have shown that 30% of patients with recurrent disease will not have an elevated CA15-3 level while 6% of patients without recurrence will show a false marker elevation.[15]

The lead time from marker elevation to clinically apparent recurrent disease ranges from 2 to 9 months but at present no curative salvage therapy exists for recurrent breast cancer hence negating any clinical benefit from early detection of recurrence.

CONCLUSION AND FUTURE DIRECTIONS

The advent of tumor markers was initially hailed as a major breakthrough in cancer research amidst great enthusiasm and hope that they will serve as the perfect tool in cancer screening, diagnosis and treatment monitoring. However subsequent research has shown that the ideal tumor marker remains an elusive goal.

Current tumor markers have no role in disease screening of low-risk populations at this point in time. As diagnostic tests they have limited utility and do not replace tissue diagnosis. Exceptions to this rule include AFP in the context of hepatocellular carcinoma, and b-HCG in choriocarcinoma and gestational trophoblastic disease.

In the work-up of metastatic cancer of unknown primary, tumor markers have limited utility due to lack of specificity. Exceptions to this include PSA in the setting of a male patient with bone metastasis of unknown primary and the use of AFP and b-HCG in the work-up of a patient with mediastinal or retroperitoneal lymphadenopathy.

The mainstay of application of currently available tumor markers is in monitoring of disease response to treatment and the detection of recurrent disease after successful completion of treatment.

Research is underway to develop algorithms using combination of tumor markers to enhance sensitivity and specificity, and new tumor markers are continuously being developed and tested.

SUMMARY TABLE

Tumor Marker	**CEA**		**AFP**		**bHCG**		**PSA**
Classification	**Oncofetal Protein**		**Oncofetal Protein**		**Hormone**		**Enzyme**
Utility	Colon	Breast	Liver	Germ Cell	Germ Cell	Gestational Trophoblastic Disease	Prostate
Screening	—	—	—	—	—	—	X
Diagnosis	—	—	X	—	—	X	—
Prognosis	X	—	X	X	X	X	—
Monitoring Response	X	X	—	X	X	X	X
Detection of Recurrence	X	—	X	X	X	X	X

Tumor Marker	**CA19-9**	**CA125**	**CA15-3**	**VMA**
Classification	Carbohydrate Epitope	Carbohydrate Epitope	Carbohydrate Epitope	Catecholamine
Utility	Pancreas, Cholangio-carcinoma	Ovary	Breast	Carcinoids
Screening	—	—	—	—
Diagnosis	—	—	—	X
Prognosis	—	—	—	—
Monitoring Response	X	X	X	—
Detection of Recurrence	X	X	—	—

REFERENCES

1. Phil G, Goldenberg NA, The carcinoembryonic antigen: past, present and future, *McGill J Medicine* **3**:46–66, 1997.
2. Bast RC Jr, Ravdin P, Hayes DF, *et al*., 2000 Update of Recommendations for the Use of Tumour Markers in Breast and Colorectal Cancer: Clinical Practice Guidelines of the American society of Clinical Oncology, *J Clin Oncol* **19**:1865–1878, 2001.
3. Fletcher RH, Carcinoembryonic Antigen, *Ann Intern Med*, **104**:66–73, 1986.
4. Ward U, Primrose JN, Firian PJ *et al*., The use of tumour markers CEA, CA-195, CA-242 in evaluating response to chemotherapy in patients with advanced colorectal carcinoma, *Br J Cancer* **67**:1132–1135, 1993.
5. Chalasani N, Horlander JC Sr, Said A, *et al*., Screening for hepatocellular carcinoma in patients with advanced cirrhosis. *Am J Gastroenterol* **94**:2988–2993, 1999.
6. Stillwagon GB, Order SE, Guse C, *et al*., Prognostic factors in unresectable hepatocellular cancer: Radiation Therapy Oncology Group Study 83-01, *Int J Radiat Oncol Biol Phys* **20**:65–71, 1991.
7. Roulston JE, Novel tumour markers: A diagnostic role in pancreatic cancer? *Br J Cancer* **70**:389–390, 1994.
8. Nakao A, Oshima K, Nomoto S, *et al*., Clinical usefulness of CA19-9 in pancreatic carcinoma, *Semin Surg Oncol* **15**:15–22, 1998.
9. Ziske C, Schlie C, Gorschlutter M, *et al*., Prognostic value of CA19-9 levels in patients with inoperable adenocarcinoma of the pancreas treated with gemcitabine, *B J Cancer* **89**:1413–1417, 2003.
10. American Urological Association: Prostate-Specific Antigen (PSA) Best Practice Policy, *Oncology* **14**:267–286, 2000.
11. Seregni E, Botti C, Ballabio G, *et al*., Biochemical characteristics and recent biological knowledge on prostate-specific antigen, *Tumori* **82**:72–77, 1996.
12. Bast RC Jr, Knapp RCL, CA125: History, Current Status, and Future Prospects, *McGill J Med* **3**:67–71, 1997.
13. NIH Consensus Statement, Ovarian Cancer Screening, Treatment, And Follow-up, **12**(3) 1994.
14. American Society of Clinical Oncologists, Clinical Practice Guidelines for the Use of Tumour Markers in Breast and Colorectal Cancer, *J Clin Oncol* **14**:2843–2877, 1996.
15. Safi F, Kohler I, Rottinger E, *et al*., Comparison of CA 15-3 and CEA in diagnosis and monitoring of breast cancer, *Intl J Biol Markers* **4**:207–214, 1989.

This page intentionally left blank

54

Metastatic Cancer of Unknown Primary

See Hui Ti and Sandeep Kumar Rajan

BACKGROUND

Metastatic cancer of unknown primary is defined as metastatic solid tumors at one or more sites for which the site of origin is not known after clinical and pathological evaluation.[1]

The incidence of metastatic malignant disease of undetermined origin in Singapore is approximately 3.1%. It is ranked 9th most commonly diagnosed cancer for male residents in Singapore, and 10th most commonly diagnosed cancer for female residents in Singapore.[2] In other parts of the world, the incidence ranges from 0.5% and 7.0% depending on the scope and the duration of the diagnostic investigation. This entity is a persistent and perplexing problem as it is difficult to define the limits of test in search for the primary tumor.

Most international series concur on the median survival to be less than 6 months for this group of patients as a whole. Despite this dismal overall survival statistics, there are subsets of patients with distinct clinical and pathological details who carry a better prognosis. These patients

could be considered for potentially curative management. The focus is therefore in identifying these patients.

PRESENTATION OF CARCINOMA OF UNKNOWN PRIMARY AND GENERAL PRINCIPLES OF MANAGEMENT

Most solid tumor cancers are described as arising from an organ and then have spread to lymph nodes or distant organs. Hence patients may present as carcinoma of unknown primary as lymph-node enlargements, bone metastases, or with malignant cells in various body fluids. If presenting complaints, physical exam and a cost-effective management strategy do not localize the primary site of cancer, the neoplasm may be called that of unknown primary. An optimum work-up would exclude a neoplasm for which curative treatment is possible or which better fits as a locally advanced site cancer rather than a metastatic one of unknown primary or a tumor amenable to hormonal palliative treatment. A simple algorithm for this work-up is described in Fig. 1.

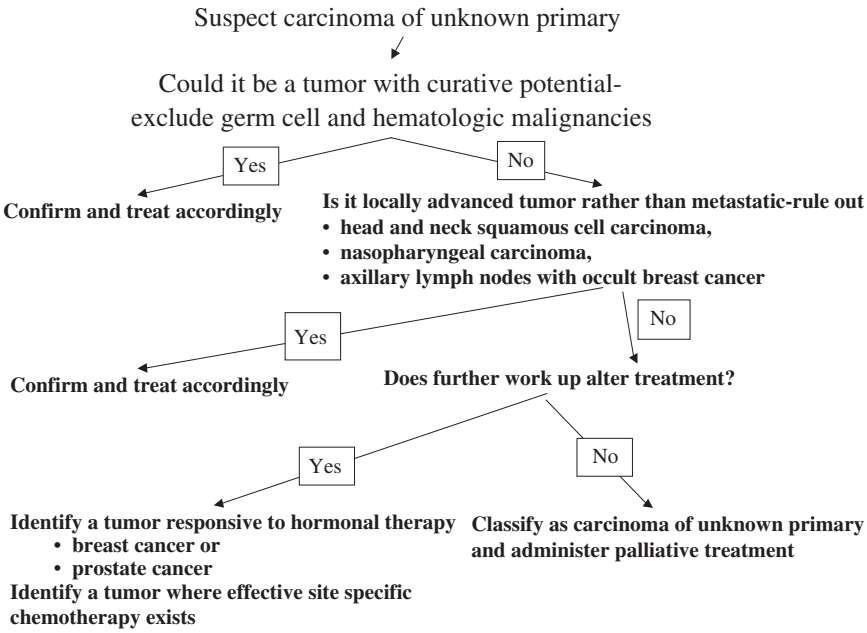


Fig. 1 Approach to carcinoma with unknown primary.

PATHOLOGIC SUBSETS OF CARCINOMA OF UNKNOWN PRIMARY

Pathological review separates various subtypes of carcinoma of unknown origin. These subtypes include adenocarcinoma, undifferentiated or poorly differentiated cancer, squamous cell carcinoma, and neuroendocrine or small cell carcinoma. The frequency of the histology is consistent from one series to another. The most common subtype being adenocarcinoma followed by undifferentiated carcinoma and squamous cell carcinoma. Direct communication with a pathologist is essential in the management of these patients. Fine needle aspiration biopsies are often inadequate for appropriate pathologic examination. Indeed, the commonest cause of reason for non-specific pathologic diagnosis is a small, inadequate biopsy. For the undifferentiated or poorly differentiated neoplasm, meticulous pathological work-up is essential to exclude better prognosis tumors like lymphoma and germ cell cancers. This may require special immunohistochemical testing as well as sometimes FISH probes. Such studies can help to distinguish the more poorly differentiated tumors.[3–4]

Usually it is important to distinguish between a poorly differentiated tumor of epithelial, hematopoietic, or neuroectodermal origin (i.e. melanoma). Key immuno-stains used for this distinction include cytokeratins (for carcinoma), leukocyte common antigen (for lymphomas), vimentin (for sarcoma) and S-100, a neuroectodermal antigen expressed in melanomas.[5]

Distinguishing the origin of an adenocarcinoma may be aided by staining for milk protein, PSA, surfactant protein, thyroglobulin. These may help direct an adenocarcinoma respectively to a specific site such as breast, prostate, lung or thyroid. Prostate specific antigen (PSA) analysis can accurately differentiate tumors of prostatic origin from other types of cancer. Immunohistochemical stains are also available for human chorionic gonadotropin (HCG) and alpha-fetoprotein (AFP) analysis. Although neither is absolutely specific for germ cell tumors (HCG and AFP) or hepatoma (AFP), they are important because germ cell tumors are effectively treated with combination chemotherapy. Undifferentiated tumors presenting as carcinoma of unknown primary can be germ cell tumor, the recognition of which would lead to a substantial clinical response or even to cure.[6]

Although currently not routinely available in Singapore, polymerase chain reaction analysis can also allow DNA amplification of Epstein–Barr virus (EBV) genomes to support the diagnosis of nasopharyngeal carcinoma in tissue provided by fine-needle aspiration biopsy. The presence of EBV in metastases from an occult primary tumor suggests the development of overt nasopharyngeal carcinoma.[7] A single study has also shown that the i(12p) marker chromosome may be used as a diagnostic tool in patients with suspected midline germ cell tumors.[8]

EVALUATING A PATIENT WITH CARCINOMA OF UNKNOWN PRIMARY

First and foremost is in obtaining histological diagnosis. Common sense tells us that probably the best way of getting a diagnosis is from the most accessible tumor; either the lymph node or the actual tumor itself. After that, one can then embark on the search for the primary from whence the tumor evolved.

Inability to identify a primary site for metastatic cancer often generates anxiety for patients and relatives. They may believe that the physician's evaluation has been substandard or that the prognosis would be improved if a primary site could be definitely established. However, we know now that that is not the case. As a result, patients are often subjected to a battery of investigations that yield distressingly little valuable information.

Routine clinical evaluation should include a complete history taking, focusing on identifying risk factors such as smoking, or a family history of cancer. A complete physical examination is mandatory, paying close attention to thyroid gland, integument, mucous membranes, breast, testicles, rectum, prostate and gynecological breast exam in females. Blood tests should include liver function tests, full blood count, urinalysis, and stool occult blood.

IMAGING STUDIES IN EVALUATION OF CARCINOMA OF UNKNOWN PRIMARY

A computed tomography of the thorax and abdomen must be done in all patients detected with a carcinoma of unknown primary. These imaging studies help rule out a primary in lungs or pancreas. When a female

patient presents with an axillary or supraclavicular lymph-node or with bone metastases, a bilateral mammogram should also be considered to exclude a breast primary. Similarly pelvic imaging with an ultrasound or CT-Scan must be considered in a patient with symptoms localizing to this area or when presentation is with malignant ascites. In addition, further imaging must become essential with emergence of abnormalities of primary work-up.

Any imaging study, however, would be futile if the overall performance status of a patient is very poor and no oncologic therapy is planned.

ENDOSCOPY IN MANAGEMENT OF CARCINOMA OF UNKNOWN PRIMARY

For a patient presenting with squamous cell or undifferentiated carcinoma in the neck region, it is imperative to perform endoscopic evaluation (often under general anesthesia) to exclude a primary in head and neck, or nasopharynx. These tumors may be curatively treated with combination of surgery and radiation therapy, radiation alone or with combined chemotherapy and radiation. Upper and lower gastrointestinal endoscopy and bronchoscopy are not generally recommended unless symptoms or signs like microcytic anemia or fecal occult blood positivity indicate a gastrointestinal neoplasm. Even when widely metastatic, it may be important to localize the primary in such cases as identification and arrest of bleeding may be mandatory for optimum management or a specific diagnosis such as colon cancer may allow a specific chemotherapy, which may be more efficacious. Routine endoscopies all cases of carcinoma of unknown primary are not indicated, except for age-appropriate screening, which may pick up a synchronous premalignancy or malignancy.

THE USE OF TUMOR MARKERS

A variety of tumor markers are available, including carcinoembryonic antigen (CEA), Ca 125, Ca 153, Ca 19.9, PSA alpha-fetoprotein, and human chorionic gonadotropin. Marker may be helpful in suggesting a primary tumor, but because of the diagnostic uncertainty posed by patients with metastatic carcinoma of unknown origin, determinants are often ordered indiscriminately. Studies have shown a significant overlap among markers. With the exception of PSA, they are seldom helpful in

Table 1 Tumor Markers in Carcinoma of Unknown Primary

Subgroup	Condition to be Excluded	Appropriate Tumor Markers
Patients with mediastinal or retroperitoneal masses	Extragonadal primary germ cell tumors	Alpha-fetoprotein, beta-hCG
Liver mass or principal liver mass with diffuse metastases	Hepatocellular carcinoma	Alpha-fetoprotein
Pancreatic mass alone or with liver metastasis	Pancreatic cancer	Ca 19-9
Men with diffuse metastatic disease to bone and/or lungs	Prostate cancer	PSA

making specific diagnosis unless patients fall into one of the favorable subgroup as shown in Table 1.

DEFINING PATIENTS WITH CARCINOMA OF UNKNOWN PRIMARY WITH BETTER PROGNOSIS

In 1994, a large series of more than 900 patients identified a possible better prognosis for patients with lymph node only involvement, and of neuroendocrine pathology.[9] Since then, researchers have been trying to define models to prognosticate patients.

The following sections discuss management of a subgroup of patients with unidentified primary tumors. These patients may have a better prognosis than other patients with metastatic disease of unknown origin, and specific therapeutic approaches may be justified.

Patients with Midline Anaplastic Carcinoma

All patients who have midline tumors, which have histology of undifferentiated cancer or anaplastic cancer, should have their alpha-fetoprotein and beta-hCG checked. Even if they are not raised, the diagnosis of

extragonadal germ cell carcinoma may be entertained. The supporting characteristics of the patient include that of young age at presentation (<50 years old), presence of a midline tumor, short duration of onset with rapid tumor growth, negative smoking history. Patients may be offered treatment with combination cisplatin containing chemotherapy regime. These patients tend to do well, and may have long-term survival. [nccn guidelines]

Patients with Neuroendocrine Tumors

A subset of patients with poorly differentiated neuroendocrine tumors has been described. Patients have rapidly growing tumors in multiple sites and involvement of mediastinal and retroperitoneal lymph nodes. They may present with diffuse hepatic or bone metastases. Results of immunocytochemical studies are positive for neuron-specific enolase. Neuroendocrine tumors have a response rate of over 50% to combination chemotherapy. Patients may therefore attain good palliation from such regimes.

Patients with Squamous Cell Carcinoma of the Cervical Lymph Nodes

Patients who present with metastatic cancer in the high cervical lymph nodes require imaging of the head and neck and chest. Even if the result of initial imaging is negative, examination using general anesthesia should include nasopharyngoscopy, laryngoscopy, bronchoscopy, and esophagoscopy should be performed. With this examination, a primary head and neck cancer is identified in 20–50% of patients.[10] Even if no primary tumor is found, blind biopsies should still be performed of the postnasal space, tonsillar fossa, and base of the tongue. Even when the primary site remains occult, oncological management strategies similar to that used for head and neck can still result in a 20–30% cure rate.[11]

Patients who present with metastatic cancer in the low cervical lymph nodes tend to have a worst prognosis as the primary site is more likely to be an occult lung or gastrointestinal tumor. These patients would require a bronchoscopy and/or a gastroscopy. If the primary remains occult, involved field radiation can be offered.

Women with Axillary Node Metastases

In this subgroup of patients, breast cancer should be suspected until proven otherwise. Following a thorough breast examination, a mammogram should be performed, and should it prove negative, an ultrasound. Even in the absence of a palpable mass in the breast or mammographic abnormality, subsequent mastectomy still reveals occult primary tumor in 65–75% of these women.[12–14]

In the absence of any primary tumor identified in the breast, axillary lymph node dissection and measurement of estrogen and progesterone receptors are performed. This not only provides information regarding the identity of the primary lesion but also allows treatment to be instituted. Patients are treated for stage II breast cancer.[15]

Women with Peritoneal Carcinomatosis

In women who have peritoneal cancer with an unidentified primary site, pathologic and diagnostic considerations are similar to those patients with ovarian cancer. Even if an oncological Whethim's surgery fails to identify a primary lesion, patients should be offered treated as primary peritoneal carcinoma or as stage III ovarian cancer.[16]

Patients with Inguinal Lymph Node Disease

Cancer found in the inguinal lymph node usually indicates a primary site in the genital or anorectal area. Women should have careful examination of the vulva, vagina and cervix; men need careful inspection of the penis. In both sexes, the anorectal area should be examined for any suspicious lesions. Patients may be treated with surgery (radical resection), chemoradiation, chemotherapy, or radiation alone. Half of the complete responders have a prolonged survival.[17]

Patients with Skeletal Metastases

Skeletal metastases, especially if osteoblastic, should raise the suspicion of metastatic prostate cancer in male. An elevated serum PSA level or a prostate biopsy specimen that stains positive for PSA by immunochemical technique confirms the diagnosis. Patients would then be treated as for metastatic prostate carcinoma. The availability of relatively non-toxic

hormonal approaches emphasize the need to evaluate this diagnostic possibility in the setting of unknown primary.[18]

Skeletal metastases in women should raise the suspicion of metastatic breast cancer. The histology specimen should be send for estrogen and progesterone receptor staining to assist in the diagnosis, and to guide in the management of such patients.

MANAGEMENT OF PATIENTS WHO DO NOT FIT INTO THE FAVORABLE GROUP

In 1995, a large study involving more than 900 patients was published. It identified diagnoses, which were associated with male sex, increased tumor load and sites of involvement, liver and adrenal metastases, and histology of adenocarcinoma, which carries a far worse prognosis.

Despite their prognosis, patients may still benefit from treatment with palliative chemotherapy. However, there are no randomized phase III studies that can prove that treatment with combination or single agent chemotherapy can result in any prolongation of survival. A truthful discussion with the patient and his or her family should be undertaken to discuss the role of any treatment with or without chemotherapy. The aim of any treatment is to maintain function, and improve quality of life.

CONCLUSION

Unfocused radiological and endoscopic evaluation of asymptomatic areas is not productive and should be avoided. Clinical and pathological evaluation should be directed at identifying the specific subgroups of patients who could potentially benefit from curative treatment.

REFERENCES

1. Leonard RJ, Nystrom JS, Diagnostic evaluation of patients with carcinoma of unknown primary tumour site, *Semin Oncol* **20**(3):244–250, 1993.

2. Chia KS, Seow A, Lee HP, Shanmugaratnam K, Cancer Incidence in Singapore 1993–1997, *Singapore Cancer Registry Report*, **No. 5**, 2000.

3. Ruddon RW, Norton SE, Use of biological markers in the diagnosis of cancers of unknown primary tumour, *Semin Oncol* **20**(3):251–260, 1993.

4. Mackay B, Ordonez NG, Pathological evaluation of neoplasms with unknown primary tumour site, *Semin Oncol* **20**(3):206–228, 1993.

5. Battifora H, Recent progress in the immunohistochemistry of solid tumours, *Semin Diagn Pathol* **1**(4):251–271, 1984.

6. Greco FA, Hainsworth JD, Cancer of unknown primary site, in: DeVita VT JR, Hellman S, Rosenberg SA, (eds.), *Cancer: Principles and Practice of Oncology*, Philadelphia, Pa, Lippincott–Raven Publishers, 5th ed., pp. 2423–2443, 1997.

7. Feinmesser R, Miyazaki I, Cheung R, Freeman JL, Noyek AM, Dosch HM, Diagnosis of nasopharyngeal carcinoma by DNA amplification of tissue obtained by fine-needle aspiration, *N Eng J Med* **326**(1):17–21, 1992.

8. Bosl GJ, Ilson DH, Rodriguez E, Motzer RJ, Reuter VE, Chaganti RS, Clinical relevance of the i(12p) marker chromosome in germ cell tumours, *J Natl Cancer Inst* **86**(5):349–355, 1994.

9. Abbruzzese JL, Abbruzzese M, Hess K, Raber M, Lenzi R, Frost P, Unknown primary carcinoma: natural history and prognostic factors in 657 consecutive patients, *J Clin Oncol* **12**(6):1272–1280, 1994.

10. Kirsten F, Chi CH, Leary JA, Ng AB, Hedley DW, Tattersall MH, Metastatic adeno or undifferentiated carcinoma from an unknown primary site: natural history and guidelines for the identification of treatable subsets, *Q J Med* **62**(238):143–161, 1987.

11. De Braud F, Al-Sarraf M, Diagnosis and management of squamous cell carcinoma of unknown primary tumour site of the head and neck, *Semin Oncol* **20**(3):273–278, 1993.

12. Baron PL, Moore MP, Kinne DW, Candela FC, Osborne MP, Petrek JA, Occult breast cancer presenting with axillary nodal metastases: updated management, *Arch Surg* **125**(2):210–214, 1990.

13. Bhatia SK, Saclarides TJ, Witt TR, Bonomi PD, Anderson KM, Economou SG, Hormone receptor studies in axillary metastases from occult breast cancers, *Cancer* **59**(6):1170–1172, 1987.

14. Ellerbroek N, Holmes F, Singletary E, Evans H, Oswald M, McNeese M, Treatment of patients with isolated axillary nodal metastases from occult primary carcinoma consistent with breast origin, *Cancer* **66**(7):1461–1467, 1990.

15. Ellerbroek N, Holmes F, Singletary E, Evans H, Oswald M, McNesse M, Treatment of patients with isolated axillary nodal metastases from an occult primary carcinoma consistent with breast origin, *Cancer* **66**(7):1461–1467, 1990.

16. Guarischi A, Keane TJ, Elhakim T, Metastatic inguinal nodes from an unknown primary neoplasm; a review of 56 cases, *Cancer* **59**(3): 572–577, 1987.

17. Yam LT, Winkler CF, Janckila AJ, Li CY, Lam KW, Prostatic cancer presenting as metastatic adenocarcinoma of undetermined origin: immunodiagnosis by prostatic acid phosphatase, *Cancer* **51**(2): 283–287, 1983.

55

Chemoprevention

Sandeep K. Rajan

Malignant tumors generally take many years to emerge, through multiple steps of carcinogenesis. This lengthy period of development provides an opportunity to interrupt a pre-cancerous condition during its march towards malignancy. Chemoprevention is defined as the use of specific natural or synthetic chemical agents to reverse, prevent or suppress the carcinogenic process to invasive cancer. It is hoped that using this strategy against pre-cancerous lesions will reduce cancer deaths; just as treatment of early cardiovascular disease process has prevented deaths from heart disease. Only few chemopreventive agents have been approved for clinical use. Many were developed for other indications and their chemoprevention property has recently been recognized. This chapter will describe briefly how chemoprevention works and discuss the tumor types where chemoprevention has firm basis supported by current literature.

HOW CHEMOPREVENTION WORKS

Carcinogenesis is a multi-step process. Initiation stage is characterized by the conversion of a normal cell into a cell altered by DNA damaging

agents. The promotion stage is marked by the transformation of an initiated cell into a population of pre-neoplastic cells, as a result of alterations in gene expression and cell proliferation. The progression stage involves the transformation of the pre-neoplastic cells to a neoplastic cell cluster with additional genetic alterations. In some tumors, an early manifestation of such neoplastic cells may be an intra-epithelial neoplasia (IEN). Ductal carcinoma *in situ*, oral pre-malignancy, cervical carcinoma *in situ* are such examples. Surveillance for these and surgical excision is one way of preventing cancer and has existed for long. However, not only does this modality cause major morbidity but also it does not treat the entire epithelial field at risk (Field cancerization). Chemoprevention is a more comprehensive approach that works at a molecular level to modify cancer risk. Such agents may bind to carcinogens, which minimizes their exposure to normal cells, hence preventing tumor initiation. Such example would be calcium, which binds bile salts that may prove to be carcinogenic for colonic epithelium. Alternatively these may act on targets that overexpress in the promotion and progression stages. Many such targets are currently poorly understood. Figure 1 briefly describes this multi-step carcinogenesis and activity of various chemopreventive compounds at various levels. These agents may eventually diminish the pre-malignant cells by either diminishing their growth, causing them to differentiate, or cause cell death (apoptosis). The multiple genetic and molecular events

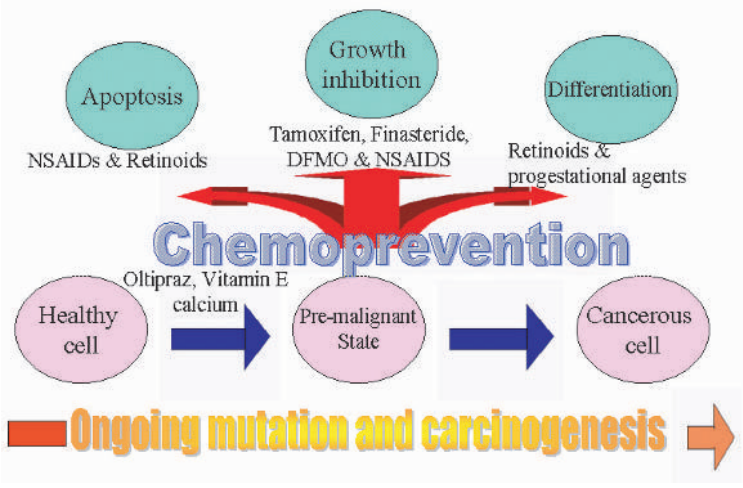


Fig. 1 Sites of action for chemoprevention strategies.

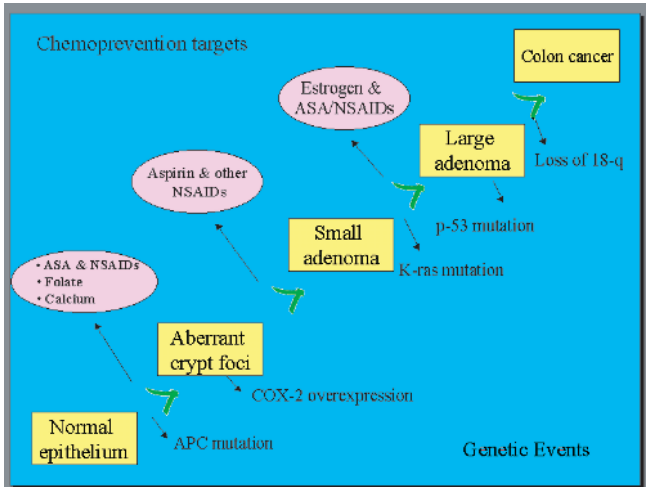


Fig. 2 Spectrum of colorectal carcinogenesis and Targets for chemoprevention.

have been best studied in colon cancer and the steps of carcinogenesis and chemoprevention are shown in Fig. 2. Hence chemoprevention agents may work at various levels: at clinical level, by reduction of cancer development, at the tissue level by reversal of pre-malignant lesions and at the cellular level by re-regulation of growth and differentiation.

WHO SHOULD RECEIVE CHEMOPREVENTION?

Chemoprevention can be offered in many settings. One is primary prevention, when no malignant or pre-malignant lesion exists. At least 3 servings of fresh fruits and vegetables, and regular exercise are some methods recommended to the population at standard risk for minimizing development of cancer. However, no chemoprevention drug has to date been proven in randomized controlled studies to prevent the risk of cancer for all standard risk groups. Evidence-based medicine for such drug interventions have mainly focused at high-risk group by way of inherited or demographic conditions; or with pre-malignant conditions (in this case, intervention would be secondary prevention); or in patients who have been diagnosed with a curable cancer, but have risk of developing a second cancer due to associated risk (second primary). In the latter it is called tertiary prevention. Some examples of secondary prevention

would be interventions in patients detected with pre-malignant lesions like dysplasia, or leukoplakia in the oro-digestive tract, colonic polyps or intra-epithelial neoplasm of breast, cervix or prostate. Other groups would be subjects with a high-risk behavior like smoking or a demographic risk like the people of Linxian province in China, who have a high incidence of esophao-gastric cancers due to paucity of essential micronutrients in the soil and deficiency of key vitamins in the diet. Early intervention with nutrients like Selenium and antioxidant vitamins in a large population intervention study in over 3000 people showed a 16% reduction in pre-cancerous lesions on endoscopy in patients with chemoprevention.[1] Most subjects receiving chemoprevention are usually healthy and without cancer. Hence the agents must have no appreciable toxicity, especially since carcinogenesis is a slow process, thus duration of use would be lengthy and often life-long.

WHAT ARE THE CHEMOPREVENTION AGENTS?

Certain infections like hepatitis B and C, and *Helicobacter pylori* are associated with certain cancers like hepatoma, gastric cancer, MALT lymphoma, etc. Though anti-viral and antibiotic therapies directed at eradicating these infections could also be considered chemopreventive for cancer, in the ensuing paragraphs, agents with a direct role will be presented. Among the agents that have been extensively studied in laboratory and clinic as chemoprevention agents are non-steroidal anti-inflammatory drugs (NSAIDs), retinoids and its analogues, and estrogen hormone antagonists like tamoxifen and raloxifen. Most elaborate studies with chemoprevention are available in efforts in preventing colon cancer, head and neck squamous cell carcinoma, and breast cancer. These will be discussed below.

CHEMOPREVENTION IN COLON CANCER

Epidemiologic, experimental (animal), and clinical investigations suggest that diets high in total fat, protein, calories, alcohol, and meat (both red and white) and low in calcium and folate are associated with an increased incidence of colorectal cancer. But cereal fibre supplementation and diets low in fat and high in fibre, fruits, and vegetables, however, did not reduce the rate of adenoma recurrence over a 3-year to 4-year period.[2] The most substantial advancements in the development of

chemopreventive regimens for colorectal cancer to date have been made using NSAIDs that are effective inhibitors of prostaglandin synthesis. Enthusiasm for this approach is based on the presence of high levels of prostaglandin E2 in colon cancers.

NSAIDS, including piroxicam, sulindac and aspirin, may prevent adenoma formation or cause adenomatous polyps to regress in the setting of familial adenomatous polyposis. Hence it can be said that "an aspirin a day keeps colon adenoma away". Most, but not all, epidemiological studies have reported a reduction in colon cancer incidence associated with the use of aspirin. In large group of over 600 000 adults enrolled in an American Cancer Society study, mortality in regular users of aspirin was about 40% lower for cancers of the colon and rectum.[3] In a report from the Health Professionals Follow-up Study of 47 000 males, regular use of aspirin (at least 2 times per week) was associated with a 30% reduction in colorectal cancer.[4] A population-based retrospective cohort study of non-aspirin NSAID use among individuals aged 65 and older was also associated with lower risk, particularly with increasing durations of use.[5] However in another study, 22 000 men aged 40 to 84 were randomized to placebo or aspirin (325 mg every other day) for 5 years. There was no reduction in invasive cancers or adenomas at a median follow-up of 4.5 years.[6] In a subsequent analysis over a 12-year period, both randomized and observational analyses indicated that there was no association between the use of aspirin and the incidence of colorectal cancer. The low dose of aspirin and the short treatment period may account for the negative findings.[7] Many studies have demonstrated the efficacy of sulindac in reducing the size and number of adenomas in familial polyposis as well.[8,9]

Another NSAID piroxicam, at a dose of 20 mg/day, reduced mean rectal prostaglandin concentration by 50% in individuals with a history of adenomas.[10] The cyclooxygenase (COX) enzymes are either continuously expressed (COX-1) or induced by inflammatory processes (COX-2). They have been associated with cancer at various sites, including the colon, gastrointestinal tract, lung, and skin.[11] Increased prostaglandin and thromboxane production in tumor cells has been linked to increased angiogenesis and proliferation, and decreased differentiation and apoptosis.[12,13] Epidemiologic, experimental, and intervention research on inhibition of COX-1 and COX-2 by NSAIDs indicate that NSAIDs may prevent tumor growth and increase differentiation and apoptosis through this

mechanism.[14,15] For example, in patients with familial adenomatous polyposis, celecoxib (a COX-2 inhibitor) significantly reduced the number of colorectal polyps by 28% and the polyp burden (sum of polyp diameters) by almost 31% after 6 months of treatment, compared with patients administered placebos.[16] Though NSAIDs are proving to be effective in reducing adenomatous polyps, there are several unresolved issues before making general recommendations for their use. These include uncertainty about the proper dose and duration for these agents, and concern whether the potential preventive benefits would balance against long-term risks such as gastrointestinal ulceration and hemorrhage.[17]

Studies in Europe and the United States are under way to assess the efficacy of aspirin, rofecoxib, and celecoxib in preventing the recurrence of sporadic adenomatous polyps after polypectomy, with a planned recruitment of more than 1000 subjects. A randomized French study of two doses of aspirin (160 mg and 300 mg) in a similar setting has finished recruiting subjects. Combination therapy with NSAIDs and folate, eflornithine, angiotensin-converting enzyme inhibitors, statins or calcium, is also under investigation. Cyclooxygenase-2 inhibitors may become the preferred chemopreventive agent owing to their favorable gastrointestinal safety profile.

A randomized placebo-controlled trial tested the effect of calcium supplementation [3 g calcium carbonate daily (1200 mg elemental calcium)] on the risk of recurrent adenoma.[18] In this study with majority male patients (72%), risk of detecting recurrent adenoma on follow up for 2 endoscopies had an adjusted risk ratio of 0.81 (a 19% risk reduction). The investigators found the effect of calcium was similar across age, sex, and baseline dietary intake categories of calcium, fat, or fibre. The study was limited to individuals with a recent history of colorectal adenomas and so could not determine the effect of calcium on risk of first adenoma, nor was it large enough or of sufficient duration to examine risk of invasive colorectal cancer.

Epidemiologic studies have found a lower incidence of colorectal cancer among those with the highest dietary folate intake,[19] whereas those with diets low in folate (and often with high alcohol intake) appear to have an increased risk of colorectal adenomas and carcinomas.[19,20] Although large amounts of folate in the diet appear to be protective against the development of colorectal adenomas (relative risk, 0.91 in women and 0.78 in men), the degree of benefit is greater among those

who take folate supplements (relative risk, 0.66 for women and 0.63 for men).[20,21] In the Nurses' Health Study, supplementation with folate (usually as part of multivitamin supplementation) was protective against colorectal cancer, with the greatest risk reduction among women taking high daily doses of folate (more than 400 μg); this reduction (relative risk, 0.25) became statistically significant only after 15 years of use.[21] The long time needed for a clinical benefit to become evident suggests that folate acts early in colon carcinogenesis.

These recent observations suggest that aspirin and other NSAIDs, supplemental folate and calcium, have a chemopreventive benefit. Since the value of such prophylactic strategies has not yet been confirmed in double-blind, placebo-controlled, randomized studies in the general population, chemoprevention cannot yet be accepted as standard medical practice. Chemoprevention should not replace periodic fecal occult-blood tests and endoscopic screening, as well as modification in known risk factors for colorectal cancer, such as reduction in the intake of red meat, appropriate exercise, smoking cessation, and weight control.

CHEMOPREVENTION IN BREAST CANCER

The risk of breast cancer is related to levels of endogenous and exogenous estrogenic hormones.[22] Tamoxifen and raloxifene are selective estrogen-receptor modulators, hence as estrogen antagonists, they can reduce the risk of breast cancer. Tamoxifen is active against advanced breast cancer. In the analysis of trials of adjuvant therapy conducted by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG), treatment with tamoxifen for five years reduced the annual odds of contralateral breast cancer by 47%, regardless of the receptor status of the initial tumor.[23]

This led to National Surgical Adjuvant Breast and Bowel Project (NSABP) Tamoxifen Prevention Trial, a randomized, placebo-controlled study, evaluating if 20 mg of tamoxifen daily for five years could reduce the incidence of breast cancer in women at increased risk.[24] Subjects with a predicted five-year risk of breast cancer that was equivalent to that of a 60-year-old woman (greater/equal 1.66%) were enrolled. Among 13 388 women followed for about four years, tamoxifen reduced the overall odds of invasive and non-invasive breast cancer by nearly 50% ($P < 0.001$). The effect of tamoxifen was exerted exclusively against receptor-positive tumors. The reduction occurred among women in all age groups, those

with a history of lobular carcinoma *in situ* (56% reduction), and those with atypical hyperplasia (86% reduction). Two smaller European trials failed to show this beneficial effect of tamoxifen, probably due to their smaller sample size and inclusion of lower risk patients or those with very high genetic risk that is not modified by tamoxifen. This reduction in invasive breast cancer does come at a higher incidence of uterine cancer, thromboembolic phenomenon and other side effects of tamoxifen like climacteric and depression. Hence raloxifene, which has no incremental risk of uterine carcinoma, was evaluated. The risk of breast cancer, have been monitored in several ongoing placebo-controlled trials of raloxifene directed at osteoporosis and other endpoints. The Multiple Outcomes of Raloxifene Evaluation (MORE) trial randomly assigned 7705 postmenopausal women with existing osteoporosis to receive 60 or 120 mg of raloxifene per day, or a placebo. After 40 months of follow-up, raloxifene was found to have reduced the annual odds of breast cancer by 65% (on the basis of 58 events) and reduced the risk of invasive breast cancer by 76% ($P < 0.001$).[25] A partially overlapping meta-analysis of nine randomized trials of raloxifene (including the MORE trial), which involved 10 575 patients, found a somewhat smaller (54%) reduction in the risk of invasive and non-invasive breast cancer. Like tamoxifen, raloxifene influences only receptor-positive cancers and there is less evidence supporting the use of raloxifene to reduce the risk of breast cancer than there is supporting the use of tamoxifen. Hence use of raloxifene for this indication remains investigational. Furthermore none of these trials have been directed to show any improvement in survival with this risk reduction of invasive breast cancer.

Nevertheless in a recent report by the American Society of Clinical Oncology the use of tamoxifen and raloxifene to reduce the risk of breast cancer was reviewed.[26] It was summarized that women at increased risk for breast cancer (defined as a risk of at least 1.7% over five years) may be offered tamoxifen (20 mg/day) to reduce their risk after an informed decision-making process, with careful consideration of risks and benefits. Since the overall benefits to health and survival have not been established, the decision to use tamoxifen for risk reduction depends on an individual woman's perception of her breast-cancer risk and her reaction to this risk. On the basis of current information, the routine use of raloxifene should be reserved for treatment of bone loss in postmenopausal women.

CHEMOPREVENTION IN HEAD & NECK SQUAMOUS CELL CARCINOMA (HNSCC)

Pre-malignant lesions like leukoplakia, erythroplakia, and dysplasia often precede head and neck squamous cell carcinoma. These provide an ideal target for testing of chemoprevention principles. Retinoids have established efficacy in reversal of early pre-malignant lesions[27,28,30] but is associated with considerable mucocutaneous toxicity. Initial reports in 1986 by Hong *et al*.[27] using high-dose 13-cis retinoic acid for 3 months versus a placebo had resolution of lesions in 2/3 of treated patients versus 10% of placebo arm ($P = 0.002$). Several other trials confirmed retinoid activity in oral pre-malignant lesions. However relapses after treatment-cessation were frequent, which indicated the need for maintenance. Also, even though the high dose was tolerated for short periods, it was unlikely that it could be tolerated for long periods of maintenance. To address these questions a follow-up trial was done to evaluate long-term maintenance of retinoid response in oral pre-malignancy. Maintenance consisted of 9 months of either low-dose 13cRA (0.5 mg/kg/day) or beta-carotene following induction with 3 months of high-dose 13cRA (1.5 mg/kg/day). Low-dose retinoid was more effective in maintaining the initial response to induction 13cRA than was beta-carotene (92% versus 45%; $P = 0.001$).[35] Another retinoid, retinamide (40 mg/day orally and 40 mg/day topically) with major response in lesions occurring in 27 of the 31 patients on treatment (87%), but only 5 of the 30 patients on placebo (16.7%) had major responses, and none were complete.[30] Toxicity was minimal, consisting of minor elevations of serum transaminase in two patients. No data on skin toxicity was reported. Stich *et al*.[28] evaluated the effects of vitamin A in 65 patients (tobacco users or betel nut chewers) with oral leukoplakia; 30 subjects received 100 000 IU of vitamin A twice weekly, and 34 were assigned to placebos. Among the 21 evaluable vitamin A patients, 12 (57.1%) had complete remissions and no patient progressed during treatment. Among the 33 subjects on placebos, only 1 (3.0%) had a complete remission, and 7 (21.2%) progressed.

Another randomized maintenance trial, conducted by Chiesa *et al*.,[30] was reported with 170 evaluable patients randomized to receive either fenretinide (200 mg/day) or no intervention (control) for 1 year following laser resection of oral leukoplakia. The failure rate in terms of local relapses or new lesions was 29% in the placebo-control arm, versus an

18% failure rate in the fenretinide group ($P = 0.01$). Other agents have also been tested in this setting, most notably beta-carotene and vitamin E.

Advanced pre-malignant lesions of the upper aero digestive tract like moderate to severe dysplasia are associated with a 36–50% risk of progression into invasive cancer. Such lesions are resistant to single agent retinoid chemoprevention, and, to provide a chemoprevention strategy for this high-risk group, a combination of α-interferon, α-tocopherol, and 13-cis-retinoic acid were evaluated.[31] The study showed that the treatment combination was active in preventing progression of laryngeal lesions but not oral lesions. After 12 months of treatment, laryngeal sites showed no complete responses. Based on these findings, a new study addressing exclusively laryngeal dysplasia incorporating a year of induction with the same regimen and 2 years of maintenance with fenretinide or placebo, has been designed and the results are eagerly awaited. The same combination for 1 year has been highly effective as a bioadjuvant approach in patients previously treated for a locally advanced head and neck cancer. The Phase II trial by Shin *et al*.[29] showed that 84% of patients were disease-free and alive at 2 years after the completion of definitive treatment for their primary tumor.

In head and neck squamous cell carcinoma (HNSCC), the overall annual second primary tumor (SPT) rates range from 1.2% to 4.7% in retrospective studies. In prospective studies, however, these rates have been reported as ranging from 4% to 7%. Based on positive retinoid data in oral leukoplakia and the lack of effective adjuvant therapy in preventing primary recurrence or SPTs, Hong *et al*.[32] conducted an adjuvant randomized, double-blinded, placebo-controlled, chemoprevention trial of high-dose 13-cRA for 1 year in 103 patients curatively treated for cancer of the head and neck. The major trial endpoints were recurrence of the primary disease, development of an SPT (different histological type or at a site more than 2 cm from the previous disease or occurring more than 3 years after initial diagnosis), and survival. With a median follow-up of 32 months, SPTs developed in significantly fewer 13-cRA treated patients (4%) than in patients receiving placebo (24%) ($P = 0.005$). As with retinoids in other preclinical and clinical carcinogenesis systems, the retinoid's impact on annual overall SPT incidence has decreased over time since completing the 12 months' intervention. A subset analysis of SPT only within the high-risk tobacco-exposed field of the head and neck,

lung, and esophagus showed chemopreventive activity persisting at the same level of significance as earlier overall results. This long-lasting retinoid activity is unprecedented in previously reported clinical or preclinical retinoid carcinogenesis studies. The synthetic retinoid etretinate was tested in 1994 in a French trial[33] designed to prevent SPT following definitive treatment of squamous cell carcinoma of the oral cavity or oropharynx. Patients randomly received either a placebo or etretinate (50 mg/day for 1 month followed by 25 mg/day for 24 months). SPT and primary recurrence rates were equivalent in both study arms. The French report provided few details regarding tobacco and alcohol usage, study adherence, or toxicity, making interpretation difficult.

Based on the positive findings of the trial by Hong *et al.*,[32] the National Cancer Institute sponsored the largest chemoprevention trial in head and neck cancer, NCI C91–002, in which patients were prospectively randomized to low-dose 13-cRA for 3 years versus placebo. This trial was closed to accrual in June 1999 with more than 1200 participants. An interim analysis shows that the annual SPT rate in active, former, and never smokers was 5.1%, 4.1%, and 3%, respectively ($P = 0.06$ for active smokers vs. nonsmokers).[34] More mature treatment results are eagerly awaited. Future agents for use in preventing cancer of the head and neck most likely will be selected from new-agent studies in the oral pre-malignancy system. Current study in these two areas involves intervention targeted toward p53 abnormalities, selective cyclooxygenase-2 inhibitors, epidermal growth factor receptor kinase inhibitors, and farnesyl transferase inhibitors.

CHEMOPREVENTION IN LUNG CANCER

All over the world, lung cancer is soon becoming the leading cause of cancer and mortality from it. Staying away from smoking is the best strategy for its prevention. Individuals who are at risk for lung cancer and were treated with beta-carotene, retinol, isotretinoin, or N-acetyl-cysteine for lung cancer prevention did not experience clinical benefits. There is also evidence that the use of beta-carotene and isotretinoin for lung cancer chemoprevention in high-risk individuals may increase the risk for lung cancer, especially in individuals who continue to smoke.[36,37] These agents should not be used outside of a clinical trial for primary, secondary, or tertiary lung cancer prevention.

SUMMARY

Chemoprevention is a useful strategy to reduce cancer incidence. Prospective studies have identified success of chemoprevention strategy in limited clinical scenario, like ladies with high risk of breast cancer (with tamoxifen), patients with familial polyposis (with COX-2 inhibitors and other NSAIDs), those with pre-malignant lesions of head and neck area and in curatively treated patients with HNSCC, for prevention of second primary tumor (retinoids). Though a reduction cancer risk is noted, it comes with definite side effects, and long-term mortality benefit and adequate length of intervention are controversial. Epidemiological studies and large observational studies have suggested benefit of some such agents in lower risk general population as well and prospective randomized studies to this effect are ongoing. Also, the same agents used for preventing some other cancer type may also paradoxically raise the risk of cancer. Hence a physician experienced with their use should recommend chemoprevention after careful discussion of pros and cons with the patient, and wherever possible eligible subjects should participate in a clinical trial when available.

REFERENCES

1. Dawsey SM, Wang GQ, Taylor PR, Li JY, Blot WJ, Li B, Lewin KJ, Liu FS, Weinstein WM, Wiggett S, *et al.*, Effects of vitamin/mineral supplementation on the prevalence of histological dysplasia and early cancer of the esophagus and stomach: results from the Dysplasia Trial in Linxian, China, *Cancer Epidemiology, Biomarkers & Prevention* **3**(2):167–172, 1994.

2. Reddy B, Engle A, Katsifis S, *et al.*, Biochemical epidemiology of colon cancer: effect of types of dietary fiber on fecal mutagens, acid, and neutral sterols in healthy subjects, *Cancer Research* **49**(16): 4629–4635, 1989.

3. Thun MJ, Namboodiri MM, Health CW, Aspirin use and reduced risk of fatal colon cancer, *N Engl J Med* **325**(23):1593–1596, 1991.

4. Giovannucci E, Rimm EB, Stampfer MJ, *et al.*, Aspirin use and the risk for colorectal cancer and adenoma in male health professionals, *Ann Internal Med* **121**(4):241–246, 1994.

5. Smalley W, Ray WA, Daugherty J, *et al.*, Use of nonsteroidal anti-inflammatory drugs and incidence of colorectal cancer: a population-based study, *Arch Internal Med* **159**(2):161–166, 1999.

6. Gann PH, Manson JE, Glynn RJ, *et al*., Low-dose aspirin and incidence of colorectal tumors in a randomized trial, *J Natl Cancer Inst* **85**(15): 1220–1224, 1993.

7. Sturmer T, Glynn RJ, Lee IM, *et al*., Aspirin use and colorectal cancer: post-trial follow-up data from the Physicians' Health Study, *Ann Internal Med* **128**(9):713–720, 1998.

8. Labayle D, Fischer D, Vielh P, *et al*., Sulindac causes regression of rectal polyps in familial adenomatous polyposis, *Gastroenterology* **101**(3): 635–639, 1991.

9. Giardiello FM, Hamilton SR, Krush AJ, *et al*., Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis, *N Engl J Med* **328**(18):1313–1316, 1993.

10. Earnest DL, Hixson LJ, Fennerty MB, *et al*., Inhibition of prostaglandin synthesis: potential for chemoprevention of human colon cancer, *Cancer Bull* **43**(6):561–568, 1991.

11. Fosslien E, Molecular pathology of cyclooxygenase-2 in neoplasia, *Ann Clin Lab Sci* **30**:3–21, 2000.

12. Wolf LA, Laster SM, Characterization of arachidonic acid-induced apoptosis, *Cell Biochem Biophys* **30**:353–368, 1999.

13. Attiga FA, Fernandez PM, Weeraratna AT, *et al*., Inhibitors of prostaglandin synthesis inhibit human prostate tumor cell invasiveness and reduce the release of matrix metalloproteinases, *Cancer Res* **60**:4629–4637, 2000.

14. Kelloff GJ, Crowell JA, Steele VE, *et al*., Progress in cancer chemoprevention, *Ann N Y Acad Sci* **889**:1–13, 1999.

15. Giovannucci E, The prevention of colorectal cancer by aspirin use, *Biomed Pharmacother* **53**:303–308, 1999.

16. Steinbach G, Lynch PM, Phillips RKS, *et al*., The effect of celecoxib, a cyclooxygenase-2 inhibitor, in familial adenomatous polyposis, *N Engl J Med* **342**:1946–1952, 2000.

17. Sandler RS, Aspirin and other nonsteroidal anti-inflammatory agents in the prevention of colorectal cancer, *Cancer: Principles and Practice of Oncol Updates* **11**(6):1–14, 1997.

18. Baron JA, Beach M, *et al*., for the Calcium Polyp Prevention Study Group: Calcium supplements for the prevention of colorectal adenomas, *N Engl J Med* **340**(2):101–107, 1999.

19. Ferraroni M, La Vecchia C, D'Avanzo B, Negri E, Franceschi S, Decarli A, Selected micronutrient intake and the risk of colorectal cancer, *Br J Cancer* **70**:1150–1155, 1994.

20. Baron JA, Sandler RS, Haile RW, Mandel JS, Mott LA, Greenberg ER, Folate intake, alcohol consumption, cigarette smoking, and risk of colorectal adenomas, *J Natl Cancer Inst* **90**:57–62, 1998.

21. Giovannucci E, Stampfer MJ, Colditz GA, *et al*., Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study, *Ann Intern Med* **129**:517–524, 1998.

22. Lacassagne A, Hormonal pathogenesis of adenocarcinoma of the breast, *Am J Cancer* **27**:217–228, 1936.

23. Early Breast Cancer Trialists' Collaborative Group, Tamoxifen for early breast cancer: an overview of the randomised trials, *Lancet* **351**:1451–1467, 1998.

24. Fisher B, Costantino JP, Wickerham DL, *et al.*, Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study, *J Natl Cancer Inst* **90**:1371–1388, 1998.

25. Cummings SR, Eckert S, Krueger KA, *et al.*, The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial, *JAMA* **281**:2189–2197, 1999.

26. Chlebowski RT, Collyar DE, Somerfield MR, Pfister DG, American Society of Clinical Oncology technology assessment of breast cancer risk reduction strategies: tamoxifen and raloxifene, *J Clin Oncol* **17**:1939–1955, 1999.

27. Hong WK, Endicott J, Itri LM, *et al.*, 13-cis-retinoic acid in the treatment of oral leukoplakia, *N Engl J Med* **315**:1501–1505, 1986.

28. Stich HF, Hornby AP, Mathew B, *et al.*, Response of oral leukoplakias to the administration of vitamin A, *Cancer Lett* **40**:93–101, 1988.

29. Shin DM, Khuri FR, Murphy B, *et al.*, Combined interferon-alpha, 13-cis-retinoic acid, and alpha-tocopheral in locally advanced head and neck squamous cell carcinoma: novel bioadjuvant phase II trial, *J Clin Oncol* **19**:3010–3017, 2001.

30. Chiesa F, Tradati N, Marazza M, *et al.*, Fenretinide (4-HPR) in chemoprevention of oral leukoplakia, *J Cell Biochem Suppl* **17F**:255–261, 1993.

31. Papadimitrakopoulou VA, Clayman GL, Shin DM, *et al.*, Biochemoprevention for dysplastic lesions of the upper aerodigestive tract, *Arch Otolaryngol Head Neck Surg* **125**:1083–1089, 1999.

32. Hong WK, Lippman SM, Itri LM, *et al.*, Prevention of second primary tumors with isotretinoin in squamous-cell carcinoma of the head and neck, *N Engl J Med* **323**:795–801, 1990.

33. Bolla M, Lefur R, Ton Van J, *et al.*, Prevention of second primary tumours with etretinate in squamous cell carcinoma of the oral cavity and oropharynx. Results of a multicentric double-blind randomised study, *Eur J Cancer* **30A**:767–772, 1994.

34. Khuri FR, Kim ES, Lee JJ, *et al.*, The impact of smoking status, disease stage, and index tumor site on second primary tumor incidence and tumor recurrence in the head and neck retinoid chemoprevention trial, *Cancer Epidemiol Biomark Prev* **10**:823–829, 2001.

35. Lippman SM, Batsakis JG, Toth BB, *et al.*, Comparision of low dose isotretinoin with beta-carotene to prevent oral carcinogenesis, *N Engl J Med* **328**:795, 1993.

36. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group, The effect of vitamin E and beta-carotene on the incidence of lung cancer and other cancers in male smokers, *N Engl J Med* **330**:1029–1035, 1994.

37. Omenn GS, Goodman GE, Thornquist MD, *et al.*, Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease, *N Engl J Med* **334**:1150–1155, 1996.

Neurology

This page intentionally left blank

56

Approach to the Patient with Neurological Disease

K. Puvanendran and Udaya Seneviratne

In no field of medicine is a good basic knowledge of anatomy and physiology along with clinical examination more necessary than in the diagnosis of neurological diseases because the nervous system is the most logically predictable. There are two essential questions when approaching a neurological diagnosis.

1) What is the site of lesion?
   This is usually determined by the physical signs.
2) What is the likely pathology?
   Usually the patient's history answers this. Many consultations are also to prove the absence of pathology in the brain for reassurance that the patient's headache is not caused by a brain tumor or internal bleeding. If time is limited, it is prudent to be selective with the examination.

HISTORY

Details of history taking are not in the scope of this chapter but it is emphasized that the history should be a chronology of events and should

tell the story of the disease. It should tell the onset, evolution to and period of maximum deficit, continuing progression or recovery to the present state. If the condition came on acutely with dramatic suddenness, reaching the maximum in a few hours, and if it improved subsequently, though not completely, the likely pathology is vascular. Disease that comes on gradually and worsens is a neoplasm or degeneration. Spontaneous remission occurs in multiple sclerosis. Periodic weakness of limbs with recovery can occur in periodic paralysis or a vascular malformation of the cord. In transient ischemic attacks, the symptom improve within 24 hours. Epilepsy, migraine, trigeminal neuralgia all cause a series of discrete events but the patient might only be interested in describing the most recent event, which might be the worst and the reason for consultation.

Often one recognizes a characteristic clustering of symptoms and signs, constituting a syndrome. This helps to localize and gives the nature of the disease and progress. The clinical picture of Parkinson's disease for example is so characteristic that the diagnosis becomes apparent at once.

At the end of the history taking the neurologist would have conceptualized the anatomic localization of the lesion and the mechanism of the disease. His disease hypothesis could be confirmed or revised after the neurological examination, and laboratory investigation.

On many occasions history can make the diagnosis. Take for example the history of a young woman who had been getting throbbing hemicranial headaches for 5 years, usually associated with her menses. The facts that her mother too had similar headaches and that the patient experiences flashes of light and vomiting during the attack confirm the diagnosis of migraine.

A previous history of mitral valve disease in a young stroke leads us to an embolic etiology for the stroke. Correlation with other symptoms of the nervous system and with other organs is often useful. For example, headache in a case of paraplegia leads to a diagnosis of a parasaggital meningioma rather than a lesion in the spinal cord. Multiorgan affection leads us to suspect a storage disease, mitochondrial disorder, or connective tissue diseases.

Information from an eyewitness rather than an EEG is more useful in diagnosing a seizure. A patient who complains of every symptom probably has a somatiform disorder.

The examiner should, at the onset, assess the reliability of the history from the patient's intelligence and whether there could be any financial

gain from the symptoms. This occurs often in pain symptoms after an accident, for example, a whiplash injury in a motor car accident. Exposure to toxins and work environment is relevant in neurological diseases. The many cases of glue sniffing presenting like the Guillain Barre syndrome in Singapore in the previous decade is a good example to note. Sexual history should be inquired discretely now, because sexually transmitted disease is reaching greater proportions.

Family history of a similar disease or consanguinity is very relevant. In dystrophia myotonica the first generation member of the family tree can be affected with a mere cataract. The subsequent generation could have profound muscle weakness (because of DNA amplification, gene expression may become more severe in each successive generation).

Ethnicity is rarely important in diagnosis, though we know that some neurological disorders are common in some races. Thyrotoxic periodic paralysis is a consideration for a young Chinese male in our population.

NEUROLOGICAL EXAMINATION

A complete neurological examination checking every muscle, every modality of sensation as taught in the clinical examination books is hardly performed in practice. However, the resident should know how to go into details as appropriate. The purpose of the examination is to demonstrate functional aberrations that can localize in a part of the nervous system. The presence of some signs or absence of other signs can be as diagnostic. We learned to perform a *focused* examination to confirm the initial hypothesis of disease that we guessed from the history. When a right-handed person presents with a right-sided hemiparesis, and if a left cerebral lesion is suspected, we should look for dysphasia, right hemianopsia, and right-sided pyramidal signs. A patient presenting acutely with a unilatered foot drop, amongst other diagnosis, could have either a common peroneal nerve palsy (trauma over neck of fibula) or a L5 radiculopathy (lumbar disc protrusion). To differentiate, check another muscle (far apart) supplied by L5 root but through a different nerve. Turn the patient on his face and check the gluteus maximus muscle (supplied by the inferior gluteal nerve) by asking him to lift the knee off the bed. Another cause of an acute unilateral foot drop is motor neurone disease.

It is sometimes difficult to decide if a physical sign is abnormal. An examiner will interpret a tuning fork vibration sensation on a patient as

abnormal by comparing to himself. Loss of ankle jerks or vibration sensation in the toe could be an expected finding for a patient aged over 60 years. Soft signs like an asymmetric face or an inverted supinator jerk may be sufficient for an expert to place a diagnosis. It was on a lock of white hair that a diagnosis of pernicious anemia as a cause of paraparesis was made, in spite of it being a rarity in Singapore. This is a rare presentation of an uncommon disease where clinical acumen has to be sharp.

When the patients' complaint is headache, a focused examination is not possible. A screening examination testing the function of limbs, gait, cranial nerves, mental status and reflexes would be helpful. If the patient can walk tandem, eyes closed, one can safely assume there is no affection of the peripheral nerves, spinal cord or cerebellum. If the patient can do a push up, arise from the floor without hand support, walk on heels and toes — one can assume normal muscle strength.

A general physical examination of other systems is always necessary. Storage disorders, mitochondrial disorders and connective tissue disease affect many systems including the nervous system. The lesion in the brain could be a secondary deposit if the patient has, for example, carcinoma of the lung. Cerebellar signs in this patient with carcinoma of the lung would indicate a paraneoplastic syndrome. A heart murmur (for example, mitral valve disease) with a cerebral lesion points to an embolic pathology.

CONCEPTUALIZING THE DIAGNOSIS

The findings from the history and clinical examination should be dovetailed to come out with a hypothesis to explain the disease. The other diseases that are to be considered, even if they are unlikely, form the basis of the differential diagnosis. Anatomical localization comes before finding the cause. One should first localize the disease to the muscle, peripheral nerve, spinal cord or brain. Next, the physician has to decide if the disease occurs in one area or more than one area of the nervous system. *The diagnosis will be more likely if all the findings can be explained by one lesion*. For example, foot drop and gluteus maximus weakness on the same side (though the muscles are far apart in location) is explained on the basis of an L5 root innervation (common to both), rather than affection of 2 separate nerves, namely the common peroneal nerve and inferior gluteal nerves. Lesions that cannot be explained by one anatomical site or blood

supply could be from multiple metastasis, embolic infarct or multiple sclerosis. A central scotoma (from optic neuritis) associated with paraparesis must have two lesions — optic nerve and spinal cord probably from multiple sclerosis (Devics disease). This is the grand over riding law of parsimony of nature, which says that entities must not be multiplied beyond what is necessary. This is done by shaving an argument to its simplest form as in Occam's Razor (after the English philosopher William of Occam who devised this principle).

When a patient presents with acute vertigo, a peripheral vestibular end organ cause is usually suspected. This cause is excluded if there are brainstem symptoms and signs. These signs are dysphagia, dysphonia, (9^{th} and 10^{th} nucleus), cerebellar ataxia (inferior cerebellar peduncle) and Horners syndrome. Dissociated sensory loss where the laterally placed spinothalamic tracts (pain and temperature) are involved but the medial lemiscus is spared, places the lesion in the lateral medulla where the vestibular nucleus is also situated. This cluster of symptoms jigsaws into the lateral medullary syndrome of Wallenberg with a reputation of a benign outcome.

Patients who present with thalamic pain syndrome, rubral tremor, ataxia of limbs and homonymous hemianopsia, all on one side, have only one thing in common. The thalamus, red nucleus, cerebellar connections and the occipital cortex are all supplied by branches of the posterior cerebral artery (thalamogeniculate, interpeduncular and cortical), thus one lesion explains all of the symptoms. A good knowledge of neuro-anatomy comes in handy to tie up the signs.

After localizing the lesion anatomically we think of the likely pathology. A hemiparesis coming on suddenly is probably a stroke. If it came on slowly over weeks or months, it could be a tumour. Multifocal affection can be progressive as in metastases or can be relapsing and remitting as in multiple sclerosis. Degenerative disease and disease due to vitamin deficiency can be progressive. Febrile history with multiple cranial nerve palsy suggests meningitis. A patient who presents initially with wasting of small muscle of the hands and develops a bulbar palsy a few months later most probably has amyotrophic lateral sclerosis (ALS). Again the principle of parsimony makes a diagnosis of bilateral ulnar nerve palsy and stroke an unlikely possibility.

While considering the diagnostic possibilities, it would do well to remember that an unusual presentation of a common disease is more

common than a common presentation of a rare disease. Syphilis though less frequently seen is a great mimic of diseases in the nervous system. Tuberculosis and HIV infection presenting in an unusual fashion should always be thought of now in the differential diagnosis.

Treatable conditions, though unlikely, should be considered and even a trial of treatment tried, even when they mimic untreatable conditions like Alzheimer's disease. Memory loss, for example, is the basis for diagnosing dementia like Alzheimer's disease. If one goes that extra effort of getting a history of snoring, obstructive sleep apnea could be considered and a trial of CPAP (continuous positive airway pressure) can give occasional success.

LABORATORY INVESTIGATION

These should be ordered with a clear objective. Cost and risk to the patient should be kept in mind. A routine investigation is not justified. One should begin with hematological and biochemical bloods tests, and then proceed with neuroimaging and neurophysiological testing before considering biopsy. A blood count, urea electrolytes is appropriate for a patient admitted for stroke. This may not be justified in a patient with migraine. A macrocytosis and low B_{12} level in blood could avoid expensive imaging in subacute combined degeneration. The skilled diagnostician arrives at the final diagnosis by artful analysis of clinical data aided by the least number of essential laboratory tests. Strategy of laboratory study should be based on therapeutic and prognostic consideration and not on the physicians curiosity or presumed medico legal exigencies. It should be remembered that diagnostic tests comprise a large part of health care expense.

LUMBAR PUNCTURE

CSF examination by lumbar puncture (LP) is necessary for diagnosing meningitis and starting the appropriate antibiotics without delay. A delay clearly worsens the outcome. Other indication for ordering an LP is meningeal malignancies, CNS vasculitis, sarcoidosis, multiple sclerosis (MS) and pseudotumor cerebri. In subarachnoid hemorrhage however LP could establish the diagnosis when a CT Scan brain is falsely negative.

When raised intracranial pressure is suspected (evidenced by headache and papilloedema) a CT Scan brain should be done before the LP in order to exclude a mass lesion or edema.

IMAGING

CT Scan and magnetic resonance imaging are the core imaging method used.

For reasons of cost, speed and availability, CT Scan brain is still widely used for screening in the acute evaluation of stroke, head injury or acute infection. CT is useful in showing the normal outline of the brain, ventricles and CSF spaces. In daily clinical practice CT remains the choice (compared to MRI) for detecting acute hemorrhage. Some early cerebral infarcts (in the first 24 hours) and subdural hematomas will be missed on CT. CT is less useful for visualization of brainstem, cerebellum and spinal cord.

MRI is the gold standard and is the investigation of choice for demonstrating lesions in multiple sclerosis (in brain and spinal cord), brain stem pathology (tumors and infarct), and intrinsic lesions in spinal cord specially syringomyelia. Gadolinium enhanced MRI (Gd-MRI) is most useful in increasing sensitivity to neoplastic and inflammatory lesions of the brain and spinal cord. Gd-MRI is very useful in detecting meningitis, encephalitis (Herpes Simplex) and myelitis. In AIDS, Gd-MRI is useful to differentiate between cerebral lymphomas, toxoplasmosis, or fungal lesions. In temporal lobe epilepsy, high-resolution imaging can show hippocampal atrophy or sclerosis. Diffusion weighted imaging (DWI) is a useful part of acute stroke imaging which can identify cerebral infarction within minutes of onset. DWI can be combined with perfusion imaging to identify the penumbra zone of potentially salvageable tissue within the area of reduced perfusion but outside the unrecoverable infarct represented by DWI high signal. Being non-invasive, MR angiography (MRA) is often used to study intracerebral arteries in strokes. Conventional angiography is still the most sensitive examination for intracranial aneurysms and AVM, though MRA can be useful.

Duplex ultrasonography of carotid arteries is routinely used to evaluate carotid artery stenosis in strokes. Though not asked for routinely, transcranial doppler is useful in detecting intracerebral artery occlusive disease.

EEG

EEG is useful in aiding diagnosis and classification of epilepsy. Characteristic interictal epileptiform discharge supports diagnosis of

epilepsy but its absence does not rule this out. The distinction between the generalized 3 Hz spike and wave EEG discharge in childhood absence epilepsy and the temporal lobe spikes in complex partial seizures is important for prognosis and the choice of antiepileptic drug. EEG has come to play a less important role in diagnosing structural lesions as neuroimaging is widely available. Virtually diagnostic EEG changes (periodic complexes) are seen in subacute sclerosing panencephalitis and Creutzfeld–Jacob disease.

Cortical evoked potentials are of great value in demonstrating clinically unsuspected lesions in diagnosis of multiple sclerosis. These potentials in the brain are evoked by stimulating the visual, auditory and somato sensory pathway peripherally.

ELECTROPHYSIOLOGIC STUDIES

EMG and nerve conduction studies are well established as an essential part in the workup for neuromuscular disorders. It is also essential for us to understand the pathophysiology of such diseases. Recently, many new neuromuscular disorders have been identified by their characteristic electrophysiologic picture. Such examples include Lambert–Eaton myasthenic syndrome (LEMS), multifocal motor neuropathy and acute axonal neuropathy. There is no standard or routine testing but rather the test depends on the clinical evaluation. The physician has to decide which studies to do, which muscles or nerves to examine. It is good to recognize the limitations of electrophysiology testing and to interpret these results in the context of clinical findings.

Nerve conduction study (NCS) is a useful test in diffuse peripheral neuritis (like GBS, diabetic neuritis) or focal pressure neuropathy like carpal tunnel syndrome (median compression) or meralgia paresthetica (lateral cutaneous nerve of thigh compression). One occasionally recognizes a disorder, multifocal motor neuropathy with conduction block on NCS. This closely resembles clinically motor neurone disease (ALS) but is treatable with immunosuppressive medication and intravenous immunoglobulin.

Peripheral neuropathy could be predominantly demyelinating or axonal, depending on the etiology. This can be characterized by nerve conduction study and needle EMG of the muscle supplied. Demyelination slows or blocks conduction where as axonal affection

causes low amplitude response on stimulating proximally and distally and causes denervation changes in EMG of that muscle.

The usefulness of electrophysiological studies is illustrated by this common clinical problem of numbness and paresthesia of the little finger with associated wasting of the intrinsic hand muscle. This could result from a spinal cord lesion, C_8T_1 radiculopathy, brachial plexus (lower trunk or medial cord) or a lesion of the ulnar nerve. If sensory nerve action potential can be recorded normally at the wrist following stimulation of digital nerve in the affected finger, the pathology is probably proximal to the dorsal root ganglion (DRG), i.e. there is a radiculopathy or more central lesion. Absence of sensory potential by contrast suggests distal pathology (distal to DRG). EMG will indicate whether pattern of affected muscles conforms to radiculopathy or ulna nerve territory or is more extensive (plexopathy). Ulna motor conduction study will distinguish between radiculopathy (normal conduction) and ulna neuropathy (abnormal conduction) and will identify site of ulna nerve lesion.

Late responses (H reflex or F waves) are able to study the proximal part of the nerve segments. Stimulating a motor nerve causes impulses to travel orthodromically (towards nerve terminal) as well as antidromically (towards spinal cord). Such anti-dromic impulses go to the Anterior Horn Cell (AHC) through the anterior root and causes the AHC to discharge producing a small motor response that occurs considerably later (late response) than the direct response elicited by nerve stimulation. This is the F wave and is delayed in proximal lesion like radiculopathies in GBS, and is useful in early diagnosis.

Needle EMG is helpful in detecting myotonia and in differentiating between denervating process and myopathy. It helps to distinguish axonal neuropathy from demyelinating neuropathy. It is useful in anterior horn cell disease like motor neurone disease. The interpretations of EMG depend a lot on the electromyographer unlike nerve conduction studies which are well standardized.

Though MRI has became the first choice in workup for cervical or lumbar radiculopathy it cannot confirm radiculopathy in all cases (such as diabetes). Judicious use of needle EMG is needed even in the era of MRI.

Repetitive Nerve Stimulation (RNS) test is useful in diagnosing neuromuscular transmission disorders like myasthenia gravis (MG) or LEMS. A decrement of the compound muscle action potential (CMAP) at 3 Hz stimulation occurs in both. An incremental response of the CMAP at

a fast rate of stimulation (50 Hz) or after exercise is characteristic of LEMS and differentiates this from MG. Tensilon test done clinically or during RNS is more dramatically positive in MG than in LEMS.

Single fiber EMG (SFEMG) is an extremely sensitive test for myasthenia gravis (MG) and other neuromuscular transmission (NMT) disorders (thus, it is non-specific for MG). A normal SFEMG study in a weak muscle rules out NMT disorder. When recording single fiber muscle action potential (SFMAP) from two single muscle fibers belonging to the same motor unit by inserting the electrode between the two, there is always a slight variability in the time interval between the two potential in such pairs. This variability is called the jitter that is measured.

ANTIBODY TESTS IN NEUROLOGICAL DISORDER

Ach R antibodies in myasthenia gravis is a fairly routine testing. Anti-voltage gated calcium channel antibodies of muscle are found in LEMS. Anti-Hu associated with small cell lung cancer and anti-Yo associated with cerebellar disorder are found in paraneoplastic syndrome. These are not routinely tested. Anti-GM_1 antibodies are often seen in axonal type of GBS. In GBS, antibodies to gangliosides is not a recommended test because many GBS patient do not have this. However most patients with Miller–Fisher Syndrome (a variant of GBS) have serum GQ_1b antibodies and this antibody appears specific. This assay is now commercially available and is useful in supporting the clinical diagnosis of Miller–Fisher Syndrome.

We will now consider how to approach two of the common neurological complaint that we often see in clinical practice. This is the patient whose presenting complaint is either weakness or a sensory complaint. The third is Nystagmus, which has anguished those just initiated and those more experienced in neurology.

SENSORY COMPLAINT AND SENSORY LOSS

Ask patient to delineate the area of pain or sensory loss. Patient normally can map it as well or better than the examiner and much more quickly.

The sensory impairment or pain can be of a dermatomal pattern. A good example is a sacral saddle shaped loss of sensation around the anus in a cauda equina lesion or a forefinger paresthesia of a cervical disc disease causing C_6 root lesion.

Neuralgic facial pain is distributed along a branch of the trigeminal nerve and is paroxysmal. In trigeminal neuralgia, the patient suffers severe paroxysmal bouts of pain distributed crisply over the distribution of a branch of the trigeminal nerve covering the face, usually the maxillary and mandibular branches. Sensory loss over this area is characteristically absent. Pain is so severe to cause grimaces, a tic douloureux. Prompt response to carbamazepine is diagnostic. Post-herpetic neuralgia follows the zoster eruption of the ophthalmic branch of the trigeminal nerve. After all causes of facial pain are excluded, there remains a patient usually female where the doctor is tempted to attribute depression or a psychological cause and call this condition atypical facial pain.

Spinal cord lesion causes a sensory level with loss of sensation below this level. A band of hyperasthesia at the level of lesion is a very useful and reproducible sign. The sensory level becomes more credible if the level is elicited also on the back of the patient. A hemisection of the cord (Brown–Sequard Syndrome) produces loss of dorsal column sensation (vibration) and limb weakness on the same side as the lesion with loss of pain and temperature sensation on the opposite side. A syringomyelic lesion (syrinx usually at around C_4–T_5 cord) usually affects the crossing spinothalamic tracts causing a cuirass like suspended sensory loss (for pain and temperature) over the chest and inner part of the arms.

Dissociated sensory loss is an interesting phenomenon resulting from anatomical distribution of sensory nerve fibers where some and not all modalities of sensation are lost, for example pain and temperature loss but vibration is intact. Other than in the above two conditions (Brown–Sequard syndrome and syringomelia), it can occur in lesions affecting the lateral medulla where the lateral spinothalamic tract runs discretely lateral to the medical lemiscus. This occurs in Wallenberg's syndrome. Some peripheral neuritis can have dissociated sensory loss in arms or trunk due to preferential affection of the small unmyelinated pain fibers in the peripheral nerve.

Sensory changes affecting the whole of the opposite side of body with simultaneous presence of pain and anesthesia is called anesthesia dolorosa. This could occur in a thalamic lesion, usually vascular (Dejerine–Roussy syndrome) but often mistaken as hysteria.

Astereognosis occurs in parietal cortex lesions. Stereognosis is the ability to recognize the size, shape and texture whilst the eyes are kept closed of a readily identifiable object, say a dollar coin placed in the hand of a patient who has no obvious sensory disturbance.

While peripheral neuritis causes a glove and stocking distribution of paresthesia and sensory loss (with generalized muscle weakness and wasting of muscles and diminished tendon jerks), entrapment neuropathies cause symptoms and signs limited to the nerve distribution. A good example is carpal tunnel syndrome where pain and sensory loss occurs over the first three digits and the lateral half of the fourth digit. As pathology is a focal demyelination of the median nerve at the wrist, tapping the nerve or stretching the nerve by wrist flexion or extension causes tingling over the affected fingers (Tinel and Phalens sign respectively). Typically the pain awakens the patient at night. Another good example is meralgia paresthetica, which is the result of lateral femoral cutaneous nerve of thigh entrapment at the groin. The patient has painful paresthesia with sensory loss over the anterolateral surface of the thigh.

Hansens disease should be suspected if there is patchy sensory loss due to intradermal affection of the nerve. This is usually associated with thickened nerve.

PATIENTS WHO PRESENT WITH PARALYSIS OR WEAKNESS OF LIMBS

Diagnosis can be made easier by considering the problem as a hemiplegia, paraplegia, quadrepligia or isolated paralysis.

Hemiplegia is weakness of an arm, leg and sometimes the face on one side of the body. This occurs in pyramidal tract lesion of the brain usually in the opposite cerebral cortex or internal capsule. A small capsular lesion can cause a more extensive weakness than a lesion in the cortex for obvious reasons. The site of the lesion is deduced from the associated neurologic finding. Aphasia or a visual field defect localizes to the opposite cortex. Pure hemiplegia localizes to the internal capsule. Crossed hemiplegia localizes the lesion to the brain stem. Examples of this are Webers syndrome (3rd nerve palsy with opposite hemiplegia, a lesion of the midbrain), Millard–Gubler syndrome (Ipsilateral 6th or 7th cranial nerve palsy with contralateral hemiplegia, a lesion in the pons). The most common cause of this is a vascular lesion. Rarely does Brown–Sequard syndrome at the cervical level cause a hemiplegia.

Paraplegia occurs in disease of the spinal cord, spinal root or peripheral nerves. Parasaggital lesions in the brain and hydrocephalus can cause leg weakness. Spinal cord lesions affect bladder control in addition.

Acute paraplegia is usually from trauma, metastatic tumor or an epidural abscess. Prompt laminectomy surgery for an epidural abscess is rewarding. An occasional case of Potts disease (tuberculosis of spine) usually affects the mid-thoracic spine, causing a gibbus deformity. Anterior spinal artery thrombosis or dissecting aortic aneurysms can infarct the spinal cord and present acutely. The more common transverse myelitis we see in the ward is post-infectious or is acute demyelination. Acute demyelinating myelitis with optic neuritis is Devics disease. This is the common form of demyelinating disorder in Singapore and is believed to be the local variant of multiple sclerosis. Cervical disc disease causing myelopathy is seen more commonly than spinal cord tumor. Cervical root pain usually preceeds cord compression by the disc. A tiny ventral pontine infarct can manifest as a locked-in state, which is a unique deefferented state where all muscles are paralyzed except the extraocular muscles. Guillain-Barre Syndrome causes an ascending lower motor type of paralysis which usually progresses over a week to become quadreparesis. This is the commonest acute peripheral neuritis seen in our wards. In general, neuropathic disorders cause distal limb weakness and disorders of neuromuscular junction or muscles cause proximal weakness.

Isolated paralysis is usually from a lesion of a peripheral nerve. The acute ulna nerve palsy or the peroneal nerve palsy is usually a pressure palsy. The affected muscles are supplied by the particular nerve. Weakness causing a wrist drop is usually from a radial nerve palsy, and rarely can be a cortical lesion. L_5 radiculopathy (lumbar disc prolapse) can cause a foot drop. This is differentiated from a peroneal nerve palsy because the tibialis posterior is also weak. This is supplied by the posterior tibial nerve originating from the L_5 root. Mononeuropathy multiplex occurs in 2 or more nerves in more than one involved extremity (eg. left ulna neuropathy and right peroneal neuropathy). This is typically found in vasculitis. Other causes are leprosy and diabetes.

Monoplegia is weakness of all the muscles of a limb, though a detailed examination may disclose weakness of another limb, making this a hemiplegia. Atrophy of the muscles makes one suspect a lower motor neurone palsy (other signs are diminished reflexes and fasciculation) where the lesion can be in the nerve, root or anterior horn cell. Poliomyelitis, brachial neuritis, brachial plexus trauma, monomelic motor neurone disease should be considered. Monoplegia without muscular atrophy is probably due to a lesion of the motor cortex (usually vascular and rarely a tumor).

Hysterical paralysis causes diagnostic difficulty, for example in paralysis of a leg. The Hoover sign is helpful: When the patient is asked to lift the paralyzed leg while lying down, the other heel pushes down. With the examiner's hands under the heels, pressure is felt by the examiner on the non-paralyzed side. In hysterical paralysis the contralateral heel does not press down as the patient makes no effort to move.

Recent advances of technology has opened up new vistas of investigation and management of neurological diseases. However, one must not forget that clinical neurology is very much an art and no amount of investigation can replace the clinical wisdom.

NYSTAGMUS

Doctors often get into difficulty in interpreting nystagmus by mixing up different classifications because the mechanism is speculative. Nystagmus is a rhythmic oscillation of the eyes, usually involuntary and conjugate. This results from dysfunction of the vestibular end organ, vestibular nerve, brainstem or cerebellum. The slow phase of jerk nystagmus is responsible for generation of nystagmus where as the fast saccade is the corrective movement bringing the fovea back to target. The leaky brainstem integrator is unable to maintain constant output to the gaze center in order to hold the eyes in an eccentric position. This causes the slow phase of the nystagmus and is often caused by the effect of alcohol, and anticonvulsants. Vestibular nystagmus is either central or peripheral end organ in origin.

Vestibular end organ problem causes imbalance of peripheral vestibular input to the brainstem gaze center. Labyrinth disease usually suppresses labyrinthine input there by simulating the nystagmus of cold water caloric test, which stimulates the horizontal semicircular canal.

The mixed vertical and torsional nystagmus seen in position-induced nystagmus suggests posterior semicircular canal irritation. This plane relates to the geometric relationship of the semicircular canal.

The doctor needs to evaluate the nystagmus critically. He will have to describe the character of the nystagmus (Table 1), differentiate a peripheral from a central vestibular cause (Table 2) and localize it anatomically (Table 3). Testing for nystagmus while the patient wears frenzel goggles to impair visual fixation brings out peripheral nystagmus. This is usually done by the ENT surgeons.

Table 1 Jargon of Nystagmus

1) Wave form
   a) Pendular
   b) Jerk
2) Plane
   a) Horizontal
   b) Vertical (upbeat and downbeat)
   c) Torsional
   d) Combination of above
   e) Is there a spontaneous alteration of direction? This requires observation over a period of time, and occurs in periodic alternating nystagmus.
3) In relation to gaze
   a) Nystagmus in primary position
   b) Gaze-evoked nystagmus
4) Positional — provoked by head posture

Etiology
1) Physiological — optokinetic, end-point (on extreme gaze), caloric
2) Pathological
   a) congenital — accompanied by poor vision, albinism
   b) acquired — commoner type than congenital

Anatomical Localization
1) Central — vestibular projections in the brainstem or cerebellum
2) Peripheral — vestibular end organ or acoustic nerve.

Table 2 Central versus Peripheral Vestibular Nystagmus

Feature	Peripheral	Central
a) Plane	Mixed horizontal-torsional	Pure vertical, pure horizontal Pure torsional or mixed
b) Direction	Unidirectional	Can be bi-directional (changes direction with gaze)
c) Abolishing fixation by Frenzel goggles	Apppears or increases	No change
d) Vertigo, tinnitus and deafness	Prominent	Less prominent
e) Other signs	No	Brainstem or cerebeller

Table 3 Axioms in Nystagmus

1)	Jerk nystagmus in central or peripheral.
2)	In jerk nystagmus, the amplitude increases with gaze in the direction of the fast phase (Alexander's Law), which is conventionally defined as the direction of nystagmus.
3)	Pendular nystagmus is either congenital or acquired from central cerebellar and brainstem disease, usually multiple sclerosis.
4)	When anticonvulsants and psychoactive drugs are excluded, gaze-evoked vertical nystagmus indicates acquired central disease.
5)	Pure vertical or pure torsional nystagmus almost never occurs with peripheral vestibular disease.

Congenital nystagmus is characteristically pendular or jerky in primary position and remains horizontal whether patient looks up or down or to a side (uniplanar). There is a "null zone" (a field of gaze where nystagmus in minimal and acuity is best) and nystagmus dampens with convergence.

Nystagmus with Specific Localizations

Downbeat nystagmus (fast phase down) in primary position or lateral gaze is characteristically seen in lesions of the cervicomedullary junction at foramen magnum region. Upbeat nystagmus occurs in lesions at the ponto-medullary junction or superior cerebellar vermis. Pure torsional nystagmus indicates brainstem disease. See-saw nystagmus is a unique type with torsional and pendular characteristics (rising eye intorts and falling eye extorts) frequently associated with lesions of the mesodiencephalitic junction (Cajal nucleus) and chiasm. Convergence-retraction nystagmus on attempted upward gaze is a component of dorsal midbrain syndrome (Parinaud's syndrome). Ocular bobbing (rapid downward movement of eyes followed by a slow drift up, occurring 2–15 times per minute) occurs in comatose patients with massive pontine lesion. A lesion in the medial longitudinal fasciculus gives rise to internuclear ophthalmoplegia characterized by ipsilateral adduction failure and nystagmus in abducting eye (described as ataxic nystagmus) on looking away from the side of the lesion. Drugs and toxins may cause nystagmus in any direction.

57

Headache

Einar Wilder Smith, Wong Meng Cheong, Lim Shih Hui and Pavanni Ratnagopal

BASIC CONCEPTS AND EPIDEMIOLOGY OF HEADACHES

Headache is one of the most common human maladies afflicting people of all societies and cultures. Studies that examined the lifetime and yearly prevalence have shown high rates in most regions where these have been conducted. In Asia, Europe and the United States, estimates of the lifetime prevalences of any form of headache, tension headache and migraine in the general adult population are approximately 90%, 70–80% and 10–25% respectively.

The head is defined as all structures above the neck, whereas the face is part of the head and is defined as the area between a line drawn between the lateral epicanthus and the superior border of the ear pinna and the chin. In this chapter, all types of aches and pains located in the head and face are discussed.

Headache Pathophysiology

The head is prone to hurt because the scalp and face have some of the body's richest supplies of pain receptors in order to protect the sensitive and important contents and attachments. The eye, nasal and oral passages all reside here. When diseased, they can all result in headache or facial pain.

The pathophysiology of headache is complex and differs with the type of headache. Overall, it can conveniently be divided into pain generation in the sensitive structures of the head and its attachments, and the subsequent transmission and processing of pain. Pain in and around the head can be generated by many different structures.

The pain-sensitive innervation of the forehead, orbits, anterior and middle fossae of the skull is largely derived from branches of the ophthalmic division of the trigeminal nerve and, to a lesser extent, from the second trigeminal division. The sphenopalatine branches of the facial nerve innervate naso-orbital region. The 9^{th} and 10^{th} cranial nerves, as well as the first 3 cervical nerves, convey pain from the inferior surface of the tentorium and the posterior fossae. Consequently, pain from the supratentorial structures is referred to the first two divisions of the trigeminal nerve and pain from the infratentorial structures is referred to the back of the head and neck. Structures that are not pain-sensitive include the brain parenchyma, the bony skull, and the ependyma and choroid plexuses.

Many types of headache are a result of the stimulation of the pain-sensitive intra- and extracranial vessels. Both migraine and cluster headache are caused by dilatation of intra- and extracranial vessels. Intracranial mass lesions cause pain through the displacement of vessels. Inflammation of arteries underlies the headache of temporal arteritis. Pain in sinus disease is due to the stimulation of pain-sensitive sinus walls and is relieved when aeration is restored and the inflammation subsides. Headaches of ocular origin (hypermetropia and astigmatism) are either the result of sustained contraction of frontal, temporal or occipital muscles or are due to raised intraocular pressure, with or without inflammation. Any disease of the neck ligaments, muscles and vertebral joints is capable of resulting in headaches. The pain in these conditions is thought to arise from secondary protective spasm of the muscles or directly from the pain-sensitive joints and ligaments.

Meningeal irritation due to infection or bleeding is due to direct stimulation of the meningeal trigeminal nerve endings. A further headache mechanism is due to low cerebrospinal fluid pressure. Here, pain is a result of the downward displacement of intracranial contents and the subsequent traction of vessels. The recent identification of the "trigeminovascular system" has helped in the developing of a new theory for the generation of the migraine headache. Cells located in the caudal trigeminal nucleus have been identified as the anatomical locus for pain transmission in migraine. Anti-migraine drugs, such as dihydroergotamine and the triptans, can directly inhibit these cells.

A CLINICAL APPROACH TO HEADACHES

History Taking

In a patient with a headache, the most important factor that leads to an accurate diagnosis is the history. In most cases, detailed history taking alone will allow an accurate etiological classification of the headache. Table 1 lists the important features to be sought during a physical examination.

Table 1 Physical Examination in Headache Patients

General physical examination
• Blood pressure (malignant hypertension as headache, diastolic of >120 mmHg)
• Meningism (this should include the Lasegue, Kernig and Brudzinski's signs)
• Temporal artery: pain on palpation, absent pulsation and prominence
• Signs indicating a local affection of the skull
– Sinusitis: percussion sensitivity over the sinuses and streaks of pus at the back of the throat
– Otitis: check ear with otoscope
– Check skull form for deformities or enlargement (hyperostosis with Paget's disease, meningiomas or certain osteosclerotic metastasis)
– Glaucoma/iridocyclitis: check eye pressure and anterior eye chamber
– Temporo-mandibular joint disease: check for asymmetrical bite and pain over the joint
– Consider dental pain
Neurological examination
Signs of raised intracranial pressure (optic discs for papilledema)
Cranial nerves
Other signs of focal neurology

HEADACHE CLASSIFICATION

Table 2 is a simplified version of the International Headache Society's classification of headaches.

Table 2 Classification of Headache and Facial Pain (modified from the International Headache Society Classification)

1) Migraine
   a) Migraine without aura (common migraine)
   b) Migraine with aura
      - Typical aura ("classical migraine")
      - Atypical aura (basilar migraine and hemiplegic migraine)
2) Tension-type headache
   a) Episodic tension-type headache
      - Evidence of scalp/neck muscle involvement
      - No evidence of muscle involvement
   b) Chronic or continuous tension-type headache
      - Evidence of scalp/neck muscle involvement
      - No evidence of scalp/neck involvement
3) Cluster Headache and chronic paroxysmal hemicrania
   a) Episodic cluster headache
   b) Chronic cluster headache
   c) Chronic paroxysmal hemicrania
4) Miscellaneous benign dysfunctional headaches
   Includes ice cream headache, cough headache, exertional headache, sexual headaches and so forth
5) Post-traumatic headaches
6) Headache associated with vascular diseases
   a) Subarachnoid hemorrhage
   b) Vasculitides
   c) Arterial dissections
   d) Intracranial venous thrombosis
   e) Arterio-venous malformations
7) Headache from other intracranial disease
   a) Altered CSF pressure
      - High, such as pseudotumor cerebri
      - Low, such as post lumbar puncture or spontaneous intracranial hypotension
   b) Intracranial mass lesion
   c) Intracranial inflammation
8) Headaches associated with substances or their withdrawal
   a) Acute use or exposure
      - Nitrates, monosodium glutamate and methylalcohol
   b) Chronic use or exposure
      - Ergotamine, analgesics and benzodiazepines
      - H2 blockers, some NSAIDs and antibiotics

c) Withdrawal
   - Alcohol, caffeine, ergotamine, analgesics and so forth

9) Headaches associated with diseases of the skull, neck, eyes and sinuses
10) Cranial neuralgias
   a) Trigeminal neuralgia
   b) Occipital neuralgia
   c) Glossopharyngeal neuralgia
   d) Herpes zoster
   e) Tolosa–Hunt syndrome
11) Others
   a) Neck-tongue syndrome

Table 3 Diagnostic Criteria for Migraine and Tension Headache

1) Diagnostic criteria for migraine without aura
   a) At least 5 attacks of headaches that fulfill criteria (b) and (d)
   b) Headaches lasting four to 72 hours each
   c) Headache has at least two of the following characteristics:
      - Unilaterality
      - Pulsatility
      - Intensity that is great enough to affect normal activities
      - Headache aggravation by exercise
   d) Headache is accompanied by at least one of the following:
      - Nausea and/or vomiting
      - Photophobia
      - Phonophobia
   e) No evidence from the history or examination of any other disorder which could be causing the headaches
2) Diagnostic criteria for migraine with aura
   a) One or more transient focal neurological aura symptoms (usually visual)
   b) Gradual development of aura symptoms after more than four minutes
   c) Aura symptoms that last from four minutes to an hour
   d) Headache follows or accompanies aura within an hour
3) Diagnostic criteria for tension-type headache
   Multiple attacks of headache with the following characteristics:
   a) Headaches must last half an hour to a week each
   b) Headache has at least two of the following:
      - Bilateral
      - Non-pulsatile
      - Moderate intensity
      - Not aggravated by physical activity
   c) The headache is not accompanied by anything or, at the most, by a little hypersensitivity to light or noise
   d) No evidence from the history or examination of any other disorder that could be causing the headaches

Table 3 lists the diagnostic criteria for migraine and tension headache.

OTHER TYPES OF HEADACHES

Indomethacin Responsive Headaches

Because of their sometimes dramatic response to indomethacin, it is worth considering this group of headaches in a differential diagnosis. Paroxysmal hemicrania and hemicrania continua are invariably responsive to indomethacin. Other short-lasting headaches, such as exertional or cough headache, ice-pick headaches, idiopathic stabbing headache and headache associated with sexual activity, are partially responsive to indomethacin.

Headache Due to Low CSF Pressure

This often occurs after dural puncture. Though less common, there may be spontaneous rupture of the dural membrane. The headache consistently appears when the body is in a vertical position and abates when it is lying down. It usually disappears after one or two days of bed rest and taking plenty of fluids. Persistent post-dural puncture headaches have been successfully treated with epidural blood patches.

HANDHELD CELLULAR TELEPHONE INDUCED HEADACHES

There was a recent epidemiological study in Singapore on central nervous system symptoms of users of handheld cellular phones and headache was the most prevalent symptom compared with non-users. It was also related to the increased duration of use of the handphone (min/day) and headache was reduced considerably among those who used hands-free equipment.

Headache Due to High CSF Pressure (Pseudotumor Cerebri)

Pseudotumor cerebri (also called benign intracranial hypertension) is an uncommon condition more commonly found in young to middle-aged overweight females. It is diagnosed in the presence of headache with bilateral papilledema and (optional) VI nerve palsy. Other features of the neurological examination, CSF and imaging of the brain (CT or MRI) must

be normal. In a proportion of cases, the condition is related to treatment with corticosteroids, vitamins or antibiotics. The treatment consists of alleviating raised intracranial pressure using diuretics (frusemide, diamox and hydrochlorothiazide) or ventriculo-peritoneal shunting. In grossly overweight patients, weight reduction alone can result in remission. Patients can become blind and therefore warrant careful follow-up to prevent this. Optic nerve sheath fenestration can be used to treat the progressive loss of vision.

INVESTIGATION OF HEADACHE

The most important part of the investigation of headache is the history and examination. Most headaches fall into the category of tension headache or migraine and do not require further investigation. Plain views of the skull are of little help in the work-up of headaches. Their usefulness is limited to the rare patient with headache due to Paget's disease or multiple myeloma. Plain views of the paranasal sinuses can identify air/fluid levels in sinusitis. Plain films of the cervical spine (+/− views in hyperextension/hyperflexion) identify cervical spondylosis or cervical bone destruction. If an intracranial lesion is suspected or needs to be excluded, brain scanning with CT is usually sufficient. CT scanning is particularly useful in the imaging of areas that consist mainly of bones, such as the orbit, sinuses, mastoid or base of the skull, as well as in identifying intracranial blood. MRI imaging is superior in situations where accurate imaging of the posterior fossa is required or more detailed views are needed.

Lumbar puncture is required to confirm the diagnosis of meningitis and pseudotumor cerebri. Although iron deficiency anemia can be associated with headache (and can respond to iron therapy), hemoglobin levels are rarely helpful in the investigation of headache. The ESR (or CRP) is of great value in helping to establish the diagnosis of temporal arteritis (TA). If TA is present, the ESR will be greatly raised (>70 mm in the first hour). Because of the long ensuing period of treatment with steroids and its associated complications, we recommend establishing a definite diagnosis of TA with temporal artery biopsy. A psychiatric referral may be helpful when diagnosis of atypical facial pain is considered for an evaluation of possible depression.

TREATMENT OF HEADACHES

The treatment of headaches encompasses the treatment of an underlying secondary cause, such as an intracranial space occupying lesion or infection, as well as symptomatic or prophylactic treatment with pharmacological agents. A wide range of analgesics can be used for symptomatic headache treatment.

Treatment of Tension-type Headache

Explanation and reassurance form the mainstay of therapy. The patient needs to know what the presumed mechanism of headache is. Hence explaining that the headache is due to fatigue, worry, stress or poor posture is important. Physical measures can be quite useful. Correction of poor body posture is helpful. Also, short neck muscle massage several times a day and a hot shower sometimes help. Chronic tension-type headache with no response to physical measures may respond to a three to six-month course of antidepressants, such as amitriptylline (10–20 mg) or fluoxetine (20 mg) at night. Other options include doxepin, desipramine or nortriptylline. The tricyclic antidepressants are generally more effective than the serotonin re-uptake inhibitors (fluoxetine), but are also less well tolerated.

Treatment of Migraine

Migraine attacks can be triggered off by a variety of factors. The avoidance of triggering factors, such as certain foods, alcohol, caffeine or smoking, should be watched out for.

Analgesics

Paracetamol, on its own, may be helpful, but often a "resistance" seems to develop with its long usage. Over-the-counter analgesics should not be used more than three times a week. This is because it can result in headache due to its abuse. If caffeine or codeine is mixed with simple analgesics, a medication induced headache is more likely to occur. If acetylsalicylic acid is used, a high enough dosage (1000 mg)

should be used and its combination with 10 mg or 20 mg of metoclopramide increases gastrointestinal absorption and effectiveness.

Specific Anti-migraine Drugs

Ergotamine or dihydroergotamine are reserved for moderate to severe attacks. Because of its powerful vasoconstrictive action, ergotamine should not be used in patients with ischemic heart disease, peripheral vascular disease or arterial hypertension. Sumatriptan was the first 5 HT agonist available for the treatment of acute migraine. At the time of writing, three different triptans are available although one more is in the process of completing clinical trials and hopefully be launched soon.

Prophylactic Treatment

All pharmacological prophylactic treatments take one or more months for their maximum effectiveness to be realized. The usual dosages of commonly used drugs are given in Table 4.

Recently, botulinum toxin has been found to be useful in relieving certain types of migraine headaches and some of these patients even underwent removal of the corrugator supercilii muscles if they responded to Botox.

Table 4 Prophylactic Anti-migraine Agents

Action	Drug	Range of Effective Dose (mg/day)	Effective Dose (mg/day)
Non-specific Beta-blockers	Propranolol	20–360 mg	80–240
Specific beta-1	Atenolol	25–150	50–100
Tricyclic Anti-depressants	Amitriptylline	25–300	75–150
Calcium channel blockers	Flunarizine	5–10	5
	Verapamil	120–160	120
Anti-epileptics	Valproate	500–1500	500–800
Vitamin	Riboflavin	400	400

REFERENCES

1. Ho KH, Ong BK, Lee SC, Headache and self assessed depression scores in Singapore University undergraduates, *Headache* **37**(1):26–30, 1997.

2. Ho KH, Chong PN, Headache characteristics in university graduates presenting to medical attention, *Singapore Med J* **37**(6):583–584, 1996.

3. Rasmussen BK, Jensen R, Schroll M, Oleson J, Epidemiology of headache in a general population—a prevalence study, *J Clin Epidemiol* **44**(11):1147–1157, 1991.

58

Meningitis and Encephalitis

Pavanni Ratnagopal, Einar Wilder Smith and Tan Ban Hock

INTRODUCTION AND HISTORY

Meningitis and encephalitis are devastating diseases of infancy, childhood and adulthood. Despite antibiotic therapy, the morbidity and mortality remains high.

Tuberculous meningitis was first described in 1768 by Robert Whytt who initially described it as "dropsy in the brain". Gaspard Vieusseux's description of an outbreak of "cerebrospinal fever" in Geneva (1805) was probably the first description of meningococcal meningitis. This was followed in 1806 by the first epidemic of spotted fever to be documented in America in Medfield, Massachusetts. In 1810, a medical student, Nathan Strong, first recognized meningism as a diagnostic sign for meningitis. Vladimir Mihailovich Kernig described his maneuver for detecting meningeal irritation in 1882 and Josef Brudzinski described his observations in 1909. Heinrich Quincke pioneered the technique of lumbar puncture in 1891.

At the turn of the century, Jochmann in Germany and Flexner in New York demonstrated the protective effect of antimeningococcal serum

in experimental meningococcal infections in animals. The discovery of the antibacterial activity of sulfonamides in the early 1930s ushered in the antibiotic era.

PATHOGENESIS AND PATHOPHYSIOLOGY

Bacterial Meningitis

The most common bacteria that causes meningitis, *Hemophilus influenzae*, *Neissaeria meningitidis* and *Streptococcus pneumoniae* initially colonize the nasopharynx by secreting IgA proteases that breakdown the mucous barrier and allow bacterial attachment to the epithelium.

Once these bacteria gain access to the bloodstream, a polysaccharide capsule helps avoid phagocytosis by neutrophils. The mechanisms by which bacteria enter the CSF are largely unknown although it is speculated that it is through the choroid plexus of the lateral ventricles and other areas of altered blood brain permeability.

Since the CSF is an area of relatively impaired host defence with few phagocytic cells, low protein concentration, no IgM and low levels of complement (in particular C_3 and C_4), bacteria initially multiply with little opposition. Meningeal inflammation develops as toxic components of the bacterial cell wall are released with cell lysis. Antibiotics that kill Gram-negative bacteria by cell wall lysis with subsequent release of endotoxin have been shown in experimental models to be associated with a dramatic increase in vasogenic brain edema. There is also a strong correlation between CSF endotoxin concentration and the development of seizures and morbidity and mortality in children with Gram-negative meningitis.

The purulent subarachnoid exudate of meningitis is mainly due to CSF neutrophilic pleocytosis. As the infective process develops alterations in cerebral blood flow can occur due to raised intracranial pressure, loss of intracerebral vascular autoregulation, vasculitis and thrombosis of cerebral arteries, veins and major sinuses. This in turn may lead to ischemic infarcts.

Viral Meningitis

The common viruses causing meningitis are from the enteroviruses group, followed by the mumps virus, arboviruses, and the herpes simplex virus.

Enteroviruses are a major subgroup of picarnoviruses and include the polioviruses, coxsackieviruses, echoviruses and more recently discovered

agents that are simply designated as enteroviruses. They can survive for prolonged periods in sewage as well as in chlorinated water, if sufficient organic debris is present. Most of these agents can be isolated by cell cultures and humans are their natural host. Infection is generally asymptomatic, but nervous system manifestations include aseptic meningitis and less commonly encephalitis, transverse myelitis or a poliomyelitis-like syndrome. Enteroviral infections of the nervous system may leave significant sequelae like mental retardation and cerebral palsy in young children under 1 year of age and fatigue and weakness in adults.

Mumps virus is a paramyxovirus spread by respiratory droplets. About half the patients with parotitis will have CSF pleocytosis without neck stiffness. Aseptic meningitis may be a presenting feature of mumps, but can also develop two or three weeks into the disease. Viral meningitis generally causes a lymphocytic pleocytosis, with the cell counts less than 1000/μL. CSF glucose is usually normal and protein is either normal or slightly elevated. Specific treatment may be withheld if the patient is well and CSF changes show those typical for a viral etiology. The differential diagnosis of lymphocytic "aseptic meningitis" includes partially treated bacterial meningitis and tuberculous meningitis and poor inflammatory response in an immunesuppressed patient. Therefore, in an ill patient, antibiotics often need to be used before the diagnosis can be finalized.

Fungal and TB Meningitis

In fungal and TB meningitis, there is a subacute or chronic basal meningitis with inflammatory foci of organisms and associated infiltrates of lymphocytes, macrophages, and multinucleate giant-cells. In cryptococcal meningitis, the subpial cortical inflammatory response is lacking, and there are subarachnoid clusters of cystlike cavities filled with organisms ("soap bubbles") in the cerebral cortex with little or no surrounding inflammation. The site of primary infection is usually from the lungs and less commonly from the skin. Primary lesions may occur in the mouth and pharynx in paracoccidioidomycosis and in the gastrointestinal tract in histoplasmosis.

Encephalitis

In general there is diffuse and widespread CNS infection as opposed to a localized inflammation. Specific viruses affect specific anatomic areas or subpopulaton of cells, a phenomenon known as "tropism". The best known examples are the herpes simplex virus, which causes lesions in the

limbic system, and polio virus, which affects the anterior horn cells of lower motor neurons.

Encephalitis can occur either by hematogenous spread after viremia or by retrograde spread along nerves like in rabies and herpes simplex.

World-wide, the important organisms usually include the herpes simplex, herpes zoster, HIV, polio, rabies, equine encephalitis and St. Louis encephalitic viruses. Microscopic changes include perivascular cuffing by lymphocytes and plasma cells, neuronal necrosis, inclusion bodies, microglial proliferation and glial nodules. Hemorrhagic necrosis is common in herpes simplex encephalitis. Calcification can be detected in some neonates infected in utero by with encephalitic producing agents like CMV and AIDS viruses.

ETIOLOGIC ORGANISMS

Bacterial Meningitis

The commonest organisms encountered are *Streptococcus pneumoniae* in adults and *N. meningitidis* in children and young adults.

Table 1

	Commonest Organisms	**Uncommon Organisms**
Neonates	Gp B Streptococcus *Escherichia coli* *Listeria Monocytogenes* Klebsiella-Enterobacter species	*Citrobacter diversus* Salmonella species
Childhood	Hemophilus influenza type B (Hib) *Neisseriae meningitides* *Streptococci pneumoniae*	*Listeria monocytogenes*
Adults (16–50 yrs)	*Strep. pneumoniae* *Neisseriae meningitides* *Hemophilus influenzae*	*Listeria monocytogenes* Borrelia Syphilis Gonococcus
Adults (> 50 yrs)	*Strep. pneumoniae* *Escherichia coli* *Klebsiella pneumoniae* *Hemophilus influenzae* Pseudomonas organisms Enterobacter species Serratia species	As above

Streptococcal pneumoniae is most common in USA and some areas of Europe but *Neisseriae meningitidis* is more common in Northern Europe and the Sahara belt in Africa.

In Singapore we have recently described an unusual outbreak of Group B *Streptococcal meningitis* in adults (E. Wilder Smith, personal communication). The usual predisposing conditions were lacking in the overwhelming majority of cases.

Meningococcal meningitis caused by *N. meningitidis* presents acutely with fever, vomiting, lethargy, neck stiffness and headache. The classical diffuse erythematous maculopapular rash may rapidly become petechial and purpuric. The fulminant illness, called the Waterhouse–Friderichsen syndrome presents with adrenal hemorrhage and a disseminated intravascular coagulopathy. There is good protection from meningococcal vaccine for A & C, W135 and Y polysaccharide.

Fungal Meningitis

This occurs primarily, but not solely, in individuals who are in a state of immunosuppression, for example, in those who have:

- the acquired immune deficiency syndrome (AIDS);
- an organ transplantation;
- immunosuppressive chemotherapy or chronic corticosteroid therapy; or
- lymphoreticular malignancies.

The most common fungi causing meningitis are *Cyptococcus neoformans* and *Coccidioides immites*, although Histoplasma capsulatum, *Blastomyces dermatitidis* and Candida species are increasingly reported. *C. neoformans* is a yeast-like fungus found in pigeon droppings, decaying fruits and vegetables, milk and soil. *C. neoformans* has a predilection for the CNS and spreads from a primary pulmonary infection to the meninges by hematogenous spread, and is the most common cause of central nervous system (CNS) infection in the transplant patient.

C. neoformans may present as an acute illness with fever, headache, photophobia and an altered sensorium, or as an indolent illness with headache and low grade fever. Fungal meningitis has a tendency to infect the basilar meninges and so may present with cranial nerve palsies. CNS coccidioidomycosis may present as an acute illness or follow a subacute

chronic cause. Meningitis is the predominant clinical manifestation of CNS histoplasmosis. Mental status abnormalities are common, including stupor, confusion, personality changes and cognitive deficits.

The CSF studies generally show a lymphocytic pleocytosis with an elevated protein content and decreased glucose concentration. The India ink stain is positive in about 50% of HIV-negative individuals but positive in 75–80% of HIV-positive individuals. Latex agglutination testing for the cryptococcal antigen is highly specific and sensitive and should be performed in all CSF specimens.

The serum cryptococcal antigen titers are also often positive in patients with cryptococcal meningitis and can be useful in monitoring the response to therapy. The organism takes about 3 to 10 days to grow. The organism in Coccidioides immitis meningitis grows rapidly in about 3 days but large volumes of CSF are needed. Histoplasma capsulatum is positive in about 25–50% of cases, typically requiring as long as 45 days to grow. The Histoplasma polysaccharide antigen can be measured in urine, blood and CSF, and has the same diagnostic significance as the cryptococcal antigen but this test is not widely available.

Tuberculous Meningitis

Tuberculous meningitis does not develop acutely from hematogenous spread of tubercle bacilli to the meninges. The neurologic complications of tuberculous meningits are initiated by a hypersensitivity reaction to the discharge of tubercle bacilli and tuberculous antigens into the subarachnoid space. A thick exudate is produced and fills the basilar cisternsand surrounds the cranial nerves and major blood vessels at the base of

Table 2 Signs and Symptoms of Tuberculous Meningitis

Symptoms	Signs
Prodromal	Adenopathy
Anorexia	Adventitious sounds
Weight loss	Choroidal tubercles
Cough	Fever
Night sweats	Nuchal rigidity
CNS	Papilledema
Headache (worse in recumbancy)	Focal neurologic signs
Meningismus	Positive tuberculin test
Altered level of consciousness	

the brain, constricting the vessels that comprise the circle of Willis. Within a matter of days, a proliferativ arachnoiditis develops. The inflammatory exudate in the basilar cisterns obstructs the flow of CSF, with resultant obstructive hydrocephalus, and blocks resorption of CSF by the arachnoid granulations as fibrous adhesions develop. Cerebral ischemia and infarction develop as a result of vasculitis due either to a direct invasion of arterial walls by mycobacteria or to compression of the blood vessels at the base of the brain from the adjacent arachnoiditis.

At the early stages of TB meningitis, the CSF may show very few white blood cells and a mild protein elevation but subsequently a repeat study shows a progressive increase in the protein concentration and a progressive decrease in the glucose concentration and a shift to a mononuclear pleocytosis.

The most important prognostic factor that is reported repeatedly in tuberculous meningitis is the level of consciousness at the initiation of therapy. The mortatility rate of patients who are comatose prior to the onset of therapy is 50–70%. Other factors affecting the outcome adversely are:

1) age (mortality is highest in the young and the very old);
2) malnutrition/debilitating disease;
3) the presence of miliary disease;
4) hydrocephalus;
5) cerebrovascular complications;
6) low CSF glucose concentration; and
7) elevated CSF protein concentration.

The diagnosis of tuberculous meningitis is made, and empiric therapy begun, based on a strong clinical suspicion and laboratory data suggest the diagnosis. The initiation of therapy should not await bacteriologic proof of tubercle bacilli by smear or culture.

UNUSUAL TYPES OF MENINGITIS

Aseptic Meningitis

Viral meningitis and aseptic meningitis are terms used interchangeably but should not be. The defining criteria of aseptic meningitis were described by Wallgren in 1925 and is as follows:

1) acute onset;
2) meningeal signs and symptoms;

3) CSF abnormalities typical of meningitis with a predominance of mononuclear cells;
4) absence of bacteria on smear and by culture of CSF;
5) no parameningeal focus of infection; and
6) self-limited benign course.

The classic CSF abnormalities are a mononuclear or lymphocytic pleocytosis with absence of bacteria. The differential diagnosis of a CSF lymphocytic pleocytosis is much broader than one disease entity and includes infectious and non-infectious etiologies.

Generally, patients whose symptoms have not resolved within a week to ten days should have a repeat lumbar puncture. In aseptic meningitis, the investigations other than that of CSF should be geared towards the most likely etiologies.

Table 3 Differential Diagnosis of CSF Lymphocytic Pleocytosis

Infectious etiology
Viral
Enterovirus
Mumps
Lymphocytic choriomeningitis
Herpes simplex virus
Human immunodeficiency virus
Arthropod-borne virus
Non-viral
Mycobacterium tuberculosis
Listeria monocytogenes
Mycoplasma pneumoniae
Rickettsia rickettsii
Treponema pallidum
Borrelia burgdorferi
Cryptococcus neoformans
Partially treated bacterial meningitis
Non-infectious etiology
Systemic lupus erythematosus
Sarcoidosis
Migraine
Traumatic lumbar puncture
Chronic benign lymphocytic meningitis
Vasculitis
Meningeal carcinomatosis
Medications (ibuprofen, isoniazid, azathioprine, trimethoprim)

Source: Connolly and Hammer (1990), and Wilhelm (1992).

Enteroviruses are the commonest cause of aseptic meningitis, accounting for >80% of cases of identified etiology. They are transmitted primarily by the fecal-hand-oral route. They have a worldwide distribution and is generally seen all year round. Mumps virus remains an important cause of aseptic meningitis and often follows the onset of parotitis by several days to weeks. The introduction of the mumps vaccine in the late 1960s has decreased the incidence markedly. The herpes simplex virus is the most frequent of the herpes viruses causing CNS disease, and it is more commonly associated with HSV type 2 genital infections. An increasingly recognized cause of aseptic meningitis over the last decade is the human immunodeficiency virus (HIV). HIV may cause an acute aseptic meningitis at the time of initial infection and seroconversion along with the mononucleosis-like syndrome of fever, malaise, rash, myalgias and arthralgias. One prospective study of high-risk adults noted symptoms suggestive of aseptic meningitis in 7 out of 12 patients at the time they underwent HIV seroconversion.

Of interest is carcinomatous meningitis with leptomeningeal spread, the clinical presentation depending on the extent of the disease along the neuraxis. Headache is the most frequent complaint, being severe and constant, either diffuse or located at the base of the skull with radiation into the neck, and frequently worse on awakening in the morning. There is often associated neck pain and stiffness, though the meningismus is much less severe than that described in purulent meningitis. Cognitive abnormalities consisting of lethargy, confusion or memory loss are also common initial complaints. Focal or generalized seizures may develop. CSF examination is the single most important test for the diagnosis of leptomeningeal metastases.

The median survival from the time of diagnosis of leptomeningeal metastases is 4–6 weeks, unless aggressive treatment is initiated. The purpose of treatment is to prolong survival and stabilize the neurologic function. The standard therapy is radiotherapy to symptomatic areas followed by intrathecal chemotherapy.

Chronic Slow Viral Infections

The Icelandic pathologist Bjorn Sigurdsson first used the term slow infection in 1954 in reference to a chronic degenerative disease of the brain in sheep known as rida or scrapie. These slow infections have 3 criteria: a

long period of latency; a regular protracted course after clinical signs have appeared; and it usually ends in serious disease or death. The infection is limited to a single host species, and anatomical lesions are restricted to a single organ or tissue. Subsequently slow viral infections were divided into two groups: those transmitted by conventional identifiable viruses; and those spongiform encephalopathies associated with a fundamentally distinct molecular pathogen, the prion. Prion stems from the word proteinaceous infectious particle.

The prion diseases may be infectious, inherited or sporadic. Infectious types (iatrogenic Creutzfeld–Jakob disease (CJD), kuru) is due to horizontal transmission of infectious prions. Inherited diseases (familial CJD, Gerstmann–Straussler–Scheinker syndrome or fatal familial insomnia) is due to a mutaion in the protein coding region of the PrP gene.

Chronic viral infections in the nervous system are associated with the persistence of all or part of the viral genome. They cause a variety of syndromes, several of which are distinctive for the specific viruses and associated neuropathology. Some of the more distinct syndromes are: subacute sclerosing panencephalitis (SSPE) due to measles, progressive multifocal leukoencephalopathy (PML) due to papovavirus, and progressive rubella panencephalitis (PRP) due to rubella virus. These typify the classic course of conventional slow virus infection: neurological deterioration leading to death within months to years.

CLINICAL PRESENTATION

The triad of headache, fever and stiff neck is the classical presentation of bacterial meningitis. In addition, photophobia, vomiting and lethargy or altered level of consciousness may also be present.

If there are further signs of confusion or coma, and/or seizures, overlapping encephalitis is likely to be present.

The symptoms and signs vary with different age groups. In addition to the previously mentioned signs and symptoms, vomiting and poor feeding are seen in very young children. The classical signs of bacterial meningitis are meningism as demonstrated by nuchal rigidity and Kernig's and Brudzinski's signs. In a recent study at Yale University by Thomas *et al.*, the diagnostic accuracy of Kernig's sign, Brudzinski's sign and Nuchal rigidity was assessed in 297 patients with CSF results and it was only diagnostic in severe meningeal inflammation. In the broad spectrum of adults with

suspected meningitis, 3 classical meningeal signs did not have diagnostic value. In the elderly this can be difficult to interpret at times. When neck stiffness is due to meningitis, the neck usually resists flexion and can still be rotated from side to side, but if due to cervical spondylosis, parkinsonism or paratonic rigidity, any movement meets with resistance.

Raised intracranial pressure (ICP) is an expected complication of meningitis, especially of bacterial meningitis. The Cushing reflex — bradycardia, hypertension and irregular respiration, dilated unreactive pupils, unilateral or bilateral sixth nerve palsies, papilledema, projectile vomiting and decerebrate posturing, may be seen in extreme cases.

In a patient with a pure encephalitic picture, neck rigidity with Kernig's or Brudzinski's may not be necessarily present. The symptoms and signs of encephalitis are rather similar to those of meningitis except for the element of confusion and coma being more distinctly seen and seizures having more prominence. Seizures may possibly be a sign of cortical venous thrombosis with hemorrhage and this should be considered as part of the differential diagnosis. If bacterial meningitis is left untreated, mortality rate is nearly 100%. In developed countries the mortality rate is stuck at 5–10% and underdeveloped countries at 40%.

INVESTIGATIONS

The following tests should be performed in all cases of meningitis +/− encephalitis:

1) *CT Scan Brain* — If clinical signs do not suggest raised ICP, the LP can be done without a prior CT Scan, but should be considered in centers where scan facilities are available.
2) *Lumbar Puncture* — should be done as soon as possible with stains and cultures for bacterial, fungal and tuberculous organisms as a

Table 4 CSF Results in Common Conditions

	Cells	Glucose	Protein	Globulin
Normal	Up to 5 Lymphocytes	Serum/gluc. ratio 0.6	0.2–0.4	Neg.
Viral	Lymphocytes	N. ratio	May be ↑	Neg.
Bacterial	Polymorphs	↓ ratio	↑	+
TB/Fungal	Lymphocytes	↓ ratio	↑ markedly	+

routine. Viral studies and additional uncommon viral and bacterial studies should be requested for if suspected.

The opening CSF pressure is generally increased in all the conditions but mostly in acute bacterial meningitis.

Although CT Scan and lumbar punctures are the main investigations, the usual investigations for a generalized septic process should still be carried out, including a blood culture, chest X-ray, urine microscopy and culture.

MANAGEMENT

The most important aspects in managing patients with meningitis and encephalitis is ascertaining whether the patient needs emergency supportive treatment. The most serious consequences of meningitis and encephalitis are seizures and raised intracranial pressure which may lead to cerebral coning and death. Once these are dealt with, then specific antiviral, antibacterial or antifungal/TB treatment can be instituted.

Currently there are very few antiviral agents available, the commonest one being acyclovir which is usually given for at least 8 to 10 days at a dosage of 10–12 mg/kg (over one hour infusion intravenously) 8 hourly.

Antibacterial therapy as a broad spectrum coverage without culture results or sensitivity include Gram-positive and Gram-negative organisms.

Table 5 Recommended Antimicrobial Therapy of Common Bacterial Meningitis

Organism	Antibiotic Total Daily Dose (Dosing Interval)
Neisseria meningitides	Penicillin G 20–24 million U/day IV (every 4 hrs) or Ampicillin 12 g/day IV (every 4 hrs)
Streptococci pneumoniae	Ceftriaxone or cefotaxime plus Vancomycin
Gram-negative bacilli (except *Pseudomonas aeroginosa*)	Ceftriaxone 2–4 g/day IV (every 12 hrs) or Cefotaxime 8 g/day IV (every 4 hrs)
Pseudomonas Aeroginosa	Ceftazidime 6 g/day IV (every 8 hrs)
Hemophilus influenzae type B	Ceftriaxone or cefotaxime
Staphylococcus aureus (methicillin resistant)	Vancomycin 2 g/day IV (every 6 hrs)
Listeria monocytoges	Ampicillin 12 g/day IV (every 4 hrs)
Enterobacteriaceae	Third generation cephalosporins

Source: Roos KL, Tunkel AR, Scheld WM, Acute bacterial meningitis in children and adults, in: Scheld WM, Whitley RJ, Durack DT (eds.), *Infections of the Central Nervous System*, Raven Press, New York, pp. 335–409, 1991.

For a long time, intravenous ceftriaxone has been the standard recommendation. With the advent of penicillin resistant streptococci pneumoniae however, recent reviews have recommended the combination of Vancomycin and Ceftriaxone as empirical treatment.

Anti-TB drugs used commonly are rifampicin, PZA, INH and streptomycin, and although the duration of treatment is for 3 to 6 months for lung lesions, most physicians would continue for 9 months to a year at least, if not longer for meningitis. Steroid therapy has been used as an adjunct to TB treatment to decrease brain edema, decrease CSF outflow resistance, decrease the production of inflammatory cytokines and number of leukocytes, thus minimizing damage to the blood brain barrier.

Antifungal treatment generally is geared towards the specific organism, with amphotericin being used commonly in the initial treatment of cryptococcal meningitis.

Chronic recurrent meningitis in immunosuppressed patients may need prophylactic coverage, for example in AIDS patients, cryptococcal meningitis may recur and so prophylaxis with fluconazole may have to be used.

This page intentionally left blank

59

Seizures and Epilepsy: Diagnosis, Investigation and Treatment in Adults

Lim Shih-Hui, Andrew Beng-Siong Pan, Einar Wilder–Smith, Puvanendran K

INTRODUCTION

Epilepsy is a common neurological disorder. It has been estimated to affect about 5 per 1000 of the population at any one time and the accumulated risk of having at least one epileptic seizure over an 80-year life span is 1 in 10.[1] The cumulative incidence of epilepsy for Singapore children by age 9 years is 3.5 per 1000.[2] There is no significant difference in the incidence among males and females, with significant lower incidence in Malays compared to Chinese and Indians. The costs to society and to the individual are considerable, reflecting the increased morbidity in people with epilepsy. There is also evidence of increased mortality in adults with epilepsy. In this chapter, diagnosis and general management of seizure and epilepsy in adult are discussed. Status epilepticus and pediatric epilepsy are excluded.

DEFINITION OF EPILEPTIC SEIZURES AND EPILEPSY

Epileptic seizures are symptoms of cerebral dysfunction, resulting from an imbalance of excitatory and inhibitory influences within a neuronal aggregate, leading to paroxysmal hypersynchronous discharges of neurons of the cerebral cortex. The clinical manifestations are extremely variable. Signs and symptoms of seizures are determined by the function of the cortical area involved. Epileptic seizures are usually self-limited, lasting a minute or two, and may be followed by a period of postictal cerebral depression manifested clinically as diffuse or localized neurologic deficits.

Not all patients with seizures have epilepsy. Epilepsy is defined as a disorder characterized by 2 or more unprovoked epileptic seizures due to primary brain abnormality. One episode of epileptic seizure, therefore, does not constitute epilepsy.

Epileptic seizures could also be reactive, known as reactive seizures, acute symptomatic seizures, provoked seizures or situation-related seizures.[3] These seizures occur at the time of systemic disturbance or in close temporal association with a documented brain insult. Systemic disturbance causing reactive seizures include electrolyte imbalance (e.g. hypocalcemia, hypomagnesimia, hyponatremia, hypoglycemia or nonketotic hyperglycemia), drug-induced (e.g. amitryptyline or aminophylline)[4] and drug or alcohol withdrawal, toxic insults (e.g. carbon monoxide poisoning or acetylsalicylic acid overdose) and eclampsia. Primary brain insults causing reactive seizures include meningitis/encephalitis, trauma, acute stoke, and tumor. These seizures usually resolve spontaneously, with symptomatic treatment, and/or with treatment of the underlying reversible cause. Patients with reactive seizures should not be labeled as having epilepsy.

ETIOLOGY OF EPILEPSY

The cause of epilepsy is believed to be multifactorial, involving both genetic and acquired factors. There are 3 potential genetic contributions to the appearance of an epileptic disorder: 1) variations among individuals in their susceptibility to epileptic symptoms given a transient epileptogenic insult or to chronic epilepsy given a specific structural lesion are in large part genetically determined; 2) some specific cerebral disturbances that give rise to secondary epilepsies are genetically transmitted, such as

tuberous sclerosis and *phenylketonuria*; and 3) the primary epilepsies reflect genetic impairment in cerebral excitability and synchronization.[5]

Acquired lesions that give rise to secondary epilepsies may be bilateral and so diffuse that seizures are generalized from the start, or the lesions may be localized to smaller areas of the brain, giving rise to partial seizures or generalized seizures with focal features. Common pathological substrates of secondary epilepsies include developmental defects, hippocampal sclerosis, neoplasms and other tumors, and cerebral damage due to trauma, vascular accidents, and infection.

DIAGNOSIS OF EPILEPTIC SEIZURE AND EPILEPSY

As many other neurologic, psychiatric and systemic disorders can have intermittent symptoms, diagnosis of epileptic seizures requires first determining whether the intermittent "ictal" event is in fact epileptic in nature. In order to do so, the physician must be aware of the various phenomenologies of epileptic seizures (Table 1). These phenomena depend on the part of the cerebrum being activated by seizure activity.

Once it is determined that the event was epileptic, the next diagnostic step is to decide whether it is a reactive seizure or the first manifestation of epilepsy. If a reversible and/or treatable cause (e.g. hypocalcemia, brain tumor) is successfully treated and seizure no longer recurs, the patient should not be labeled with the diagnosis of epilepsy. When seizures persist or recur, either because an underlying treatable cause has not been found or treatment of it has not resulted in the remission of seizures, a diagnosis of epilepsy is then made.

Subsequent diagnostic evaluation is aimed at determining what type(s) of seizures are occurring and whether the clinical features constitute a recognized epileptic syndrome. The epileptic seizure classification describes the types of seizures only (Table 2), while the epilepsy syndrome classification describes the syndromes in which seizures occur (Table 3). Epileptic seizures should be classified because it allows appropriate selection of antiepileptic drug (AED) treatment (Table 1) while classification of epilepsy syndromes help to determine the prognosis (e.g. response to treatment) as well as type of investigation.

The International League Against Epilepsy (ILAE) classifies seizures using seizure semiology and EEG findings.[6] In clinical practice, EEG information may not be available or diagnostic (vide infra). Thus one has

Table 1 Seizure Type, Phenomenology and AEDs of Choice

Adapted from ILAE Seizure Classification	Seizure Phenomenology	Potentially Effective AEDs	
Therapy		1st Choice Monotherapy	2nd Choice Monotherapy or as Add-on
Generalized			
Tonic-clonic	loss of consciousness, vocalization, bilateral sustained contraction followed by rhythmic twitching	VPA, CBZ, PHT	LTG, PB, PRM, CZP, CLB
Clonic	bilateral rhythmic twitching/jerking		
Myoclonic	brief, sudden bilaterally synchronous muscle jerks without impaired consciousness	VPA	LTG, CZP, CLB, PB, PRM
Tonic	bilaterally synchronous sustained contractions, with or without impairment of consciousness, often causing patient to fall		LTG, CZP
Atonic	bilaterally synchronous loss of tone / posture of head, limbs and/or body, with or without impairment of consciousness; often result in falls		LTG, CZP
Absences (Typical and Atypical)	lapse of consciousness, at times associated with minor movements such as eye blinking or facial twitching		LTG, ESM, CZP, CLB
Partial			
Simple Partial (SPS):	Consciousness preserved	CBZ, PHT, VPA	LTG, GBP, VGB,
• Motor	• Rhythmic jerking/twitching (clonic) or sustain contraction (tonic) of certain part of body including eye/head (with deviation)		TPM, VGB, PB, PRM, CZP, CLB
• Somatosensory	• Tingling, numbness, pain burning sensation of certain part of the body		LEV

• Visual	• Flashing lights, visual hallucination	
• Autonomic	• Flushing, pallor, epigastric sensations, sweating, changes in pupillary sign, tachycardia	
• Psychic	• hallucination/illusions (visual, auditory or olfactory), disturbance of memory (déjà vu or jamais vu, memory flashback, amnesia), emotional (fear, sadness, pleasure, anger) and others	
Complex Partial	loss of consciousness, may follow a SPS, may be associated with oro-alimentary (e.g. lip smacking, chewing, swallowing) and/or limb (e.g. fumbling, scratching) automatism, and vocalization	
Partial Seizures with Secondary Generalization	Above partial seizure evolving into generalized, clonic or tonic-clonic seizures	
Unclassifiable	Histories from patient and/or witnesses are not detailed or specific enough to allow a differentiation between partial or generalized epileptic seizures; seizure during sleep; no witness	VPA or CBZ

CBZ = Carbamazepine; CZP = Clonazepam; CLB = Clobazem; ESM = Ethosuximide; GBP = Gabapentin; LTG = Lamotrigine; PB = Phenobarbital; PHT = Phenytoin; PRM = Primidone; TPM = Topiramate; VPA = Valproic Acid; VGB = Vigabatrin; LEV = Levetiracetam.

Table 2 International Classification of Epileptic Seizures (abbreviated)

1) Partial (*arising from a focal or local cortical lesion, most commonly the temporal lobe*)
- Simple partial *(no loss of consciousness)*
- Complex partial (loss of consciousness; may start with loss of awareness or may follow a simple partial seizure; may be with or without automatisms, e.g. lip smacking, rubbing hands, walking, running with no recollection)
- Partial evolving to secondary generalized with tonic, tonic-clonic or clonic features

2) Generalized (*with bilateral discharges involving subcortical structures — convulsive or non-convulsive; EEG shows bilateral discharges; consciousness is lost at the onset except in myoclonic seizures; motor features bilateral*)
- Absence (last seconds; +/− minor automatisms)
- Myoclonic (may be simple or multiple jerks, often upper limbs)
- Tonic
- Tonic-clonic
- Clonic
- Atonic *(sudden loss of posture of head, limbs and/or body):*

3) Unclassified (*usually used when an adequate description is not available, e.g. often in seizures from sleep; no eyewitness*)

Adapted from Commission of Classification and Terminology of the International League against Epilepsy (Epilepsia 1981; **22**: 489–501) and Scottish Intercollegiate Guidelines Network (SIGN) on Diagnosis and Management of Epilepsy in Adults (1997).

to depend on good history taking to classify seizures (Table 4).[7] The International Classification of Epilepsies and Epileptic Syndromes[8] divides epileptic disorders into those due to inherited epileptogenic cerebral dysfunction (primary or idiopathic epilepsy) and those due to specific structural abnormalities, which may be genetic (such as tuberous sclerosis) or acquired (secondary or symptomatic epilepsy). For practical purposes, an abbreviated form of the International Classification of Epilepsy Syndromes is recommended (Table 3). Most adults presenting with epileptic seizures have a symptomatic localization-related epilepsy.

In reactive seizures, ictal events are usually generalized. However, they may be partial when pre-existing localized cerebral disturbances make one area of the brain more epileptogenic than others (e.g. in the case of seizures secondary to alcohol withdrawal, this may be partial if the alcoholic patient had a previous cerebral contusion; presence of a previous stroke will make hypoglycemic seizure partial).

Table 3 International Classification of Epileptic Syndromes (abbreviated, pediatric syndromes excluded)

1) Localization-related epilepsies
- Idiopathic *(age-related; common — benign Rolandic Epilepsy, rare — primary reading epilepsy, benign occipital epilepsy)*
- Symptomatic *(most commonly temporal lobe epilepsies, but also frontal parietal, or occipital epilepsies)*

2) Generalized epilepsies
- Idiopathic *(age-related)*
  – juvenile absence epilepsy
  – juvenile myoclonic epilepsy
  – epilepsy with tonic-clonic seizures on awakening
  – others not defined above
  – epilepsies with seizures precipitated by specific modes of activation, e.g. photic stimulation
- Idiopathic or symptomatic
  – *epilepsy with myoclonic absences*
- Symptomatic *(e.g. inherited metabolic or congenital disorders)*

3) Epilepsies undetermined whether focal or generalized

4) Situation-related
Seizures occurring only when there is an acute metabolic or toxic event (e.g. alcohol, drugs, eclampsia, non-ketotic hyperglycemia)

Adapted from Commission of Classification and Terminology of the International League against Epilepsy (Epilepsia 1989; **30**: 389–399) and Scottish Intercollegiate Guidelines Network (SIGN) on Diagnosis and Management of Epilepsy in Adults (1997).

A clear history from the patient and a witness (when possible) gives the most important information leading to a confident diagnosis. Primary care physicians (e.g. general practitioners) and doctors in Accident & Emergency Departments can play a vital role in obtaining detailed witness accounts while the events are still fresh in witnesses' minds.

DIFFERENTIAL DIAGNOSIS

The diagnosis of epilepsy is not always straightforward. On one hand, epilepsy may be under- or misdiagnosed and not treated appropriately. On the other hand, patients with non-epileptic disorders may be wrongly diagnosed as epilepsy and treated with antiepileptic drugs. Among the most common disorders which need to be distinguished from epileptic seizures are vasovagal syncope and psychogenic seizures.

Table 4 Taking a History in Patients With or Suspected of Having a Seizure Disorder

1)	**History of attacks from the patient**
	• frequency of attacks
	• circumstances and trigger factors
	• symptoms before and during the attacks
	• duration of symptoms
	• symptoms following the attack
	• injury, tongue biting and incontinence
2)	**History of attacks from a witness**
	• frequency of attacks
	• detailed description of observations before and during the attacks (including level of responsiveness, motor phenomena, vocalization, color, breathing, pulse)
	• detailed description of behavior following the attacks
3)	**History should include the following diagnostically relevant factors**
	• age
	• sex
	• past medical history, including head injury and febrile convulsions
	• past psychiatric history
	• social history
	• family history
	• alcohol and drug use

Vasovagal syncope is a more common cause of sudden loss of consciousness in patients presenting to a hospital emergency department (40%) than epilepsy (29%).[9] It should be considered when there has been a potential trigger such as fear, pain, sudden standing from squatting or supine posture, micturition, and cough. Its motor, sensory and/or autonomic features may simulate epileptic seizures. For example, myoclonic jerks lasting less than 10 seconds, head turning, automatisms (lipsmacking, chewing), upward deviation of the eyes, vocalization and incontinence may occur in syncope.[10] Sensory symptoms, including visual hallucinations and even out-of-body experiences can also occur. Although there may be a transient throbbing headache after syncope, postictal stupor, prostration or malaise are usually absent.

Some patients with a firm diagnosis of "epilepsy" actually have *psychogenic seizures* (also known as pseudoseizures, non-epileptic seizures or hysterical seizures). The clinical features suggestive of psychogenic seizures are listed in Table 5.[11–14] Studies show that experienced epileptologists are wrong in distinguishing between psychogenic

Table 5 Epileptic versus Psychogenic Seizures

1) **Features suggestive of psychogenic seizures**
   - asymmetrical thrashing movements of the limbs
   - side-to-side head movements (rather than bilaterally symmetrical tonic-clonic movements)
   - pelvic thrusting (especially forward)
   - opisthotonic posturing
   - lack of stereotypic pattern with repeated events
   - talking or screaming throughout the seizure
   - long duration
   - bilateral involvement without impairment of consciousness
   - sudden return to consciousness following a prolonged generalized seizure
   - resistance to eye opening in an otherwise unresponsive patient
   - aversion of the eyes away from the examiner when the lids are elevated
   - recurrent episodes of prolonged unresponsiveness lasting more than a few minutes
   - induced and/or stopped by suggestion
2) **Epileptic phenomena that could be seen in psychogenic seizures**
   - pupillary dilation
   - urinary and fecal incontinence
   - tongue-biting
   - other injuries
   - Babinski responses
3) **Seizures that appear to be "bizarre" but could still be epileptic if the followings are present**
   - turning to a prone position during seizure
   - nocturnal occurrence
   - short ictal duration
   - stereotyped patterns of movements
   - associated with MRI and/or EEG abnormality
4) **Predisposing factors for psychogenic seizures**
   - female sex
   - previous childhood sexual abuse
   - previous psychiatric history
   - sexual maladjustment
   - depression
   - morbid anxiety

and epileptic seizures from videotapes 20% to 30% of the time.[15] On one hand, psychogenic seizures can have features of epilepsy.[16] On the other hand, ictal events that appear not to make neurophysiological sense or have bizarre features, can occasionally be epileptic.[17–20] Patients with both epileptic and psychogenic seizures pose the most difficult diagnostic

dilemma. In general the diagnosis of psychogenic seizures should be considered when an organic disorder is deemed unlikely, and psychological as well as psychiatric assessments provide supportive evidence. The most important information in this regard is documented secondary gain.

Other conditions which may be mistaken for epilepsy include normal phenomena (e.g. déjà vu may be normal), transient ischemic attack (need to differentiate from simple partial seizures), paroxysmal movement disorders (could be mistaken for partial motor seizures), transient global amnesia (amnesia lasting from 30 minutes to a few hours), migrainous aura (a march of less than a minute suggests partial seizures; over several minutes suggests migraine), sleep-related episodes (e.g. hypnic jerks and periodic movement of sleep), cardiac arrhythmias, postural hypotension (e.g. autonomic neuropathy), hyperventilation, and other psychiatric disorders (panic attacks, episodic dyscontrol, dissociative states).

In patients presenting with transient loss of consciousness, the history and physical examination may be sufficient for diagnosis in 85% of patients in whom a diagnosis is established. However, diagnosis can sometimes be difficult and specialist investigation with long-term EEG and/or video monitoring, prolactin assays and cardiac studies may be necessary (vide infra).

INVESTIGATIONS

Electroencephalogram (EEG)

The EEG is the only available investigation to record and evaluate the paroxysmal electrical discharges of cerebral neurons. Properly performed by experienced technicians and carefully studied and interpreted in the cortex of a well-described clinical context by experienced neurologist, it is the most important investigation for epilepsy. It is also harmless and relatively inexpensive.

It is a common misconception that diagnosis of epilepsy can only be made by EEG. Very often clinicians dismissed a diagnosis of epilepsy because EEG did not show epileptiform discharges (spikes or sharp waves). It is important to emphasize that EEG is only diagnostic of epilepsy when ictal epileptiform discharges are recorded, especially if accompanied by behavioral changes (either observed by experienced

medical/paramedical personnel or captured during video-EEG monitoring). However, it is rare to record a seizure in 20–30 minute routine EEG recording. Nevertheless, even without an ictal event, EEG can support a clinical diagnosis of epilepsy if interictal epileptiform discharges were recorded. The pattern and location of interictal epileptiform abnormalities also help to characterize the type of epileptic disorder or epileptic syndrome.

An interval of 20–30 minutes of EEG may record epileptiform discharges in 29–38% of adult patients who have epilepsy.[21] With repeated EEG recording this increases to 59–77%, with little extra gain obtained after five routine EEGs.[22] The chance of recording epileptiform discharges can be enhanced by prolonging each EEG recording period and/or obtaining a sleep recording (especially after prior sleep deprivation). If and when possible, EEG should be obtained shortly after a seizure, as epileptiform discharges are more likely to be found hours or days after an epileptic seizure. It is important to emphasize that epileptiform discharges may be recorded in 1.8–4% of individuals who never have or develop epilepsy.[23] Thus, "not all fits spike and not all spikes fit".

EEG can be used to predict seizure recurrence. Patients with epileptiform discharges on EEG performed within the first few weeks after a first seizure are significantly more likely to have a second seizure if untreated, compared with those showing non-epileptic abnormalities or with a normal EEG. If performed later, the EEG does not appear to be predictive of recurrent seizures.[24,25] Patients with generalized epileptiform EEG abnormalities may have a higher risk of seizure recurrence.[26] Such findings should be considered when counseling the patient who has initially declined treatment.

When a clinical diagnosis of epilepsy is doubtful and routine EEG is non-diagnostic or supportive, it may be necessary to admit the patient for long-term EEG recording with video monitoring. The purposes of this are to increase the yield of recording interictal epileptiform discharges and to capture the habitual seizures during the period of monitoring. The recording of habitual seizures help to differentiate psychogenic from epileptic seizures, characterize different seizure types so that AED treatment can be optimized, and identify suitable candidates for epilepsy surgery.[27] In selected patients, intracranial EEG monitoring is necessary to define the location and extent of epileptogenic zone before planning for further definitive epilepsy surgery.[28,29]

Structural Neuro-imaging

Computed tomography (CT) scanning will identify a focal lesion in a clinically relevant proportion of patients presenting with partial and generalized tonic-clonic seizures, including a significant proportion of patients with no focal abnormalities on neurological examination and EEG. Some of the abnormalities identified by CT scanning may be treatable by surgery. However, CT scanning may fail to show a focal abnormality in a significant proportion of patients with focally abnormal neurological examination and/or EEG.[30,31]

Magnetic resonance imaging (MRI) of the brain detects lesions that are not detected by CT scanning. The number of such lesions detected has increased as sophistication of MRI techniques employed and the quality of magnetic resonance images have improved. MRI abnormalities that can be missed on CT scan include mesial temporal sclerosis (hippocampal atrophy and signal change), cortical dysgenesis, tumors and cavernous hemangiomas. A significant proportion of patients with these lesions will be potentially treatable by surgery.[32,33] Therefore, if MRI is available and affordable, this will be the neuro-imaging of choice in the evaluation of clinically focal epilepsy.

As seizures of idiopathic generalized epilepsy rarely occur for the first time after age 25 years, patients presenting with first seizure over the age of 25 years will nearly always have a localization-related epilepsy. Thus brain imaging should always be performed in these patients to identify the cause of epilepsy and to elucidate potentially treatable lesions. MRI brain is preferable to CT scan for the reasons explained above. In patients under the age of 25 years, brain imaging is unnecessary if a firm diagnosis of an "idiopathic generalized epilepsy syndrome" has been made on the basis of the clinical history and EEG findings. If a firm diagnosis of idiopathic generalized epilepsy cannot be made, brain imaging is necessary.

Neuro-imaging, especially MRI, should be performed in patients with chronic epilepsy to determine the cause of epilepsy as well as to identify suitable surgical candidates. In a local study, approximately 3 quarters of patients with epilepsy for more than 3 years had a focal structural abnormality demonstrated on MRI. Of these, about 75% had either mesial temporal sclerosis or other temporal lobe abnormalities that might potentially be treated by surgery.[34]

Other Investigations

In patients presenting with first seizure, investigations such as full blood counts, renal function tests, liver function tests, serum glucose, calcium, magnesium and resting ECG are indicated. Depending on the clinical presentation, 24-hour ECG holter monitoring, drug toxicology, and cerebrospinal fluid examination may also be warranted. Tests like serum calcium and magnesium should not be routinely performed in patients who are known to have epilepsy admitted for seizure recurrence due to poor compliance. Instead, these patients should have blood AED levels taken routinely (vide infra).

Serum prolactin determination is useful as the levels are elevated following generalized, complex partial, and some simple partial seizures. However, a normal serum prolactin level during a seizure without impairment of consciousness does not rule out epilepsy. In patients being evaluated for surgery, neuropsychological assessment (useful for characterizing functional deficits for psychosocial prognosis and rehabilitation), intracarotid amobarbital procedure (to lateralize language area and assess memory function) and ictal Single Photon Emission Computed Tomography (study blood flow during seizures and correlate with EEG and MRI localization of epileptogenic zone) would be useful.[35]

DRUG TREATMENT OF SEIZURES AND EPILEPSY

The main aim of treating epilepsy is to render the patient completely seizure-free with appropriate use of AEDs. If seizures could not be completely suppressed with AEDs, the aim is to reduce seizure frequency and severity without unacceptable medication side effects.

Ideally all patients with definite or suspected epilepsy should be referred to a neurologist for further investigation and treatment, except in exceptional circumstances, e.g. elderly incapacitated patients with stroke. However, general physicians and other non-neurological internists could initiate the initial treatment of epilepsy, if they are confident of the diagnosis, have easy access to EEG and neuro-imaging information and are of the opinion that the patient needs treatment. However, when seizures are not adequately controlled within a reasonable period of time (weeks for frequent seizures or a few months for less frequent seizures), these patients should be referred for neurological consult.

When to Start Treatment

The issue of whether to commence AED treatment after a first seizure remains controversial. Studies of selected hospital populations report seizure recurrence rates after a first seizure of around 20–30% after two years.[36,37] After five years the rate may be as high as 30% in idiopathic seizures and 45% in patients with remote symptomatic seizures. However, three studies in general populations in which patients were seen within a week of the first seizure give much higher estimates of seizure recurrence of 67–80% after first untreated seizure.[38–40] Treatment of patients with tonic-clonic seizures without a clear metabolic or structural cause reduces risk of seizure recurrence from 51 to 25%.[41]

Patients with first unprovoked generalized tonic-clonic seizures should be treated with an AED if history reveals that he already has myoclonic and/or absence seizures, which was previously undiagnosed. The decision to treat simple and complex partial seizures will depend on the seizure frequency, severity (including the impact of seizures on work, studies and quality of life), and patient preference. Seizures arising from alcohol withdrawal, metabolic or drug-related causes should not be treated with AEDs. Treatment may be considered if there are recurrences suggestive of epilepsy. Patients should not be treated if there is uncertainty about the diagnosis.

The risk of a second seizure rises when the EEG is epileptogenic, the neurologic examination is abnormal, or the MRI reveals a structural abnormality, with the risk typically $>50\%$ in these patients. In patients with partial seizures and abnormal EEG and imaging studies, the risk rises further to near 80%. After two unprovoked seizures, the risk of a third unprovoked seizure is 73%.[42] These individuals should receive AED therapy. Patients with single seizure, simple partial seizure, and seizures resulting from medications, metabolic disturbances or alcohol withdrawal may not need treatment.

Which Antiepileptic Drug (AED)?

Choice of drugs depends on drug (efficacy, side effect profile, ease of use, availability and affordability) as well as patient factors.[43]

Efficacy of an AED refers to the effectiveness of the AED in preventing or reducing the recurrence of a particular seizure type (Table 1). For seizures that cannot be classified based on history, sodium valproate is

the drug of first choice for those presenting under the age of 25 years, and carbamazepine is the choice for those presenting over the age of 25 years.

Potential AED side effects and their appearance not only affect the physicians' choice but also determine the acceptance of the drug by the patient. For example, potential teratogenicity may make physicians avoid using valproate in female patients who are planning to have children. Not all patients will develop side effects and not all side effects are unacceptable. For example, weight gain from valproate and hirsutism from phenytoin are probably less acceptable in female patients than in males. Preparation of AEDs also affect the severity of dose related side effect. AED in control-release formulation is preferable as the occurrence of dose related side effect is less likely and higher dosages are possible.

One of the most important factors in deciding which drug to prescribe is the cost of AEDs and their affordability to patients. Although newer AEDs might be more effective and less toxic than older AEDs, their usage in this part of the world is limited by their relatively high cost. In Singapore, older AEDs such as phenytoin, phenobarbitone, carbamazepine and valproate are subsidized by our government whereas newer AEDs such as lamotrigine, gabapentin, topiramate and vigabatrin are not. It has been estimated that the cost of having one newer AED in a 2-drug combination is 4 times the cost of 2-drug combinations without a newer AED.[44] Patients who require 2 or more AEDs are likely to be medically refractory, unemployed or under-employed, and dependent on other family members for financial support. Thus, taking newer AED on a long-term basis can be quite costly for many patients and their families. Newer AEDs are usually used as add-on treatment for patients with refractory partial seizures not responding to conventional AEDs.

Ease of use of an AED affects the patient's compliance to medication. For example, AEDs that have a long half-life and can be given once per day are preferable. In addition, the availability of other treatment options such as epilepsy surgery might affect the pattern of AED usage. If presurgical evaluation facilities and surgical expertise are available, candidates with good surgical potential are likely to be offered surgery as an alternative to long-term AEDs.

Monotherapy

Whichever AED is chosen, it is important to maintain patients on monotherapy as compliance is better, side effects are less and there is no

problem of drug-to-drug interaction. Even in a tertiary referral hospital where epilepsy patients are expected to have difficult-to-control seizures, monotherapy is achievable in the majority of patients. Approximately two-thirds of epilepsy patients seen at the Singapore General Hospital were on monotherapy.[43] Many patients being referred for management of "failed" monotherapy actually had not been given maximum tolerated doses of AED. Some patients on polytherapy could also be converted to monotherapy. Regardless of the blood level, by increasing the dosage of the same AED gradually to the maximum tolerated dose, seizures could be controlled in many patients. It is important to emphasize that there is no standard dose for any AED. Every patient has his/her own necessary dose.

Screening laboratory studies (e.g. baseline hematology, liver function test and serum electrolytes) should be obtained before initiation of AED treatment. These studies can identify patients with special risk factors that could influence drug selection, and can be used as a baseline for future monitoring of hematological, biochemical and liver functions.

AED Serum Level Monitoring

The 2 main indications for AED serum level monitoring are assessment for compliance in patients who do not respond to medication and assessment for drug toxicity. Serum level monitoring is not necessary for adjustment of AED doses in otherwise healthy and asymptomatic patients. Exceptions are those patients who do not complain (e.g. mentally retarded) of side effects and when phenytoin dose needs to be increased. The latter is because phenytoin has variable rates of hepatic metabolism, and because a small increase in dose can result in a large increase in serum level (saturation kinetics).

Published "therapeutic ranges" should only be used as approximate guides. While some patients achieve seizure control at serum levels below the therapeutic range, others require and tolerate a serum level above the upper limit of the "therapeutic range". Serum levels may be useful when there are possible interactions between AEDs and/or other drugs. Carbamazepine, phenytoin and barbiturates are hepatic enzyme inducers while sodium valproate is an enzyme inhibitor. Changes in the dose of one may alter the serum levels of others. Serum level estimations should not be done so soon after a change in dose as the new plateau level may not have been reached.

Alternative Monotherapy and Polytherapy

Patients who fail one conventional AED because of inadequate efficacy or unacceptable side effects could respond well to another conventional AED given as monotherapy. Before considering alternative AED(s), one should review the patient's diagnosis, current AED dosage, and compliance. If the patient has been given the maximum tolerated dose of one AED and if his diagnosis and compliance are not doubted, an alternative monotherapy should be tried before attempting combination therapy. This is usually done by introducing a second AED specific for the patient's seizure type. The side effect profile of the second AED must be acceptable. Once the patient's seizures are adequately controlled, the first AED can be gradually withdrawn. If seizures recur with repeated attempts to withdraw the first AED, it would be better letting the patient continue with the combination therapy.

It needs to be emphasized that even when combination therapy is unavoidable, the patient should be given the minimum number and dosage of AEDs that are required to achieve adequate seizure control without unacceptable medication side effects. There is no guidelines as to which combination is superior to another. In patients with refractory partial seizures, a large meta-analysis of 29 randomized placebo-controlled trials on 4091 patients showed no significant differences in the efficacy or tolerability of lamotrigine, gabapentin, topiramate, and vigabatrin as add-on treatment.[45] Patients with absence seizures not adequately controlled by valproate can have lamotrigine added. For myoclonic and atonic seizures, lamotrigine or clonazepam can be added to valproate. For primary generalized tonic-clonic seizures, lamotrigine would be a good add-on AED.[46]

Epilepsy Surgery

The probability of responding to another AED progressively declines with each AED tried and failed. The percentage of successful seizure control falls from approximately 70% with the first AED used to about 40% with the second AED.[47] When more than two AEDs have failed, the probability of achieving complete or nearly complete seizure freedom on a long-term basis is nil.[48–51] Seizure control in this very small fraction of patients is often short-lived and no longer than a year. In studies using investigational

AEDs, it is very uncommon for patients to become seizure-free for more than a year. Furthermore, combination AED therapy is associated with a higher risk of side effects. For many of these patients, especially those with a focal structural pathology (lesional epilepsy), surgery presents a much better treatment option with a probability of excellent outcome at 75–95% and risk of major complications at 1% or lower.[52,53] In addition, another 10% of patients can be expected to have at least 75% reduction in their seizure frequency. Whether the more modest outcome in this group of patients will result in meaningful improvement in their quality of life depends on the occupational, social and cultural expectations for each patient.

Patients should be considered for epilepsy surgery evaluation if their epilepsy is intractable despite having: 1) a correct seizure-type and epilepsy syndrome diagnosis; 2) optimal use of at least two AEDs appropriate for the seizure-type; 3) good compliance; and 4) no major medical or psychosocial disturbance that interferes with seizure control. Seizure control or improvement that is likely to reverse disability or improve quality of life is the other important consideration.

Patients who do not respond to different monotherapy and have tried various combination therapies must be referred for video-EEG monitoring to confirm that seizures are indeed epileptic in nature, as well as to identify potential epilepsy surgery candidates. Localized resections are particularly effective in temporal lobe epilepsy. Experience at SGH showed that 80% of patients with medically refractory temporal lobe epilepsy either became seizure-free or had more than 90% reduction in seizure frequency after temporal lobectomy.[54] AED could be stopped in about 40% and reduced in another 40% of patients after surgery. Focal resection can also be beneficial for many forms of neocortical epilepsy due to well-circumscribed lesions. Corpus callosum sections can abolish disabling drop attacks, and hemispherectomies or large multilobar resections can yield gratifying results in patients, usually infants and small children, with catastrophic secondary generalized epilepsies. In surgically remediable epilepsies, early intervention provides the best opportunity to avoid psychosocial disability.[55]

Vagus Nerve Stimulator (VNS)

This consists of leads wrapped around the vagus nerve in the neck tunneled and connected to a pulse generator embedded in a subcutaneous

pocket created infra-clavicularly. The VNS has been shown to effect a mean seizure reduction of about 25% to almost 30% in short-term follow-up and over 40%[56] in long-term follow-up of adults and children with intractable partial onset seizures. More recent reports have suggested efficacy in generalized seizures[57] in adults and in children.[58] It has also been shown to be efficacious in patients with Lennox Gastaut Syndrome.[59] Its main indication currently is in patients with intractable partial onset or generalized seizures where AEDs either do not afford adequate control or inflict intolerable side-effects.

Discontinuing Drug Therapy

The Medical Research Council Antiepileptic Drug Withdrawal Study Group conducted a large multicenter randomized trial of continued antiepileptic treatment versus slow withdrawal in adults and children with epilepsy who had been seizure-free for at least two years.[60] Patients who had their AED therapy slowly withdrawn were found to be more likely to have a seizure recurrence over the next 12 months than patients on continued therapy.

Seizures are more likely to recur after AED discontinuation if a patient has a history of generalized tonic-clonic seizures, is taking more than one AED at time of withdrawal, is experiencing one or more seizures after the start of treatment, has a history of myoclonic seizures, has partial seizures that have "never generalized" and has epilepsy for ten years or more. Patients in whom seizure recurrence is less likely are those who have been seizure-free for five or more years.[61]

Discussion of whether to withdraw AEDs should take into account the patient's need to work and/or study, fear of seizures, attitude to prolonged AED therapy and benefits and risks to both patient and society from a recurrent seizure. Withdrawal can be attempted if patients have been seizure-free for 2–5 years.

REFERENCES

1. Hauser WA, Hesdorffer DC, *Epilepsy: Frequency, Causes and Consequences*, Demos Pr, New York, pp. 378, 1990.

2. Lee WL, Low PS, Murugasu B, Rajan U, Epidemiology of epilepsy among Singapore children, *Neurol J Southeast Asia* **2**:31–35, 1997.

3. Hauser WA, Annegers JF, Epidemiology of acute symptomatic seizures, in Engel JJ, Pedley TA (eds.), *Epilepsy: A Comprehensive Textbook*, Lippincott–Raven Publishers, Philadelphia, pp. 87–91, 1997.

4. Jain KK, Drug-induced seizures, in Gilman S, Goldstein GW, Waxman SG (eds.), *Neurobase*, 2nd ed., Arbor Publishing, San Diego, 1999.

5. Engel J, Epilepsy, in Gilman S, Goldstein GW, Waxman SG (eds.), *Neurobase*, 2nd ed., Arbor Publishing, San Diego, 1999.

6. Commission on Classification and Terminology of the International League Against Epilepsy, Proposal for revised clinical and electroencephalographic classification of epileptic seizures, *Epilepsia* **22**:489–501, 1981.

7. Scottish Intercollegiate Guidelines Network (SIGN), Diagnosis and management of epilepsy in adults, 1997.

8. Commission on Classification and Terminology of the International League Against Epilepsy, Proposal for revised classification of epilepsies and epileptic syndromes, *Epilepsia* **30**:389–399, 1989.

9. Day SC, Cook EF, Funkenstein H, Goldman L, Evaluation and outcome of emergency room patients with transient loss of consciousness, *Am J Med* **73**:15–23, 1982.

10. Lempert T, Bauer M, Schmidt D, Syncope: a videometric analysis of 56 episodes of transient cerebral hypoxia, *Ann Neurol* **36**:233–237, 1994.

11. Gates JR, Ramani V, Whalen S, Loewenson R, Ictal characteristics of pseudoseizures, *Arch Neurol* **42**:1183–1187, 1985.

12. Meierkord H, Will B, Fish D, Shorvon S, The clinical features and prognosis of pseudoseizures diagnosed using video-EEG telemetry, *Neurology* **41**:1643–1646, 1991.

13. Betts T, Boden S, Diagnosis, management and prognosis of a group of 128 patients with non-epileptic attack disorder. Part II. Previous childhood sexual abuse in the aetiology of these disorders, *Seizure* **1**:27–32, 1992.

14. Boon PA, Williamson PD, The diagnosis of pseudoseizures, *Clin Neurol Neurosurg* **95**(1):1–8, 1993.

15. King DW, Gallagher BB, Murvin AJ, Smith DB, Marcus DJ, Hartlage LC, Pseudoseizures: diagnostic evaluation, *Neurology* **32**:18–23, 1982.

16. Lesser 1985 (pg.5)

17. Williamson PD, Spencer DD, Spencer SS, Novelly RA, Mattson RH, Complex partial seizures of frontal lobe origin, *Ann Neurol* **18**:497–504, 1985.

18. Sussman NM, Jackel RA, Kaplan LR, Harner RN, Bicycling movements as a manifestation of complex partial seizures of temporal lobe origin, *Epilepsia* **30**:527–531, 1989.

19. Saygi S, Katz A, Marks DA, Spencer SS, Frontal lobe partial seizures and psychogenic seizures: comparison of clinical and ictal characteristics, *Neurology* **42**:1274–1277, 1992.

20. So NK, Mesial frontal epilepsy, *Epilepsia* **39** (Suppl 4):S49–61, 1998.

21. Doppelbauer A, Zeitlhofer J, Zifko U, Baumgartner C, Mayr N, Deecke L, Occurrence of epileptiform activity in the routine EEG of epileptic patients, *Acta Neurol Scand* **87**:345–352, 1993.

22. Salinsky M, Kanter R, Dasheiff RM, Effectiveness of multiple EEGs in supporting the diagnosis of epilepsy: an operational curve, *Epilepsia* **28**:331–334, 1987.

23. Zivin L, Marsan CA, Incidence and prognostic significance of epileptiform activity in the EEG of non-epileptic subjects, *Brain* **91**:751–778, 1968.

24. Van Donselaar CA, Schimsheimer RJ, Geerts AT, Declerck AC, Value of the electroencephalogram in adult patients with untreated idiopathic first seizures, *Arch Neurol* **49**:231–237, 1992.

25. Hopkins A, Garman A, Clarke C, The first seizure in adult life. Value of clinical features, electroencephalography, and computerised tomographic scanning in prediction of seizure recurrence, *Lancet* **1**:721–726, 1988.

26. Hauser WA, Anderson VE, Loewenson RB, McRoberts SM, Seizure recurrence after a first unprovoked seizure, *N Engl J Med* **307**:522–528, 1982.

27. American Academy of Neurology, Assessment: intensive EEG/video monitoring for epilepsy, *Neurology* **39**:1101–1102, 1989.

28. Lim SH, Pillay P, Lüders HO, Boenigk H, Use of intracranial neurophysiologic recording techniques in the evaluation for epilepsy surgery in children, *Singapore Med J* **33**(2):131–138, 1992.

29. Pillay PK, Lim SH, Chee M, Awad I, Luders H, Epilepsy surgery in children and adults, *Ann Acad Med Singapore* **22**(3 Suppl):501–512, 1993.

30. Ramirez–Lassepas M, Cipolle RJ, Morillo LR, Gumnit RJ, Value of computed tomographic scan in the evaluation of adult patients after their first seizure, *Ann Neurol* **15**:536–543, 1984.

31. Schoenenberger RA, Heim SM, Indication for computed tomography of the brain in patients with first uncomplicated generalised seizure, *BMJ* **309**:986–989, 1994.

32. Kuzniecky RI, Jackson GD (eds.), *Magnetic resonance in epilepsy*, Raven Press, New York, 1995.

33. Cascino GD, Jack Crjr (eds.), Neuroimaging, in *Epilepsy: Principles & Practice*, Butterworth–Heinemann, Newton, MA, 1996.

34. Tan EK, Lim SH, Lim W, Tan KP, Magnetic resonance imaging in patients with chronic partial epilepsy: identification of potential epilepsy surgery candidates, *Neurol J Southeast Asia* **2**:37–43, 1997.

35. Goh SWA, Lim SH, Post-ictal single photon emission computed tomography (SPECT) in lateralization of epileptogenic zone, *Neurol J Southeast Asia* **2**:110, 1997.

36. Cleland PG, Mosquera I, Steward WP, Foster JB, Prognosis of isolated seizures in adult life, *BMJ* **283**: 1364, 1981.

37. Hauser WA, Rich SS, Annegers JF, Anderson VE, Seizure recurrence after a first unprovoked seizure: an extended follow-up, *Neurology* **40**:1163–1170, 1990.

38. Hart YM, Sander JWA, Johnson AL, Shorvon SD, National general practice study of epilepsy: recurrence after a first seizure, *Lancet* **336**:1271–1274, 1990.

39. Goodridge DM, Shorvon SD, Epileptic seizures in a population of 6000. II: Treatment and prognosis, *BMJ* **287**:645–647, 1983.

40. Elwes RD, Chesterman P, Reynolds EH, Prognosis after a first untreated tonic-clonic seizure, *Lancet* **2**:752–753, 1985.

41. First Seizure Trial Group, Randomised clinical trial on the efficacy of antiepileptic drugs in reducing the risk of relapse after a first unprovoked tonic-clonic seizure, *Neurology* **43**:478–483, 1993.

42. Hauser WA, Rich SS, Lee JR, Annegers JF, Anderson VE, Risk of recurrent seizures after two unprovoked seizures, *N Engl J Med* **338**(7):429–434, 1998.

43. Lim SH, Tan EK, Chen CLH, Pattern of anti-epileptic drug usage in a tertiary referral hospital in Singapore, *Neurol J Southeast Asia* **2**:77–85, 1997.

44. Tan EK, Lim SH, Chen CLH, Cost of drug and surgical treatment for medically refractory epilepsy in Singapore, *Neurol J Southeast Asia* **1**:76, 1996.

45. Marson AG, Kadir ZA, Hutton JL, Chadwick DW, The new antiepileptic drugs: a systematic review of their efficacy and tolerability, *Epilepsia* **38**:859–880, 1997.

46. Brodie MJ, Yuen AW, Lamotrigine substitution study: evidence for synergism with sodium valproate, *Epilepsy Res* **26**:423–432, 1997.

47. Schmidt D, Richter K, Alternative single antiepileptic drug therapy for refractory epilepsy, *Ann Neurol* **19**:85–87, 1985.

48. Dasheiff R, McNamara D, Dickinson L, Efficacy of second line antiepileptic drugs in the treatment of patients with medically refractive complex partial seizures, *Epilepsia* **27**:124–127, 1986.

49. Ojemann L, Dodrill C, Natural history of drug-resistant seizures: clinical aspects, *Epilepsy Res* (Suppl 5):13–17, 1992.

50. Hermanns G, Noachtar S, Tuxhorn I, *et al*., Systematic testing of medical intractability for carbamazepine, phenytoin, and phenobarbital or primidone in monotherapy for patients considered for surgery, *Epilepsia* **37**:675–679, 1996.

51. Camfield P, Camfield C, Antiepileptic drug therapy: When is epilepsy truly intractable, *Epilepsia* **37**(Suppl 7):S60–S65, 1996.

52. So E, Radhakrishnan K, Silbert P, *et al*., Assessing changes over time in temporal lobectomy: Outcome by scoring seizure frequency, *Epilepsy Res* **27**:119–125, 1997.

53. Radhakrishnan K, So E, Silbert P, *et al*., Predictors of outcome of anterior temporal lobectomy for intractable epilepsy. A multivariate study, *Neurology* **51**:465–471, 1998.

54. Lim SH, Pillay P, Effect of temporal lobectomy on seizure frequency and antiepileptic drug requirement in patients with medically refractory focal epilepsy, *Neurol J Southeast Asia* **2**:110–111, 1997.

55. Lüders HO (ed.), *Epilepsy Surgery*, 2000 (in press).

56. Morris GL, Mueller WM, Long-term treatment with vagus nerve stimulation in patients with refractory epilepsy. The Vagus Nerve Stimulation Study Group E01-E05, *Neurology* **53**(8):1731–1735, 1999.

57. Labar D, Murphy J, Tecoma E, Vagus nerve stimulation for medication-resistant generalized epilepsy. E04 VNS Study Group, *Neurology* **52**(7):1510–1512, 1999.

58. Pardwardhan RV, Stong B, Bebin EM, Mathiesen J, Grabb PA, Efficacy of vagal nerve stimulator in children with medically refractory epilepsy, *Neurosurgery* **47**(6):1353–1358, 2000.

59. Ben–Menachem E, Hellstrom K, Waldton C, Augustinsson LE, Evaluation of refractory epilepsy treated with vagus nerve stimulation for up to 5 years, *Neurology* **52**(6):1265–1267, 1999.

60. Medical Research Council Antiepileptic Drug Withdrawal Study Group, Randomised study of antiepileptic drug withdrawal in patients in remission, *Lancet* **337**:1175–1180, 1991.

61. Medical Research Council Antiepileptic Drug Withdrawal Study Group, Prognostic index for recurrence of seizures after remission of epilepsy, *BMJ* **306**:1374–1378, 1993.

SUGGESTED READING

Engel JJ, Pedley TA (eds.), *Epilepsy: A Comprehensive Textbook*, Lippincott–Raven Publishers, Philadelphia, 1997.

This page intentionally left blank

60

Neuromuscular Disorders

Lo Yew Long and Pavanni Ratnagopal

INTRODUCTION

The peripheral nervous system motor pathway consists of the anterior horn cell, motor roots, plexi, peripheral nerve, neuromuscular junction and muscle. This section will discuss conditions involving each level. In general, neuromuscular diseases have predominant lower motor neuron affectation, but it should be understood that many have concomitant upper motor or central nervous system involvement.

Some of these conditions will be elaborated in other sections of this publication.

APPROACH

The accurate localization of the levels of lower motor neurone involvement requires a good history, thorough physical examination and appropriate investigations. The presence of cortical features should alert one to the presence of an upper motor lesion and bladder/bowel disturbances and a sensory level often suggest cord pathology.

1) *History*

The tempo of onset of signs and symptoms often lends a clue in finding the underlying etiology. Sudden onset of deficits often point to a vascular mechanism, while a more gradual onset over some hours may signify an inflammatory or infective process. Metabolic process like hypokalemic periodic paralysis often present with rapid onset of weakness over hours as well. More protracted presentation may point to neoplastic or degenerative processes. Although it is frequently assumed that neuromuscular disorders often present with distal patterns of weakness, some conditions, e.g. acquired demyelinating polyneuropathies, can show proximal loss of power.[1]

The presence of cramps is often non-specific and can be a result of anterior horn cell, peripheral nerve or muscle disorders. They must be distinguished from clonus which is a upper motor feature. Similarly, pain can also be due to thalamic sensory stroke syndromes. Neuralgic or root pain often conforms to a peripheral nerve or root distribution, in contrast to plexus lesions, which can have a more diffuse pattern. Although the bulk of plexus lesions are of traumatic and inflammatory origin, they can have surprisingly few sensory symptoms. A history of preceeding trauma, infections, vaccinations and familial tendencies is often helpful in the management of inflammatory plexopathies, as well as neuropathies in general. Numbness is a frequent complaint that should not be taken to mean hypoesthesia. Varied sensations ranging from pain, dysthesia, and hyperesthesia can be interpreted as numbness by the patient.

Autonomic symptoms should not be forgotten in the assessment of the peripheral nervous system.[2] Postural giddiness, dry mouth, impotence and loss of sphincter control must be elicited.

The importance of a well-documented drug history, past medical history, especially of endocrine conditions, as well as family history, is often helpful in establishing the etiology.

Finally, a systemic review should probe the presence of cardiorespiratory symptoms that may have an underlying neuromuscular etiology.

2) *Physical Examination*

Observation of general appearance should be done systematically to detect any Cushingnoid appearance, cachexia or syndromic facies. Examination of pupils, speech, cranial nerve abnormalities and

fundoscopy can be instituted as appropriate. The presence of ptosis is often an important lead to extend the examination to assess fatigability of various muscle groups. Bulbar palsy and tongue fibrillation may suggest an anterior horn cell disorder.

Peripheral inspection must include checking for muscle atrophy, trophic changes, burns, deformities, fasciculations and involuntary movements.

Most peripheral nerve disorders have diminished tone in the limbs when tested. However, it is to be borne in mind the occurrence of underlying cord or upper motor lesions may mask this feature. Although the most peripheral motor neuropathies manifest distal weakness, some acquired forms of polyradiculoneuropathies can present with predominantly proximal distribution of weakness. The hallmark physical finding of root lesions is the demonstration of weakness in muscles from different peripheral nerve innervation sharing the same root levels. Reflexes corresponding to the affected root levels can also be diminished. Lesions involving the plexus are often difficult to differentiate from multiple root or peripheral nerve lesions, particularly if they are mild. As a rule, plexopathies often but not invariably have sensory involvement of a less well defined distribution as would be expected of radicular lesions. It may also be difficult to separate root from plexus pathologies distinctively, and both can sometimes coexist. In such situations, electrophysiological techniques are often required, as an extension of clinical testing, in their localization.

The complete examination should also include testing joint position sense, vibration, gait, cerebellar function and other systems.

3) *Investigations*

It is important that the clinician understands the diagnostic capabilities and limitations of various neurophysiological investigations. It would be inappropriate to address cord abnormalities with nerve conduction and electromyography, which are contributory to the localization of peripheral nerve, plexus and root lesions. The astute attending neurophysiologist or neurologist would assess each case clinically and tailor relevant investigations to the patient's needs. Electrophysiological studies play a complementary role to neuroimaging by determining if structural lesions seen with CT scan or MRI are of any functional relevance. The findings of active denervation and axon loss are usually indicative of a more severe lesion as

opposed to findings of focal demyelination. Moreover, in certain non-compressive or inflammatory pathologies, imaging for structural lesions may be normal despite evidence of denervation elicited with neurophysiological techniques.

Conventional nerve conduction studies are useful in confirming peripheral sensory or motor neuropathies as well as mononeuritic entrapment neuropathies. They are often inadequate in addressing proximally-situated lesions in the plexus and less conventional nerve conductions are often required for more comprehensive localization. Mixed nerve conduction studies may be useful in these situation.[3] The finding of normal sensory conduction in the presence of sensory root symptoms localizes the lesion proximal to the dorsal root ganglion (preganglionic).[4]

Needle electromyography provides direct evidence of denervation in muscles sampled, each of which have fairly unchanged levels of root innervation. Hence, it is mandatory in all cases where localization of root or plexus pathologies are suspected. Electromyographic sampling is also helpful in differentiating a lesion undergoing active denervation from one of longstanding duration, as well as demonstrating unique changes in radiation induced nerve injuries.[5,6]

For conditions with co-existing peripheral nerve and spinal cord pathologies, transcranial magnetic stimulation and somatosensory evoked responses are useful diagnostic adjuncts. Delays in central motor conduction time can be demonstrated with magnetic stimulation and is electrophysiological evidence of corticospinal tract dysfunction. Similarly, conduction abnormalities pertaining to the somatosensory evoked response pathways addresses dysfunction in the dorsal columns.

The techniques of repetitive nerve stimulation and single-fibre electromyography are highly specialized investigations for assessing neuromuscular transmission, particularly if it is performed close to the site of maximal weakness.[7] However, these sensitive tests are not specific for myasthenia gravis and may be abnormal in other conditions with neuromuscular junction defects.

The autonomic nervous system is often involved in peripheral neuropathic conditions and various indirect autonomic function tests are increasingly becoming standard test panels. The sympathetic skin response addresses the sudomotor autonomic component while

vasomotor reflex testings assesses autonomic reactivity in the vascular system.

Muscle and nerve biopsy is sometimes necessary if less invasive investigations prove unfruitful for diagnosis. They are sometimes helpful in the exclusion of rarer conditions with pathognomonic histological features.

This systematic discussion outlines the diagnostic approach followed by neurologists when faced with a likely neuromuscular disorder. A reasonable list of differentials can be considered according to infective, neoplastic, toxic, metabolic, autoimmune, degenerative, vascular, iatrogenic or congenital etiologies.

ANTERIOR HORN CELL SYNDROMES

Motor Neuron Diseases

The anterior horn cell syndromes, more commonly known as motor neurone disease (MND), were first recognized as distinct clinical entities by Aran's report in 1850 of 11 patients with "progressive muscular atrophy". There is no completely satisfactory classification but a working group of the World Federation of Neurology in 1990 in El Escorial, Spain was proposed, shown in Table 1.

Epidemiology

The incidence of MND is 1–2 per 100 000 and the prevalence is 4–6 per 100 000 in most parts of the world.

Clinical picture

Weakness is the most common presenting complaint and is usually painless. The atrophy may be noticed even before functional loss. The weakness may be bulbar or pseudobulbar or limb involvement initially but will eventually become generalized. It is usually a mixed picture of upper and lower motor neurone involvement. Fasciculations, previously thought to be the hallmark of the disease, may be fairly florid initially but disappears or subsides with progression.[8]

No actual sensory deficit and sensory symptoms are not infrequently reported by patients. Dementia is not a common feature in amyotrophic

Table 1 Diagnostic Criteria for MND (ALS)

The diagnosis of ALS requires the presence of:
- LMN signs (including EMG features in clinically normal muscles)
- UMN signs

Progression of the disorder

Diagnostic categories:
- Definite ALS: UMN signs in 3 regions
  Probable ALS: UMN plus LMN signs in two regions with UMN signs rostral to LMN signs
  Possible ALS: UMN plus LMN signs in one region, or UMN signs in two or three regions, such as monomelic ALS, progressive bulbar palsy, and primary lateral sclerosis
  Suspected ALS: LMN signs in two or three regions, such as in progressive muscular atrophy, and other motor syndromes

The diagnosis of ALS requires the absence of:
- Sensory signs and sphincter disturbances
- Visual disturbances
- Autonomic dysfunction
- Parkinson's disease
- Alzheimer type dementia

The diagnosis of ALS is supported by:
- Fasciculations in one or more regions
- Neurogenic change in EMG studies
- Normal motor and sensory nerve conduction
- Absence of conduction block

lateral sclerosis (ALS) and if found, should raise the possibility of the Western Pacific types like those from the Marianas and Kii Peninsular. ALS from Guam also can be considered and is attributed to the cycad nut, which is a known neurotoxin.

Diagnostic criteria

The World Federation's criteria set at El Escorial in 1990 were based on clinical evidence and do not take into account the vagaries of clinical practice.[9]

Pathology

The essential pathologic feature is a loss of neurons in the ventral horn cells and degeneration and atrophy of the remaining ones.

Table 2 Differential Diagnosis of MND

Cervical Myelopathy
Autoimmune neuropathies
Thyrotoxicosis
Diabetic amyotrophy
Radiation induced neurogenic disorders
Post poliomyelitis progressive muscular atrophy
Myopathies
Multifocal motor neuropathy

Investigations

The purpose of investigations is to exclude other diagnosis, and to support the diagnosis of MND.

There are no biochemical or pathological markers of the disease. Nerve conduction tests and electromyography help support the diagnosis.

Prognosis

The median survival for all sporadic MND patients is about 3–5years from onset of symptoms.

Management strategies

Riluzole, a glutaminergic modulator, has been tried but is unsuccessful.

Ventilatory failure is usually a terminal event.

Multifocal Motor Neuropathy

This is a rare disorder of the peripheral nerves that was first described in 1982 by Lewis *et al*. These patients had predominantly motor symptoms and signs but sensory abnormalities were well delineated.

Clinical features

This is characterized by progressive weakness and muscular atrophy. The atrophy is usually severe and nearly always present in some muscles at the time of diagnosis. Reflexes are present or mildly weak. This is a disease that usually runs an indolent course over several years, and

possibly even decades from case reports, but rapid progression has also been documented.

Antiganglioside antibodies, especially anti-GM1 have been sometimes associated with this.[10]

Electrodiagnosis

This is the most important test to differentiate between MND and MMN. Conduction block is a very important feature to try and distinguish between the two conditions.

Treatment

In almost all the cases some form of immune modulation has been tried ranging from very high dose steroid therapy to plasmaphereisis and intravenous immunoglobulins.

Spinal Muscular Atrophy

Spinal muscular atrophy (SMA) is a genetic motor neurone disease characterized by wasting of the skeletal muscles caused by progressive degeneration of the anterior horn cells of the spinal cord. The disorder produces weakness and atrophy of the voluntary muscles. It involves the legs more often than the arms. There are many types. It is an autosomal recessive disorder linked to chromosome 5q11-q13.

Investigations and management

The main diagnostic test is in the electromyography where there would be signs of acute or subacute denervation changes.

Myasthenia Gravis

Myasthenia gravis (MG) is a distinctive disease, being an autoimmune disease.

Mary Walker in the 1930s initiated the use of acetylcholinesterase drugs after learning that MG had similar clinical manifestations to curare poisoning. This remained the mainstay of treatment of the disease until the mid-1960s.

Clinical picture

The clinical hallmark of MG is muscle weakness. The weakness is distinctive due to its fatigable nature.

The course of MG is that of a monophasic illness generally. It can be ocular or generalized and after a few months, will usually remain so in one course or the other. It was found that if symptoms were purely ocular for 7 months, there was a 60% chance that they would remain so indefinitely. The likelihood of remaining ocular purely rose to 84% at 1 year, 88% at 2 years and 92% at 3 years.

Subsequently, the patients become generalized beyond two years of symptoms in only 15% of patients.

The bulbar muscles are often involved with the ocular muscles and later the limb muscles are also affected.

Osserman classification

Grade 1: Focal disease (usually ocular)
Grade 2: Generalized disease that is mild (IIa) or moderate (IIb)
Grade 3: Severe generalized disease
Grade 4: A crisis with life-threatening impairment of respiration

Intramuscular temperature can influence neuromuscular transmission, with it being optimal in a temperature that is a few degrees Centigrade lower than that of body temperature.

Clinically it can be seen that patients become weaker when they become infected and febrile. The temperature change is also more clearly seen in countries with marked seasonal changes.

Diagnosis

It is imperative that the diagnosis be established unequivocally. History and physical examination are the most important, but in addition to this, diagnostic testing is also necessary.

Diagnostic testing

The following are suggested:

- tensilon test (Edrophonium test);
- repetitive nerve stimulation;

- assay for anti-acetylcholine receptor antibody; or
- single fiber electromyography.

The Tensilon test is useful only if there is an objective clinical sign that responds. Repetitive nerve stimulation is done by repeated electric shocks at a rate of 3 per second with action potentials recorded over the muscle. A rapid reduction in the amplitude of the evoked muscle action (decremental response of 10–15%) is considered a positive response.

Single fiber electromyography detects delayed or failed neuromuscular transmission and is present in 88–92% of patients but its specificity is limited.

A radioimmunoassay for the anti-acetylcholine-receptor antibody is specific for myasthenia gravis and detectable in only about 85% of all patients and in about 50% of purely ocular patients.

Management

The outlook for patients has improved tremendously dramatically since 1958, especially in recent years. Previously the mortality was 30% but now with optimal care, the rate is practically zero.

In general, four methods of treatment are in use:

- anticholinesterase agents;
- surgical thymectomy;
- immunosuppression; and
- short-term immunotherapies.

DISEASES OF THE NERVE ROOT, PLEXUS AND PERIPHERAL NERVES

DISEASES OF THE NERVE ROOT

These conditions are often termed radiculopathies and can coexist with peripheral neuropathies. For example, Guillain–Barre syndrome is often thought to be a disease of peripheral nerves but the radicles are involved in a great majority of cases. The motor or sensory root can be selectively involved or in combination. Most root problems are due to degenerative causes (spondylolythic radiculopathy) or traumatic avulsions, followed by space occupying conditions.

Cervical and Lumbar Radiculopathy

Although this condition often presents with insidious onset of motor or sensory symptoms in a root distribution due to root impingement by bony overgrowths, it can rarely present with rapid onset of symptoms due to a disc prolapse. In such situations, it is prudent to exclude myelopathy as well as vascular causes. Cervical spondylolithic radiculopathy most commonly affects the C7, C6 and C5 root levels. Traumatic cervical root avulsions are most commonly the result of shoulder or neck trauma occurs in tandem with significant brachial plexus injury.

In the evaluation of lumbar radiculopathies, two types of lesions should be excluded. A conus lesion may be the result of tumors cysts or other space occupying lesions. It presents with sensory disturbances in the perianal region, impotence and loss of sphincter control. Loss of ankle jerks and S1 myotomal weakness can result from lesion extension. A cauda equina lesion involving T12 or L1 and lower levels can lead to simultaneous upper and lower motor features. This and the presence of pain, relative asymmetry and higher root levels of involvement help to distinguish it from a conus medullaris lesion.

Electromyography is required in addition to nerve conduction studies in the diagnosis of these proximally-situated lesions and is complementary to neuroimaging.

DISEASES OF THE PLEXUS

Brachial plexus lesions can be due to traumatic or non-traumatic causes. The upper plexus can be forcefully stretched by a heavy backpack or traction forces. The lower plexus is often a site for invasive neoplastic disease from the lung apex (Pancoast's tumor). Traumatic root avulsions of the lower plexus often result in an ipsilateral Horner's syndrome. The brachial plexus is also susceptible to radiation injury and metastatic disease, especially from the breast.

Electrophysiological studies are often required in the management of plexopathies as many conditions, other than space occupying lesions, cannot be clearly imaged. The findings of grouped repetitive potentials (myokymia) with electromyography is highly correlated with radiation plexitis.

The lumbosacral plexus is most commonly affected by neoplasms either from direct invasion or metastases. Other causes of lumbosacral

plexopathy are compression from hematomas, especially with anticoagulation, radiation and inflammatory neuritis which occurs to a much lesser extent than in the brachial plexus. The site of lesion in diabetic amyotrophy is uncertain, but is thought to possibly involve the lumbar roots, femoral nerve and lumbar plexus.[11]

Brachial Neuritis (Neuralgic Amyotrophy)

This is a well-defined classically, presenting with severe acute pain of the shoulder region followed by rapid weakness and wasting of the shoulder girdle and upper limb muscles, which become prominent features as the pain begins to subside. The distribution of weakness is variable, although muscles around the shoulder girdle are most commonly involved. Sensory symptoms are not marked, but may occur in a root or nerve distribution. Lesions conforming to a peripheral nerve distribution or a patchy plexopathy may occur. Recurrent and familial cases are well-described. The diagnosis is often incorrect initially as patients are often thought to have a frozen shoulder, cervical spondylosis, entrapment neuropathy or even a cerebrovascular event due to the acuteness of onset.

The underlying pathology is often thought to be of axon loss, but demonstration of focal demyelination has been made using electrophysiological techniques and autopsy samples.[12] Brachial neuritis is linked to several precipitating factors like preceeding viral infection, trauma, vaccination and surgical procedures.

The disease carries a relatively good prognosis with 80–90% of patients expecting to fairly completely recover by 3 years. Recovery can begin as late as 6 months into the disease and can proceed beyond 3 to 4 years. Early and severe muscle wasting carries an unfavorable prognosis. However, with supervised rehabilitation, most patients have a high degree of functional recovery.

DISEASES OF THE PERIPHERAL NERVES

These can be broadly and conveniently classified into immune-mediated neuropathies, neuropathies secondary to medical conditions, inherited neuropathies and entrapment neuropathies. Diseases of the motor neuron will be discussed separately.

Guillain–Barre Syndrome

This is an acute form of neuropathy in which two-thirds of cases have an identifiable form of infection. These can range from minor upper respiratory infections to more severe forms like HIV infections and Borreliosis. A pure motor axonal form is associated with distal weakness, preceeding *Campylobacter jejuni* infection and anti-GM1 antibodies. A sensory demyelinating form with predilectation for respiratory and cranial nerve involvement is linked to anti-GM2 antibodies. The Miller Fisher variant of ataxia, areflexia and ophthalmoplegia is associated with anti-GQ1b antibodies.

Classically, this condition has an acute or subacute onset and runs a monophasic course over 4 weeks. Relapses are extremely rare. Presenting symptoms are usually motor weakness and occasionally distal sensory complaints. It is not uncommon to find mainly proximal rather than distal weakness. Early loss of limb reflexes and facial weakness are common findings. Sensory findings are often minimal and could be limited to loss of position or vibration sense. It is important to watch for respiratory insufficiency in patients who have severe weakness, ambulatory difficulty or a forced vital capacity below 1 liter. Autonomic neuropathy can manifest in variation of heart rate and blood pressure fluctuations but many cases probably have subclinical involvement only.

Nerve conduction studies usually show slowing of conduction velocities or conduction block suggesting demyelination and F wave abnormalities indicating proximal segment or root involvement. Less commonly, small motor amplitudes suggest axon loss which carries a less favorable prognosis. Transcranial magnetic stimulation studies have demonstrated subclinical prologation of central motor conduction time, which normalizes in tandem with clinical recovery.[13] Brain sulcal changes, likely due to an underlying immunological reaction, has also been shown on MRI recently. Lumbar puncture typically shows elevated protein and few or no cells.

Intravenous immunoglobulin (0.4 g/day for 5 days) has been shown to be equally efficacious as plasma exchange (1.5 times plasma volume for 4 exchanges) in the treatment of this condition. It is important to recognize early patients who are likely to progress and institute intensive monitoring for respiratory and autonomic instabilities.

Chronic Inflammatory Demyelinating Polyneuropathy

This is a progressive or relapsing type of immune-mediated neuropathy which generally runs a course of at least 2 months presenting with insidious onset of proximal or distal limb weakness and sensory symptoms, particularly affecting large fibre modalities of touch and vibration.[14] Facial nerve pathologies, facial weakness and respiratory insufficiency are extremely rare. Central nervous system demyelination has been reported. In addition to weakness and sensory deficits, sensory ataxia and enlarged nerves are infrequent findings.

An elevated cerebrospinal fluid protein above 45 mg/dL is found in at least 80% of patients with less than 10 cells/mm^3. Nerve conductions show diminished motor conduction velocities, conduction block and abnormal sensory studies. Clinically, it is important to exclude conditions that causes neuropathy presenting in a similar fashion: HIV infection, autoimmune disease, diabetes mellitus, monoclonal gammopathies, lymphoproliferative diseases, drug intoxication, toxins ingestion and paraneoplastic neuropathy.

The condition has been shown to respond to steroids, plasma exchange and intravenous immunoglobulin. Azathioprine, cyclophosphamide, cyclosporine and interferons has been reported anecdotally to be effective. However, these reports have not been validated in randomized, placebo-controlled trials. Collective experience suggest that although 90% respond to treatment if instituted early, as high as 50% relapse in the subsequent 4 years.

Diabetic Neuropathy

Diabetic neuropathy is the most common form of neuropathy secondary to an underlying medical condition. Diabetic affectation of the nervous system can be in various forms. Most commonly, patients present with symmetrical large fibre type sensorimotor peripheral neuropathies involving large fibre modalities. Less commonly seen is small fibre neuropathy with prominent pain and autonomic features in insulin dependent longstanding diabetics. Painful, asymmetrical wasting of proximal thigh muscles constitute diabetic amyotrophy, which may represent a form of femoral mononeuropathy, although the exact site of involvement is uncertain. Diabetic patients are prone to entrapment neuropathies involving median, ulnar, common peroneal and even cranial nerves.

Other Secondary Forms of Neuropathy

Chronic alcohol consumption damages peripheral nerves from a combination of direct toxicity, dietary insufficiency and impaired absorption. Alcoholic neuropathy presents with insidious onset of symptoms, but acute onset over days is also seen.[15] Initial complaints are distal pain and paresthesia, with motor weakness and muscle atrophy seen mainly in advanced cases. Nerve conduction studies show an axon loss pattern of denervation and sensory symptoms may be improved with vitamin B1 administration.

Peripheral neuropathy often develops in patients with severe chronic renal impairment or undergoing long-term hemodialysis. Symptoms usually develop abruptly with an initial lower limb predilection. Loss of vibration sense and proximal muscle weakness are typical features. The neurological deficits have been shown to improve with regular hemodialysis.

Amyloidosis is known to cause sensorimotor axonal peripheral neuropathy with dysautonomia and conspicuous lower limb involvement. Sarcoidosis should not be forgotten as a rare cause of distal sensorimotor polyneuropathy. Critical illness neuropathy is known to develop in severely ill, ventilated patients with sepsis and organ failure.

Inherited Neuropathies

This is a group of diseases showing sensorimotor peripheral nerve involvement and increasing evidence of genetic basis in their pathogenesis.

Charcot–Marie–Tooth (CMT) or Hereditary Motor and Sensory Neuropathy presents with distal muscle weakness and atrophy, impaired sensation and diminished tendon reflexes. Most cases will manifest by the late twenties but the clinical presentation and severity is wide. Type 1 disease generally denotes individuals with a hypertrophic demyelinating neuropathy while Type 2 refers to individuals with axon loss neuropathy. Genetic mapping to chromosome 17 (CMT1A), chromosome 1 (CMT1B), and the X chromosome (CMTX) has been demonstrated in CMT Type 1. CMT Type 2 subtypes has been linked to chromosomes 1 and 3 but as well worked up as Type 1.

Hereditary Sensory and Motor Neuropathy Type III (HSMN III) or Dejerine–Sottas disease is a severe, generalized form of sensorimotor neuropathy with infantile onset. Marked thickening of affected nerve is

a characteristic feature, as well as pes cavus, kyphoscolioliosis, incoordination and even facial palsies.[16]

HSMN IV or Refsum's disease is characterized by deafness, anosmia, night blindness, retinitis pigmentosa, cerebellar signs and raised serum phytanic acid levels.

Hereditary neuropathy with liability to pressure palsies is an autosomal dominant disorder with focal sausage-like thickenings of the peripheral nerves due to recurrent bouts of remyelination. Sensory symptoms and pressure-induced reversible weakness of the ulnar, radial, peroneal nerves and even a brachial plexopathy occurs.

Friedreich's ataxia presents with sensory deficits while acute intermittent porphyria affects motor fibres more commonly than sensory fibres.[17] Demyelinating neuropathies are known neurological deficits in Fabry's disease and leukodystrophies.

Entrapment Neuropathies

Entrapment neuropathies usually result from recurrent damage to peripheral nerves but it should be differentiated from an underlying peripheral neuropathy. For example, patients who have diabetic neuropathy may have underlying carpal tunnel syndrome, which can be easily overlooked. Neurophysiological investigations will be helpful as an extension of clinical examination in such circumstances.

Carpal tunnel syndrome is the most common of all entrapment neuropathies. It is more common in females, presenting with dysthesia or paresthesia in the median nerve distribution. Symptoms may be referred retrogradely to the elbow or shoulders. They are made worse by maneuvers that narrow the carpal tunnel diameter. Treatment is aimed at abolishing symptoms and preventing motor wasting of the abductor pollicis brevis muscle.

The ulnar nerve is usually compressed at the elbow region in tardy ulnar nerve palsy or between the 2 heads of the flexor carpi ulnaris in the cubital tunnel syndrome.[18] It should be differentiated from distal ulnar neuropathies, most commonly at the canal of Guyon's and the palm. Electrophysiology, using a combination of segmental mixed nerve conduction and side to side comparison, helps establishes the diagnosis.

In the lower limbs, common peroneal nerve injury at the head of fibula may result from leg crossing or injury at this site. The deep branch

is most commonly affected with minimal sensory loss over the web of skin between first and second toes. It must be distinguished from foot drop arising from lumbosacral root lesions, sciatic nerve injury and upper motor neuron causes.

Meralgia parasthetica presents with pain and numbness over the anterolateral thigh and is due to entrapment of the lateral femoral cutaneous at the anterior superior iliac spine region. It should be distinguished from L2L3 radiculopathy and a good guide is the presence of the knee jerk. Nerve conduction tests are helpful in its diagnosis.[19]

DISEASES OF THE MUSCLE

This section will be discussed under the sections of muscular dystrophies, metabolic myopathies, inflammatory myopathies and myotonia.

Muscular Dystrophy

Both Duchenne muscular dystrophy and Becker muscular dystrophy are classified as dystrophinopathies. They result from the absence, deficiency or presence of altered forms of dystrophin, a structural protein found in normal muscle. The responsible gene is on the short arm of the X chromosome at locus Xp21. Most cases are due to duplication or deletion of segments within the gene, the rest being attributed to point mutations.

Duchenne muscular dystrophy represents the severe form with complete absence of dystrophin. Presentation is as early as 2 years old with clumsiness, difficulty walking and calf hypertrophy. The disease is progressive and patients are wheelchair-bound by 10 years. Severe contractures develop in addition to scoliosis and the usual cause of death is cardiomyopathy or respiratory failure. In Becker muscular dystrophy, which is a milder form, presentation is usually in the first decade but can be as late as the fourth. The clinical features are similar to Duchenne's but much milder and many can walk even after 20 years of age. Leg cramps and muscle pains are prominent features. Treatment of these conditions comprises physiotherapy, corrective surgery and bracing. The use of steroids is controversial and has never been conclusively shown to be of benefit in trials.[20]

The limb girdle dystrophies are presently grouped as sarcoglycanopathies. There are 4 known forms of sarcoglycans (alpha, beta,

gamma and delta), a structural muscle protein, each coded by a separate gene that may be defective. Patients present with truncal and limb girdle weakness and occasional calf hypertrophy. A deficiency of gamma sarcoglycan is observed to be associated with severe weakness while other forms have variable severities. Diagnosis is usually made clinically, aided by electromyography, raised creatinine kinase and biopsy.

Facioscapulohumeral dystrophy is an autosomal dominantly inherited condition with the responsible genetic defect localized to the long arm of chromosome 4 in many but not all cases. Presentation is usual in the teen years with mild facial weakness, weakness of scapular fixators despite relative preservation of the deltoids. Hip flexor, ankle dorsiflexor weakness, lumbosacral lordosis and impaired hearing is common. Treatment is supportive with surgical scapular stabilization, orthosis and tendon transfers.

Oculopharyngeal dystrophy is also an autosomal dominant condition presenting with mild ptosis and weakness of eye muscles from the third decade. It is progressive and eventually facial weakness and dysphagia develops. Patients die from emanciation or pneumonia unless aspiration is prevented with a permanent gastrostomy. The condition should be differentiated from myasthenia gravis, which shares the same clinical features.

Metabolic Myopathy

This encompasses myopathic conditions secondary to endocrine conditions as well as those from inborn errors of metabolism.

Thyrotoxic myopathy is probably the most common of the endocrine myopathies and tends to affect males more frequently. Typically, patients complain of muscle twitching and shoulder weakness. Hyperthyroidism is also associated with myasthenia gravis, hyperkalemic periodic paralysis and ophthalmoplegia. Hypothyroidism also causes proximal muscle weakness, hypertrophy and painful spasms. Delayed relaxation of contracted muscle, as well as ankle jerks, is a common occurrence in myxoedema.

Systemic administration of steroids causes myopathy usually affecting the pelvic girdle and thigh. Diseases of the adrenal and pituitary gland also causes non-specific muscle weakness as in Cushing's syndrome, acromegaly and Addison's disease.

In hyperparathyroidism, pelvic girdle weakness, brisk reflexes and even extensor plantar responses have been reported. Distinction from

motor neuron disease can be made electrophysiologically. Finally, hypercalcemia secondary to metastases, myeloma or chronic renal disease can also cause neuromuscular weakness.

Muscle weakness can also occur secondary to glycogen storage disease, disorders of lipid metabolism and mitochondrial disease.

Inflammatory Myopathy

Dermatomyositis can present at any age group and its diagnosis depends on the combination of the typical skin rash and muscular involvement. Associated collagen vasculopathy and underlying malignancy are well-known associations. Pain and tenderness of weak muscles are often absent and the typical heliotrope rash over the eyelids can become generalized.

Polymyositis is more common in adults and it is unclear if it represents a spectrum of diseases which includes dermatomyositis. Weakness is the usual presenting feature and may run a fulminant course at its onset. Neck flexion and diaphragmmatic weakness can occur. Sphincter and extraocular muscles are often spared. Tendon reflexes are normal until late in the disease. In cases of extensive muscle atrophy, serum creatinine kinase levels may be normal despite active disease. Electromyography is helpful in the diagnostic work-up, documentation of treatment response and diagnosing steroid myopathy. Most cases respond to high dose and subsequent tapering steroid regimes.

Myotonia

This refers to a clinical state of painless delayed muscle relaxation, which can be distinguished from the painful state of muscle cramps. However, electromyographic myotonic discharges without overt clinical myotonia can result from hyperkalemic periodic paralysis, hyperthyroidism, acid maltase deficiency, malignant hyperpyrexia and cholesterol-lowering drugs.

In myotonic dystrophy (dystrophia myotonica), an autosomal dominant condition which presents in early adulthood, a hatchet-faced appearance from wasted temporalis and masseter muscles is typical but not invariably seen. Sternomastoid atrophy, ptosis, frontal balding, cataracts, gynecomastia and testicular atrophy are associated features. Cardiac conduction abnormalities, respiratory symptoms, bowel disturbances and low intelligence can occur. The myotonia is typically distal,

diminishes with exercise and responds to membrane stabilizing agents like phenytoin, quinine and mexiletine. A congenital form presents neonatally with hypotonia, club feet, mental retardation but no overt clinical myotonia.

Myotonia congenita occurs in 2 forms: Thomsen's (autosomal dominant) and Becker's (autosomal recessive). The former appears in infancy and remains mild throughout life, compared with the latter which shows more severe lower limb myotonia. Some patients develop severe muscle hypertrophy reaching a Herculean appearance. Life expectancy is normal and involvement of other systems does not occur. The defect is in the chloride channel and the affected gene resides on chromosome 7q35.

Paramyotonia congenita of Eulenberg is a autosomal dominant condition closely associated with hyperkalemic periodic paralysis which can occur simultaneously. The myotonia intensifies rather than improve with exercise and cold exposure results in prominent eyelid, tongue and limb muscle myotonia. Proximal and distal muscles are involved equally. The genetic defect lies in chromosome 17q resulting in mutation of the sodium channel and is dominantly inherited. Acetozolamide may be useful in prophylaxis of paralysis and the myotonia responds symptomatically to the agents mentioned before.

REFERENCES

1. Barohn RJ, *et al.*, Chronic inflammatory demyelinating polyradiculopathy. Clinical characteristics, course and recommendations for diagnostic criteria, *Arch Neurol* **46**:878–884, 1989.
2. Bannister R, Autonomic Failure, Oxford University Press, Oxford, 1992.
3. See SJ, Lo YL, Mixed nerve potentials in brachial plexopathy, *Clin Neurol Neurosurg* **104**:16–19, 2002.
4. Aminoff MJ, *et al.*, Relative utility of different electrophysiologic techniques in the evaluation of brachial plexopathies, *Neurology* **38**: 546–551, 1988.
5. Gorkhaly M, Lo YL, Segmental neurogenic muscle hypertrophy associated with radiation injury, *Clin Neurol Neurosurg*, in press, 2002.
6. Lau DPC, Lo YL, Wee J, Tan NG, Low WK, Vocal fold paralysis following radiotherapy for nasopharyngeal carcinoma. Laryngeal electromyography findings, *J Voice*, in press, 2002.
7. Stalberg E, Trontelj J, *Single Fibre Electromyography,* Mirvalle Press, Old Woking, Surrey, 1979.

8. Leigh PN, Ray–Chaudhuri K, Motor Neurone Disease, *J Neurol Neursurg Psychiatry*, **57**:886–896, 1994.

9. World Federation of Neurology Research Group on Neuromuscular Diseases and Subcommittee on Motor Neurone Diseases/Amyotrophic Lateral Sclerosis, *J Neurol Sci* **124**(Suppl):96–107, 1994.

10. Adams D, Steck AJ, Perruisseau G, *et al*., Predictive value of anti GM1 antibodies in Neuromuscular diseases, *Neurology* **40**(Suppl 1): 299, 1990.

11. Chokroverty S, *et al*., The syndrome of diabetic amyotrophy, *Ann Neurol* **2**:181–186, 1977.

12. Lo YL, *et al*., Motor root conduction in neuralgic amyotrophy: evidence of conduction block, *J Neurol Neurosurg Psychiatry* **66**: 586–590, 1999.

13. Lo YL, Ratnagopal P, Transcranial magnetic stimulation studies in the Miller Fisher syndrome: evidence of corticospinal tract abnormality, *J Neurol Neurosurg Psychiatry* **71**:210–214, 2001.

14. Gorson KC, *et al*., Chronic inflammatory demyelinating polyneuropathy: clinical features and response to treatments in 67 consecutive patients with and without a monoclonal gammopathy, *Neurology* **48**: 321–328, 1997.

15. Behse F, *et al*., Alcoholic neuropathy: Clinical, electrophysiological and biopsy findings, *Ann Neurol* **2**:95–100, 1977.

16. Chance PF, Survey of inherited peripheral nerve diseases, *Electroenceph Clin Neurophysiol* **103**(1):12, 1997.

17. Lo YL, Dan YF, Lee MP, Ratnagopal P, Segmental mixed nerve conduction studies in ulnar neuropathy, *J Clin Neurophysiol* **18**:456–459, 2001.

18. Lo YL, Ratnagopal P, Leoh TH, Dan YF, Lee MP, Yong FC, Clinical and electrophysiological aspects of distal ulnar neuropathy, *Acta Neurol Scand* **105**:390–394, 2002.

19. Lo YL, *et al*., Electrophysiological features in the management of meralgia paraesthetica, *Ann Acad Med Singapore* **27**:530–532, 1998.

20. Dubowitz V, 47th ENMC International workshop: treatment of muscular dystrophy, *Neuromuscular Disorders* **7**:261–267.

This page intentionally left blank

61

Stroke

Chang Hui Meng and Christopher Chen

INTRODUCTION

Stroke is a common condition, resulting from a pathological disruption of cerebral blood flow, and this produces diverse and variable clinical stroke syndromes. The magnitude of the problem is reflected not only by the mortality that results, but also the morbidity and disabilities that follow. Additional stroke attacks compound the initial physical and cognitive disabilities already inflicted. The costs on family and society is enormous, for instance in America, the estimated annual cost of stroke care is about US$40 billion.

EPIDEMIOLOGY AND LOCAL FIGURES

Stroke is the second leading cause of death worldwide, with an estimated 4.6 million deaths. The risk of stroke increases exponentially with age, with the genu of curve at 55 years. The incidence is more common in men (1.25 times), even though more women die annually. In Singapore, it accounts for 10–12% of all deaths annually, and is one of the leading

4 causes of mortality. The average life expectancy of Singaporean has increased from 65.8 years in 1970 to 77.3 years in 1998. Coupled with an aging society, the magnitude of the problem will increase in Singapore.

At the Dept of Neurology, Singapore General Hospital, approximately 1400 stroke patients (excluding subarachnoid hemorrhage) were hospitalized in 1998, with a slight male predominance (54%) and a mean age of 65 years (range 23–96 years old). Most of these were ischemic strokes (92.6%) as hemorrhagic strokes were usually admitted to Neurosurgery Dept. The ethnic distribution approximated our population constitution, with 83.8% Chinese, 8.8% Malays, 5.8% Indians and 1.7% other races. Racial differences in ischemic and hemorrhagic stroke have been described, with hemorrhagic strokes occurring more frequently in Orientals compared with Caucasians.

APPROACH TO STROKE

Stroke refers to any damage to the central nervous system (brain or spinal cord) caused by disturbances to the blood supply. This disruption of blood supply may be due to occlusive/stenotic lesions (causing ischemic infarcts) or hemorrhage (causing either intracerebral hematoma or subarachnoid hemorrhage). A cerebral artery or a vein may be involved, when a cerebral vein is occluded, the syndrome is referred to as Cerebral Vein Thrombosis (CVT). CVT is an uncommon cause of stroke, and the etiologies and management differ from arterial occlusive strokes. They will not be discussed in this chapter. The term "Stroke" or "Cerebrovascular Accident" is usually used when the event occurs suddenly, and implies that permanent brain damage has occurred. On the other hand, "Transient Ischemic Attack" (TIA) is used when the symptoms are transient, and has been arbitrarily defined as neurological deficits that resolve within 24 hours. In actuality, studies have shown that true TIAs last less than an hour. In the carotid circulation, they last an average of 15 minutes, and in the vertebrobasilar system, less than 10 minutes usually. Transient ischemic attacks that last beyond 1 hour often are due to small strokes with rapid resolution of clinical signs. The term Cortical Infarct with Transient Signs (CITS) has been used to describe this condition. Ischemic stroke accounts for 80–85% of all strokes, with hemorrhagic stroke accounting for the remaining.

Hemorrhagic strokes may present as parenchymal/intracranial hemorrhage (ICH) or subarachnoid hemorrhage (SAH). It accounts for about

10% of all strokes. Common causes of ICH include hypertension, congenital or acquired coagulopathies (of which anticoagulant use is important), drugs (amphetamine, cocaine), amyloid angiopathy, arteriovenous malformations and aneurysms. SAH are most commonly caused by arteriovenous malformation and aneurysms.

Stroke and TIA are syndromes of diverse mechanisms and etiologies. The clinical syndrome is a reflection of the underlying vascular abnormality, and cerebral structures that have been damaged. Identifying the stroke type, mechanism and etiology aids in identifying patients at increased risk of early neurologic or medical complications, including death and long-term disability, and in preventing recurrent strokes. Management of stroke requires basic knowledge of the nervous system and the underlying vascular supply. A brief review of the cerebral supply is given below, (anatomical variations exist), followed by mechanisms of stroke, etiologies and stroke syndromes.

BASIC CEREBRAL SUPPLY AND ANATOMY

The vessels to the brain come off the aortic arch, from right to left in the following order: right brachiocephalic trunk (innominate artery), the left common carotid artery (CCA) and left subclavian artery. The right CCA is the first vessel off the brachiocephalic trunk, followed by the vertebral artery (VA), thyrocervical and costocervical trunk, after which the trunk continues as the right subclavian artery. Both CCA run up the neck, where they usually bifurcate at the level of the 4^{th} or 5^{th} cervical vertebrae into the external carotid artery (ECA) and internal carotid artery (ICA). The ECA gives off early branches to supply the thyroid gland, and musculature of the face (superior thyroid, lingual, facial, occipital, auricular, maxillary and superficial temporal artery), while the ICA courses intracranially through the temporal bone, until it reaches the cavernous sinus. This portion of the ICA, within the carotid sinus, is known as the carotid siphon because of its curvature. The ophthalmic artery exits anteriorly from here to supply the eye, while the ICA continues into the supraclinoid portion where the anterior choroidal and posterior communicating arteries are given off. The ICA finally bifurcates into the anterior (ACA) and middle cerebral artery (MCA). The MCA courses laterally in the brain, giving off penetrating vessels known as the lenticulostriates that supply the basal ganglia. The MCA

finally bifurcates or trifurcates into anterior and inferior trunks that supply most of the temporoparietal lobe. The ACA courses medially and anteriorly to supply the medial frontal hemisphere, as well as the caudate nucleus (via the recurrent artery of Heubner). It gives off the callosomarginal artery, which courses along the sulcus of same name, while the artery continues as the pericallosal artery.

The VA originate from the subclavian artery, and courses up the neck until it enters the transverse foramina of the 5th or 6th cervical vertebra, coursing through the foramina of the cervical vertebra till it reaches the 2nd or axis vertebra. It loops up and laterally around the transverse process of the 1st or atlas vertebrae courses medially, where it gives off muscular branches (which may anastomize with branches of the occipital artery). It pierces the dura, and joins the opposite VA to form the basilar artery (BA) at the medullopontine junction. The VA is sometimes divided into 4 segments: V1 refers to the first portion before it enters the cervical vertebra foramen, V2, the segment within the transverse foramen, V3, the segment around the "atlas loop," and V4 the segment after. A major cerebellar artery, the Posterior Inferior Cerebellar Artery (PICA) if given off the intracranial VA. The BA traverses the dorsal surface of the pons, giving off a second major cerebellar artery, the Anterior Cerebellar Artery (AICA). Small medial and circumferential penetrating vessels arise off the BA to supply the pons. The BA bifurcates into the Posterior Cerebral Arteries (PCAs) at the midbrain. The Superior Cerebellar Artery (SCA), the last major cerebellar artery is given off just before this.

The PCAs supply the midbrain, occipital lobe, medial inferior temporal lobe and thalamus. The polar arteries branch off the posterior communicating arteries to supply the anterior thalamus, the thalamic-subthalamic arteries (also known as thalamoperforaters) arise from medial PCA to supply the posteromedial thalamus, branching distally from the PCAs, the thalamogeniculate and posterior choroidal arteries supply the ventrolateral thalamus, and the anteroposterior (including part of the lateral geniculate body) respectively. The cerebellar arteries provide the major vascular supply to the cerebellum, and portions of the brainstem as well.

The carotid circulation is also referred to as the anterior circulation, and the vertebrobasilar circulation, as the posterior circulation. Connections exist between the anterior and posterior circulation, which allow for collateral supply when one or more vessel becomes compromised. Within the intracranial vessels, the anterior communicating artery

(A comm) connects both ACAs. The posterior communicating artery (P comm) connects the carotid siphon with the PCA. Connections also exist within the extracranial vessels and intracranial vessels; from the ECA, the frontal and supratrochlear arteries may provide anastomic link with the ophthalmic artery, while the occipital artery of the ECA anastomizes with the VA. Under pathological conditions, the cerebral vessels may also supply blood to the upper extremities as in the case of the subclavian steal syndrome where blood flows retrogradely from VA to the subclavian artery. A summary of these potential channels is given below.

Table 1 Potential Sources of Collateral Flow to the Brain

With Anterior Circulation Occlusive disease
Frontal and supratrochlear branches of ECA with ophthalmic artery of ICA
Meningeal branches of ECA, via the lacrimal branches to the opthalmic artery
Meningeal branches of ECA to cortical branches of the hemisphere
Anterior communicating artery between left and right ACA
Posterior communicating artery between carotid siphon and PCA
With Posterior Circulation Occlusive disease
Branches from costocervical and thryocervical trunk with VA (V2) in the neck
Occipital artery from ECA with VA (V3)
Posterior communicating artery between carotid siphon and PCA

STROKE MECHANISMS

Traditionally, stroke mechanisms are divided into thrombotic, embolic or hypoperfusion.

Thrombosis

Thrombosis refers to vascular occlusion by local disease within the vessel. Most commonly, this is due to atherosclerosis, where fatty and fibrous plaques form within the vessel wall, and either slowly occlude the lumen from increasing size, or suddenly rupture and occlude the lumen. When the surfaces of the plaque are ulcerated, small platelet and fibrin clumps may form, break off and flow further downstream to occlude a smaller caliber vessel (artery-to-artery embolism). Atherosclerotic plaques, common in an elderly population, are usually found at arterial bifurcation, where the hemodynamic stress is believed to have a contributory role. Other conditions that may occlude the vessel lumen include arterial dissection and fibromuscular dysplasia, which occur more often in the young.

Embolism

Embolism refers to the sudden obstruction of a blood vessel from an occluding material that has originated upstream in the circulatory system (either from the heart or blood vessels) and traveled downstream into the cerebral circulation. The occluding material may be a blood clot, air embolus, small pieces of plaque, bacterial containing, tumor or injected particulate material. If it originates from the heart, the mechanism is described as cardioembolism, otherwise it originates from a vessel (e.g. aortic and carotid artery plaques), where it is referred to as artery-to-artery embolism. In paradoxical embolism, a clot from the venous circulation has entered the arterial circulation to finally occlude a cerebral artery. This occurs when there are existing right to left cardiac shunts (e.g. atrial or ventricular Septal Defect) concomitantly with a venous source of thrombosis (such as deep vein thrombosis in the legs).

Hypoperfusion

Hypoperfusion occurs when cerebral perfusion pressures fall. Global hypoperfusion occurs in massive cardiac failure, cardiac arrest, and massive blood loss. In such cases, global damage will be inflicted, with the watershed areas of vascular territories most affected. Focal hypoperfusion may occur if severe occlusive disease exists in one vessel. This stenotic vessel impedes adequate blood flow, resulting in symptoms referable only to this vascular territory. A severe carotid arterial plaque hence may be symptomatic by embolising small clots distally (artery to artery emboli) or result in hemodynamically significant decrease in perfusion pressure (focal hypoperfusion). A single arterial lesion may cause symptoms via different mechanisms.

STROKE ETIOLOGIES

The most common cause of stroke is atherosclerosis and cardiac diseases in the elderly. Many non-atherosclerotic vascular diseases exist (see Table 2).

Atherosclerosis

Atherosclerotic plaques are intramural depositions of fatty and fibrinous materials, with predilection for certain sites: the aortic arch, origin of the

Table 2 Causes of Ischemic Stroke

1) Atherosclerosis
2) Cardiac diseases: Arrhythmias (atrial fibrillation, Sick Sinus Syndrome), poor left ventricular function, dilated ischemic and non-ischemic cardiomyopathies, intracardiac clots, left ventricular aneurysm, valvular diseases (mitral stenosis), mitral annulus calcification, mitral valve prolapse, right to left shunts (e.g. Patent foramen ovale, atrial or ventricular septal disease), atrial septal aneurysm, bacterial or non-bacterial endocarditis, prosthetic valves, atrial myxomas, etc.
3) Hematologic conditions: deficiencies of Protein C, Protein S and antithrombin III, resistance to activated factor V, plasminogen activator inhibition or deficiency, polycythemia, thrombocytosis, leucocytosis, sickle cell disease, paroxysmal nocturnal hemoglobinuria, disseminated intravascular coagulopathy, thombotic thrombocytopenic purpura, cryoglobinuria, etc.
4) Drugs: cocaine, heroin, L-arginase, amphetamines, ergot drugs, etc.
5) Inflammatory: syphilis, tuberculosis, HIV, systemic lupus erythematosus, anticardiolipin syndrome, giant cell arteritis, polyarteritis nodosa, Crohn's disease, rheumatic arthritis, scleroderma, Takayasu's disease, isolated angiitis of the central nervous system, etc.
6) Arterial dissection
7) Migraine
8) Related to female hormones: high-dose estrogen contraceptive pills, pregnancy, puerperium
9) Genetic/metabolic: Marfan's syndrome, pseudoxanthoma elasticum, homocystinuria, MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and strokes), familial hyperlipidemias, etc.
10) Others: fibromuscular dysplasia, Moya Moya disease, Behcet's disease, CADASIL, etc.

ICA, origin of VA, carotid siphon, main MCA, distal intracranial VA and basilar artery. It is a common disease of the elderly. Intracranial disease is more common in Chinese, Blacks and Japanese. Extracranial ICA bifurcation disease is more common in Whites. Symptomatic extracranial ICA stenosis occurs within 2 cm of its origin from CCA and often presents with warning TIAs. These may be retinal TIAs (also known as Amaurosis Fugax or Transient Monocular Blindness) where usually a painless, dark shade descending over the ipsilateral eye is described. Symptoms for hemispheric TIAs would depend on which side was affected. For the dominant hemisphere (usually the left brain), aphasia maybe present, besides visuomotor-sensory deficits. It is rare for patients to describe symptoms of neglect, seen with involvement of the non-dominant lobe. These TIAs

are often due to small broken-off plaques that embolise distally to the ophthalmic artery or ICA branches, though in the presence of severe stenosis, a hemodynamic component may be present. In TIAs secondary to ICA stenosis, the ipsilateral eye, and contralateral limbs are affected. The risk of stroke with TIA is 24–29% over 5 years, with the highest risk described in the first month (4–8%) and 12–14% within the first year. Hemispheric TIAs are also associated with a higher stroke risk than retinal TIAs. A 5-year retrospective review of all Doppler examinations in our department confirmed that extracranial ICA disease is not common in a predominantly Chinese population. We retrospectively reviewed 5235 carotid Doppler studies over a 5-year period from 1994. Ninety three percent of these patients had presented with symptoms and signs of TIA or stroke. Only 6.6% of patients studied had severe (more than 70% stenosis) or occluded extracranial carotid disease. 4.3% had symptoms of stroke within the last 6 months, the remaining one-third had symptoms beyond 6 months. This has important bearings as large clinical trials have confirmed the superior benefit of carotid endarterectomy over best medical treatment in the management of severe recently (within last 6 months) symptomatic carotid stenosis, provided disabling strokes have not occurred.

Cardiac Sources of Stroke

Approximately 15–20% of all ischemic strokes are cardioembolic. The prevalence of cardioembolic stroke in young strokes is high (23–36%), partly because of the lower prevalence of atherosclerotic disease. Atrial fibrillation is the most common cause, myocardial infract, rheumatic valvular disease and prosthetic valves are also common. In the evaluation for cardiac sources of stroke, arrhythmias, valvular abnormalities and cardiac wall abnormalities need to be excluded. The clinical history may give clues, palpitations, rheumatic fever or myocardial ischemia may be elicited. The presentation is often sudden, with maximal neurological deficits at onset, and very rarely preceding TIAs. Clinical examination may give clues to the underlying cardiac disease, splinter hemorrhages, Roth's spots on fundoscopy are signs of systemic embolisation. Echocardiography is useful for structural imaging of the heart, valves and intracardiac clots. Transesophageal echocardiography (TEE) is superior though more invasive than transthoracic echocardiography (TTE). The aortic arch and left atrial appendage are studied better. However, it is comparatively more invasive. The use of echo contrast

agents, or bubbles (agitated saline) improves the detection of intracardiac shunts. A 24-hour cardiac rhythm monitoring is useful when the suspicion for paroxysmal arrhythmias is high. Cerebral imaging with either CT or MRI may reveal multiple vascular territorial infarcts.

Arterial Dissection

Arterial dissection occurs when blood extrudes into the vessel wall. Blood may enter the subintimal plane or media-adventitial plane. Symptoms may arise because of luminal compression by the subintimal hematoma, clot formation within a stenosed lumen with distal embolisation or aneurysmal dilatation, especially in subadventitial hematoma, with rupture and subarachnoid hemorrhage. Dissections are common in young strokes.

Extracranial arterial dissection is more common than intracranial dissection. In the neck, the ICA is most often involved, usually some distance after its origin (unlike atherosclerotic plaques) and may extend rostrally for a variable distance, though it rarely extends intracranially. The first and third portions of the VA are next most commonly involved sites. Intracranially, the ICA and MCA stem are usually involved. Intracranial vertebrobasilar artery dissections are very rare.

The most common symptom is pain, in the neck or head. Antecedent trauma may be absent or mild and a temporal relationship is not always established. Angiography usually establishes the diagnosis. A tapering occlusion is characteristic (rat's tail sign). Axial Magnetic Resonance Imaging (MRI) of the involved vessel may reveal the false and true lumen, and is a less invasive procedure.

STROKE SYNDROMES

Various terminologies have been used to describe stroke syndromes and subtypes. These give clues not only to the underlying cerebral and vascular abnormality, and mechanism, e.g. small (or large) vessel stroke and cardioembolic stroke, but help prognostic neurologic outcomes, and recurrence. Obstruction of branch or large cerebral vessels results in a constellation of signs and symptoms that allows the physician to identify the offending vessel. Such strokes have also been described as "Large Vessel Strokes or Disease." The table below describes the clinical manifestation that may occur.

Table 3 Clinical Signs and Symptoms

Vascular Territory	Cerebral Structures Supplied	Signs & Symptoms
Middle Cerebral Artery (MCA)	Frontal, parietal and temporal lobes	Hemiplegia involving face and arm, more so than leg Hemianesthesia Hemianopia Global aphasia or neglect
Upper division MCA	Frontal-parietal lobe	Predominantly brachiofacial weakness and sensory loss Expressive (Broca's) aphasia or neglect
Lower division MCA	Temporo-parietal lobe	Minimal weakness and sensory loss Upper quadrantnopia or hemianopia Receptive (Wernicke's) aphasia or constructional apraxia
Anterior Cerebral Artery (ACA)	Medial frontal lobe, Corpus callosum	Predominantly leg weakness Apathy and abulia Transcortical motor and sensory aphasia Left limb apraxia (Anterior disconnection syndrome) Incontinence
Recurrent Artery of Heubner	Caudate nucleus Anterior limb of internal capsule	Behavioral changes: abulic and apathetic or agitated and hyperactive Mild dysphasia or neglect Mild motor weakness
Anterior Choroidal Artery	Globus pallidus, lateral geniculate body, posterior limb Internal capsule, medial temporal lobe	Hemiplegia of face, arm and leg Hemianesthesia Hemianopia with sparing of middle sector (30° to 60°)
Posterior Inferior Cerebellar Artery (PICA)	Inferior cerebellum, Dorsolateral medulla	Lateral Medullary Syndrome
Anterior Inferior Cerebellar Artery (AICA)	Middle cerebellum, pons	Hemiataxia Characteristic VIII cranial nerve palsy
Superior Cerebellar Artery (SCA)	Superior cerebellum, Upper pons	Hemi or quadriataxia Characteristic IV cranial nerve palsy

Table 3 Continued

Vascular Territory	Cerebral Structures Supplied	Signs & Symptoms
Basilar Artery (BA)	Pons	Quadriparesis Multiple cranial nerves palsies
Thalamic-subthalamic artery	Posteriormedial thalamus	Vertical gaze abnormalities Abnormal/decrease conscious level Memory and cognitive abnormalities
Polar artery	Anterolateral thalamus	Apathy, abulia Memory abnormalities Mild-moderate aphasia or neglect
Thalamogeniculate artery	Ventrolateral thalamus +/− Posterior limb of internal capsule	Hemianesthesia Mild hemiparesis Hemiataxia or hemichorea Dejerine Roussy syndrome (above features, with delayed disturbing pain in hemianesthetic limbs)
Posterior choroidal artery	Pulvinar, habenula, anterior thalamus, lateral geniculate body	Uncommon, characteristically, sectoranopia (from 60° to 120°)
Posterior Cerebral Artery (PCA)	Midbrain, thalamus, occipital lobe, medial inferior temporal lobe	Hemiparesis, hemianesthesia, hemiataxia Vertical gaze abnormalities Hemianopia, visual disturbances Alexia without agraphia (Posterior disconnection syndrome) Memory abnormalities Mild-moderate aphasia and neglect

Occlusion of small penetrating vessels in the brain produces “Small Vessel Stroke/Disease”, which have also been called “Lacunar strokes.” As the damage is small and restricted, patients with lacunar strokes would not have concomitant language, spatial or visual impairments, unless these were sustained from previous strokes. Five clinical lacunar

syndromes have been described: pure motor, pure sensory, motor-sensory, ataxic hemiparesis and clumsy-hand-dysarthria. These are clinical syndromes, which do not identify the underlying area of brain damaged. However, lacunes have a predilection for certain sites, the putamen, pallidum, pons, thalamus, caudate nucleus, internal capsule, corona radiata and cerebral peduncles.

Another classification, commonly used in our institution is the Oxfordshire Community Stroke Project (OCSP). The OCSP is a simple clinical (based on history and physical examination) classification for acute stroke which predicted functional recovery, mortality and recurrent strokes. In OCSP, strokes are classified into primary intracerebral hemorrhage (PICH), total anterior circulation infarct (TACI), partial anterior circulation infarct (PACI), lacunar infarct (LACI), and posterior circulation infarct (POCI). Radiological findings are included in the final assessment. Patients with TACI have the full constellation of signs seen in large vessel occlusive disorder, e.g. in the dominant Middle Cerebral Artery, this would be dysphasia, motor and sensory weakness, and hemianopia. At the other spectrum, patients with LACI would have one of the lacunar syndromes described with no language, spatial or visual deficits. Patients with PACI have intermediate signs.

The natural history of stroke is influenced by the stroke subtype and comorbidities. The 30-day mortality is highest in intracerebral hematoma (48–82%) and lowest in ischemic infarct (8–15%), with subarachnoid hemorrhage inbetween (42–46%). Within the ischemic strokes, large vessel occlusive infarcts, large volume infarcts, TACIs and cardioembolic strokes have been shown to fare much worse than LACIs or small vessel strokes. Stroke progression is a difficult parameter to assess. One database showed that about a quarter of strokes admissions had unstable course, of which three-quarters were attributable to stroke progression. Progression was most often in large vessel occlusive disease. Progression in lacunar strokes had the best prognosis for recovery. The risk of recurrent stroke is similiarly affected by the stroke subtype and co-morbidities. The risk of recurrent stroke is highest in the first 30 days of stroke (up to 4%) and averages about 5–25% at one year. It is highest in symptomatic severe carotid stenosis, up to 30% within 2 years, and lowest in lacunar strokes. Other predictors of stroke recurrence hemispheric stroke, hypertension, atrial fibrillation and heart failure.

INVESTIGATIONS

Investigations are planned to confirm the diagnosis of stroke and the area damaged, and further elucidate the underlying mechanisms and etiologies. Imaging of the nervous system should be aimed at the following issues:

1) identification of the lesion — that it is a stroke, and not other cerebral lesions such as tumor;
2) differentiation between ischemic and hemorrhagic stroke;
3) localization and quantification of the size of stroke;
4) determination of the age of stroke; and
5) identification of other previous strokes.

Vascular imaging, on the other hand, helps to confirm the mechanism and etiology of stroke, and also prognostic the risk of recurrent stroke:

1) the vessel involved, and if an occlusive lesion is present;
2) the severity of the occlusive disease, and what further brain tissues are at risk for damage;
3) the cause of the occlusive lesion; and
4) the state of the other cerebral vessels: are there occlusive disease in other parts, what are the collateral support like, are there any vascular variants present.

Structural Imaging

Computed tomography

Computed Tomography (CT) head is a simple, readily available test that allows a rapid and accurate differentiation of infarcts from hemorrhage. It is also relatively inexpensive, compared with some of the newer imaging modalities available.

Early CT infarct signs are mainly attributable to a combination of vasogenic and interstitial edema:

- loss of the gray-white differentiation due to edema of the gray matter
- sulcal effacement due to mass effect of the edema
- compression of the ventricles due to mass effect of the edema
- "insular ribbon sign" where there is diminished attentuation of the gray matter in the insular cortex and claustrum, again due to edema

- the definition between the lentiform nucleus and internal capsule is lost or obscured
- a hyperdense middle cerebral artery may be seen, due to an acute thrombosis

Chronic infarcts show up on CT scans as

- atrophy (a clearly delineated defect) with dilatation of the adjacent ventricles and sulci
- Wallerian degeneration, where there is denervation atrophy of damaged tracts (e.g. motor tracts from internal capsule, to cerebral peduncle, pons, medulla and pyramidal tracts), which is rare
- dystrophic calcification, which is rare

CT scan does suffer from certain disadvantages. It is not as sensitive as MRI in detecting early ischemic infarcts, especially if they are small (lacunes or spinal cord infarcts) or adjacent to bony structures (such as the posterior fossa).

Magnetic resonance imaging

MR imaging depends on the reaction of body tissue nuclei (the hydrogen nuclei is most commonly manipulated) to radio frequency pulses in the presence of a powerful magnetic field. The collected data is manipulated by a computer to provide the final image. Hydrogen nuclei are the most common protons in the body, found in water, fats and many other tissues. Water and fat protons have been most extensively imaged with MRI. Other nuclei can also be studied, using MR spectroscopy. These include fluorine, carbon and phosphorus. However, because these nuclei are less abundant in the human body, they produce weaker signals, and the final image collected has poorer spatial resolution. Manipulation of various MR parameters allows images to be obtained with differing qualities. T2-weighted and especially diffusion weighted imaging allows early visualization of acute lesions by demonstrating the increase in tissue water that accompanies acute infarcts. T1-weighted sequences provide good anatomic definition, and are useful for detecting subacute-chronic infarcts. Perfusion imaging (also referred to as hemodynamically-weighted MRI) gives a reflection of the cerebral blood flow, and may allow the identification of ischemic tissue at risk for further infarction. Advancement with technology has allowed the time for MR imaging to

be shortened (echoplanar imaging). MRI is more sensitive than CT, in particular with small and early ischemic strokes. However it does not define hemorrhage as well, though there are sequences (e.g. fluid attenuated inversion recovery or FLAIR) that are sensitive to the presence of blood.

Neurovascular Imaging

Ultrasound imaging and Doppler sonography

Ultrasound is a frequency that is beyond the range perceptible to human ears. It can make the motion of blood audible by the Doppler effect, and also allow a visual display of the vascular tissues imaged.

Transcranial Doppler (TCD) ultrasound is based upon the use of a range-gated pule-Doppler ultrasonic beam of 2 MHz frequency to assess the hemodynamic characteristics of major intracranial arteries. The TCD takes advantage of ultrasonic "windows" in the skull, where the bone is sufficiently thin to allow penetration of the ultrasound waves. The sonographer is unable to see the vessels being insonated, however, using set criteria, the sonographer is able to identify major arteries imaged. Cerebral veins can also been studied. New techniques available now allow for direction visualization of the vessels imaged. Its applications include the detection of severe stenosis intracranially, vasospasm in subarachnoid hemorrhage, assessing collateral circulation, assessing brain death and detecting vascular malformations.

Computed tomographic angiography and magnetic resonance angiography

Both CTA and MRA study the cranial vessels non-invasively. There is a limit to the spatial resolution of smaller vessels, which is still best studied on angiography.

Cerebral angiography

This has long been held to be the gold standard for diagnostic imaging of the cerebral vasculature. There is a small but significant risk of major stroke or death, during or soon after angiography. This risk is about 0.5%, and is associated with advanced age, elevated serum creatinine,

hypertension, and a lengthy angiographic time. The risk for minor stroke and TIA is usually about 1%. However it gives more precise information on the cerebral vasculature, the nature, severity and extent of underlying vascular lesions, the identification of other lesions present and the extent of collateral flow present.

Other Imaging Modalities

Single-photon emission computed tomography (SPECT)

Single photon gamma-emitting radionuclides are used in SPECT imaging to provide information about perfusion (and to a smaller extent neuroreceptor distribution) in cerebrovascular disorders. SPECT can demonstrate local hypoperfusion in acute stroke, and identify the cerebral perfusion reserves in the presence of an extracranial occlusive disease. Current SPECT perfusion techniques utilize either the cerebral clearance of a radionuclide or the retention of a radionuclide-tagged-ligand, which is fixed intracranially according to blood flow.

Positron emission tomography (PET)

Like SPECT, PET is another functional neuroimaging technique. Positron-emitting radionuclides are tagged to physiologically active compounds, and given to patients. Cerebral metabolism and blood flow is assessed with CT studies on distribution of these radionuclides with time. Unfortunately, these radionuclides have very short lives, and a dedicated cyclotron is required on-site for synthesis of these compounds. The equipment and maintenance makes this very costly, which is why PET is still a mainly research tool.

Others

Other investigations should be dictated by the patient's presentation and physician's clinical suspicions. In particular, young patients with stroke should be aggressively investigated for treatable conditions. Lifestyle factors should be analyzed, as the co-existence of many contributory diseases increases the risk of stroke and its recurrence. For instance, in a young woman with acute stroke and a history of migraine, smoking and the use of contraceptive pills should be stopped. Laboratory tests should

include screening for hematologic, inflammatory and infective causes. Cardiac diseases should be actively excluded.

TREATMENT OF STROKE

Stroke is an acute neurological emergency. Currently, there are effective strategies that help to reduce neurologic disabilities and mortality. All patients with acute stroke should be admitted for management and investigations.

Unfortunately, most stroke patients present late to medical institutions. At our institution, only 26.6% presented within 6 hours of symptoms onset to the Emergency department, with an additional 13% presenting between 6–12 hours after symptom onset. Men tended to seek medical attention slightly earlier than women do (41.3% of men presented within 12 hours of symptom onset, compared with 37.7% women). These figures are worrying as for effective management of stroke should be instituted as early as possible to minimize stroke complications.

Management of acute stroke can be divided into an acute phase, and a chronic management cum rehabilitation phase. These phases overlap, in the acute stage, the thrust of management is stabilization of stroke and medical conditions, though simple rehabilitation is usually initiated at this early stage if the patients are medically stable.

In the chronic management phase, the goals are to prevent further strokes (secondary prevention), treat co-existing diseases, and rehabilitate the patient's functional status to the best possible. Frequently this phase is initiated in the acute care hospital and continued in a step-down facility such as a rehabilitation hospital or a community hospital. It does not end with discharge but merges into home care and day care rehabilitation.

MANAGEMENT OF ACUTE STROKE

General

An acute stroke pathway can be used to guide the management and investigation of stroke in a timely fashion. In our institution, a multidisciplinary stroke clinical pathway, involving physicians, nurse, therapists, dieticians and social workers has been developed (Fig. 1). Unstable patients such as those with progressive strokes, large or brainstem strokes are better

managed in the Acute Stroke Unit, where the neurological and medical parameters are monitored at close intervals. The Acute Stroke Unit is a geographical area for the multidisciplinary care of Stroke patients. Such Stroke Units have been shown to reduce hospitalizations and improve the outcomes of stroke patients by decreasing medical complications, neurological dependence and mortality. Studies have shown a 20–30% reduction in death, dependency and discharge to institutional care.

Frequent neurological assessments allow for early detection of neurological deterioration, which should be a trigger for additional evaluation. Deterioration could be attributable to neurological factors (such as progression of infarct, hemorrhagic transformation of the infarct, a new infarct or cerebral edema from existing infarct) or non-neurological factors (such as sepsis, metabolic abnormalities like hyper or hyponatremia). These causes should be rapidly identified and corrected if possible.

Protection of the airway and maintenance of oxygenation remains of paramount importance. The routine use of oxygen supplementation is of unproved benefit, but determination of oxygen saturation by pulse oximetry or arterial blood gas measures is recommended. Oxygen should be given if the patient is hypoxic. The cause of the hypoxia also should be sought. Airway integrity should be monitored, especially if the patient has depressed consciousness or evidence of brainstem dysfunction. Intubation and ventilatory assistance may be needed.

Patients frequently have an elevated blood pressure after stroke and the management of hypertension is controversial. Arterial hypertension can be seen with either hemorrhagic or ischemic stroke. After the initial stress of the event is over, the blood pressure tends to decrease gradually without any medical therapy. To date, there are no data about any specific level of systolic or diastolic blood pressure that is harmful to a patient with stroke. Hypertension in the acute phase should not be treated unless there indications for aggressive treatment, which includes hypertensive encephalopathy, renal or myocardial compromise, or an acute aortic dissection. Thrombolytic therapy necessitates a more energetic response. Whether cerebral hemorrhage is an indication for aggressive control of arterial hypertension to prevent continued bleeding is uncertain.

Pyrexia should be managed quickly. Data suggests that hyperthermia exacerbate brain damage. Pneumonia secondary to aspiration is the most common cause of fever appearing during the first 24 hours after stroke. Subsequently, fever usually is secondary to pulmonary or urinary tract

infections. A fever should not be ascribed to the stroke until an infectious etiology is eliminated. Fever also can point to a cause of stroke such as infective endocarditis. Hypothermic treatments for stroke are experimental, but early use of antipyretics (or even cooling blankets) in the febrile patient seems advisable while the search for the cause of the fever is underway.

Fluid and electrolyte management is an important aspect of early care. Many patients are dehydrated from poor intake or use of diuretics. Early intravenous fluid support is indicated, especially if oral intake is restricted. Achievement of a normal blood volume seems appropriate in most patients. A mild fluid restriction is advised for patients who are at risk for increased intracranial pressure. The usual intravenous fluid used is normal saline as experimental and epidemiological evidence suggests that high levels of blood glucose may be detrimental for outcomes after stroke. There is no data to show that the aggressive control of the blood glucose improves neurologic recovery. Treatment of an elevated or low level of blood glucose is advised, but no specific protocol can be recommended for treatment after stroke.

Food intake sometimes is avoided during the first 24 hours after stroke, especially in the presence of a reduced level of consciousness, dysarthria, impaired gag reflex, or wet cough. Swallowing evaluation should be used liberally. If adequate oral intake is not feasible, enteral feeding can be started, either via a nasogastric tube or via a percutaneous gastrostomy.

Early mobilization is advised in order to reduce the risk of aspiration and the development of deep vein thrombosis. Sitting or standing determinations of blood pressure will help detect a postural drop, which could induce neurologic worsening when the patient begins mobilization.

SPECIFIC MEASURES

Antithrombotics

Both the International Stroke Trial (IST) and Chinese Acute Stroke Trial (CAST), which combined, enrolled over 40 000 patients demonstrated a small but beneficial effect of early antiplatelet with aspirin in acute stroke, less than 48 hours. IST used an aspirin dose of 300 mg for 14 days, while CAST applied 160 mg for 4 weeks. Acute antiplatelet therapy was also

studied in one arm of the Italian Streptokinase study. Together, these results show that with early aspirin therapy, for every 1000 patients treated, there would be about 7 recurrent strokes prevented, 5 less deaths and 9 less deaths/fatal strokes with a slight increase in hemorrhagic strokes of 2. Although the effect is modest, the ease of administration, wide availability and low cost argues for its use in acute ischemic stroke without any contraindications. A dose range of 160 to 300 mg is recommended. The Antiplatelet Trialists Collaboration determined that the lowest, effective aspirin dose for the prevention of recurrent ischemic stroke is 75 mg. Hence it may be reasonable to convert patients to a lower dose of aspirin after 2 to 4 weeks.

Anticoagulation

IST also studied the effect of heparin in all types of stroke. Heparin subcutaneously administered was not shown to be of any benefit in decreasing the risk of recurrent stroke in the immediate post stroke period. Heparin did decrease the risk of ischemic stroke, but was associated with a greater risk of hemorrhagic stroke, which negated its beneficial effect.

The role of low molecular weight heparin in acute stroke is still not clear. An earlier study in Hong Kong revealed some beneficial effect, which has not been reproduced in subsequent studies.

Anticoagulation is proven treatment in prevention of cardioembolic strokes. However, the time of initiation from the onset of an acute stroke, to prevent a recurrent stroke is not clear. The risk of recurrent stroke in the first 2 weeks after an initial stroke in the setting of atrial fibrillation has been described to be as high as 21%, however, recent large clinical trials showed that this might not be true. There is no consensus on the appropriate timing to start anticoagulation soon after an acute infarct to prevent another stroke. A balance needs to be made between the risk of hemorrhagic transformation and recurrent stroke. The role of anticoagulation in patients with high grade stenoses of either carotid or the vertebral-basilar vascular systems, and progressive strokes remains theoretically attractive but unproven.

While the role of heparin in the setting of acute ischemic stroke is limited, subcutaneous, low-dose heparin to prevent deep venous thrombosis and pulmonary embolism is well proven.

Thrombolytics

The advantages of clot-busting agents have always been attractive. The removal of an offending clot would allow tissue salvation, and prevent further neuronal death. Its effectiveness in acute coronary syndromes lends further support for studies of its usage in strokes. However, considerations for its usage must be tempered by knowledge of its safety profile. In a meta-analysis of about 12 thrombolytic trials by Wardlaw, it was demonstrated that the odds ratio of increased dependency or death was reduced at 6 months to 0.75. Unfortunately there was a real increase in early deaths (odds ratio 1.99) and symptomatic or fatal hemorrhage (odds ratio 3.62 and 4.44 respectively). Thrombolytic agents are efficacious, but only if the correct dose and route is administered to appropriate patients.

Thrombolytic agents studied are either fibrinolytic (induce lysis of the fibrin clot via activation of plasmin) or fibrinogenolytic (reduce fibrinogen, the substrate for formation of the fibrin clot). Streptokinase and Urokinase are early fibrinolytic agents used with low fibrin specificity, while tissue-Plasminogen Activator (t-PA) and Pro-urokinase are newer generation agents. Ancrod, synthesized from viper venom, is a fibrinogenolytic agent, with some fibrinolytic activity as well.

The NINDS t-PA trial was the first to demonstrate the efficacy of intravenous tissue-Plasminogen Activator (t-PA) in the setting of ischemic stroke if administered to appropriately selected patients. The main exclusion criteria are treatment beyond three hours, intracranial hemorrhage on the CT, a blood pressure greater than 185 systolic or diastolic greater than 110 (or aggressive antihypertensive therapy to reach these blood pressure limits), and a minimal neurologic deficit. All stroke types, less than 3 hours, were recruited. The dose of t-PA was 0.9 mg/kg bodyweight, up to a maximum of 90 mg. The initial 10% dose was given as a bolus, and the remaining 90% infused over 1 hour. The results were in favor for thrombolytic treatment, where treated patients were more than 30% more likely to have mild or no neurological deficits at 3 months, with no increase in mortality or severe strokes, despite a 10 times increase in hemorrhagic transformation. Unfortunately, this positive result has not been reproduced in other thrombolytic trials with t-PA. However, these trials had slightly differing criteria for recruitment. In June 1996, the FDA approved the use of recombinant tissue plasminogen activator (rt-PA) for

acute (< 3 hrs) ischemic stroke in the US. Tissue-Plasminogen Activator is not licensed outside of the US for acute stroke.

One t-PA trials, the European Co-operative Acute Stroke Study (or ECASS I study) suggested that patients with large areas (greater than 1/3 of the middle cerebral artery territory) of ischemic damage have a higher mortality and hemorrhage risk when given t-PA. Based upon the results of the ECASS study, many stroke physicians would not recommend treatment of patients with large areas of hypodensity on the baseline CT scan, even if the patients have a clearly defined time of onset of less than three hours.

Streptokinase has been shown to be associated with a high morbidity and parenchymal hemorrhage in 3 recent studies. It has been postulated that the negative results of streptokinase might have been related to higher relative doses of streptokinase used, the associated use of aspirin and heparin, the later time of treatment from onset, and a lack of a protocol for managing blood pressure.

Two other thrombolytic trials, intra-arterial Pro-urokinase in acute MCA occlusive strokes less than 3 hours, and Ancrod in acute ischemic strokes less than 3 hours have reported beneficial outcomes in 1999.

Neuroprotective Agents

Many neuroprotectant agents have been tested in Phase II and III studies. None have currently been shown to be effective.

MANAGEMENT OF STROKE RELATED COMPLICATIONS

Early deaths from stroke are usually due to neurologic causes, such as herniation from a large hemispherical infarct or hemorrhage, or brainstem ischemia or hemorrhage. Subsequent deaths are typically due to medical complications, such as infection or pulmonary embolus.

Cerebral Herniation

It is a common cause of early death in the severe stroke patient, and prevention and treatment of cerebral edema remains an elusive goal.

Although fluid overload is to be avoided in the early phase, there is little evidence that fluid restriction reduces the incidence of edema, indeed the resulting dehydration and hyperviscosity may exacerbate stroke. Hypo-osmolar fluids such as 5% dextrose should be avoided. Cerebral edema typically presents on the third to fifth day post-stroke. There is no evidence it will respond to corticosteroids, and their use is discouraged. Treatment of symptomatic edema remains controversial. Mild fluid restriction may be helpful. Elevation of the head of the bed to 30° is advised. Antihypertensive agents, that produce vasodilatation, such as nitroprusside, should be avoided. In an emergency, elevated intracranial pressure can be quickly lowered by intubation and hyperventilation. Reduction of CO_2 by 5 to 10 mmHg lowers intracranial pressure 25% to 30%, but the effect is short-lived. Hyperventilation is usually followed by additional therapy, such as osmotic diuresis or surgery. Mannitol (0.25–0.5 g/kg IV) can be administered. Monitoring of electrolyte status and osmolarity is needed if several doses of mannitol are given. Osmolarity should not exceed 310 mOsm. Furosemide can be used as an additional diuretic with mannitol, especially if cardiac function is compromised.

Intracranial pressure monitoring with a subdural catheter or intraventricular catheter can help guide therapy, but its value has not been established in the stroke patient. Placement of a ventricular catheter for cerebrospinal fluid drainage can relieve pressure if hydrocephalus is present. Resection of large ischemic lesions can be life-saving, but may leave the patient with severe residual deficits. Successful avoidance or reduction of cerebral edema remains an elusive goal for experimental stroke therapy.

Cerebellar infarction with edema represents a special problem requiring early diagnosis as it is eminently treatable with surgery, including relief of hydrocephalus by ventriculostomy and removal of cerebellar tissue via suboccipital craniectomy.

Seizure

Seizure activity tends to occur early in acute stroke, with an incidence of 4–8%. Most seizures are partial in onset, and status epilepticus is rare. Control of seizures is usually easy following typical guidelines. Prophylactic use of anticonvulsants is not helpful.

GENERAL MANAGEMENT OF STROKE

Death during the acute-phase hospitalization for stroke is usually from medical complications, including heart problems, pneumonia, and pulmonary emboli. Preventing these complication is one of the main contributions of a dedicated stroke unit.

Venous thromboembolic disease appears to be uncommon in our population compared to the West. Prophylactic measures include compression stockings, heparin or the low molecular weight heparins. Effective prevention of deep vein thrombosis may lower the risk of developing pulmonary embolus. Pulmonary embolus should be suspected in the stroke patient with sudden hypoxia, breathlessness and respiratory alkalosis in the setting of a normal chest film.

Infections of the lungs and urinary tracts are common. Aspiration and consequent chest infections are often found in post-stroke patients and have been shown both to worsen the outcome of stroke as well as contribute to mortality after a stroke. This may be due to bulbar weakness, poor conscious level, or both.

Patients at high risk for or with dysphagia should not be fed orally. Alternative routes of feeding, such as intravenous fluids, nasogastric or feeding gastrostomy should be employed. There is no demonstrated advantage of any feeding route over the other. A 3-month study at our institution revealed that dysphagia was common in acute strokes, 25% of patients had dysphagia, decreasing to 12% the seventh day. Large strokes (TACIs) were most likely to develop dysphagia, and the presence of dysphagia independently predicted an increase risk of poor neurologic outcome, and recurrent chest infections.

Bladder and bowel movements need to be regulated. Urinary catheters should be avoided if possible, and intermittent catheterization is preferred if bladder drainage is needed.

Depression may hinder rehabilitation. Although antidepressants may not reach effectiveness during the short stay of the acute hospitalization, early institution of antidepressants (especially serotonin re-uptake inhibitors such as fluoxetine) should be considered in the high-risk patient. Treatment of confusion and agitation should be approached with caution, because many psychoactive drugs may retard cerebral recovery. Low doses of benzodiazepines may be appropriate.

Proper body positioning and alignment are important aspects of care. Careful attention should be paid to the effects of immobility. These include joint stiffness and the potential for contractures and bed sores in the stroke patient.

STROKE PREVENTION

Risk factors for ischemic stroke can be divided into modifiable and non-modifiable factors. Non-modifiable risks include age, genes, sex and race. Although these factors cannot be modified, their identification helps to predict those at risk, and more aggressive treatment of risk factors that can be modified.

Secondary Prevention of Stroke

Antiplatelet agents

There are a number of antiplatelet agents available. Aspirin acts on the enzyme, Cyclo-oxygenase, irreversibly inhibiting the production of the potent vasoconstrictor and platelet activator, Thromboxane A. Ticlopidine and clopidogrel are 2 chemically related thienopyridine derivatives. Both drugs selectively, and irreversibly inhibit ADP-induced platelet aggregation. These drugs have long lasting effect, platelet aggregation and bleeding time require 4–8 days to return to baseline. Dipyridamole is a phosphodiesterase inhibitor with vasodilatory effect as well.

Aspirin is the most well-studied drug, and its effectiveness well-demonstrated. The optimal dose is still not clear, doses from 30 mg to 1300 mg have been shown to be effective, though at higher doses, side effects involving especially the gastrointestinal system was common. Patients with stroke or TIA have an annual risk of stroke of 5–10%. In the Antiplatelet Trialist Collaboration, the risk of non-fatal stroke is reduced by 23%, and vascular death by 14%. However, cumulative evidence from aspirin trials alone, suggest that the benefit on vascular death, non-fatal strokes and myocardial infarcts offered by aspirin may be lower, about 13%. Ticlopidine has been compared with aspirin, and even though there was a 6% relative risk reduction in favour of ticlopidine, the confidence intervals were large. Furthermore, patients on ticlopidine experienced more adverse effects, in particular diarrhea

and skin rashes. Neutropenia was also seen in up to 2%, and the usage of ticlopidine mandates regular blood count monitoring in the first 3 months.

The second European Stroke Prevention Study (ESPS-2) demonstrated that 400 mg daily of dipyridamole was as effective as low dose (50 mg) aspirin daily. The combination of these 2 drugs had a synergistic effect in stroke prevention. These results have been questioned as dipyridamole was not shown to be an effective drug in earlier studies, however, such high doses as those in ESPS-2 were not studied. The CAPRIE Study showed that clopidogrel 75 mg was slightly more effective than 325 mg of aspirin a day in preventing the combined outcome of ischemic stroke, myocardial infarction and vascular death. An absolute difference of 0.5% was found, i.e. for every 1000 patients treated, clopidrogrel would prevent 5 events more than aspirin. Clopidogrel has a good safety profile as compared against aspirin and ticlopidine, however, it is quite costly.

The newer glycoprotein IIb/IIIa receptor blockers block the final common pathway of platelet aggregation. The safety and efficacy of these drugs in ischemic strokes have still to be established.

The choice of secondary stroke prevention with aspirin alone, aspirin and dipyridamole, ticlopidine or clopidogrel would be dictated by patient's tolerance to these drugs, the availability and costs of these drugs.

Blood pressure control

While there are no large trials on management of blood pressure in the setting of an acute stroke, several recent trials have demonstrated a beneficial effect for treating blood pressure in patients with previous stroke, even in those who do not have hypertension. Rashid *et al*. performed a meta-analysis on the effect of blood pressure lowering in stroke patients. Seven randomized trials with over 15 000 patients using mainly angiotensin converting enzyme inhibitors (ACE), diuretics or beta-blockers were analyzed. There was a significant benefit in reducing recurrent strokes and myocardial events, with no benefit seen in vascular mortality. Two trials, Post-stroke Antihypertensive Treatment Study (PATS) and Perindorpil Protection Against Recurrent Stroke Study (PROGRESS), contributed two-thirds of the data. These trials used either an ACE or a diuretic. In PROGRESS, treatment also appeared more effective in Asian patients and patients with hemorrhagic strokes.

Lipids

Previous trials using non-statin drugs (such as fibric acid derivatives) did not show a reduction in the incidence of stroke. The recent statin trials in patients with cardiovascular disease revealed that long-term statin therapy substantially reduced the risk of recurrent cardiovascular events, including first ever stroke (primary prevention of stroke).

The beneficial effect of statin in stroke patients was confirmed when the Heart Protection Study was published in 2000. Twenty thousand patients were recruited, 65% of patients had established coronary heart disease and 16% had prior stroke. Patients received up to 40 mg of simvastatin daily. Statin treatment led to a 24% reduction in relative risk of major vascular events and a 25% reduction in relative risk of ischemic stroke. The benefit was seen in stroke patients with and without any coronary heart disease. However, this reduction was all attributed to lowered coronary events without an apparent decrease in stroke recurrence (secondary prevention of stroke). There were less carotid revascularization procedures (endarterectomy or carotid stenting) with statin therapy. This landmark paper showed that long-term statin therapy was beneficial in stroke patients. There are ongoing trials to determine if statin therapy in stroke patients will lead to a reduction in recurrent strokes.

Anticoagulation

For patients with a stroke or TIA who have atrial fibrillation, congestive heart failure, or another clear cardiac cause of ischemic stroke, warfarin remains the first choice unless contraindicated. An INR (International Normalized Ratio) of 2–3 is aimed for, except for prosthetic valves, where 3–4 is more appropriate.

The role of anticoagulation in patients with ischemic stroke of non-cardiac source has not been demonstrated.

Carotid endarterectomy

The efficacy of carotid endarterectomy for patients with non-disabling stroke or TIA (within the last 6 months) and an ipsilateral carotid stenosis that is greater than 70% is well documented in 2 large trials. The risk of stroke over 2 years, in the surgical was 9%, compared with 26% in the medical group. Mild to moderate carotid stenosis (0–50%) had no benefit,

the risk of stroke was low and any benefits of surgery was outweighed by the early peri-operative risk. For those with 50–69% stenosis, there was moderate benefit. Over 5 years, the surgical arm had 15.7% stroke while the medical arm had 22.2%, giving an absolute risk reduction of about 6.5%. The risk of peri-operative complications are dependent upon the surgeon's skills, in the North American study (NASCET), the overall peri-operative stroke and death was 6.5%, 1.1% were deaths, 1.8% disabling strokes and 3.7% non-disabling stroke. The effect of carotid endarterctomy was durable over 8 years. The risks of surgery are increased in the presence of hemispheric TIA, contralateral carotid occlusion, ulcerated plaque, lesion imaged on CT or MR on the symptomatic side and left sided procedure. The most common peri-operative medical complication was myocardial infarct. Endarterectomy was 1.5 times more likely to trigger medical complications if patients had history of myocardial infarct, angina or hypertension. The benefit is seen more for men than women. For women there is no benefit if the stenosis is less than 70% and for men when it is less than 50%.

Angioplasty and stenting

Percutaneous transluminal angioplasty (PTCA) and intravascular stenting of the extracranial vessels are being examined for safety, efficacy, and long-term durability as an alternative to carotid endarterectomy and other revascularization procedures. PTCA of intracranial vessels (carotid, middle cerebral artery stem, vertebral artery, and basilar artery) has also been performed in a small number of symptomatic patients refractory to maximal medical treatment. While technically feasible, procedure-related stroke and death rates have been substantially higher than for patients undergoing extracranial procedures. Both extracranial and intracranial PTCA (with or without stenting) are considered investigational, as they have not been fully evaluated in clinical trials.

Primary prevention of stroke

Ideally, the prevention of a first stroke is better than the prevention of a recurrent stroke. The National Stroke Association issued a consensus in 1999, identifying hypertension, myocardial infarct, atrial fibrillation, diabetes, hyperlipidemia, symptomatic severe carotid stenosis, and

life-style factors (smoking and excessive alcohol use) as modifiable primary preventive risk factors which would make an impact on first ever stroke if controlled because of their prevalence.

Hypertension is the most prevalent and modifiable risk factor for stroke. It is more prevalent in the elderly. A 1998 survey in Singapore revealed that 27% of the population aged between 30–69 years old are hypertensive by WHO criteria (blood pressure $\geqslant 140$ systolic or $\geqslant 90$ diastolic). The relative risk of hypertension for stroke is as high as 4 times. While the prevalence of hypertension increases with age, the impact of hypertension may decrease with age. The odds ratio for stroke is 4 at 50, and decreases to 1 at 90 years. Studies have shown that with appropriate treatment, risk of strokes may be reduced by about 40%, irrespective of age. Screening for hypertension in elderly patients, and regular monitoring for good control is recommended.

Up to 3% of acute myocardial infarcts will develop stroke, the risk increases to 6% in those with anterior wall infarctions. The risk is greatest in the first month (30%). Most of these strokes are thought to be embolic, from intracardiac clots, but some strokes are probably due to co-existing cerebral atherosclerotic disease and cardiac hemodynamic factors. The risks for developing intraventricular clots are higher in large myocardial infarcts and congestive cardiac failure. Anticoagulation, to a target INR (International Normalized Ration) of 2.0 to 3.0 has been shown to reduce embolic strokes in patients at high risk for cardiac embolism (poor left ventricular function, atrial fibrillation and intracardiac clots).

Atrial fibrillation may occur in the setting of acute myocardial infarction, and is an independent risk factor. Unlike hypertension, the impact of atrial fibrillation persists with increasing age. Atrial fibrillation in association with mechanical valves, mitral stenosis, intracardiac clots and dilated left ventricles predicts a high risk of stroke. Non-valvular atrial fibrillation (NVAF) is also an important risk factor for stroke. The SPAF studies also identified elderly patients (>75 years) with previous thromboembolism, hypertension and congestive cardiac failure as high risks, with stroke rates at approximately 20% per year. Paroxysmal atrial fibrillation appears to have as great a risk as constant atrial fibrillation. Anticoagulation is generally more effective than antiplatelet therapy in preventing stroke, warfarin reduces stroke by up to 70%, and aspirin by 21%, even after considering the hemorrhagic risks involved. An INR of 2.0 to 4.0 (higher ranges for mechanical valves) is recommended. In

younger patients, less than 65 years of age with no high risks, the annual stroke rate is low, about 1%. Warfarin or aspirin may be considered for such patients.

In the past, the relationship between lipids and stroke was not clear. Many primary prevention trials of cholesterol lowering with statins (hydroxyl methylglutaryl coenzyme A reductase inhibitors −HMG CoA reductase inhibitors), mainly in patients with coronary arterial disease, have revealed risk reduction in stroke about 30–40%. This was not shown with other cholesterol lowering agents. It is postulated that the beneficial effects of statins may involve non-lipid mechanisms, such as plaque stabilization, modification of inflammatory responses, and thrombus formation. The beneficial effect of statins to prevent stroke recurrences in stroke patients without coronary disease has not been demonstrated yet.

In a 1998 survey, 9% of Singaporeans were found to have diabetes mellitus, and 15% with impaired glucose intolerance by WHO criteria. The older population (above 40 years old) and Indians were more commonly affected. Epidemiological data suggests that the relative risk of stroke in diabetics is increased 1.5 times, and studies have shown that tight glycemic control reduces microvascular complications such as neuropathy, retinopathy and nephropathy, but the effect on macrovascular complications like stroke is not so clear. Guidelines have been established for management of diabetes.

The risk of stroke in asymptomatic carotid stenosis ranges from 2–5%. In the Asymptomatic Carotid Atherosclerosis Surgery (ACAS) trial, patients with 60–99% stenosis who were randomized to endarterectomy had an absolute risk reduction of stroke of 5.9% over 5 years compared with medical treatment. The trial used ultrasound screening only to detect carotid disease, to avoid the 0.5–1% risk of stroke with cerebral angiography. The trial surgeons also had very low operative complication rates. The prevalence of asymptomatic carotid stenosis in Singapore is not known, though our experience in this hospital has been that symptomatic carotid stenosis is uncommon, and one may extrapolate (though it may not be accurate) to presume that asymptomatic disease is uncommon. Mass screening for asymptomatic carotid disease is not cost-effective currently. Furthermore, because of the low annual risk of stroke with asymptomatic disease, and the mild benefit with endarterectomy, which could easily be negated with higher surgical complication rates (in ACAS, stroke and deaths was 1.5%), surgery for asymptomatic disease is not

routinely advocated. With symptomatic severe carotid stenosis, 6 endarterectomies can prevent 1 stroke in 2 years. In asymptomatic patients, 67 endarterectomies are required to prevent 1 stroke.

Life-style factors such as smoking, excessive alcohol use and sedentary life have been shown to be associated with stroke. Studies reveal that the relative risk of stroke for smokers is 1.5 times, and increases with heavy smokers. Even passive smokers are affected. A J-shaped relationship exists between alcohol use and stroke, with the risk of ischemic and hemorrhagic stroke being lowest in mild to moderate drinkers. Alcohol may exert a protective influence on coagulopathy and lipids. Heavy consumption is associated with higher risks of stroke. Regular exercise (at least 3–5 times weekly) has well-established benefits for prevention of cardiovascular disease.

Table 4 Risk Factors for First Ischemic Stroke

Well-documented, modifiable risk factors
Transient ischemic attack
Asymptomatic carotid stenosis
Hypertension
Cardiac disease
- Atrial fibrillation
- Infective endocarditis
- Mitral stenosis
- Recent large myocardial infarct

Cigarette smoking
Sickle cell disease

Less well-documented risk factors
Cardiac diseases
- Cardiomyopathy
- Segmental wall motion abnormalities
- Non-bacterial endocarditis
- Mitral annular calcification
- Mitral valve prolapse
- Valve strands
- Aortic stenosis
- Patent foramen ovale and atrial septal aneurysm
- Spontaneous echocardiographic contrast

Well-documented, potentially modifiable risk factors
Diabetes mellitus
Hyperhomocysteinemia
Left ventricular hypertrophy

Non-modifiable risk factors
Age
Gender
Race
Hereditary factors

- Hyperlipidemia
- Oral contraceptives
- Excessive alcohol
- Physical inactivity and obesity
- Elevated hematocrit
- Hyperinsulinemia
- Migraine

REHABILITATION AND DISCHARGE PLANNING

Early institution of rehabilitation efforts, including language, swallowing, physical, and occupational therapy is advised by expert consensus, rather than based on proven data. Early mobilization of patients can reduce the incidence of deep vein thrombosis and aspiration pneumonia. Whether a delay in the institution of rehabilitation therapies reduces the level of recovery has not been proven. Patient and family education about stroke, recovery, and prevention are important components of the rehabilitation process.

Ideally, discharge planning should begin on the day of admission and should incorporate a consensus from all involved persons. The specifics of discharge planning will depend on the extent of neurologic deficit or involvements the patient experiences.

Accurate assessment of cognitive and functional abilities is important in determining how dependent the patient will be at discharge and what resources will be necessary to employ at discharge. Cognitive assessment should include these key aspects: attention, orientation, memory, language, reasoning, judgment, spatial skills, motor coordination, and social skills. Depression is the major emotional disturbance seen post-stroke. Early recognition of depression is important when assessing a patient's motivation for rehabilitation and recovery. Functional status includes aspects of cognition, but also includes physical motor function, mobility, speech and language function, and measures of activities of daily living. All of these aspects are crucial to stroke survivors to help them become integrated into a functional family and social life.

62

Parkinson's Disease: Diagnosis and Treatment

Tan Eng King and John Thomas

INTRODUCTION

Parkinsonism is manifested by a combination of the following five main features, namely rest tremor, rigidity, bradykinesia, loss of postural reflexes and freezing phenomenon.[1,2] Parkinson's Disease (PD), an idiopathic and commonest form of parkinsonism is a progressive neurodegenerative disease characterized by loss of dopaminergic cells in the substantia nigra pars compacta and presence of Lewy bodies. It affects predominantly the elderly population. The mean age of PD patients seen in tertiary referral centers in Singapore was about 65 years old, similar to reports in many countries. With the advancing age of the population, the prevalence of the disease will increase. The mortality is 2 to 5 times as high among affected persons compared to age-matched controls. In fact, neurodegenerative diseases (PD, motor neuron disease and dementia) are projected to surpass cancer as the second most common cause of death among the elderly by the year 2040 in the United States.[3] This projection is relevant to many countries in Asia due to rapidly aging population.

In the West, the prevalence of PD is estimated to be approximately 1 : 1000, and annual incidence around 1 : 10 000.[4] The exact prevalence of PD in Singapore is not known, but estimated to be around 1 : 2000. In a survey in Singapore General Hospital, we found about 800 admissions involving 500 PD or parkinsonism patients over a 2.5-year period.

GENE/ENVIRONMENTAL ETIOLOGY

The relative role of genetic and enviromental factors in the pathogenesis of Parkinson's disease (PD) has been the focus of research and debate.[5] The discovery of MPTP (methyl-4-phenyl-1,2,3,6-tetrahydropyridine)-induced Parkinsonism in intravenous drug users[6] led to numerous epidemiologic studies evaluating potential enviromental causative agents, though no specific agent has thus far been identified. Studies of genetic polymorphisms of candidate genes have thus far been inconclusive.[7–9] Eight gene loci have been identified by linkage analysis on human chromosome 4q21-23 (PARK 1), 6q25-27 (PARK 2), 2p13 (PARK 3), 4p15 (PARK 4), 4p13 (PARK 5) and more recently 1p35-p36 (PARK 6), 1p36 (PARK 7) and 1p (PARK 8).

In 1997, Polymeropoulos *et al*.[10] found a missense mutation A53T in the *alpha-synuclein* gene in a family with autosomal dominant (AD) mode of inheritance for parkinsonism. This mutation was subsequently found in other families, all were of Greek and Southern Italian origin. In 1998, Kruger *et al*.[11] described a second mutation A30P in the same gene in a German family. However, these mutations were not found in sporadic and most familial PD patients, indicating that these are extremely rare. A mutation in the ubiquitin carboxy-terminal hydroxlase gene on 4p13 has also been identified in a family with AD inheritance.

Autosomal recessive parkinsonism was initially described in Japanese families. While they manifest typical signs of PD, they tend to be associated with young age of onset, diurnal fluctuations, dystonia, early but severe levodopa-induced dyskinesias, and slow progression of disease. Neuropathological features thus far did not show any Lewy bodies, a hallmark of PD. In 1997, investigators mapped the locus to chromosome 6q25.2-27 (PARK 2) in consanguineous Japanese families. Subsequent linkage analysis found non-Japanese PARK 2 families in Europe, the United States, and the Middle East. In 1998, Kitada *et al*.[12] from Juntendo University, Tokyo, identified homozygous deletions of exon 4 or

exons 3-7 of the *Parkin* gene in 4 Japanese families with autosomal recessive parkinsonism.

DIAGNOSIS

A definite diagnosis of PD can only be made pathologically. A clinical diagnosis can frequently be made if a patient presents with ipsilateral rest tremor, bradykineisa and rigidity, and these symptoms and signs progress to involve the contralateral side, and the patients respond well to levodopa. However, even amongst specialists, misdiagnosis rate can be as high as 12.5%.

In addition to motor dysfunction, PD patients may develop neuropsychological complications such as depression and dementia. Hallucinations or confusion are usually associated with later stages of the disease and with use of high doses of levodopa or other anti-parkinsonian medications. In a preliminary study, we found 26% and 16% of our PD patients exhibit depression and impaired cognition respectively. Other seldom emphasized problems include autonomic dysfunction manifested by urinary or bowel disturbances and impotence.

DIFFERENTIAL DIAGNOSIS

A high index of suspicion is required to exclude treatable secondary causes in all patients who present with parkinsonism (Table 1). It is important to take a history of encephalitis, exposure to heavy metals and toxins, taking of anti-psychotic medications (e.g. neuroleptics). On clinical examination, one should particularly assess for the presence of Kayser–Fleischer Ring (Wilson's disease), supranuclear gaze abnormality (supranuclear gaze palsy), postural hypotension, pyramidal tract signs, cerebellar signs (multiple system atrophy), dementia (dementia of Lewy body), limb apraxia (cortico-basal ganglionic degeneration). If there is predominantly lower body parkinsonism, no tremor, brisk reflexes, and imaging showed microvascular ischemic changes in the subcortical region or basal ganglia, "vascular parkinsonism" is a likely diagnosis. Recently, familial spinocerebellar ataxias have been found to present with pure parkinsonism. Essential tremor (ET) is manifested by postural tremor involving the limbs, head, trunk and voice, with no bradykinesia and rigidity.

Table 1 Causes of Parkinsonism

1) Idiopathic	Parkinson's disease
2) Secondary parkinsonism	Drugs (e.g. neuroleptics) Heavy metals and toxins (e.g. copper, iron, manganese, carbon dioxide) Infections (e.g. encephalitis) Vascular (e.g. multi-infarcts) Others (e.g. head trauma)
3) Parkinson's-plus syndromes	Progressive supranuclear gaze palsy Multiple system atrophy Cortico-basal ganglionic degeneration Dementia of Lewy body Others (e.g. familial spinocerebellar ataxia, etc.)

INVESTIGATIONS

There is generally little need for investigations in typical PD patients to establish the diagnosis. However, if there are suggestions of secondary causes of parkinsonism in history and examination, checking for ceruloplasmin, thyroid function, toxicology studies, imaging (Magnetic Resonance Imaging), electrophysiologic and neuropsychological testings may be warranted. Functional imaging studies (e.g. Positron Emission Tomography) may be useful to detect asymptomatic cases, monitor progress after PD transplant surgery, and assist in diagnosis in some patients.

MEDICAL TREATMENT

The replacement of dopamine deficiency forms the fundamental basis of drug treatment in PD. Pharmacologic therapy is only initiated if activities of daily living are affected. Levodopa remains the cornerstone of drug treatment in PD.[1,13] It is usually prescribed in combinations with a decarboxylase inhibitor (which prevents the peripheral breakdown of levodopa so that more can be available in the brain) in the form of *Madopar®* or *Sinemet®*. Slow release formulations such as *Madopar HBS®* and *Sinemet CR®*, and fast release formulation, *Madopar Dispersible®* are also available. Due to the disabling complications as dyskinesias and motor fluctuations associated with long-term levodopa use, it is prudent to delay the use of

Table 2 Drugs for Parkinson's Disease Patients

1) Levodopa	
2) Dopamine agonists	Ergots Bromocriptine Pergolide Carbergoline Lisuride Non-ergots Pramipexole Roprinole Piribedil Apomorphine
3) Catechol-O-methyltransferase inhibitors	Entacapone Tolcapone
4) Monoamine oxidase inhibitors	Selegiline
5) Anti-cholinergics	Benzhexol
6) NMDA receptor anatagonist	Amantadine
7) Anti-oxidants	Vitamins A, C, E
8) Anti-depressants	Ttricyclics (e.g. amitrptyline) SSRI (e.g. fluoxetine)
9) Muscle relaxants	
10) Laxatives	

this drug. This is especially so in the younger group of patients. Should levodopa be started, it is advisable to step up the dose gradually to a minimal dose which can produce maximal results with little side-effects. For instance, *Madopar*® is prescribed at 62.5 mg twice a day and subsequently over days or weeks. A course of levodopa up to 1–2 g/day may be needed before one can conclude that patient has no response to the drug. Regular checking for postural hypotension and adjusting the timing of levodopa to reduce nausea and vomiting may be required. Domperidone, given half an hour before levodopa is useful in overcoming gastrointestinal symptoms associated with levodopa. Prescribing levodopa at a lower dosage but at more frequent intervals can alleviate levodopa-induced dyskinesias. Wearing off periods can be reduced with the longer-lasting levodopa formulations.

Dopamine agonists (DA) have been shown to be effective as monotherapy in early stages of PD and as an adjunctive treatment to levodopa in advanced PD.[13] While DA are less effective than levodopa in the symptomatic treatment of advanced PD, they are increasingly used early, particularly in young-onset PD because of their levodopa sparing effects

and their putative role as neuroprotective agents. Whether DA are comparable to levodopa in early PD is currently being investigated. Recent published trials comparing the ropinirole and pramipexole monotherapy with levodopa suggest that the risk of dyskinesias is less with the DA than with levodopa, but motor improvement is more robust with levodopa monotherapy than with the DA agonists.

DA agonists directly activate dopamine receptors, bypassing the presynaptic synthesis of dopamine. There are two main classes of dopamine receptors: the D1 class (comprised of subtypes D1 and D5), linked to the enzyme adenylate cyclase, and the D2 class (comprised of subtypes D2, D3, and D4), coupled to G proteins that inhibit adenylate cyclase.

Bromocriptine, an ergot compound, was the first to be introduced more than 25 years ago. Currently there are a number of DA available including ergots such as pergolide, lisuride and cabergoline, and non-ergots such as piribedil, apomorphine, ropinirole, and pramipexole. Bromocriptine and pergolide are frequently referred to as "old" DA, and pramipexole and ropinirole as "new" DA, simply because the latter two DAs have only been introduced in the market in recent years, and are still not available in many countries, particularly in Asia. Recently, hypersomnolence in the form of "sleep attacks" was a concern for patients on the non-ergots (e.g. pramipexole). However, subsequent studies suggest that this adverse effect is likely a class effect.[14] Other potential complications with DA therapy include nausea, postural dizziness, hallucinations, retroperitoneal fibrosis, and peripheral edema.[15]

Tolcapone and Entacapone, both catechol-O-methyltransferase (COMT) inhibitors, extend the action of levodopa by preventing its breakdown. They are used in conjunction with levodopa and effective for patients who develop motor fluctuations. Side-effects include worsening of the dyskinesias associated with peak doses of levodopa, hypotension, constipation and urine discoloration. Fatal cases of liver toxicity have been reported with Tolcapone and hence monitoring of liver function is recommended for this drug.

Amantadine, a NMDA receptor antagonist has been demonstrated to be effective for levodopa-induced dyskinesias. Selegiline, a monoamine oxidase inhibitor is frequently used in early stages of the disease. It has been suggested that it may have neuroprotective potential and helps delay the introduction of levodopa. Anti-cholinergics (such as benhexol) are used for tremor and rigidity. However, they have to be used with

caution, especially in the elderly as they can cause confusion, cognitive impairment, and hallucinations. They can also aggravate constipation, urinary retention and lead to an increase intraocular pressure.

Anti-depressants such as selective serotonin reuptake inhibitor (SSRI) and tricyclics compounds are useful to treat depressive symptoms in PD. However side-effects like acute dystonia, akathisia, and aggravation of parkinsonism may occur in some patients. Putative neuroprotective agents such as vitamin E or other anti-oxidants are also frequently used by some physicians, though their therapeutic effect in PD has not been proven.

It cannot be overemphasized that physiotherapy, speech therapy, psychotherapy are important in a number of patients. Support from family members, caregivers and friends, and participation in educational programs provide the necessary social framework to support the medical care by the physicians.

SURGICAL TREATMENT

Surgery for Parkinson's disease is enjoying a significant renaissance around the world. We are now in a position to significantly improve the lives of the patients who are worst afflicted by the disease, in whom medical therapy has reached its limits, who would previously have been confined to their beds and left to die. There are several reasons for this renewed interest.

1) The widespread recognition that there are limits to the benefits of medical treatment in many patients with the disease. Problems arise from declining response to medication with time, from unpredictable responses such as on-off phenomena and from side-effects of medication such as dyskinesias.
2) The vast improvement in our understanding of the pathophysiology in the circuitry of the basal ganglia that underlie many of the symptoms of Parkinson's disease.
3) The fantastic improvement in MRI brain imaging that allows us to visualize the many component subnuclei in the basal ganglia that are involved in the disordered circuits. This makes it possible to target these subnuclei.
4) The significant advancement in neurosurgical technical accuracy and safety that arose with the introduction of computer workstations for surgical planning.

There are two standard approaches in the surgical treatment of Parkinson's disease. The traditional approach is to silence the subnuclei that are hyperfunctioning by destroying them with heat (thermocoagulation). The alternative approach achieves the same result by depolarizing the relevant subnuclei with high frequency stimulation (deep brain stimulation).

The main targets of surgery are the subthalamic nucleus (STN), the medial globus pallidus (GPi) and the ventrointermediate nucleus (Vim) of the thalamus, which are very closely related anatomically and physiologically. Intervention in the STN and GPi improves the symptoms of tremor, rigidity and bradykinesia contralaterally. Drug induced dyskinesias are also significantly improved. Interruption of the Vim nucleus of the thalamus mainly reduces contralateral tremor by 80%. The surgical target is decided upon after careful study of the patient's symptoms by the movement disorder team. Bilateral problems will require bilateral surgery.

Post-surgery the patient ideally will achieve his best on-drug performance with less medication, spend about 70–80% of the day at this maximum performance status without significant on-off phenomena and be free of drug-related dyskinesias. This usually translates to a very significant improvement in their independence and quality of life.

In choosing between thermocoagulation and deep brain stimulation (dbs), the usual considerations are:

1) Thermocoagulation lesions are irreversible once made. The effects of dbs are scalable with the amount of current delivered. This applies to the desired effects as well as the unwanted effects of surgery.
2) Bilateral surgical destructive lesions (thermocoagulation) can pose special problems because of the increased incidence of unwanted side-effects such as speech, swallowing and cognitive deficits. These problems can be avoided by using dbs.
3) Thermocoagulation lesions are maintenance-free once made. They are relatively "low-tech". Optimization of stimulation parameters after dbs placement is labor intensive and requires frequent doctor visits. About 10–20% of dbs patients may face problems related to wire-breakage or disconnection. DBS patients will also face the cost of the implants and the need to replace the batteries every 3–5 years.

For patients in Singapore deep brain stimulation is usually the better choice. Thermocoagulation best serves patients who live far away from the tertiary medical centers and those in third-world countries.

Patients selection for surgery is a delicate task. The patients will have been on follow-up for some time with the movement-disorder clinic and will have been tried on a number of drugs. Those selected for surgery will, in general, satisfy the following criteria:

1) They will have clinically confirmed Parkinson's disease.
2) They will have disabling symptoms despite maximal medical therapy. However they should be medication responsive and not bedridden.
3) They will not be suffering from severe dementia.
4) They will have normal MRI brain studies.
5) They will not have any significant concurrent medical illness.
6) They will have given informed consent to surgery.

Prior to surgery there will be a period of detailed inpatient assessments where the patient's functional status while on medication and off medication is objectively documented on validated scoring systems. Video-recordings of these assessments are made to allow pre- and post-op comparisons as well as blinded assessments. The surgery is performed under local anesthesia using stereotactic neurosurgical techniques. The surgical team will include the movement disorder neurologist treating the patient. The patient will have been off medication for at least 12 hours prior to surgery. The patient will be mildly sedated, but able to speak and perform movements during the operation. The placement of the surgical probe in the correct nucleus is confirmed by recording the electrical activity of the nucleus in response to joint movement (both active and passive) and by the clinical effect of electrical stimulation delivered to the target.

The surgery is safe with an overall risk of adverse outcome in the region of 2–4%, which includes both minor and major complications (which include paralysis, vegetative survival and death). The complications can be divided into those related to general risks of stereotactic procedures as well as the general medical risks of patients undergoing surgery.

Most patients tolerate surgery extremely well. The usual patient will be ambulant the day after surgery. The hospital stay is usually 3–4 days. The beneficial effects of lesion surgery are apparent immediately post-operation. If deep brain stimulators are placed the stimulation is usually turned on a couple of weeks after surgery to allow objective assessment in the outpatient clinic. The patient's medical therapy will be modified once the stimulation parameters have been optimized.

Patients with advanced Parkinson's disease are now in a much better position than they were even 10 years ago. The combination of new medications and effective surgical therapy can deliver them from the fate of immobility, invalidism and an early death.

REFERENCES

1. Jankovic J, Marsden CD, Therapeutic Strategies in Parkinson's Disease, in Jankovic J, Tolosa E (eds.), *Parkinson's Disease and Movement Disorders*, 3rd ed., Williams & Wilkins, Baltimore, pp. 191–220, 1998.
2. Tan EK, Parkinson's disease surgery: advances and future strategies, *Int J Clin Pract* **53**:623–626, 1999.
3. Lilienfield DE, Perl DP, Projected neurodegenerative disease mortality in the United States, 1990–2040, *Neuroepidemiology* **12**:219–228, 1993.
4. Ben–Shlomo Y, Sieradzan K, Idiopathic Parkinson's disease: epidemiology, diagnosis and management, *Brit J Gen Pract* **45**:261–268, 1995.
5. Langston JW, Epidemiology versus genetics in Parkinson's disease: progress in resolving an age old debate, *Ann Neurol* **4**(3suppl1): S45–52, 1998.
6. Langston JW, Ballard P, Tetrud JW, Irwin I, Chronic parkinsonism in humans due to a product of meperidine-analog synthesis, *Science* **219**:979–980, 1983.
7. Tan EK, Khajavi M, Thornby I, Nagamitsu N, Jankovic J, Ashizawa T, Variability and Validity of Polymorphism Association Studies in Parkinson's disease, *Neurology* **55**:533–539, 2000.
8. Tan EK, Nagamitsu S, Matsuura T, Khajavi M, Jankovic J, Ondo W, Ashizawa T, Alcohol dehydrogenase polymorphism and Parkinson's disease, *Neurosci Lett* **305**(1):70–72, 2001.
9. Tan EK, Matsuura T, Nagamitsu S, Khajavi M, Jankovic J, Ashizawa T, Polymorphism of NACP-Rep1 in Parkinson's disease: an etiologic link with essential tremor? *Neurology* **54**:1195-1198, 2000.
10. Polymeropoulos MH, Lavedan C, Leroy E, *et al*., Mutation in the alpha-synuclein gene identified in families with Parkinson's disease, *Science* **276**:2045–2047, 1997.
11. Kruger R, Kuhn W, Muller, *et al.,* Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease, *Nat Genet* **18**:106–108, 1998.
12. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, *et al*., Mutations in the Parkin gene cause autosomal recessive juvenile parkinsonism, *Nature* **392**:605–608, 1998.
13. Tan EK, Jankovic J, Choosing dopamine agonists in Parkinson's disease, *Clin Neuropharmacol* **24**:247–253, 2001.

14. Tan EK, Lum SY, Fook–Chong SM, Teoh ML, Yih Y, Tan L, Tan A, Wong MC, Evaluation of somnolence in Parkinson's disease: comparison with age- and sex-matched controls, *Neurology* **58**:465–468, 2002.

15. Tan EK, Ondo W, Clinical characteristics of pramipexole-induced peripheral edema, *Arch Neurol* **57**:729–732, 2000.

This page intentionally left blank

63

Alzheimer's Disease and Other Dementias

Alexander P. Auchus and Christopher Chen

DEMENTIA

Dementia is a clinical syndrome defined by an acquired and persistent loss of intellectual function severe enough to interfere with social or occupational functioning. Dementia generally involves global or multifocal impairment of higher brain functions, affecting memory (amnesia), language (aphasia), visuospatial and visuoconstructive abilities, skilled motor coordination (apraxia), abstraction, judgment, insight, personality, and other functions to variable extents. Dementia is to be distinguished from mental retardation in that it is acquired, and from delirium in that it is usually persistent and accompanied by a normal level of alertness. Dementia is also distinguished from syndromes of focal higher brain dysfunction, such as aphasia or agnosia.

Although certain cognitive changes occur with advancing age, dementia is not a normal consequence of aging. It is estimated that about 5–10% of population over age 65 in the United States are demented. This figure rises to at least 25% to 30% by age 85, and some studies even suggest a prevalence of almost 50% by this age. Dementia is an expensive

condition. The annual cost for care of dementia victims in the US tops US$30 billion. The increasing percentage of elderly in the population, coupled with the high costs of caring for demented patients, will make dementia an even more devastating medical, social, and economic problem in the near future.

The etiologies of dementia are many. Alzheimer's disease accounts for over half of cases in most series from developed countries (USA, Canada, United Kingdom, Western Europe, Scandinavia). Vascular dementia (formerly termed "multi-infarct dementia") is usually the next most common etiology. Vascular dementia may be more common in Asia than in the West. Other etiologies include metabolic disorders (chronic renal, hepatic, pulmonary, or cardiac failure), toxic disorders (especially iatrogenic), nutritional disorders (B-1, B-12, folate deficiencies), alcoholism, movement disorders (Parkinson's, Huntington's, Wilson's), psychiatric disease (especially depression), hydrocephalus, chronic subdural hematoma, traumatic brain injury, anoxic brain injury, CNS infections (including HIV), and CNS neoplasms (metastases, glioma, meningioma). Treatable causes of dementia are found in 20% of patients under age 65, as opposed to only 5% over age 65. Depression is also common among the elderly, and disturbances of thinking and memory frequently accompany depression.

Detecting dementia may be difficult, especially in the higher functioning patient. Though some families may deny even gross cognitive changes in their relatives, history from family and close friends usually provides the most reliable data on the patient's cognitive status.[1,2] Inquiries into how patients spend their day, whether any previously independent tasks have had to be subsumed by others (such as shopping, managing personal finances, and driving) are essential aspects of the history in a patient suspected of having dementia. Consideration should also be given to complaints of anxiety, low mood, and to vegetative signs of depression (loss of appetite, loss of libido, and sleep disturbance). A careful neurologic examination and mental status examination are also essential and may uncover relevant findings. For example, a dementia characterized by impairments in memory, language, and praxis, without any abnormal motor findings is most commonly due to Alzheimer's disease. In contrast, a patient with forgetfulness, slowing of mental processes, as well as the presence of extrapyramidal signs (e.g. rigidity, bradykinesia) suggests Parkinson's disease, another movement disorder,

or hydrocephalus. Dementing illnesses due to metabolic and toxic processes are often accompanied by neuropathy. Vascular dementia is usually associated with asymmetric pyramidal findings (such as increased deep tendon reflexes, hemiparesis, Babinski signs) and a stepwise deteriorating course. Rapidly progressive dementias suggest the possibility of infectious etiologies, including Creutzfeldt–Jakob disease. Since precise diagnosis is essential for guiding treatment options and family counseling, referral to specialty clinics should be considered.

ALZHEIMER'S DISEASE

Alzheimer's disease is the most common cause of dementia worldwide. The age of onset in AD is frequently the early seventies. The memory disturbance begins insidiously, affecting short-term or recent memory (e.g. orientation, news events, etc.) and, to a lesser extent, remote memory. The patient has increasing difficulty with activities of daily living, and changes in personality often develop — apathy being most common. Impairment of higher cortical function occurs either early or late in the course, though each modality is not affected equally. Language dysfunction begins with word-finding and naming difficulties, and progresses to fluent aphasia with impaired comprehension. Visuospatial dysfunction is manifest by spatial disorientation. Motor signs develop late, with the variable appearance of extrapyramidal rigidity, gait impairment, frontal release signs, and spasticity with hyperreflexia. Myoclonus, or rarely seizures, may develop late in the course. Behavioral problems commonly arise during the course of AD and include delusions, hallucinations, affective disturbances, personality changes, anxiety, agitation, aggression, wandering, sleep-wake cycle disruption, and catastrophic reactions. Electroencephalography (EEG) may be normal but will frequently reveal a diffuse slowing down. The CT or MRI also may be normal or may show diffuse cortical and hippocampal atrophy.

The brain weight is usually low with marked atrophy of the temporal, parietal, and frontal lobe cortices, and particularly the hippocampal formation. Involvement of the association areas is most marked, with sparing of the primary motor, sensory, and visual cortex. The ventricles are usually dilated (hydrocephalus "*ex vacuo*"). Hallmark pathological findings in AD include neurofibrillary tangles (NFT) and senile plaques (SP) in association with loss of neurons and synapses. NFT are abnormal

accumulations of twisted neurofibrils known as paired helical filaments composed of a hyperphosphorylated form of the tau protein. SP are spherical structures composed of degenerating neurites surrounding a central core of amyloid. The amyloid core consists primarily of insoluble deposits of a 42 amino acid peptide, derived from a larger amyloid precursor protein. In addition, amyloid deposition is found in the blood vessels ("amyloid angiopathy") of the meninges and cortex. While small numbers of SP and NFT are seen in the hippocampus and cortex in normal aged brains, in AD they are more numerous and widespread. They occur predominantly in the frontal, temporal, and parietal cortex and in the hippocampus and amygdala. Outside of cortex and hippocampus, select subcortical nuclei are also critically affected. The nucleus basalis of Meynert (nbM), a basal forebrain nucleus, and the major source of cholinergic innervation for the neocortex and hippocampus, almost always displays significant cell loss, NFT and SP. The dorsal raphe nucleus and to a lesser extent, the locus ceruleus, are also affected in many cases.

Tremendous interest has focused on the neurochemical alterations in AD, motivated in part by the prospects for neurotransmitter replacement therapies as have been successfully employed in Parkinson's disease. Brain acetylcholine loss in AD appears to be to pathological involvement of the subcortical neurons in the nucleus basalis of Meynert. The loss of cortical cholinergic innervation is of interest because most evidence has favored a key role of the cholinergic system in memory dysfunction. For example, the degree of depletion of cortical cholinergic markers correlates with the amount of cognitive impairment. Also, anticholinergic medications can cause cognitive dysfunction that resembles AD, and lesions of the basal forebrain area produce memory and related deficits in experimental animals and humans. Moreover, cholinomimetics are proven efficacious therapy for AD, as further discussed below.

Inheritability of at least some forms of Alzheimer's disease has been recognized for decades. The term "familial Alzheimer's disease" (FAD) is used primarily to refer to families in which several members show the disease phenotype, usually with a segregation pattern that suggests autosomal dominant inheritance. The age of onset varies among different families, with symptoms in some families presenting earlier than age 60, and in fact, often in the forties and early fifties (early-onset FAD), and other families with onset after 60 years (late-onset FAD). Myoclonus, seizures, a more rapid course, and occasionally other clinical features

have been suggested to be more common in early-onset FAD, but otherwise FAD is clinically and histologically similar to sporadic cases. Sporadic cases of AD, with no evidence of familial clustering or autosomal dominant pedigrees, are responsible for the bulk ($>95\%$) of AD cases. Extraordinary advances in understanding the genetic and molecular basis of AD have occurred. Several chromosomal loci have been identified which are linked to various forms of AD. For example, early onset FAD can occur with mutations of the amyloid precursor protein gene on chromosome 21, the presenilin 1 gene on chromosome 14, or the presenilin 2 gene on chromosome 1. Loci on chromosome 21 are associated with late-onset FAD and sporadic AD (see below).

An important advance has been the discovery of Apolipoprotein E (ApoE) as a major susceptibility gene for late-onset FAD and sporadic AD. There are three ApoE alleles, e2, e3, and e4, which differ only in the positions of two amino acid. The e4 allele of the ApoE gene has been found to be strongly associated with both sporadic AD and with late-onset FAD, such that the inheritance of the e4 allele appears to be a powerful risk factor for AD. The presence of the e4 allele is not only associated with an increased chance of ultimately developing AD, but is also associated with an earlier age of disease onset. However, the fact that about 20% of the late-onset AD patients have no e4 suggests that other factors are also involved in the pathogenesis of the disease. The predictive quality of the e4 allele has been confirmed in a number of studies in Canada, Europe and Asia. Although ApoE appears to play a major role in the neurobiology of AD, and the e4 allele is the most important risk factor yet identified, its use as a genetic marker for diagnostic or predictive testing is still debated. As mentioned earlier, there are many cases of AD without the e4 allele, and many individuals who carry the e4 allele without developing AD.

Alzheimer's disease is accompanied by a reduction in brain levels of the neurotransmitter acetylcholine. Pharmacological agents (cholinesterase inhibitors) that inhibit the enzyme responsible for metabolizing acetylcholine, produce an effective increase in brain acetylcholine neurotransmission, and have been shown clinically to improve both cognitive and global function in AD patients. Four cholinesterase inhibitors (tacrine, donepezil, rivastigmine, and galantamine) have been approved by regulatory bodies for use in AD. Efficacy has been demonstrated in multicenter, randomized, double-blind, placebo-controlled clinical trials for each of

these cholinesterase inhibitors. The magnitude of cognitive effect is similar for the various cholinesterase inhibitors, and corresponds roughly to the amount of cognitive decline expected to occur in AD patients over approximately nine months. Some studies have also shown cholinesterase inhibitors to improve certain behavioral problems and functional performance on activities of daily living in AD patients. The toxicity of cholinesterase inhibitor therapy varies considerably between the various agents. Tacrine can produce severe hepatotoxicity, and its use is no longer recommended. Gastrointestinal side-effects (nausea, vomiting, diarrhea) are the most common side-effects and are dose related. Cholinesterase inhibitors are started at a low dose, and doses are subsequently increased as tolerated.

AD is a neurodegenerative disease in which additional neurons continue to die throughout the course of the illness. Interventions designed to slow or stop this neurodegeneration might retard the progression of AD. One intervention using the antioxidants selegiline and alpha-tocopherol (a form of vitamin E) has received formal scientific study.[3] In a two-year randomized clinical trial, manifestations of severe dementia (loss of basic activities of daily living, institutionalization, death) were significantly less likely to develop in AD patients receiving selegiline (5 mg BD), alpha-tocopherol (1000 IU BD) or the combination of the two compared to AD patients receiving placebo. The median delay in appearance of one of the manifestations of severe dementia in the active treatment groups was approximately 8 months compared to the placebo group. These treatments did not improve patients' cognition; rather, treatment appeared to delay the appearance of manifestations of severe dementia. Treatment with selegiline or alpha-tocopherol was well tolerated. These preliminary findings, together with the favorable safety and cost profile of alpha-tocopherol, suggest that it can be recommended for most patients with AD.

One randomized clinical trial using a purified extract of the ginkgo biloba plant has shown a very small beneficial effect (approximately one-third of the effect size generally obtained with cholinesterase inhibitors) on cognitive function in AD patients. However, this study did not show any benefit on patients' global function (in contrast, cholinesterase inhibitors do improve global function). These facts, together with the troubling issues of uncertain purity and potency of health food store acquired ginkgo products, prohibit any recommendation of ginkgo biloba in the treatment of AD patients.[4]

VASCULAR DEMENTIA

Vascular dementia (VaD) is a controversial entity, as it is often difficult to ascertain if the cerebrovascular lesions seen at autopsy or radiologically are responsible for the patient's dementia. A patient's dementia cannot be ascribed to VaD merely because one or more strokes are present in any site or size. In addition, there is much confusion over the type and extent of cerebrovascular disease that results in dementia. It appears that dementia may be seen only when a certain amount of brain volume loss occurs in certain critical locations; usually this involves multiple bilateral lesions. VaD generally affects a younger population compared to AD. The following clinical features suggest the presence of a vascular dementia: 1) abrupt onset; 2) stepwise deterioration; 3) focal neurological signs; and 4) history of stroke. In addition, at autopsy, a mixed picture of both AD and VaD can be seen. Subcortical ischemic vascular dementia (SIVD) is a common form of VaD (especially in Asia) occurring in patients with persistent hypertension and systemic vascular disease.[5] Patients may lack a history of acute strokes and instead, present with progressive dementia and prominent gait impairment. Pathologically, there is ischemic demyelination in the cerebral white matter and associated lacunar infarctions in the basal ganglia, thalami, and corona radiata. One form of vascular dementia is that due to recurrent cerebral emboli. This usually presents as sequential strokes, often involving multiple arterial distributions. When multiple cerebral emboli are suspected, identification and treatment of the embolic source (valvular heart disease, mural thrombus, etc.) will halt the progression of the dementia, and some spontaneous recovery may follow.

OTHER DEMENTIAS

Parkinsonian Dementias

Dementia complicates Parkinson's disease (PD) in 35–55% of patients. The dementia of PD is characterized by slowed mental processing, impaired problem solving, decreased spontaneity, and visuospatial dysfunction. Aphasia and severe amnesia are uncommon in the dementia of PD. Other diseases that may produce dementia with associated parkinsonian features include progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and dementia with Lewy bodies (DLB).

DLB deserves special mention, as it is often quoted to be the second most common degenerative dementia (after AD). DLB is recognized by the presence of prominent visual hallucinosis, prominent fluctuations in alertness, and spontaneous motor signs of parkinsonism.[6,7]

Frontotemporal Dementias

Pick's disease is the most well known of the frontotemporal dementias (FTD). The FTDs are neurodegenerative disorders preferentially involving the frontal and temporal lobes that present in the fifth or sixth decade (earlier than AD). They generally produce one of three clinically recognizable patterns: 1) a disinhibition syndrome characterized by prominent changes in personality and judgment; 2) progressive non-fluent aphasia; and 3) "semantic" dementia, in which patients lose their knowledge of words and their meanings. Short-term memory, calculations, and visuospatial skills are often initially spared. FTD may be suspected when the CT or MRI shows characteristic atrophy selectively involving the frontal and temporal lobes.[8]

"REVERSIBLE" DEMENTIAS

Although AD and VaD may be the most common causes of dementia, there are other causes which deserve special mention since dementia due to these etiologies may sometimes be reversible. Most reversible dementias can be identified with a thorough history and examination coupled with the following ancillary studies: a full blood count, an electrolyte and metabolic panel, a thyroid stimulating hormone level, and a vitamin B-12 level. A brain imaging study (either CT or MRI) is also recommended to exclude mass lesions, hydrocephalus, and to identify multiple brain infarctions. When clinically indicated, additional laboratory testing (e.g. HIV, EEG, LP) or empirical therapeutic trials (e.g. antidepressants) may be useful.[1,2]

Metabolic Disorders

The brain is very sensitive to disruptions in metabolic homeostasis. Rapid changes in the body's metabolic parameters commonly produce an acute confusional state (delirium). When metabolic changes occur

more gradually, a dementia-like state may develop. As in acute confusional states, metabolic dementia is characterized by prominent inattention, and by fluctuations in cognitive performance. Other clues suggesting a metabolic (or toxic) cause are prominent disturbances of motor function including postural tremor, myoclonus, and asterixis. Common metabolic derangements that can produce dementia include hypoxia, uremia, hyponatremia, hypercalcemia, hyperammonemia, Wernicke–Korsakoff syndrome (B-1 deficiency), vitamin B-12 deficiency, hypothyroidism, and cerebral hypoperfusion (e.g. severe congestive cardiac failure). Correction of the underlying medical illness responsible for these metabolic derangements often halts the progression of cognitive impairment and sometimes reverses the impairment completely.

Toxic Dementia

Numerous toxins can impair brain function and produce dementia; however, pharmacotherapeutic agents and chronic alcohol abuse are the most common causes of toxic dementia. Dementia produced by drug toxicity is characterized by reduced concentration, poor attention, fluctuating course, and motor signs (tremor, asterixis, myoclonus), and thus resembles a typical delirium. Although excessive doses of nearly any pharmacologic agent can impair cognitive function, some patients will develop cognitive toxicity on "therapeutic" doses of medication. Again, this is particularly true in the elderly (and especially when the patient is taking several medications). Fortunately, reduction or discontinuation of the offending agent(s) often produces a reversal of the cognitive problems.[9]

Toxic dementia is most likely to complicate treatment with psychotropic medications. The psychotropic medications which are most commonly associated with a reversible dementia include sedative/hypnotics (especially benzodiazepines and barbiturates), antipsychotics, tricyclic antidepressants, and lithium. Sedative/hypnotics with long half-lives (e.g. flurazepam and diazepam) are especially prone to producing toxic dementia in elderly patients receiving daily doses. Antipsychotics often produce toxic dementia as a result of both their anticholinergic properties and via secondary parkinsonism with bradyphrenia (slowed thinking). The secondary amine tricyclics (amitriptyline and imipramine) possess higher anticholinergic activity than do the tertiary amine tricyclics (nortriptyline and desipramine) and are consequently more likely

to impair cognition. The selective serotonin reuptake inhibitors (fluoxetine, paroxetine, others) have been less commonly associated with toxic dementia. Toxic dementia from pharmacotherapeutic agents is probably the most important reversible dementia. It is critical to emphasize again that significant drug-induced cognitive impairment can occur in patients receiving "therapeutic" doses, and even in patients with "therapeutic" serum drug levels. A trial period of drug discontinuation is often necessary to determine whether the dementia is due to drug intoxication. Reversal of the dementia syndrome through such a simple therapeutic maneuver is greatly rewarding for both patient and clinician.

Depression

Major depression is commonly associated with cognitive impairment. Many depressed patients experience impaired concentration and forgetfulness. Reduced motivation and overall reduced functional ability are also commonly seen in patients with depression. Consequently, it is not uncommon for patients with major depression to present with symptoms of dementia, and approximately 10% of patients seen at dementia clinics have depression as the primary cause of their cognitive difficulties. The "dementia of depression" is characterized by slowness of response, reduced attention, poor concentration, and poor memory (although recognition is usually much better than is spontaneous recall). Aphasia, agnosia, and apraxia are uncommon in the dementia of depression. Associated symptoms of sleep disturbance, appetite change, low energy levels, and depressed moods, are often encountered. A past history of depression or a family history of depression is also common in these patients. Treatment of major depression with antidepressant medication or with electroconvulsive therapy often relieves the mood disturbance and improves or reverses the cognitive deficits.

REFERENCES

1. Mayeux R, Foster NL, Rossor M, Whitehouse P, The clinical evaluation of patients with dementia, in Whitehouse PJ (ed.), *Dementia*, F A Davis, Philadelphia, 1993.

2. Barclay L, Evaluation of dementia, in Barclay L (ed.), *Clinical Geriatric Neurology*, Lea and Febiger, Philadelphia, 1993.

3. Sano M, Ernesto C, Thomas RG, *et al*., A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease, *New Engl J Med* **336**:1216–1222, 1997.

4. LeBars PL, Katz MM, Berman N, *et al*., A placebo controlled randomized trial of an extract of ginkgo biloba for dementia, *JAMA* **278**: 1327–1332, 1997.

5. Erkinjuntti T, Subcortical vascular dementia, *Cerebrovasc Dis* **13** (Suppl 2):58–60, 2002.

6. McKeith IG, Galasko D, Kosaka K, *et al*., Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop, *Neurology* **47**:1113–1124, 1996.

7. McKeith IG, Perry EK, Perry RH, Report of the second dementia with Lewy body international workshop: diagnosis and treatment, *Neurology* **53**:902–905, 1999.

8. Neary D, Snowden JS, Gustafson L, *et al*., Frontotemporal lobar degeneration. A consensus on clinical diagnostic criteria, *Neurology* **51**:1546–1554, 1998.

9. Salzman C, *et al*., Cognitive improvement following benzodiazepine discontinuation in elderly nursing home residents, *Int J Geriatr Psych* **7**:89–93, 1992.

This page intentionally left blank

Obstetrics and Gynecology

This page intentionally left blank

64

Medical Disorders in Pregnancy

Tan Hak Koon

Medical disorders in pregnancy or obstetric medicine is becoming increasingly important for both obstetricians and physicians. Pregnancy is a unique state where the physiology of the mother is greatly altered to accommodate the newly developing "organ" — the fetus. The objectives are to provide the fetus with adequate nutrition for its growth and development, and to allow the mother to survive the process of reproduction. Understandably, pregnancy has a significant impact on the well-being of a mother with an underlying medical condition. At the same time, the fetus is also vulnerable to changes in the maternal condition.

Care of the pregnant woman with a medical condition requires knowledge of how the disease may be affected by the pregnancy and vice versa. Concern for the welfare of the fetus may influence the management of maternal conditions and particularly the choice of drugs.

Both physicians and obstetricians must be familiar with the normal physiological adaptation to pregnancy and the common medical disorders which may be encountered in order to ensure the safety and well-being of both mother and the fetus.

The field of obstetric medicine is vast since virtually any medical condition may complicate a pregnancy. Medical problems may complicate pregnancy during the pre-conception, antenatal or postnatal periods.

In recent years, the incidence of medical conditions in pregnancy is increasing steadily. Women today are delaying childbearing as a consequence of marriage at a later age and pursuit of professional goals. Pregnancy after the age of 35 is becoming more common. Age-related medical disorders, such as diabetes mellitus and hypertension, are encountered more frequently in pregnancies. Advances in medicine have enabled some girls who would have died prematurely from their medical or congenital disorders in the past to survive into reproductive age and bear children today. This is particularly true for girls with congenital heart disease who had undergone complex heart surgery. Other successful medical treatments like renal dialysis or transplantation have made childbearing a reality for patients suffering from chronic renal failure. These high-risk pregnancies however, is often associated with changes beyond the limit of the normal physiological response and cause decompensations, which lead to severe morbidity or mortality of either the fetus or the mother, or both.

Some of the more commonly seen medical problems in pregnancy are discussed.

GESTATIONAL DIABETES MELLITUS (GDM)

Pregnancy is a state of physiological insulin resistance and relative glucose intolerance. This is largely due to anti-insulin hormones secreted by the placenta, particularly human placental lactogen, glucogen and cortisol. The insulin production of normal women approximately doubles during pregnancy. Glucose intolerance increases progressively with increasing gestation.

Gestational diabetes is defined as "carbohydrate intolerance" of variable severity with onset or first recognition during the present pregnancy. It includes women with pre-existing but previously unrecognized diabetes.

GDM is the most common medical complication and metabolic disorder of pregnancy. Local incidence of GDM has been reported to be from 5–13%.

The World Health Organization (WHO) has proposed criteria equivalent to those for diagnosis of impaired glucose intolerance in the

non-pregnant state. A woman is diagnosed as having GDM if either their fasting or 2-hour level is > 7.8 mmol/L following a 75 g oral glucose tolerance test. There is no consensus in advocating universal biochemical screening for all pregnant women for GDM. Most centers still screen only women with clinical risk factors such as advanced maternal age, family history of diabetes, previous GDM, previous macrosomic baby, previous unexplained stillbirth, obesity, glycosuria, polyhydramnios or macrosomia in current pregnancy.

Poor diabetic control during periconceptual period is associated with early pregnancy losses and recurrent miscarriages. It is also related to an increased incidence of congenital malformations in the fetus, especially cardiovascular and neural tube defects. Intrauterine growth retardation, sudden intrauterine death, fetal hyperinsulinemia and macrosomia, polyhydramnios, perinatal morbidity and mortality are significantly increased in infants of diabetic mothers with poor blood sugar control. Neonatal morbidity such as fetal macrosomia, respiratory distress, hypoglycemia, polycythemia and jaundice are recognized sequale of diabetic pregnancy. Besides, diabetic mothers, especially those with nephropathy and hypertension, have an increased risk of pre-eclampsia. Women with history of GDM have up to 50% lifetime risk of developing gestational diabetes mellitus.

The major objective of management is to attain normoglycemia, pregnant women with gestational diabetes should be managed in joint pregnancy-diabetic clinics attended by obstetricians, diabetologists, dieticians, trained nurses and midwives who are experienced in the care of those women.

All women with established diabetes who seek fertility should attend pre-pregnancy counselling sessions for risk assessment. It also allows for optimization of diabetic control prior to conception and treatment of possible complications such as hypertension, nephropathy and retinopathy.

Regular antenatal follow-up and frequent home glucose monitoring since early pregnancy are essential. Strict adherence to low sugar, high fiber and low fat diet is important in pregnancy to achieve normoglycemia. Insulin therapy is indicated in the presence of persistent fasting or post-prondial hyperglycemia. A combination of short acting and intermediate acting insulin are often necessary in more severe cases. Regular checks of maternal blood pressure and urinalysis as well as regular ultrasound assessment of fetal growth and well-being are advised.

Pregnancies with well-controlled GDM should be allowed to continue until 38–40 weeks in the hope of a spontaneous labor and vaginal delivery. Cesarean section is only performed for usual obstetric indications. Monitoring of blood sugar level and insulin requirement during the intrapartum period is important.

For those in whom the diagnosis of the GDM is made during pregnancy, a formal oral glucose tolerance test should be performed at 6 weeks post-natally to assess the degree of impairment of glucose tolerance.

HYPERTENSIVE DISEASE IN PREGNANCY

As a result of vasodilatation, blood pressure begins to decrease in early pregnancy. It reaches a nadir at around 20–22 weeks before rising to pre-pregnancy level until term. Blood pressure should be taken with the pregnant woman sitting or lying on her lateral side rather than supine position. Phase V (disappearance) rather than IV (muffling) of korotkoff sound should be taken as the diastolic reading as it is more reproducible and correlates better with intra-arterial measurement of diastolic blood pressure.

Hypertension, which complicates up to 10% of all pregnancies, is one of the most common medical problems encountered in pregnancy. Pre-eclampsia, which affects up to 10% of primiparous women, remains one of the leading causes of maternal mortality and continues to contribute significantly to perinatal morbidity and mortality.

Hypertension is defined as a diastolic blood pressure > 90 mmHg on 2 occasions, a rise of systolic blood pressure of 30 mmHg and diastolic increase of 15–25 mmHg above the earliest recorded value. It may be divided into essential hypertension, pregnancy-induced hypertension and pre-eclampsia.

Pregnant women suffering from essential hypertension are usually diagnosed as hypertensive prior to pregnancy or in the first trimester. Secondary causes such as renal or cardiac disease, Cushing's Syndrome, Conn's Syndrome or pheochromocytoma must be excluded. These women have a significantly increased risk of developing superimposed pre-eclampsia. Pregnancy-induced hypertension may be defined as hypertension occurring in the second half of pregnancy but in the absence of proteinuria. It usually appears in the second half of pregnancy and resolves within 6 weeks of delivery.

Chronic hypertension in pregnancy should be treated to reduce the risk of severe hypertension and cerebral hemorrhage, and possibly to prolong the pregnancy, even though there is no conclusive evidence to suggest that it reduces the risk of superimposed pre-eclampsia. Methyldopa is still the drug of choice in pregnancy as it does not give rise to any serious adverse effect on the fetus.

Pre-eclampsia is a pregnancy specific multi-systemic disorder of protean manifestation. It occurs as a result of a cascade of events involving inflammation, endothelial dysfunction and unbalanced oxidation set into motion by a pathogen released by a dysfunctional placenta. Its risk factors include primiparity, multiple pregnancies, pre-existing hypertension, past history of pre-eclampsia, renal impairment, diabetes, anti-phospholipid syndrome and hydatidiform mole.

Although the classic signs of pre-eclampsia are hypertension, proteinuria of $> 0.3\,g/24$ hrs and edema, their absence does not exclude the diagnosis. Diffuse vascular endothelial dysfunction may cause widespread circulatory disturbances involving the renal, hepatic, cardiovascular, central nervous and coagulation systems. The severity, timing and order of onset of different clinical features vary enormously. Several possible crises may develop in severe cases, including eclampsia or grand mal convulsion, renal failure, hepatic rupture, HELLP syndrome (hemolysis, elevated liver enzymes, low platelet), cerebral hemorrhage, disseminated intravascular coagulation and pulmonary edema. Severe pre-eclampsia may lead to intrauterine growth retardation (IUGR), placental abruption, and intrauterine death.

The cornerstone of clinical management of pre-eclampsia remain early detection, control of blood pressure, close monitoring of the progression of the disease and fetal surveillance, and ultimate delivery when either maternal or fetal well-being is compromised or when fetal maturity is achieved.

Pregnant women suffering from pre-eclampsia require close monitoring of symptoms, blood pressure, proteinuria, renal and liver functions, platelet count and clotting factors. Patient with severe pre-eclampsia and impending eclampsia should be transferred to intensive care unit and managed jointly by obstetrician, anesthetist, physician and neonatologist. Magnesium sulphate which acts as a cerebral vasodilator is the drug of choice for primary and secondary prophylaxis in eclampsia.

The only cure for pre-eclampsia is delivery. In the absence of coagulopathy, women with pre-eclampsia are encouraged to have epidural

analgesia in labour or for cesarean section. Intensive monitoring of blood pressure, fluid balance, hematology and biochemistry are necessary postpartum.

Low dose aspirin remains the only prophylaxis for pre-eclampsia. It should be started preferably at 12 weeks gestation.

VENOUS THROMBOEMBOLISM (VTE) IN PREGNANCY

During pregnancy, increased venous stasis due to the compressive action of an increasingly gravid uterus on major veins and the hypercoagulate state as a result of physiological increase in circulatory clotting factors like fibrinogen, factors VIII, IX and X, and a decrease in fibrinolytic activity, all predispose to an increased risk of VTE.

Venous thromboembolism is the leading cause of maternal mortality in the developed world. Pulmonary embolism was the main cause of death in the last two triennial reports on Confidential Enquiries into Maternal Deaths in the United Kingdom. Similarly, an audit of autopsies of 7 cases of maternal mortality in Singapore from 1992 to 1995 found that there were 3 cases of pulmonary embolism, giving an incidence of 4.9 fatal pulmonary embolism for 100 000 maternities.

The risk factors associated with VTE development include advanced maternal age (>35 years), operative delivery, prolonged bed rest, obesity and thrombophilia.

The presence of deep vein thrombosis (DVT) may be suspected in the presence of swelling, redness, pain and tenderness of the calf. The gold standard for diagnosis of DVT remains to be venography, which can be performed with abdominal shielding to ensure minimal radiation to the fetus. However, duplex doppler ultrasound is a more convenient and less invasive method, which is more widely used to accurately detect DVT. Pulmonary embolism can present in a myriad of ways ranging from chest pain and tachycardia to cardiopulmonary collapse. A high index of suspicion must be employed in women who present with chest symptoms or atypical dyspnea. Basic investigations like chest X-ray, arterial blood gases and electrocardiography are established initial diagnostic tests. The diagnosis can be confirmed with a ventilation-perfusion scan (V/Q), which will show areas of ventilation/perfusion mismatch. Pulmonary angiography is usually reserved for pre-embolectomy cases.

Heparin has been the standard treatment of VTE as it does not cross the placenta and has no teratogenic effects, unlike warfarin. However, the monitoring of conventional unfractionated heparin (UH) with activated partial thromboplastin tissue (APTT) is often poorly performed, leading to the possibility of introgenic overdosage of heparin which may give rise to potential hemorrhage problems in pregnancy. Prolonged use of UH also carries the risk of osteoporosis and thrombocytopenia. There is a recent move away from using intravenous UH towards low molecular weight heparin (LMWH) for the treatment of deep vein thrombosis in both acute and chronic phase. LMWH can be given subcutaneously in fixed dose, thus minimizing or avoiding the need for monitoring. The simplified therapeutic regimen is more convenient and allows outpatient treatment. LMWH also has a lower risk of hemorrhage and osteoporosis.

In clinical practice, UH is preferred over LMWH in the acute phase of pulmonary embolism because of its rapid onset of action and wide expertise in its use. Pregnant women with pulmonary embolism should be closely monitored in intensive care units and co-managed by hematologist and intensive care specialists.

All pregnant women at high risk of recurrent thrombosis should receive thromboprophylaxis during the antenatal period, intrapartum as well as up to 6 weeks postpartum.

CARDIAC DISEASE IN PREGNANCY

Pregnancy and peripartum period are associated with important cardiovascular changes. Endogenous hormones cause peripheral vasodilatation leading to a fall in systemic arterial blood pressure during the first trimester. This reaches a nadir in mid pregnancy and returns to pregestational level before term. The blood volume increases substantially from early pregnancy to mild pregnancy by as much as 50%. Cardiac output also increases by around 40% during pregnancy due predominantly to an increase in stroke volume, but also by a lesser increase in heart rate. Blood volume expands further immediately after delivery as a result of autofusion of 300–500 ml of blood due to uterine contraction and relief of aortocaval compression.

Pre-pregnancy counselling of women with heart disease and detailed assessment of the cardiac status are of utmost importance. The hemodynamic significance of any lesion and functional class (New York Heart

Association, NYHA), the presence of cyanosis and pulmonary hypertension must be properly asssessed. Poor pregnancy outcome is more likely if the women is in poor functional status, (NYHA Class III and IV) and in the presence of pulmonary hypertension. Maternal morbidity is 30–50% in the presence of Eisenmenger's syndrome. Close collaboration between the attending cardiologist and obstetrician is important to ensure maternal well-being and satisfactory fetal outcome.

Congenital Heart Disease

The incidence of congenital heart disease in pregnancy is increasing. This is because more women with severe defects who underwent successful corrective surgery as children are now able to have children themselves. Majority of congenital heart disease seen are ventricular septal defects (VSD), patent ductors arteriosus (PDA) and atrial septal defects (ASD).

Maternal risks are usually small and fetal outcome satisfactory for simple acyanotic defects because there is usually minimal hemodynamic change of significance. However, women suffering from cyanotic heart disease are at a significantly higher risk. Maternal hypoxemia increases risks of miscarriage, intrauterine growth retardation and preterm labour. Women with Eisenmenger's syndrome should be managed in an intensive care unit by cardiologist, anesthetist, hematologist and obstetrician. Mortality usually occurs during the peripartum and postpartum period. The fetuses must be screened after delivery as they are at a higher risk of having congenital heart defects (2–5%).

Rheumatic Heart Disease (RHD)

Rheumatic heart disease in pregnancy is still a common problem despite improved socio-economic conditions and widespread use of antibiotics for the treatment of streptococcal pharyngitis. By far the most common type of RHD in pregnancy is mitral stenosis, which makes up about 80%.

Most patients with moderate to severe mitral stenosis demonstrate a worsening of functional status in pregnancy, especially during late pregnancy, labor and postpartum period. Restriction of physical activity and even hospitalization are needed for symptomatic patients to prevent hemodynamic deterioration.

Intercurrent infection must be treated promptly. Cardiac medication like diuretics for pulmonary edema or beta-blockers to slow the heart

may be needed or increased. Antibiotic prophylaxis should be prescribed for high-risk patients with prosthetic valves or history of endocarditis. Latrogenic premature delivery may be indicated for severely decompensated cases. Close hemodynamic monitoring from the onset of labour to 24-hour postpartum is necessary for symptomatic patients.

Balloon valvotomy and closed mitral valvotomy are associated with high risks in pregnancy and only indicated for severe cases.

Peripartum Cardiomyopathy

This rare condition is specific to pregnancy. It may occur 3 months before term and up to 6 months postpartum. Patients present with signs and symptoms of congestive cardiac failures and echocardiography shows an enlarged heart with global dilatation of all four chambers with markedly reduced left ventricular function. Early elective delivery is indicated if the condition is diagnosed antenatally as the maternal mobility may be as high as 50%. Women should be counselled against further pregnancy since there is a significant risk of recurrence.

ASTHMA IN PREGNANCY

As a result of increased metabolic rate and consumption of oxygen, there is a significant increase in oxygen demand in normal pregnancy. The minute ventilation increases up to 50%, mostly due to an increase in tidal volume. There is also a reduction in arterial carbon dioxide partial pressure $PaCO_2$ and a mild compensated respiratory alkalosis due to maternal hyperventilation.

Asthma is one of the most common medical conditions seen in pregnancy in Singapore. The diagnosis is often obvious by eliciting a history of typical symptom of breathlessness or wheezy breathing, worse at night and in the early morning. There may be provoking trigger factors like upper respiratory infection, exercise, pollen, etc. The degree of reversible bronchoconstriction can be measured with a peak expiratory flow rate (PEFR) or forced expiratory volume in one second (FEV 1).

Asthma may improve, deteriorate or remain unchanged during pregnancy. Those with only mild disease are unlikely to experience any serious problems in pregnancy. However, patients with severe asthma are at greater risk of deterioration. This is especially so for those who reduce or even completely stop the medication due to fears of its safety. Asthma has

no adverse effects on pregnancy outcome on most women. Fetal growth and well-being are only affected in severe, poorly controlled cases with chronic maternal hypoxemia.

The treatment of asthma in pregnancy is essentially no different from the treatment of asthma in non-pregnant women. Current emphasis is in the prevention, rather than the treatment of acute attacks. The main aim of treatment is to achieve virtual total freedom from symptoms so that the lifestyle of the pregnant women is not affected.

All common drugs used to test asthma appear safe in pregnancy and lactation. Corticosteroids, both inhalational and oral, are safe in pregnancy. There is no evidence for an increased incidence of congenital malformation or poor fetal outcome as a result of inhalational steroids. The risk of abortion, stillbirth, fetal anomaly or neonatal death are not increased with the use of oral steroids, even though oral steroids can increase the risk of gestational diabetes mellitus and cause a deterioration of blood sugar control. Likewise, Beta 2 agonists are safe in pregnancy.

Acute severe asthmatic attacks are potentially dangerous and should be treated vigorously. The treatment is no different from severe asthma in the non-pregnant state.

Cesarean section is only performed when there is an obstetric indication.

RENAL DISEASE IN PREGNANCY

Renal plasma flow rises early in pregnancy and may increase by up to 60–80% throughout pregnancy. Similarly, glomerular filtration rate (GFR) increases and creatinine clearance rises by 50%. This results in a fail in the serum levels of creatinine. Protein is increased and the upper limit in pregnancy is 300 mg per 24 hours.

As a result of a rise in the etiological factors like uncontrolled diabetes mellitus, chronic hypertension, glomerulonephritis and collagen vascular disease, the incidence of chronic renal disease has increased significantly in Singapore in recent years. Advancement in treatment modalities like dialysis has improved the reproductive function of women with chronic renal diseases. However, these women face significant risks during pregnancy. There may be accelerated decline in renal function, escalating hypertension and proteinuria. Patients are also more likely to suffer

from early miscarriages, pre-eclampsia, intrauterine growth retardation, preterm delivery and fetal distress.

The outcome of pregnancy and any adverse effect on underlying renal disease are both influenced by the degree of renal impairment the presence and severity of hypertension and protenuria, as well as the underlying etiological factors. In general, women with mild impairment without hypertension usually have successful pregnancy outcome.

Women with chronic renal disease should be managed jointly by an obstetrician and a renal physician. Pre-pregnancy assessment of renal function and optimization of blood pressure control should be made. Women with severe renal impairment should ideally be counselled against pregnancy as they face a real risk of end-stage renal disease.

As soon as pregnancy is diagnosed, a renal function test should be performed to provide a baseline with which to compare trends in pregnancy. In view of the increased risk of pre-eclampsia, treatment with low dose aspirins should be considered especially for those with hypertension or poor obstetric history. Careful monitoring and control of blood pressure is important. Regular assessment of renal functions, and regular ultrasound assessment of fetal growth and well-being should be done. Admission should be considered if the women develops worsening hypertension, deteriorating renal function or superimposed pre-eclampsia. Early elective delivery is indicated in severe cases.

Pregnancy in Dialysis Patient

Women on hemodialysis or chronic ambulatory peritoneal dialysis are often sub-fertile. Even if they are pregnant, the chance of a successful pregnancy outcome is low (20–30%). Pregnancy is associated with marked increase in requirement for dialysis. The problems of fluctuation in fluid balance, electrolyte imbalances, bleeding due to heparinization, anemia and infection occur commonly and these patients must be managed in a tertiary center with full facilities.

Pregnancy in Renal Transplant Recipients

Successful renal transplantation significantly improves the reproduction function of women with end-stage renal failure. Pregnancy probably had no adverse long-term effect on renal allograft function or survival.

However, renal impairment does occur in about 15% of pregnancies and therefore, close monitoring and surveillance are necessary. Pregnancy outcome is optimal in those with near normal renal function and those without hypertension or recent episodes of graft rejection. The doses of immunosuppresive drugs should be maintained at pre-pregnancy levels. Cesarean sections are only performed for obstetric indications.

THYROID DISEASE IN PREGNANCY

The changes in thyroid function in pregnancy are complex. The levels of thyroid binding globin (TBG) increase due to an increase in hepatic synthesis, as well as an increase in the half-life of TBG. This results in the elevation of both total thyroxine (total T4) and total triiodothyronine (T3). The free thyroxine (free T4) index, however, tends to yield low values in pregnancy. Hyperemesis gravidarum in early pregnancy may be associated with a biochemical thyrotoxicosis or gestational transient thyrotoxicosis (GTT) with transient elevation of free T4 levels. Serum concentrations of thyroid stimulating hormone (TSH) fall in early pregnancy as the concentration of human chorionic gonadotrophin (hCG) increase. HCG has thyrotropic activity. The levels of free T4 and free T3 decline by about 10–20% during the second and third trimester.

Hyperthyroidism

Hyperthyroidism is more common in women than men (ratio 10:1). Majority of cases of hyperthyroidism in pregnancy are due to Graves' disease (95%) and have already been diagnosed or on treatment. Toxic nodular goiter or toxic adenoma are less commonly seen. Thyrotoxicosis must be differentiated from GTT, which resolves spontaneously.

Many typical features such as heat tolerance, tachycardia, palpitations, vomiting and goiter are common in normal pregnancy. However, the most distinctive features are tremor, weight loss, lid lag and exophthalmos. The diagnosis is made by finding a raised free T4 or free T3 and suppressed TSH levels.

Severe and untreated thyroxicosis is associated with anovulation and infertility. It also increases the risk of miscarriage, intrauterine growth

retardation, premature labour and perinatal mortality significantly. Thyroid storm may occur in severe thyrotoxicosis especially during the peripartum period. Grave's disease often improves during the second and third trimester as the levels of TSH receptor stimulating antibodies fall.

The aim of treatment is to control the thyrotoxicosis as rapidly as possible with the lowest possible dose of anti-thyroid medication. Pregnant women should be clinically euthyroid with a free T4 level at the upper end of normal range. Both carbinizole and propylthiouracil (PTU) can be used in pregnancy even though PTU is probably preferable for newly diagnosed cases in pregnancy as it crosses the placenta and into breast milk. Women already on maintenance dose need not be switched to PTU in pregnancy. Beta-blockers are safe for short-term use in pregnancy for control of thyrotoxic symptoms.

Fetal or neonatal thyrotoxicosis is uncommon. This is due to transplacental passage of thyroid stimulating antibodies. Newborns of thyrotoxic mothers should be examined by neonatalogist.

Hypothyroidism

Just like hyperthyroidisis, hypothyroidism is more common in women than men. The incidence of hypothyroidism in pregnancy is estimated to be between 0.05% and 0.7%. Most cases are due to previous thyroidectomy, Hashimoto's thyroiditis or post-ablative Grave's disease.

Untreated hypothyroidism during pregnancy is associated with intrauterine growth retardation, premature labour, low birth weight babies, abruptio placenta and pre-eclampsia. Children born to hypothyroid mothers are also showed to have significantly lower IQ scores. Pregnancy itself probably had no effect on hypothyroidism.

Clinical features of hypothyroidism include lethargy, tiredness, weight gain, cold intolerance, bradycardia and delayed relaxation of tendons. Serum-free T4 level is low.

Levo-thyroxine is the drug of choice for thyroid replacement therapy. Hypothyroid pregnant women also are already on thyroxine replacement require on upward adjustment of this daily thyroxine dosage in order to maintain eythyroidism, sometimes up to as much as 40%.

Neonatal hypothyroidism is a rare condition that must be promptly treated because of its serious consequences.

Postpartum Thyroiditis

Postpartum thyroiditis which usually presents between 3 to 4 months' postpartum is caused by a destructive autoimmune lymphocystic thyroiditis. It may present with either transient hypothyroidism or hyperthyroidism or first hyperthyroidism and then hypothyroidism. Most patients recover spontaneously and treatment is not always necessary. It is, however, a significant predictor of future hypothyroidism.

EPILEPSY IN PREGNANCY

Epilepsy is the most common chronic neurological disorder that can complicate a pregnancy. Most cases of epilepsy are idiopathic with no underlying cause. However, secondary epilepsy may occur with women with previous brain surgery, intracranial mass lesion, pre-eclampsia, cerebrovascular events, cerebral infection or after metabolic conditions.

Most women suffering from epilepsy in pregnancy have already been diagnosed but all women suffer from a first fit occurring in pregnancy should be thoroughly investigated for secondary causes.

The frequency of fits is not altered in most pregnant women. However, about a third of women may experience more fits as a result of poor control. Poor compliance with anticonvulsants due to fears regarding teratogenesis, decreased serum drug levels due to nausea and vomiting in early pregnancy or increased volume of distribution and increased drug clearance in pregnancy are possible reasons for deterioration of the condition.

The incidence of miscarriage in epileptics is not increased, unless there is associated abdominal trauma during the fit. Similarly, episode seizure is unlikely to result in intrauterine growth retardation or intrauterine deaths as the fetus is relatively resistant to short episodes of hypoxia. However, status epilepticus is dangerous for both mother and fetus and should be treated promptly.

Commonly used anticonvulsants like phenytoin, primidone, phenobarbitone, carbamazepine and valprocate all cross the placenta and are teratogenic. Possible fetal anomalies associated with these drugs include neural tube defects, orofacial defects, congenital heart defects and musculoskeletal defects. There is very little difference in the level of risk between individual drugs but the risk increases with the number of drugs used.

Ideally all women suffering from epilepsy should be counselled pre-pregnancy about the risks. The control of epilepsy should be optimized and the anticonvulsants used should be reviewed. The control should be with only one drug alone if possible to reduce the risk of teratogenesis. All women taking anticonvulsants should be given folic acid daily for at least 12 weeks prior to conception. This should be continued throughout pregnancy.

All pregnant women suffering from epilepsy should be reviewed regularly by both obstetrician and neurologist. There is no need to change the anticonvulsants in pregnancy if the woman is well controlled on either phenytoin, carbamazepine or valprocate. Phenobarbitone should be slowly weaned if possible to reduce the risk of neonatal withdrawal convulsions. Serum drug levels should be monitored in women suffering from regular fits. Screening for congenital anomalies with maternal alpha fetoprotein and detailed ultrasound scan at 20 weeks should be performed.

All neonates should receive vitamin K. Breast-feeding while taking anticonvulsants is safe as only a small fraction of anticonvulsants is secreted into breast milk. All mothers should also be informed of the increased risk of the child developing epilepsy later in life.

IDIOPATHIC THROMBOCYTOPENIC PURPURA (ITP) IN PREGNANCY

During a normal pregnancy, the platelet count tends to fall progressively, although usually within the normal limits. In practice, a pregnant woman is not considered thrombocytopenic unless the platelet count is below 100×10^9/L. Thrombocytopenia detected in the initial stages of pregnancy is less likely to be due to pregnancy itself, and the possibility of an underlying pathology must be considered.

Idiopathic thrombocytopenic purpura (ITP) occurs as a result of autoantibodies against platelet surfaces antigens, causing peripheral platelet destruction by the reticuloendothelial system. The diagnosis of ITP in pregnancy is one of exclusion, and should only be made after excluding other causes of thrombocytopenia such as infections, autoimmune disease, pre-eclampsia and drugs.

Pregnancy does not affect the course of ITP. Theoretically, maternal antiplatelet IgG can cross the placenta and cause fetal thrombocytopenia,

and it is difficult to accurately assess fetal risk as there is no relation between the maternal platelet count or antibody levels to fetal platelet count. Most agree now that the fetal risk is small and the incidence of neonatal intracranial hemorrhage in women with ITP is 1% or less.

For pregnant women suffering from ITP, regular monitoring of platelet counts is necessary. Capillary bleeding and purpura are unlikely to occur if the maternal platelet count is $> 50 \times 10^9$/L and spontaneous mucosal bleeding usually does not occur if platelet count is $> 20 \times 10^9$/L. Treatment is only indicated for symptomatic cases or when the platelet is $< 50 \times 10^9$/L.

Corticosteroids are the first-line therapy. Intravenous gamma-globulin is not contraindicated in pregnancy and can be used for resistant cases. It is useful in instances where a rapid response is required. Splenectomy is not advisable in pregnancy and is performed in extreme cases. Platelet transfusion is usually given as a last resort for symptomatic cases or pre-operatively.

There is no conclusive evidence to suggest that cesarean section is less traumatic for the fetus and reduces the risk of neonatal intracranial hemorrhage compared to vaginal delivery. Therefore, cesarean section is only required for the other obstetric indications.

Monitoring of neonatal platelet levels is necessary and treatment is indicated for neonates with severe thrombocytopenia.

SYSTEMIC LUPUS ERYTHEMATOSIS IN PREGNANCY

Systemic lupus erythematosis (SLE) is a chronic systemic connective tissue disease with a reported prevalence of 5–100/100 000. Women are affected much more commonly than men (ratio 9:1), particularly those in childbearing age (15:1). The incidence in women of childbearing age is 1 per 1000 women and may be increasing.

The exact cause of SLE is unknown, but it involves both a genetic predisposition and environmental trigger factors. Its clinical features include joint involvement, skin involvement, serositis, neurological and hematological manifestations. The diagnosis of SLE can be made by using specific clinical and laboratory criteria by American Rheumatic Association. Antinuclear antibody (ANA) is the most common autoantibody found in SLE but anti-double standard DNA is the most specific. Serum compliment levels are useful in assessing disease activity. Other anti-phospholipid antibodies may be detected.

SLE in pregnancy poses serious clinical challenge as it is associated with both maternal and fetal morbidity and mortality. It is associated with increased risks of spontaneous miscarriage, intrauterine growth retardation, intrauterine death, preterm delivery and pre-eclampsia. This is especially so for active disease, in the presence of anti-phospholipid antibodies and significant renal involvement with hypertension. Pregnancy itself may exacerbate SLE and increases the likelihood of a flare, even though flares may be difficult to diagnose as features like facial erythema, hair loss, fatigue, edema also occur in normal pregnancy.

Pre-pregnancy counselling is of utmost importance in the management of women suffering from SLE. The disease should be adequately controlled as pregnancy outcome is better if conception occurs during remission. These women should be managed in joint clinic, with rheumatologist and obstetrician, where regular monitoring of disease activity and fetal growth can be performed. Regular monitoring of blood pressure, renal function, antibody titer and complement levels allow early diagnosis of disease flares, which must be actively treated.

Corticosteroids are the drugs of choice in controlling the disease. It is not associated with an increased risk of stillbirth and congenital malformation. Azathiopine is probably safe in pregnancy with no significant adverse effect in fetus. NSAIDs and aspirin are not teratogenic, although NSAIDs is associated with fetal oligohydramnios and premature closure of the ductus arteriosus. NSAIDs should be discontinued at 34–36 weeks of pregnancy.

All newborns must be assessed by neurologist for neonatal lupus syndromes which occur in 5% of pregnancies with SLE. The risk of congenital heart block is about 2–5%, especially in anti-Ro positive mothers.

HUMAN IMMUNODEFICIENCY VIRUS (HIV) IN PREGNANCY

The incidence of HIV infection is fast increasing worldwide. It is estimated that about 50 million people are infected with HIV, of which at least 40% are women. The prevalence rates of HIV in pregnancy, though vary greatly geographically, are showing an upward trend in all reported series.

HIV may be transmitted through unprotected vaginal or anal intercourse, sharing of contaminated needles, use of unscreened blood products or vertical transmission either antepartum, intrapartum or postpartum.

Early HIV infection is characterized by a high viral load. The main target of HIV is the CD4 lymphocyte and lymphocytes are gradually lost during the latent phase. Both cell-mediated immunity and humoral immunity are reduced, leading to development of opportunistic infections and malignancy.

HIV infection can be diagnosed by detecting HIV antibody to part of the viral membrane or envelope. The test usually becomes positive within 6 weeks to 3 months after exposure as the levels of P24 antigen are falling. Viral DNA may be detected with polymerase chain reaction (PCR) and the main predictor of disease progression is the CD4 lymphocyte count. Transplacental transfer of maternal HIV antibody may persist for up to 18 months, making the HIV status of the infant difficult to determine without PCR.

Pregnancy probably does not have a major adverse effect on the progression of HIV infection in asymptomatic women. Similarly, HIV infection has not been shown to adversely influence pregnancy outcome, increase miscarriage rate or the rate of congenital abnormality. However, recurrent opportunistic infections may occur in women with advanced disease and poor nutritional status.

Vertical transmission of HIV from mother to fetus may occur either *in utero*, during the time of delivery or by breast-feeding postnatally. The rate of vertical transmission ranges from 20–40%. Two-thirds of the vertical transmission seems to occur during delivery, especially in cases of advanced maternal disease and prolonged rupture of membranes. Breast-feeding is also shown to double the transmission rate.

Pregnant women infected with HIV should be managed jointly by an obstetrician, a HIV specialist, a pediatrician and other medical personnel with experience in this area. In view of the reduced life expectancy of the women and the increased risk of vertical transmission, the option of termination of pregnancy should be offered. The use of Zidovudine (AZT) monotherapy has been shown to reduce the risk of vertical transmission and should be encouraged. CD4 counts should be checked regularly and prophylactic antibiotics for opportunistic infection should be used if CD4 counts drop below $200/mm^3$.

Although a blanket policy of cesarean section is not universally acceptable, women infected with HIV should be informed of the benefit of reduction of transmission with elective cesarean section.

Universal precautions should strictly be observed by all medical staff during delivery. All babies born should be immediately seen by a neonatologist for further management. Breast-feeding, which significantly increases the risk of vertical transmission, should be discouraged as formula feeding is readily available in Singapore.

Many physicians are less than familiar with the physiological adaptation to pregnancy and medical disorders that are encountered in obstetric practice. Similarly, obstetricians may not feel confident to manage all medical complications of pregnancy since many of these conditions are seen infrequently in general obstetric practice.

A multi-disciplinary approach and close collaborative effort by obstetricians, physicians, anesthetists and neonatologists with strong support from para-clinical services are essential to ensure that patients receive optimal care.

FURTHER READING

1. Nelson–Piercy C, *Handbook of Obstetric Medicine*, Isis Medical Media, 1997.
2. deSwiet M, *Medical Disorders in Obstetric Practice*, 3rd Blackwell Scientific Publications, Oxford, 1994.
3. Medical Disorders in Pregnancy. *Annals, Academy of Medicine*, Singapore **31**(3):8, 2002.
4. Redman CWG, Roberts JM, Management of Pre-eclampsia, *Lancet* **341**(8858):1451–1454, 1993.
5. Royal College of Obstetricians & Gynaecologists, *Thromboembolic Disease in Gynaecology and Pregnancy: Recommendation for Prophylaxis*, RCOG, London, 1995.
6. American Diabetes Association, Gestational Diabetes Mellitus, *Diabetes Care* **22**:574–576, 1999.
7. Mir N, Bingham JS, The Management of HIV Infection during Pregnancy, *Maternal Child Health* **21**:64–68, 1996.
8. *Confidential Enquiries into Maternal Deaths in the United Kingdom 1994–1996*, HMSO, London.

This page intentionally left blank

Pediatrics

This page intentionally left blank

65

Practical Genetic Counseling

Ho Lai Yun

DEFINITION

Genetic counseling is a process whereby an individual or family is given information about a real or a possible genetic problem. It is a dynamic communicative and educational process. The objective is to assist them to:

1) understand the medical facts, including the diagnosis, probable course of the disorder, and the available management;
2) appreciate the way heredity contributes to the disorder, and the risk of recurrence;
3) understand the options for dealing with the risk of recurrence;
4) choose the course of action that seems appropriate for them in view of their risk, their family goals, and their ethical and religious standards, and to act in accordance with that decision; and
5) come to terms with the issues they face and make the best possible adjustment to the disorder.

Genetic counseling is directly concerned with human behavior, so it must be based on an understanding of the psychological meanings of

health and illness, procreation, and parenthood. The responsibility of the counselor must include helping the family adjust psychologically and socially to their genetic condition. Achievement of these goals requires proper training and an understanding of the genetic counseling process so that the information will not be misunderstood, misinterpreted, or misused, with potentially tragic consequences.

INDICATIONS OF GENETIC COUNSELING

Anybody who has reason to suspect that there might be an increased risk of producing a child with a birth defect should at some stage receive formal genetic counseling. This would include:

1) Couples who have previously had an unexplained stillbirth or neonatal death;
2) Couples who already have given birth to a child with a birth defect or medical problem;
3) Couples who have a child with mental retardation or developmental disabilities;
4) Couples who have relatives with known genetic disorders;
5) Couples who have a family history of any of the above;
6) Mothers who have had recurrent pregnancy loss;
7) Mothers exposed to possible teratogenic drugs, infections, radiation or other occupational hazards during or before pregnancy;
8) Mothers with pre-existing diseases, where either the disease or its treatment may have adverse effects on the infant;
9) Advanced parental age: maternal age over 35 and paternal age over 40;
10) Couples in consanguineous marriages;
11) Couples who have a previous child with a chromosomal abnormality or are themselves a translocation carrier; and
12) Family history of psychiatric disorder.

ACCURATE DIAGNOSIS

The cornerstone of genetic counseling is to establish and confirm the diagnosis. Because even the best counseling cannot make up for an inaccurate diagnosis, a great degree of diagnostic precision is required. For

example, a patient with muscular dystrophy can be correctly managed with the use of splints, wheelchairs and physiotherapy, without knowing precisely the type of muscular dystrophy. Genetic counseling in such a patient or family will require a precise diagnosis, confirmed by appropriate blood, EMG and muscle biopsy studies. The number of genetic disorders described is increasing exponentially and the disorders may be very rare. A specialist cannot rely solely on his own personal experience, but must be familiar with the literature. He may require the help of specialists in other fields to differentiate and confirm a diagnosis.

Genetic Heterogeneity

Genetic heterogeneity should be recognized. At the clinical level, two or more disorders may have the same phenotype, e.g. both Marfan syndrome and homocystinuria have dislocations of the lenses with a similar physical habitus, but Marfan syndrome is autosomal dominant, whereas homocystinuria is autosomal recessive. At least 20 different forms of muscular dystrophy that are clinically similar have been identified. Both sex-linked and autosomal recessive forms are known. At least two, Becker and Duchenne dystrophy, are sex-linked conditions that are allelic but clinically quite distinct, the latter being much less severe. Genetic heterogeneity can now be demonstrated at the molecular level, even when there appears to be no differences in phenotype or mode of inheritance. Non-genetic factors can mimic genetic factors (phenocopies); a good history and clinical and laboratory studies may help in differentiation. The possibility of non-paternity must also be considered.

Review of Medical Records

Establishing a diagnosis involves reviewing the relevant medical records, which can be a difficult task to accomplish completely. The rate of progression of symptoms and signs, the development of the physical findings, and the results of prior laboratory tests, are all best determined from the medical records. The status of the family members can sometimes be assessed from examination of their medical records. A relative may seem to have the same problem but the medical records could reveal an entirely different problem.

Medical History

The medical history should include a thorough account of any untoward reproductive outcome, including spontaneous abortions, stillbirths, and abnormal infants. Issues to be addressed in the pregnancy history include problems conceiving, bleeding, maternal illnesses or exposures, medications, cigarette smoking, amounts of alcohol intake, and possible exposures to teratogens. Information such as onset and intensity of fetal activity may help to differentiate a prenatal from a postnatal problem. It is often helpful to ask if this pregnancy was different from others in any way. A perinatal history should be obtained. A family history should always be taken in a pedigree form with a minimum of three generations. Parental ages should be recorded. Besides the known association between advanced maternal age and chromosomal abnormalities like Down syndrome, advanced paternal age is known to confer increased risk for new dominant mutations. The possibility of consanguinity should be explored. It is necessary to probe deeply into the medical background and development of all first-degree relatives (siblings, parents, and offspring), second-degree relatives (uncles, aunts, nephews, nieces, and grandparents), and third-degree relatives (first cousins), and, in all cases, objective documentation should be sought when possible. Information about ethnic background may be valuable.

Physical Examination

The physical examination should document features that vary from the usual or normal structure. Minor abnormalities are stressed as well as more obvious defects. Clinical impressions may be misleading and accurate measurements are required to confirm features such as widely spaced eyes or disproportionate short stature. These observations should be interpreted with regard to normal standards and familial variant developmental patterns. Photographs may be taken to document significant findings. Recording of the fingerprint pattern (dermatoglyphics) may provide important clues to the diagnosis, e.g. a predominance of whorls in Turner syndrome and ulnar loops in Trisomy 21. It may also be necessary to examine parents, siblings, and even more distant relatives to make a diagnosis.

Laboratory Investigations

The need for specific laboratory tests varies with the situation. Because these tests are expensive, they should be ordered with care, but certainly done when important. Chromosomal analysis is indicated in any child with multiple major or minor anomalies that is not etiologically well-defined, particularly if associated with mental retardation or developmental delay. It is also important to karyotype patients who have apparently clear-cut diagnoses if there are any atypical features or if the patient has features of two or more distinct syndromes. Karyotyping is also recommended for fetuses, stillbirths, and neonatal deaths that only have one obvious anomaly because of the possibility of finding other birth defects during a more thorough evaluation. Skeletal radiographs are obtained on any patient with disproportionate short stature and on many patients with proportionate short stature. Craniofacial radiographs may be useful for anomalies of the skull or face. Other special studies like ultrasonography, or computed tomography may be indicated in selected cases.

Post-mortem Examination

Post-mortem is one way of establishing a precise diagnosis. This is of particular importance in a child with malformation or retardation when a clinical diagnosis is not possible. Fetuses, stillbirths, and neonatal deaths should have thorough post-mortem examinations. Post-mortems need to be planned in advance so that specific tissues can be obtained for biochemical study, electron microscopy or histochemistry. It is best to raise the issue of a post-mortem before death occurs, especially if death can be anticipated. It may work better than asking permission immediately after the death when families are in grief. When permission has been granted, the parents should return to discuss the autopsy findings.

MOLECULAR GENETIC ANALYSIS

Advances in DNA technology that can provide precise definitions of the mutation or utilize linkages to a specific genetic marker has revolutionalized genetic counseling. In the past, the chromosomal location of specific genes was inferred from pedigree information for the X chromosome and

linkage to specific protein markers for autosomes. Now, the location of many genes on the chromosomes is known, linkage to specific DNA markers has been established, and many genes of clinical importance have been cloned and sequenced. The number of conditions that shows linkage to known genetic markers or to DNA probes grows rapidly. Some of the conditions for which DNA-based diagnosis has been accomplished include cystic fibrosis, hemophilia A, Huntington disease, thalassemia, phenylketonuria, etc. It is now possible to store samples of DNA from patients who are likely to die and relatives such as grandparents so that subsequent family members may benefit from the advancing technology.

WHEN DIAGNOSIS IS UNCERTAIN

Unfortunately, despite the most intensive efforts, a definitive diagnosis cannot always be established because the appropriate investigations might not have been obtained while the relevant person was alive or there is inadequate information. Of the children presenting with retardation or dysmorphic features, about one-half will not have a precise diagnosis. In such cases, advice can only be given on the basis of a probable diagnosis and a large body of empirical data used for counseling. While an inability to label a child with a specific diagnosis is disappointing and frustrating, it should not disadvantage the child as management can be based on periodic assessment and planning. Parents should be told that all reasonable steps have been taken to look for known disorders.

The evolution of knowledge and new techniques make re-study of old problems necessary.

ESTIMATION OF RISKS

Risk is a numerical estimate of the probability of a particular genetic disorder occurring in a subsequent child and in members of the family. In some counseling situations, calculation of the recurrence risk is relatively straightforward and requires little more than a reasonable knowledge of Mendelian inheritance. The risks of recurrence in this group are high but most of the diseases are rare. Sometimes, these strictly genetic risks are modified by Bayesian methods of calculation to take into account information about specific family members, age of onset, penetrance, and laboratory tests results. Such calculations may be more complex but are

particularly useful to estimate the likelihood that a person is the carrier of an X-linked condition or of an incompletely penetrant autosomal dominant trait. Nowadays, however, molecular diagnostic technology has greatly refined the ability to determine heterozygosity in many situations. The risk with chromosomal translocations varies according to the chromosome involved. With better chromosome staining technique, many more translocations are identified; some involve only small pieces of a chromosome.

There are many common disorders in which there is a genetic component and where the inheritance pattern cannot be explained simply in terms of Mendelian inheritance or chromosomal rearrangement. These disorders are due to the cumulative action of a number of genes together with environmental influences. This is called polygenic or multifactorial inheritance. Examples of multifactorial inheritance include neural tube defects, congenital heart diseases, cleft lip and palates, congenital pyloric stenosis, and diabetes mellitus. It is the most common pattern of inheritance responsible for the family tendency or predisposition to various disorders. Fortunately the risks of recurrence are usually small. When risks cannot be predicted by formal genetic methods, empirically derived estimates of risks are used. The empirical risk is defined as the probability of occurrence of a specified condition based upon prior experience and observation, and is determined by estimating the frequency of the condition in the relatives of affected persons. When empirical figures are used, however, the populations used to calculate the risks must be representative of the population from which the particular family is drawn.

COMMUNICATION OF INFORMATION

The ability to share threatening or unpleasant information incorporates both the science and the art of clinical medicine. All interactions must be open and honest, reflect a genuine sense of caring, and built on a firm foundation of trust. Such relationships cannot be developed instantaneously; they must be nurtured and allowed to evolve over time.

Inform the family as soon as possible

Most families prefer to be told as soon as a problem is suspected, even if it is incomplete and uncertain. Although this often refers to bad news, it is equally important that good news should not be withheld when available.

Parents also need to be reassured that they do not have, or are not at risk for, a certain disorder after investigations.

Choose the right time and the right place for the right information

The counseling session should be held in a suitable environment that provides protection and security. A private and quiet setting that is free of distractions or interruptions, such as supervising young children during the session, will afford an opportunity for reflection and the provision of meaningful support. Information should not be transmitted over the telephone, in the corridor, or crowded waiting room, or when strangers are present. The family must have the clinician's undivided attention.

Choosing the right time for counseling is very important. The couples who have just produced a child with a birth defect or whose child has just died from a major congenital malformation will usually want some time to come to terms with their loss before considering a possible recurrence in future pregnancies. On the other hand, leaving counseling until a pregnancy is advanced may produce an emotional crisis, which could have been avoided by anticipation. In families where there is a risk of producing a child with a birth defect, they should be informed of the availability of counseling service, and counseling should certainly be offered at the time they become concerned about the family history or are contemplating having children.

The key people must be present

Every effort should be made to ensure that both parents are present when difficult information about their child is discussed. When appropriate, other important people (e.g. grandparents) may be allowed at the discretion of the parents. The inclusion of more than one parent provides an additional source of support and relieves one family member of the responsibility of transmitting information and answering questions for the others.

Prioritize the message

Before talking to the family, the clinician must identify the most important message to convey — what is the most important message for the family to understand? When multiple issues are to be addressed, it is

important that the core message relate closely to the primary concerns that have prompted medical intervention. The most salient information should be delivered at the beginning in a brief, direct, and sensitive manner. It is not unusual that the family will stop listening after being told the traumatic news.

Proper language should be used

It is important for the family to trust that the clinician undergoing the counseling knows them and their child, and is genuinely interested in helping them. Always refer to the child by name and acknowledge specifically his or her personal characteristics. It would be a disaster if the sex of the child is mentioned wrongly, and it would be even worse if the fetus is referred to as "it".

The communication of information should be versed in language that is straightforward and understandable. Technical terms should be avoided where possible. Unfamiliar words and medical terms should be written out so that the parents can subsequently search for relevant information about that condition. Visual aids such as diagrams, pictures, photographs, sample pedigrees, and other written materials should be freely referred to. Translation, if necessary, should be verbatim. It is not unusual to find that the interpreter is digesting the information and passing it on in such a way that it is inaccurate and that any decisions made reflect the interpreter's beliefs and wishes.

Appreciate parents' reaction

During the counseling session, the clinician must be able to assess the cognitive and emotional dimensions of the parents' verbal and nonverbal reactions, and be prepared to stop, to listen, and to wait for a reaction and guidance from the family on the direction of the dialogue. Families vary dramatically in the extent to which they feel comfortable sharing their inner feelings. It is important for the clinician to encourage the parents to express and explore their own feelings, and to allow and support a range of emotions, including tears, anger, and withdrawal. Ask the question, "How do you feel about what I have just shared with you?" rather than trying to be sympathetic by saying, "I understand how you feel".

Answer concerns and questions honestly and openly

Sufficient time should be allocated for parents to ask questions. All questions must be answered directly and openly. Some persons come "armed" with questions they want answered. They would have visited the library or searched websites in an attempt to find literature about the disorder or asked medical friends to do so. Sometimes, this results in wrong or dated information. However, most families are fairly naïve with regard to biology and medicine and they do not have sufficient background to ask all of the questions that may have significant impact on their decisions. It is useful to place oneself in the position of the family member by asking, "if I were in this position, what kinds of information would I want to know?"

When the answer is unclear, it is important to distinguish between responses to questions that cannot be answered because of the limitations of the clinician's knowledge (e.g. "I do not know the answer to your question, but I will find out for you") and responses to questions that are essentially unanswerable (e.g. "It is difficult to say what your child will be like 5 years from now"). In the latter case, it is helpful to explain why the question cannot be answered and to discuss the process by which the professional and family together can try to reach a greater understanding, if not a definitive conclusion. It is also important to raise and answer questions that the parents do not identify (e.g. whether the child's condition is in any way their fault).

Balance the message

It is important for the clinician to assess at the end of a counseling session the family's understanding of the child's problem, and whether the perceptions are reasonably balanced, given the seriousness of the child's condition. Some parents may react with a sense of despondency and hopelessness. Others may seem not to hear the gravity of the situation. A skilled and sensitive clinician will listen to the parents and make a judgement about whether they are overly pessimistic or excessively optimistic. In the former, it is necessary to underscore a sense of hope with specific examples of potential positive outcomes (e.g. "Your child will always continue to gain new skills"). In the latter case, it may be necessary to acknowledge the grounds for optimism but gently remind the family about concerns and their probable sequelae (e.g. "Your child will always

need special help in school"). No family should leave the counseling session without any sense of hope.

Do not make assumptions

We also sometimes assume that we know what the parents will want to do, because of their ethnic group, social class, or religious beliefs. Yet, in practice, we are often surprised. We must all guard against making assumptions and thereby making decisions for people.

Discuss the next step

At the end of each counseling session, there should be no ambiguity with respect to the division of responsibilities between parents and the professional regarding subsequent management and care. Specific plans for further consultations, referrals, evaluations, and arrangements for appropriate therapeutic, educational, and supportive services should be discussed. Requests for second opinion should be honored and facilitated, and not interpreted as threatening and distrust.

Check whether the message has been heard

It is useful to determine whether the family has heard and understood the important content of the counseling. The counseling session may be so intimidating and so full of factual information that the amount and accuracy of information retained by many persons on follow-up at a later date is often disappointing. The process of providing information is usually an ongoing one. Whilst it is possible for a family to be counseled in a single visit, a repeat counseling session is often required to reinforce the points covered and to answer queries from the family. Providing written material about the disorder is often helpful. Pamphlets, booklets, and other literature may provide additional information. It may be helpful to follow the sessions with a written report summarizing the important points discussed in simple language.

Ongoing process

Providing information to a family is always an ongoing process. Often, the clinician will need to meet with the family on more than one occasion,

initially because of critical medical circumstances and then later to obtain additional information or to discuss test results. However, even when all the available information has been provided, it is appropriate for the family to keep in touch on a regular basis because of the possibility that there is a new development in diagnostic and management techniques, new methods of prenatal diagnosis or carrier detection, and so forth.

NON-DIRECTIVE APPROACH

The clinician is accustomed to issuing therapeutic directives, and patients are invariably dependent on such instructions in order to improve their health status. However, there is universal agreement that genetic counseling should be non-directive and non-coercive. This approach does not tell families what they should do, but rather what they could do.

Each disorder has a unique set of complications, diagnostic possibilities, and natural history. The family brings to the counseling session a unique set of personal and cultural experiences. The counseling clinician also brings to the case his or her own set of values. These values may affect the way in which counseling is provided if great care is not taken to avoid judgemental or directive advice. The intrinsic danger of using a directive approach is the opportunity for the counselor to insinuate his or her own religious, cultural, eugenic, or other beliefs or dictates of conscience into the counseling that is offered, subconsciously or inadvertently. An optimistic counselor may tell anxious individuals not to worry, whereas a pessimistic one might unwittingly exaggerate the significance of even small risks.

In the non-directive approach, the counselor should recognize his own biases and try to remain impartial and objective when providing information without dictating a particular course of action. Of course, completely non-directive counseling is probably unrealistic, particularly if the counselor is familiar with the long-term burden of a genetic disorder and he may have concerns about society at large. The natural tendency for counselors to interject their own biases by either verbal or non-verbal messages is always present.

HELP WITH DECISION MAKING

Counseling is a process that cannot be hurried and it is not appropriate to try and elicit a response of what options a couple might choose at the

session but to allow them time to reflect. The counselor should attempt to provide a balanced presentation allowing the couple to make intelligent and reasoned decisions for their future and to support them in their choice. There is a well-established maxim that it is the couple and not the counselor who have to live with the consequences of the decision.

While the principle of a non-directive approach should be upheld, it is incorrect to provide a mass of medical facts and leave the couple adrift without specific guidance. The counselor may need to help them try different alternatives "on for size," imagining how each would feel with a particular decision and how the possible outcomes of each choice might affect their lives. If the responses of different family members — particularly the husband and wife — are different and would lead to incompatible decision, the counselor needs to help them resolve the conflict. This can be done by working with them on values clarification. Which of the alternatives is most consonant with what they believe? Which would be most harmful to their relationship (family, status, lifestyle, self-image, etc.)? Sometimes it may help to have each party articulate what he or she perceives as the "best case" and "worst case" outcomes of each decision and why. The counselor may also have to guide the couple towards what is acceptable not only medically but also morally and socially, and what is within the law.

PERCEPTION AND ACCEPTABILITY OF GENETIC RISKS

The provision of a recurrence risk does not simply involve conveying a stark risk figure in isolation. Recurrence risks should not only be quantified, but also be qualified and placed in context. The perception of risk may be of more importance in family decision making than the actual numerical value of the risk.

Many families may not be able to appreciate the concept of probability and need a careful discussion to give meaning to any risk estimate. Many people may think that a 1 in 4 risk means that the next three children will be unaffected. It is therefore important to emphasize to families that chance has no memory and the chance of their having an affected child is the same with each pregnancy, irrespective of whether or not they have already had any affected children. A simple illustration like tossing a coin may help to explain the concept: it cannot be expected to remember whether it came down heads or tails at the last throw and cannot

therefore be expected to know what it should be at the next throw. Continuing the coin analogy, the good side of the coin should also be emphasized. For example, a couple faced with a probability of 1 in 4 that their next baby will be affected should be reminded that there are 3 chances out of 4 that their next baby will not be affected.

How different people perceive the same genetic risk varies widely, from overly cautious to reckless. What may seem high to one person may be interpreted as low or moderate by another. A useful point of reference is to compare the risk with the background risk in the general population. If the risk of any child born with any major birth defect is 1 in 30 in a population, it will be the baseline figure that any family in that population either accepts or ignores. It is reasonable to tell the families that risks of 1 in 2 and 1 in 4 can be considered high; risks of less than 1 in 100 are low; and other risks are intermediate.

THE BURDEN OF THE CHOICE

Several studies have indicated that the factor that most influences parents when deciding whether or not to have another child is the nature of the long-term burden associated with a risk rather than its precise numerical value. Burden is a mathematically undefined quantity that combines elements such as severity, chronicity and duration, availability of therapy, religious attitudes, financial costs, social discomfort, education, and the like. Parents of a baby recently diagnosed to have a genetic condition may have little idea of what lies ahead for the child and themselves. It is necessary to give the parents an understanding of what is going to be involved in the care of the child including the length of survival, the quality of life, treatment and complications, and the cost. The judgement of burden is a very personal one. A physical handicap may be a severe burden for one family, whereas another may find that tolerable but a mental handicap unacceptable. Some parents are prepared to accept a 1 in 4 risk of producing a child with a lethal condition like Pompes disease, knowing that the child will not survive long or be normal. Therefore, individuals differ significantly when estimating the burden and acceptability of a condition. It is particularly important for the couples to realize that in general there is no “right” or “wrong” decision to be made, but that the decision should be the right one for their own particular situation. It is also important that the counselors do not judge

"success" or "failure" in terms of a particular outcome, and that they give support to families whatever their decisions may be.

LIFE-PLANNING AND REPRODUCTIVE DECISIONS

There are also alternatives that may improve the family-at-risk's chances of having healthy children. In general term, adoption is a perfectly reasonable option, although in practice the number of couples wishing to adopt far exceeds the number of babies and children available for adoption. Some form of reproductive assistance may be applicable to some families. Artificial insemination by donor is genetically appropriate if a disease is recessive or if the father is the carrier of an autosomal dominant condition, but it is of limited appeal and many couples and ethnic groups may find it unacceptable. *In vitro* fertilization with donor ovum may be possible if the mother is the carrier. In both situations, the risk of the same genotype in the donor must be excluded.

Prenatal diagnosis with selective termination of an affected fetus is now an accepted option for families at high risk of having a child with a serious birth defect that may or may not be genetic in nature. The availability of prenatal diagnosis nowadays gives many high-risk couples the courage to try again for a healthy baby.

PRENATAL DIAGNOSIS

The ability to diagnose fetal abnormality has become increasingly sophisticated, and the once reclusive fetus has become the unborn patient. Normal and abnormal fetal anatomy can now be accurately delineated by powerful non-invasive imaging techniques. The fetal genome can be probed for defects by chorionic villus sampling, amniocentesis and fetal blood sampling. As technology improves, the indications for and accuracy of prenatal diagnosis can only increase. Nowadays, a large proportion of pregnant women undergo at least one ultrasound study during pregnancy, and more congenital abnormalities are being detected before birth even in low-risk women. Pregnancy screening tests for specific disorders like neural tube defects, Down syndrome, and hemoglobinopathies will help to select out at-risk pregnancies for more elaborate investigations. Prenatal diagnosis not only gives women the option to interrupt an abnormal pregnancy but also allows the physician to give

optimal care during and immediately after delivery. When intrauterine therapy and even fetal surgery become more feasible, prenatal diagnosis will become even more important in the process of genetic counseling.

Couples must know the difference between prenatal genetic screening (e.g. maternal serum alpha-fetoprotein screening) and prenatal genetic diagnosis. The former is generally not designed to detect all affected fetuses; there must be an acceptable balance between false-positive and false-negative results. On the other hand, prenatal genetic diagnosis indeed attempts to detect each affected fetus.

When offering prenatal diagnosis, it is critical that extremely clear counseling should be given to the couple, spelling out the expectations and limitations of the testing prior to any procedures. The focus of the diagnostic procedure must be explained. It is very easy for the woman to conclude that a normal test result shows that the baby will be "normal" when in fact only a short list of conditions has been excluded. It is therefore important to point out that a condition or conditions for which the unborn child had a risk higher than that of the general population have been focused upon. After these conditions have been excluded, the pregnancy still stands at the same risk for many other potential problems as others in the population. Couples must understand that undergoing prenatal screening or diagnostic testing cannot guarantee a normal child.

Although the possibility of an abnormal test result should be discussed beforehand and the options considered, it may not be necessary for the woman to make a decision prior to learning the test results. The implications of the diagnosis must be reviewed with care, sensitivity, and accuracy. The options for the woman are to terminate the pregnancy or to carry it to term. The decision to terminate must be made in collaboration with the obstetrician who will perform the procedure so that the process can be described and possible complications reviewed. The choice of procedure depends on the stage of pregnancy, and the complications are specific to the particular procedure.

A recent extension of prenatal diagnostic techniques is pre-implantation diagnosis. An ovulated egg is removed and fertilized *in vitro* with the husband's sperm. It is then cultured in the laboratory up to the eight-cell stage. One cell is carefully removed and the DNA sequence is amplified by polymerase chain reaction. The resultant amplified DNA is then studied using appropriate gene probes to see if the pre-embryo carries a particular genetic disorder. If not, it is implanted in the mother's uterus and

the pregnancy is allowed to continue. Since the techniques of reimplantation diagnosis obviate the possibility of abortion, it could well become the preferred methods of prenatal diagnosis in future although such methods may be limited to certain specialized centers because of the technical problems.

CARE OF THE FAMILY

The emotional impact of having a child with birth defects on the family is considerable and should be carefully explored. Psychological defences underlie all genetic counseling sessions. If not appreciated, these defences can impede the entire counseling process.

Families and individuals will usually go through predictable stages of emotional response when a child is found to have a genetic disorder: first there is denial, then a sense of loss and grieving, then anger and seeking someone to blame, followed by resignation, acceptance, and the search for meaningful and useful action. Depending on which stage they are experiencing, information will be received in very different ways. In addition, different family members will often be at different stages. They will need an opportunity to ventilate their fears and anxieties. The counselor or members of the counseling team must understand the process of mourning and contribute to the emotional healing process in the family.

Many families have difficulty in knowing how to let friends and other family members know about the problems the affected child has. It is important to help the family find the language to deal with their particular situation. It is also important to help the family find ways to break the silence that their friends often maintain because of their own discomfort regarding the problems of the child or their concern over embarrassing the family. If the child has obvious physical abnormalities, the family needs to prepare themselves to deal with the curiosity and rudeness of strangers. They should be encouraged to smile and be interactive on their "good days", while they should feel free to respond to rudeness by identifying it as rudeness on their "bad days". Discussing these feelings and experiences with other parents who have a similarly affected child can be helpful to a family in normalizing and validating their feelings and in finding alternative ways to deal with difficult situations.

It is important to remind the family that the decisions and plans they have made need to be discussed regularly among the family members,

since many things can change over time. The family of a child with Down syndrome may find it helpful to sit down at appropriate time and ask how each member is coping. Are the normal children being neglected? Does all the activity of the family center around the child with a disability? Is the marital relationship suffering because the parents have little time alone together? By recognizing such problems, a family can plan changes for the coming year, making sure that time is set aside for the entire family, and that each family member receives special attention. This reassessment will help maintain the cohesion of the family and allow members to cope with the stresses of disability without sacrificing their own needs.

The parents of a child born with an abnormality need to become experts on that abnormality. Frequently, their experience will provide them with more information and knowledge than most physicians have, particularly in the case of a rare condition. They are encouraged to develop their knowledge through appropriate reading and to seek out other parents or a support group for families of children with their particular disorder. In reading medical information, a family will probably need a certain amount of interpretation so that they do not go off on tangents. In the long run, such knowledge and support will be useful in obtaining the best possible care for their child, as well as helping to realize the child's full potential.

FURTHER READING

1. Rimoin DL, Connor JM, Pyeritz RE, Korf BR (eds.), *Principles and Practice of Medical Genetics*, 4th ed., Churchill Livingstone, London, 2002.
2. Harrison MR, Evans MI, Adzick NS, Holzgreve W (eds.), *The Unborn Patient: The Art and Science of Fetal Therapy*, 3rd ed., WB Saunders, Philadelphia, 2001.
3. Rodeck CH, Whittle MJ (eds.), *Fetal Medicine: Basic Science and Clinical Practice*, Churchill Livingstone, London, 1999.
4. Twining P, McHugo JM, Pilling DW (eds.), *Textbook of Fetal Abnormalities*, Churchill Livingstone, London, 2000.
5. Milunsky A (ed.), *Genetic Disorders and The Fetus: Diagnosis, Prevention, and Treatment*, 4th ed., The John Hopkins University Press, Baltimore and London, 1998.
6. Stevenson RE, Hall JG, Goodman RM (eds.), Human Malformations and Related Anomalies, *Oxford Monographs on Medical Genetics No. 27*, Oxford University Press, New York and London, 1993.

7. Mueller RE, Young ID (eds.), *Emery's Elements of Medical Genetics*, 10th ed., Churchill Livingstone, London, 1998.

8. Brock DJH, Rodeck CH, Ferguson–Smith MA (eds.), *Prenatal Diagnosis and Screening*, Churchill Livingstone, London, 1992.

9. Harper PS, *Practical Genetic Counselling*, 5th ed., Butterworth– Heinemann, Oxford, 1998.

This page intentionally left blank

66

The Child with Learning Problems

Ho Lai Yun

INTRODUCTION

The child with learning problems has a marked disadvantage in a society that places an extremely high value on achievement, rewarding individuals who are successful and ignoring or punishing those who are not. Children who deviate from "normal" pathways of academic achievement and progress usually create concern, anxiety, and stress for parents, teachers, and those who are attempting to help the child reach his full potential. Many will resort to seeking medical assistance. Yet medical practitioners, including pediatricians, are ill equipped to offer adequate help. Past medical apathy may have been the result of little or no formal training in his area, the time-consuming nature of the problem, and the lack of assignment of a specific role. In addition, uncertainty surrounds terminology, classification is still in a state of evolution, manifestations may be deceptive, and causation remains essentially hypothetical. As each discipline tends to perceive problems in its own context, management therefore varies with seasonal fashions. Appropriate educational resources are both stretched and limited and eventual outcome is doubtful.

Learning problems are common, with a conservative estimate of some 5–15% of school-aged children struggling with learning or requiring special help. The prevalence varies with definitions and diagnostic criteria. These problems affect more boys than girls, the most commonly quoted ratio being approximately 4:1.

Although learning problems are fundamentally educational problems, the complex nature of disorders of learning has prompted increased involvement of medical professionals. There are important roles for the pediatricians in the care of these children. These include:

1) The early screening and detection of the developmental dysfunction and exclusion of neurological and other health problems.
2) Careful evaluation to elucidate the nature and extent of the underlying dysfunctions, the possible etiological factors, the behavioral complications, and other important environmental factors.
3) Participation in multidisciplinary management teams.
4) Application and supervision of medical therapies.
5) Coordination of evaluative, therapeutic, consultative, educational, and counseling services.
6) Long-term monitoring and management.
7) Advocacy for the child and family in the community, and particularly within the school system.

EMOTIONAL SEQUELAE OF LEARNING DISORDERS

Problems in learning often occur at critical points in a child's development. During the interval from three through nine years, children initiate attempts at independent functioning and definition of self-identity. The developmental changes formulate the basic structure of their future personality. A complex relationship exists between the physical, cognitive, and emotional components of a child's development. The child with learning problems experiences numerous obstacles that prevent a healthy interaction between these factors. Considerable emotional damage can result from protracted periods of anxiety and stress emanating from a tense or disappointing learning experience. Frequently, the emotional components of the learning disability represent a more serious threat than does the specific learning disorder *per se* (Fig. 1).

Parental responses to the underachieving child are often unrealistic and harsh, expressing their feelings of guilt and disappointment in the

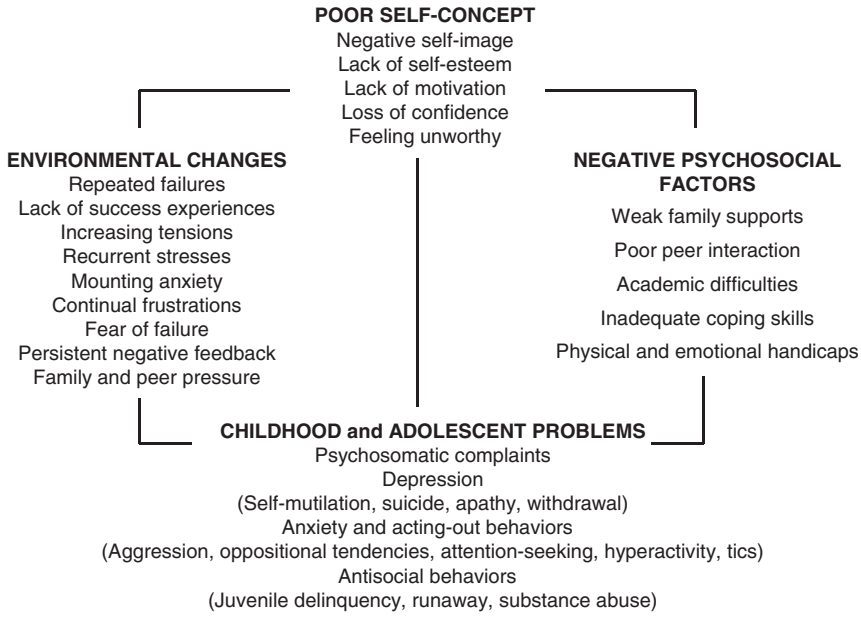


Fig. 1 Emotional sequelae of school failure.

form of anger and hostility. The bottom line of repeated frustrations, anxieties, and negative reinforcements is a loss of self-esteem and a lack of self-confidence. The disrupted balance can ultimately result in disturbed behaviors at home and in the classroom. The child often enters the examination room with his eyes and head down, realizing that once again he is expected to perform for someone who will reveal his lack of ability. He frequently exhibits defensive behaviors, attempting to excuse what he anticipates to be a poor performance. The child becomes increasingly frustrated with his inability to perform a task successfully, exhibiting aggressiveness toward the examiner in the form of resistance, poor cooperation, lack of attention, or a flippant attitude. Lack of success experiences generates additional emotional stress. The child anticipates difficulty in almost every situation and becomes emotionally paralyzed with fear of failure and expected rejection. Because school attendance is mandatory and the child cannot leave physically, he compensates by withdrawing psychologically, daydreaming and thinking more pleasurable thoughts of enjoyable times. He avoids what he regards to be aversive stimuli: the teacher, the writing board, and classmates. Occasionally, he may manifest psychosomatic complaints and symptoms of school phobia. If the academic environment

remains uncomfortable, the child may participate in school truancy or, in some instances, school vandalism.

Hyperactivity is a very disturbing and disrupting classroom behavior for teacher and peers. It has been assumed to be associated only with attention deficit hyperactivity disorder. However, hyperactivity may also be situational, induced by an emotional state such as anxiety, tension, and frustration, and is sometimes overlooked in haste to define an organic etiology. The child who reacts aggressively to the teacher frequently craves attention, gaining recognition by acting-out behaviors.

Eventually, the effects of poor self-concept extend beyond the school environment to involve behaviors at home. Chronic feelings of worthlessness, lack of confidence and motivation, and depression, if unattended to, may result in serious adolescent and adult personality disorders. Numerous investigations have revealed that an overwhelming number of youth offenders are severely retarded in reading. The reading disabilities are not caused by antisocial and rebellious attitudes that made them difficult to teach, but by underlying specific learning disabilities. Perceptual handicaps causing poor academic skills are also found in a large number of youth offenders. The severity of the reading disability is directly proportional to the seriousness of the perceptual deficits, confirming a relationship between learning disorders and delinquent behavior. In addition, many juvenile delinquents present a history in which there are prominent characteristics of attention deficit hyperactivity disorder such as short attention span, poor impulse control, and distractibility. The child who is continually frustrated may attack the system that causes his discomfort, that is, the school and society in general.

FACTORS AFFECTING LEARNING

Many children with learning problems do not have a single identifiable cause. A number of factors are likely to contribute to a child's problems. These can include constitutional or intrinsic factors within the child or factors within the environment. It is the interaction of these factors, which over time leads to the dysfunction, that presents as learning difficulties

Within the Child

- *Perinatal stress*.
  While any form of perinatal stress may be a potential contributor to subsequent learning problems, there is no clear one-to-one relationship.

Prematurity, intrauterine growth restriction, hypoxia, and other perinatal factors are sometimes associated with subsequent developmental difficulties. However, the majority of children with learning problems do not have any abnormal perinatal history. Conversely, a significant number of children with perinatal stress subsequently have normal academic achievement. The beneficial effects of a favorable environment may modify the potential of a poor outcome resulting from perinatal stress significantly.

- *Chronic disease and health problems*.
  Children with any form of chronic disease may have learning problems. The disease itself may be responsible (e.g. seizure disorder) or in some instances the treatment of the disease may contribute (e.g. side-effects of medications). There may be frequent absences from school because of hospitalizations or episodes of illness. A child with chronic disease may have low self-esteem and self-confidence, resulting in decreased motivation. Peer relationships may be affected, as his peer group may perceive the child as being different in some way. The child may be less active and not able to participate in regular classroom and physical activities. Parents and teachers may sometimes unwittingly reinforce the sick role in these children. However, there are many children with chronic and disabling illnesses who do not appear to suffer any learning problems, and, in fact, a number seem to be spurred on to later achievement.

- *Genetic factors*.
  It is commonly said that cognitive ability, intelligence and other personality traits are inherited. A family history of learning disabilities may be present and in some children there does seem to be an inherited basis for their problems. However, it is only in a minority of children that a specific genetic diagnosis or a recognizable syndrome can be made and are associated with decreased cognitive ability and learning difficulties.

- *Sensory impairment*.
  Children with subtle hearing problems are at increased risk of learning problems. This applies not only to those with sensorineural deafness, but particularly to those with repeated ear infections, leading to chronic otitis media and fluctuating conductive hearing loss, which often is unsuspected and undetected. These children may develop

subtle language and academic problems, as well as problems with attention and behavior. Similarly, children with any form of visual problems that affect acuity or eye movements will be at risk.

- *Temperamental styles and personality traits.*
  Children are born with unique temperamental characteristics that define their particular behavioral styles. These characteristics are derived from constitutional, intrauterine, central nervous system, and postnatal environmental factors. These temperamental characteristics do not, in and of themselves, cause problems. Rather, the "Goodness of Fit" between the child's temperament and the demands and expectations of the parents, teachers, or caregivers is important in avoiding and resolving conflicts that may lead to suboptimal learning experiences and difficulties.

- *Developmental differences or weaknesses.*
  The major contributing factor to the learning problems in many children is weaknesses in one or more areas of development. These are often subtle and not apparent during physical or neurological examination, and only elicited by detailed neurodevelopmental assessment. It is only when the child is asked to perform specific age appropriate developmental tasks that these weaknesses become apparent. Areas of development that are important for achieving learning competence include: language (receptive and expressive), auditory and visual sequencing, perceptual skills, motor skills (gross and fine motor functions), and attention. Each child possesses a unique profile of strengths and weaknesses in these areas of development. The concept of "differences in learning styles" is introduced to emphasize the different approaches to academic and other tasks that children naturally use according to their available strengths and areas of relative weaknesses. These differences are the manifestations of complex profiles of distinct but interdependent biologically based capacities.

Within the Home

- *Stability and supportive qualities of the family.*
  These factors, as well as the physical and emotional health of the child, all determine the degree to which the child, can mobilize his intrinsic abilities for learning. Children living in deprived socio-economic circumstances are at greater risk of reduced learning outcome. There are

multiple factors responsible, including suboptimal housing, inadequate medical care and nutrition, family disruptions, lack of early stimulation and learning experiences, poor role models, low parental education and expectations, and exposure to different combination of abuse and neglect.

- *Cultural differences*.
  Children from different cultural backgrounds may have problems with language, with adapting to a new culture, and coping with the different values and aspirations of their peer group on the one hand and family on the other. Maternal social isolation and depression are known to be associated with poor language acquisition in children and subsequent poorer learning.

Within the School

Schools themselves exert powerful pressures and expectations, which may or may not match or take into account an individual child's strengths and weaknesses. Therefore, the educational experience may be positive or negative.

Class size, teacher personality, staff stability, cultural mix, resource availability, methods used, disciplinary attitudes, and peer contacts and relationship — all may influence children's responses to learning, and their reception of education offered. Children may vary in their school performance from year to year, depending on who is their teacher or the degree of acceptance with peer groups. Fear, anxiety, uncertainty and diminished self-esteem can then block learning.

RECOGNIZABLE CLINICAL TYPES OF LEARNING DISORDERS

The following broad clinical types of learning disorders have been recognized:

- *Global developmental delays* or retardation in learning capacity refer to the learning delay and difficulties associated with overall reduced intellectual capacity. Many children in this group always function below normal levels academically. They struggle because their ability is in the borderline range, and they have particular difficulties with language skills.

- *Associated learning disorders* refer to the learning difficulties children have because of another disorder, such as cerebral palsy, seizure disorders, or a hearing or visual defect. Some of these disorders are behavioral in nature; such as autistic spectrum disorder, conduct disorders, and disorders associated with emotional and social adjustment.
- *Attention deficit*. A number of children have problems in focusing attention and maintaining concentration. Many are also described as being overactive, fidgety, impulsive and distractible. All these behaviors may contribute to problems in learning, as well as social difficulties. Some of these children have attention deficit hyperactivity disorder, which is part of their intrinsic make-up. In others, the behaviors are situational, resulting from emotional problems or anxiety, which are the sequelae of underlying developmental weaknesses such as auditory sequencing or language deficits, or hearing defects.
- *Specific learning disability*.

SPECIFIC LEARNING DISABILITY

Learning disability is a generic term that refers to a heterogeneous group of disorders manifested as significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual and are presumed to be due to central nervous system dysfunction. Even though a learning disability may occur concomitantly with other disabling conditions (e.g. sensory impairment, mental retardation, social and emotional disturbance) or environmental influences (e.g. cultural differences, insufficient/inappropriate instruction, psychogenic factors), it is not the direct result of those conditions and influences.

This definition emphasizes the heterogeneity of learning disabilities, recognizes that it extends beyond childhood, and acknowledges that there is comorbidity with other developmental disabilities and cultural disadvantages.

Learning disabilities may represent a deficit in one aspect of learning but more commonly affect multiple areas of function and in some cases are global. Seven areas in which a child might exhibit learning disabilities have been specified: basic reading skills, reading comprehension, oral expression, listening comprehension, written expression, mathematical

calculation, and mathematical reasoning. Reading disability or dyslexia is the most common of these disorders.

The concept of specific learning disabilities has moved from minimal brain damage to dysfunction. More recently, it has been proposed that the condition be considered more as a matter of "developmental differences". Developmental variation with respect to learning and school performance is determined by the maturity and efficiency of such neurodevelopmental modalities as cognition, speech and language, memory, social functioning, motor and sensory capabilities, and attention. Specific neurodevelopmental weaknesses will in turn prevent learning in one or more academic areas. It has been a defining principle that such weaknesses are unexpected, given the overall intellectual functioning of the child. Learning disabilities thus are not simply the result of global delays in learning capacities, of major sensory impairments, or a consequence of major social or emotional stressors.

CLINICAL MANIFESTATIONS

The specific patterns of academic performance and behavior represent the final common pathways and the convergence of multiple forces, including interacting cognitive strengths and deficits, environmental or cultural factors, temperament, educational experience, and associated deficits in the areas of executive function, attention, and emotional and social adjustment.

Learning problems emerge when external expectations require performance in the area(s) of the child's vulnerability or weaknesses. There are four distinct periods of challenges:

Preschool Level

For the child in the preschool age, *school readiness* requires the ability of the child to "settle", i.e. to inhibit motor impulses to allow attention and focus on demand. Basic language and communicative skills, as well as numerical concepts, are necessary. Sensory capacities to discriminate auditory and visual differences in spoken sounds and letter shapes respectively are also required. Fine motor competency must be sufficient for pencil use. Social demands require the child to be able to take turns and express his own needs verbally. Children may adapt poorly to their

first year or two of school because of specific weaknesses in any of the previously mentioned areas or because of global immaturities, commonly seen among boys and in children who are chronologically young at the start of the school.

Among the characteristics displayed by preschool children with learning disabilities are inadequate motor development, language delays, speech disorders, and poor cognitive and concept development. Common examples are the 3-year-old who cannot catch a ball, hop, jump, or play with manipulative toys; the 4-year-old who does not use language to communicate, has a limited vocabulary, and cannot be understood; and the 5-year-old who cannot count to 10, name colors, or work puzzles. In addition, preschoolers often exhibit behaviors of hyperactivity and poor attention.

Early School Years

The period in the early school years is heavily focused on the *acquisition of basic academic skills* in reading, writing, spelling, and arithmetic. The child at this stage is absorbing tremendous amount of rote knowledge with respect to symbol systems of letters, words, numbers, and the rules by which they are combined. Children are asked to use a broad array of facilities in concert. Delays in language understanding and usage will put early learners at risk of failure. Memory is heavily relied upon in all its forms, such as active working memory, recent recall, long-term information storage and retrieval. Motor output in writing may be frustrating and rate-limiting for children with fine motor delays. These basic learning skills should be almost automatic and require minimum efforts by the end of primary school.

Therefore, children with learning disabilities often fail not only in reading, but also in mathematics, writing, or other school subjects. The behaviors frequently seen in the early primary school years are inability to attend and concentrate, poor motor skills, as evidenced in the awkward handling of a pencil and in poor writing, and difficulty in learning to read.

Late Primary and Early Secondary School Levels

Toward the later part of primary school and throughout secondary school years, the child is faced with a new set of challenges: *self-organization of tasks and the processing of greater volumes of information*. Task requirements

now emphasize an increased volume of received information, both written and spoken, and increased demand for written products. The child is also expected to be substantially more self-reliant and responsible, both at home and at school. Therefore, the child is tested on his executive functions (e.g. planning, organizing, monitoring, efficiently performing tasks and activities), and his capacities to process larger volumes of information, reading, homework, and written work, which require self-monitoring and self-modulation.

Emotional problems become more of an impediment after several years of repeated failure, and the children become more conscious of their poor achievement in comparison with that of their peers. For some students, social problems and the inability to make and keep friends increase in importance at this age level.

Upper Secondary and Pre-university Levels

At these levels, academic skills are increasingly used as a means of information gathering and understanding and less as an end in themselves. Literature, mathematics, and both social and natural sciences all rely on basic skills acquired earlier. *Higher cognitive skills and abstract thinking* are then required.

Therefore, a radical change in schooling occurs, and adolescents find that learning disabilities begin to take a greater toll. The tougher demands of the curricula and teachers, the turmoil of adolescence, and the continued academic failure combine to intensify the learning disability. Adolescents are also concerned about life after completing schools. They may need counseling and guidance for college, career, and vocational decisions. A few adolescents may be drawn into acts of juvenile delinquency. As they tend to be overly sensitive, some emotional, social, and self-concept problems often accompany a learning disability at this age.

COMPREHENSIVE ASSESSMENT

Comprehensive assessment of a child with learning problems should be multidisciplinary and broad-based to take into account all the possible constitutional and environmental factors that may contribute to the child's problems. Close communication with the child's school is

essential. The evaluation process is time-consuming and usually takes several sessions.

A *complete review of the history* is essential to gather a description of the child's current function from multiple perspectives (parents, teachers, and, in case of older children, the children themselves). The focus is on information related to the various factors affecting learning. It is also important to review the child's early history to document developmental milestones, relevant health factors, early temperament, and the onset and course of the learning and behavioral difficulties. Family history should include the composition and social circumstances of family, the education and expectations and other relevant environmental factors. A review of the child's school records and reports should be done through early liaison with the school. It is equally important to have some knowledge on the school's perception of the child's academic and developmental strengths and weaknesses, peer relationships and classroom behaviors. Questionnaires for parents and teachers have been used to provide a standard structure for obtaining a history and also to save time.

A *complete physical, sensory, and neurodevelopmental examination* should be conducted to rule out any causal or associated medical problems. There are many medical conditions that can have a bearing on the learning process. These include seizure disorders, allergies, anemia, hypothyroidism, and sensory impairments. Unusual patterns of physical growth and maturation, the state of nutrition, and signs and symptoms of abuse and neglect should be actively pursued. It is also important to rule out genetic or chromosomal abnormalities, such as Turner syndrome, Fragile X syndrome, and Williams syndrome.

The traditional neurological examination is usually normal, although there may be mild asymmetries of tone, qualitative differences in reflexes, or several beats of clonus. The developmental examination is intended to demonstrate the child's neurological maturity and the qualitative efficiency of motor, sensory, and position-sense functions. Much has been written about "*soft neurological signs*" and learning disabilities. These signs are developmental in nature and do not have the same localizing value as the more traditional neurological signs. Soft signs may be found in motor, sensory, or cognitive modalities. Motor soft signs include age-inappropriate performances when standing on one foot and excessive overflow movements with rapid alternating movements and finger nose pointing. Cortical sensory soft signs include

difficulty with extinction, left-right orientation, graphesthesia, and finger identification. Disappearance of these signs is linked to central nervous system maturation but their presence is also not considered abnormal prior to seven years of age. Clusters of these signs are more significant than a single sign in isolation. Persistence of such signs beyond the ages at which they usually disappear is associated with learning disabilities, behavioral problems, and other manifestations of neurodevelopmental dysfunction.

Investigations are done only with clear indications, either from the history or from the physical or neurological examination. Special investigations such as EEG, CT scan, or MRI are rarely helpful.

Neurodevelopmental assessment should be administered to compile a profile of the child's developmental strengths and weaknesses. Any identified weaknesses should be interpreted in light of the child's academic difficulties. The most common approach for determining the existence of learning disabilities is to demonstrate a substantial discrepancy between educational achievement and intellectual potential in one or more areas of learning. Usually, this entails administration of standardized tests to measure both IQ and educational achievement. An IQ test may be helpful, but it should never be assumed to be a comprehensive audit of academically required cognitive function, nor should it represent the final word regarding a child's academic potential.

An *educational assessment* should be performed, preferably by a skilled educator or an educational psychologist, to determine and document the child's level of academic functioning, his preferred learning styles, patterns of errors, motivation and self-esteem, concentration and sensitivity to his performance. The information can then be used to develop specific educational plans and remedial programs for the student.

Other assessments should be individualized for each child according to the findings in the above evaluation components. When there are indications of the presence of significant emotional issues, either within the child or in the family context, a referral to a psychiatrist for a more in-depth mental health assessment may serve to clarify these issues. It is also appropriate for a child with language weaknesses to be referred to a speech and language therapist for a formal assessment. The child who fails a screening test of auditory acuity should be referred for a formal audiological assessment. The social workers should be involved in connecting the child and the family to the relevant supports in the community.

The evaluation process should consequently be able to describe a child on five levels:

- *Neurodevelopmental*, with an analysis of constitutional, maturational, and developmental factors interfering with function and underlying strengths.
- *Psychosocial*, elucidating factors in the environment, in the family, and in the past experience or present emotional health of the child that either enhance or compromise current performance.
- *Secondary psychological*, with an account of the emotional effects of failure and the present level and stability of self-esteem.
- *Supportive*, with a description of how the family, the school, and the community can help in the child's learning problems.
- *Strategic*, with an analysis of the child's strategies for dealing with failure.

GOALS OF TREATMENT

The goals of treatment are:

- achievement of academic competence;
- treatment of associated deficits; and
- prevention of adverse mental health outcomes.

The individual child's unique learning profile must be appreciated, and teaching must aim to challenge the child in ways in which he can successfully respond. Where weaknesses are significantly disabling, teachers and parents must help the child to compensate for them by capitalizing on the available strengths and skills as much as possible. Despite clear goals, it is the means by which they are met that provide the ultimate challenges. Optimal management of learning disabilities requires the collaboration of teachers, parents, and pediatricians. Teachers should not limit their treatment goals solely to the need for achievement of academic competence. Pediatricians need to treat associated disorders, such as inattention, to facilitate academic therapy.

The principal components of effective management include:

- multidisciplinary approach in formulating an individualized educational plan;

- counseling of parents and children;
- educational therapy;
- developmental therapies;
- strengthening the strengths;
- medical treatment and pharmacotherapy;
- advocacy; and
- periodic re-evaluation.

COUNSELING PARENTS AND CHILDREN

It is important to establish a good continuing relationship with the children and the family. Counseling may be individual or group. Family-centered interaction should focus on relieving guilt about the child's problems, diffusing blame, and delivering basic information about the nature of the disability, including his strengths and weaknesses. Other issues to be discussed include homework, behavior management techniques, discipline, parental expectations, and the child's self-esteem. The language used should be simple, non-technical, optimistic and upbeat, non-accusatory, and with appropriate analogies. It is important to be careful not to moralize about the child's problems. Such terms as "bad", "lazy", and "poorly motivated" never motivate and may intensify negative self-image. Reassurance may be comforting and beneficial, but its excessive use without proper basis can prevent a child from receiving appropriate services early or impede further evaluation. The parents should be encouraged to create an atmosphere at home that is conducive to learning, and they can be counseled regarding the most appropriate way to motivate and encourage their child.

Children should also be included in discussion about their learning problems; otherwise, they are likely to fantasize uncomfortably about private conversations between professionals and their parents, and fear that they are defective. Counseling should aim at explaining to the children their specific problems and strengths in understandable language, with encouragement to talk about such problems and to report how they are coping. Finally, the family should be provided an ongoing source of information and relevant support groups in the community.

Sometimes, psychotherapeutic counseling may be indicated in family disorganization and conflict, or when significant psychopathology is

evident in the child, whose emotional needs stand in the way of learning. Such counseling may need to involve the parents and sometimes the entire family.

EDUCATIONAL THERAPY

Educational therapy is the cornerstone of treatment of learning disabilities. There is no single "best" approach and therapy must be tailored to the individual needs of the child. Health professionals need to know more about the educational systems and the alternatives, and educators must become more familiar with the medical and developmental aspects of school failure.

It is important to know that many educational interventions have not been adequately evaluated and are based on anecdotal reports and uncontrolled trials. The individual teacher's philosophies, training and skills, and particularly the teacher's conceptualizations of children with learning disabilities will influence the professional approach and the design of programs for these children. There are three basic concepts; each has implications for curriculum content and teaching methods.

1) *Developmental delays*
   The fact that a child's development in one or more areas of functioning is delayed in comparison with "normal" children has led to the assumption that it is appropriate to treat the child as younger than his chronological age. Consequently, teaching materials and methods that have been shown to be useful with the young normal children are considered appropriate for the "delayed" child. There is also the tendency to teach developmentally sequenced skills without questioning their relevance for each child. Therefore, it can be seen in classes of children with a specific reading disability who are given reading materials that are of no interest to them, or again in classes of teenage students with global cognitive disabilities who are being taught body parts through action songs commonly used in preschool settings.

   The positive aspect of this concept is that it implies that development is possible and that it may even be possible to catch up with normally developing peers. Consequently, the professionals

who conceive of children in this way are prepared to work hard to promote the children's development.

2) *Developmental deficiency*
The concept that the child is considered deficient in some way and "not having what it takes" to succeed is doubtlessly the most negative way of viewing a child with a learning disability. The consequence of this viewpoint is therefore to protect "the poor child" from more of life's hard knocks and educationally to provide a watered-down version of that usually offered to the normal child.

3) *Developmental differences*
To consider the child with a learning disability as different is to recognize that he is a unique individual with specific strengths and weaknesses. Consequently, the professionals do not try to make the child "normal" in the sense of being the same as his peers, but do assume that the child has the normal rights to develop to his maximum potentials, and be respected as an individual with personal preferences. The challenge to the professionals caring for these children is that they must make both value and professional judgements about what the needs of these children are and how they can best be met.

Educational therapy is a combination of direct remediation of skills through specific educational programs, bypass strategies, and curriculum modifications.

Remediation of Skills

Specific educational therapies are used to bolster weak academic skills. Trained professionals, such as educational therapists, reading specialists, and mathematics tutors, devise methods and techniques based on knowledge of the child's neurodevelopmental strengths to help him overcome academic lags and to improve decoding skills, writing ability, or mathematical computation. Remediation need not focus exclusively on specific academic areas. They must assist the child acquire and develop study skills, cognitive strategies, and productive organizational habits. In addition, there must be a balance between the amount of time spent in attempting to remediate weaknesses and the amount of time spent on instruction of content.

Specific educational programs will vary considerably according to the resources available to the school and the severity and nature of the child's problems. Many children with milder forms of learning disabilities can be maintained in a regular classroom with accommodations made for areas of weakness. Therefore, the teacher must not only accept such a child in the class, but must also be fully aware of the child's needs. In other cases, they may require specialized assistance outside of regular classrooms. There is a trend to "mainstream" these children with educational problems by not segregating them into "special" classes. The goals are to reduce prejudice, to promote self-esteem, to provide peer models for social and academic skills, and to reduce the likelihood of labeling. Students in need can leave the regular classroom for set periods per week for intensive remediation either on a one-to-one basis or in small group, in special settings such as in learning centers or resource rooms. Each child will receive individualized assistance from trained teachers, whose efforts are aimed both at strengthening basic skills and at overcoming specific dysfunction, through specific exercises directed at visual-perceptual motor function, sequencing, fine motor problems, or language disabilities. In some instances, a remedial teacher or teaching assistant may be allowed to work in the child's classroom. As children grow older, more emphasis is placed on direct tutorial help in subject areas, with less emphasis on readiness skills. For some children, the measures described above are still not sufficient because of the severity and nature of the learning problems, and they may need to attend a special school.

Homework is frequently an area of conflict. It may not be a good idea for parents to coach their children at home or to spend much time going over homework with them in rote fashion because of tensions that may result. The matching of the temperamental styles between the parents and the child will decide on the feasibility of this home tuition approach. Homework should be closely monitored as attempts to improve the child's academic ability may result in assigning too much homework. This may be because the child has to make up classwork that is not completed or an inability of the child to do the same amount of homework that others find reasonable. Techniques to facilitate homework performance may include reading and reviewing difficult materials with the child, minimizing the need for peripheral activity such as copying of mathematics problems or writing only spelling words rather than full sentences. Homework should also be limited to a pre-established time allotment. Ongoing

homework difficulty may indicate that knowledge of the topic areas has not been achieved and curricular adjustment is required.

Bypass Strategies and Accommodations

Despite the fact that certain academic skills are impaired, it is important that the child must continue to be exposed to appropriate cognitive stimulation. To provide this stimulation, teachers have to make certain accommodations in instructional methods, assignments, and methods of evaluation. The healthcare team should communicate with the school the results of the child's assessment, interpreting the findings for the teachers, and making clear the child's developmental strengths and weaknesses.

Bypass strategies are techniques that enable a child to circumvent neurodevelopmental dysfunctions. They do not "cure" but minimize the negative academic and non-academic impacts, and allow the child to go on acquiring knowledge and skills.

The child who has problems with attention will need a structured classroom with little distractions, preferential seating towards the front of the class, or secret signalling from the teacher when he is "tuning out". The child who has difficulty with auditory sequencing will benefit from having verbal instructions kept simple, given in small sequences, and repeated as necessary. Wherever possible, instructions should be written down as well as spoken so that the child may have the opportunity of visual reinforcement.

Some accommodations for problems in written expression include: minimize copying from the board and teacher will hand homework assignments to the child; use peer notes for studying (a study "buddy" and photocopying may be helpful); allow dictating assignments onto tape; substitute oral presentations for written assignment and oral tests for written tests, or allow shorter reports; use objective format rather than essay format in tests (e.g. multiple choice); note spelling and punctuation errors but do not penalize the child for them; and use of word-processor.

Children with reading difficulties will benefit from oral presentation of material, reading to the child, and books on tape. Similarly, calculators can be helpful in solving mathematics problems, especially when the difficulty is with computation but not concepts. Assessment and examination procedures may need to be modified, such as additional time, oral testing, or multiple choice; so are the evaluation criteria.

Curriculum Modifications

Many children with learning disabilities require alterations in the school curriculum to accommodate the individual developmental status and learning styles in order to succeed. Modifications in course content may help, e.g. children with gross motor lags may do well in other areas if excused from regular physical exercises or enrolled in an adaptive physical education program; and students with memory weaknesses may need to select courses that do not require an inordinate cumulative memory load.

Other curricular issues will become more critical as the students progress into higher grades. They include the introduction of second and foreign languages, the special demands of the sciences, and the provision of specific vocational training.

Health professionals are often asked about the benefits of a child repeating a year, receiving a particular educational program, switching to a different school, having private tutoring, or other non-medical questions. A cautious approach should be adopted in answering these requests. Such decisions are generally best made at school level, with the medical team contributing to the discussion as appropriate and according to the level of his expertise and interest.

Repetition of grade or retention does not itself alter the natural history of learning disabilities and may have a negative effect on self-esteem and socialization. The vast majority of these children do not grow out of their problems. An assertion that the child is immature and will catch up if left alone is inappropriate. Repetition may buy time to allow the child to consolidate certain skills. To be of maximum value to the child, a structured intensive remedial education program to improve on the apparent weaknesses must be planned in order to help the child to be more ready to proceed to the next level.

DEVELOPMENTAL THERAPIES

The efficacy of direct interventions to enhance weak developmental functions is still controversial. Nevertheless, some forms of developmental therapy are widely accepted. *Speech and language therapy* offers interventions for various forms of language disabilities. *Occupational therapy* strives to improve the motor skills such as writing problems and motor

clumsiness. *Social skills training* has its main objective of maintaining self-esteem and development of social skills, in order to prevent adverse mental health outcomes. Efforts to build self-esteem may include special jobs within the classroom, sports, scouts, music, drama, arts and crafts, and other supervised activities. In *behavioral therapy*, children learn about their own learning problems and are given specific exercises aimed at enhancing the weak areas. For example, a child with attention problems may be taught about his impulsivity and then provided with exercises that encourage reflection, planning, and a less frenetic tempo.

STRENGTHENING OF STRENGTHS

While the focus of management has been on correcting the child's dysfunctions, the child's strengths must not be lost. It is important that any intervention strategy should include a strong emphasis on the ongoing identification of the child's strengths, his affinities, potentials, and talents in which he can achieve a sense of mastery and triumph. Athletic skills, artistic inclinations, creative talents, and mechanical aptitudes are among the potential assets of certain students who are underachieving academically. These children need help and opportunities to develop their talents, to build on their natural and acquired proclivities, and to achieve respect and praise for their efforts. Such efforts are likely to be critical in working toward the enhancement of self-esteem. These personal assets can have tremendous long-term implications for the child's transitions into young adulthood, including career choices.

MEDICAL MANAGEMENT

The major responsibility of the medical team is to ensure appropriate treatment of any medical problems that may interfere with the child's function, such as sensory impairments, neurological problems, seizures, or chronic medical conditions. Children with allergies may have learning problems complicated by recurrent middle ear effusion, chronic nasal congestion, and fatigue. Antihistamines may cause chronic fatigue and inattention in classroom. Many children with learning problems also have associated symptoms, such as recurrent abdominal pain, headaches, enuresis, encopresis, or other psychosomatic complaints. Proper treatment of these problems can improve function in other areas.

Certain psychopharmacological agents are important adjuncts to treatment, especially in behavioral control. Most commonly, stimulant medications are used in the management of children with attention deficits. Whether stimulants improve learning abilities has not been established; but in the child with learning disabilities and attention deficits, stimulants may improve classroom performance. When depression or excessive anxiety is a significant component of the clinical picture, antidepressants may be helpful. Children receiving medication must be closely monitored.

It is understandable that many alternative treatment options have been offered for learning disabilities because there is no one best treatment or cure. Although most of these therapies have not been proved to be ineffective, their quality and validity are otherwise unproved. In addition, the cost in dollars, time, and frustration may not be inconsequential. There are two broad categories of alternative treatment. The *orthomolecular* methods include allergic hyposensitization, additive-free diets, megavitamins, sucrose-free or low-carbohydrate diets, thyroid medication, and so on. The *neurophysiological* methods include alpha-wave conditioning, motor patterning exercises, sensory integration training, optometric training, eye muscle exercises, colored lenses, and so on. Other miscellaneous therapies include anti-motion sickness therapy, chiropractics, meditation, and so on.

As advocates for parents and children, pediatricians can help to minimize their susceptibility to irresponsible claims. Many parents seek help elsewhere when they feel abandoned by the health and educational systems. Adequate continuing support and appropriate interventions with families will lessen their need to seek miracle cures.

ADVOCACY

Advocacy means that the medical team takes the child's part and pleads his cause with others. A pediatrician can play a critical role in serving as a staunch advocate for a child with learning problems.

It may first need to take place with the parents who, out of concern, misunderstanding or frustration, may have been blaming or pressuring the child, with the risk of damage to his self-esteem. Parents must be helped to reshape, rebuild and to adjust their view of the child.

A pediatrician can be especially helpful in advocating for a child in school. Through judicious letter writing on behalf of the child and by

close interaction with the school, he can provide much-needed advocacy for an underachieving youngster. Teachers and principals should be helped to understand the child's learning difficulties. There is a need to represent the rights of the child to ensure that he is not overexposed to criticism and humiliation in front of peers. He may also need to argue strongly for the child to receive and benefit from certain remedial and educational services in the school.

Pediatricians can also perform advocacy by becoming vocal citizens of their communities. Their roles include educating the community on the implications and the special needs of children with learning problems, taking part in policy making and resource allocation, and initiating appropriate multidisciplinary programs for the child and the family.

PERIODIC RE-EVALUATIONS

Children with learning problems represent a heterogeneous group. No two children require the same management plan, and future needs are unpredictable. New problems may emerge as the child's dysfunctions evolve and academic expectations undergo progressive changes. Treatment programs are complex and require the cooperation of health and education systems. Many potential gaps exist and need careful monitoring. Therefore, affected children and family require vigilant follow-up and continued individualized objective advice, preferably under the guidance of a case manager.

The goals of monitoring visits are to foster parent-child relationships, continue mental health efforts towards improving self-esteem and socialization, convey new information, and facilitate the use of existing resources. Such visits need not be more frequent than one to two per year.

OUTCOME

The outcome in a child with learning problems is determined by the nature of the disability, its severity, associated deficits, and environmental supports. With most neurodevelopmentally-based learning problems, it is common for early weaknesses to persist to varying degrees into adulthood, regardless of interventions. However, most children with learning disabilities eventually attain the academic skills required for everyday function. In other words, those who do not achieve functional literacy in

reading and mathematics during conventional education will do so as young adults, and are seldom severely handicapped. This is not to say that the "gaps" have closed. The deficits persist, but the functional thresholds are achieved. Consequently, individual differences in skills should be accommodated and accepted, and conventional education may need to be supplemented with vocational training, other functional curricula, computers, or work study.

Factors that determine whether weaknesses in learning become disabling in the long run include:

- the child's ability to compensate and cope through his unaffected strengths;
- appropriate teaching and accommodations in school programs;
- the development over time of self-valued and socially respected competencies; and
- the degree to which the child comes to understand the disability as a real but limited part of the self, and not a feature that devalues the whole.

CONCLUSION

Children who are lacking in athletic talent, music talent or mechanical capabilities are seldom punished or subjected to sarcasm or ridicule, nor are their parents accused of neglecting or overprotecting them. Should we therefore consider children with difficulties reading, writing, and so on, as having a "lack of literary talent"? Perhaps, there may be a greater need for talented mechanics than literary critics and essay writers.

The adult world should concentrate less on disease, defect, damage or disability, but to recognize the diversity of styles, a wide range of strengths and weaknesses, and consequently the multiplicity of end products or pathways representing mastery. The educational philosophy and services should be broadened to allow for the dignity and development of such differences.

The pediatrician is uniquely positioned to encourage the schools and parents to understand and be responsive to the child as a whole.

FURTHER READING

1. Silver LB, *The Misunderstood Child: Understanding and Coping with your Child's Learning Disabilities*, 3rd ed., Times Books, USA, 1998.

2. Lerner JW, *Learning Disabilities: Theories, Diagnosis, and Teaching Strategies*, 8th ed., Houghton Mifflin Company, Boston, 2000.

3. Mercer CD, *Students with Learning Disabilities*, 5th ed., Merrill and Prentice Hall, Columbus, Ohio, 1997.

4. Batshaw ML, *Children with Disabilities*, 5th ed., Paul H Brooks Publishing Company, Baltimore, Maryland, 2002.

5. Levene MD, *A Mind at a Time*, Simon and Schuster, New York, 2002.

6. Levene MD, *The Myth of Laziness*, Simon and Schuster, New York, 2003.

7. Levene MD, Brooks R, Shonkoff J, *A Pediatric Approach to Learning Disorders*, John Wiley & Sons, New York, 1980.

8. Cratty BJ, Goldman RL (eds.), *Learning Disabilities: Contemporary Viewpoints*, Harwood Academic Publishers, Australia, 1997.

9. Oberklaid F, Children with school problems, *Aust Paediatr J* **20**:271–275, 1984.

10. Parry T, Fletcher J, The child with learning problems, in Vimpani G, Parry T (eds.), *Community Child Health: an Australian Perspective*, Churchill Livingstone, Melbourne, 1989.

11. Shapiro BK, Gallico RP, Learning disabilities, *Pediatr Clin North Am* **40**:491–504, 1993.

This page intentionally left blank

67

Pitfalls in Developmental Diagnosis

Ho Lai Yun

INTRODUCTION

The major thrust of developmental assessment is early identification and treatment of children with developmental and behavioral problems and disabilities so as to correct dysfunctions if possible, minimize the impact of a child's disability or of prevailing risk factors, strengthen families, and establish the foundations for subsequent development.

Developmental assessment is fraught with difficulty, largely because all children are different, and also because so many factors may affect the course of development. Detection of abnormality is not always simple and great care needs to be taken in diagnosing developmental problems at an early stage. Because development is so rapid, there is the danger that diagnosis may be made inappropriately, especially where deviations from normal are slight. What may initially be thought to be developmental delay may turn out to be individual variability in development, or the manifestation of a particular temperament or environmental circumstances. Errors may lead to unnecessary investigations or treatment, parental anxiety, and disadvantage for the child. By the same token, it is

just as dangerous to falsely reassure parents that a child will "grow out of it" when insufficient data are available to make this assertion. Unfortunately, it is still a common occurrence for treatment or interventions to be delayed unnecessarily because of a delay in making a diagnosis, causing bitterness and resentment. In assessment and intervention, there is a fine dividing line between doing too little and doing too much.

Considering the range of developmental problems in infancy and childhood, one confronts a series of traps into which parents and health professionals are prone to fall into. All diagnosis should be based on the history, full physical and developmental examination, special investigations where relevant, and the proper interpretation of the results.

NORMAL DEVELOPMENT AND NORMAL VARIATIONS

A thorough knowledge of the normal patterns of development is a necessary preliminary to the study of disease. Children differ widely in all aspects of development and the range of normality is quite difficult to delineate, if not impossible. A line can never be drawn between normal and abnormal. The so-called "norms" of development, on which most of the commonly used developmental tests are based upon, are the established average of a highly selected population, instead of on the population as a whole. It is important to appreciate that there is a wide variation between different populations and within a given population. Therefore, all that can be said is that the further away from the average a child is in any field of development, the less likely he is to be "normal". Failure to recognize this leads many to declare that a child should reach certain milestones by a specified age, as if all normal children are the same, with no normal variations.

Variation in the Rate of Development

There are normal variations in the age of acquiring certain skills. This can be demonstrated by considering the age range for "walking unsupported" (Table 1).

At 15 months, the majority of the 10% of children not yet walking unsupported will be normal but, in a small proportion, there will be an underlying medical problem. At 18 months, 97.5% (two standard deviations from the average age) of children will be walking independently. Of

Table 1 Age Range of Children Walking Unsupported

Age (Months)	Percentage (%) Walking Unsupported
11	25
12	50
13	75
15	90
18	97.5

those who are not, many will be normal late walkers, but a higher proportion will have an underlying medical problem, such as cerebral palsy, a muscle disorder or global developmental delay. Hence, any child who is not walking by 18 months should be carefully examined. Thus, 18 months can be set as a limit age for children not walking. Setting the limit age earlier may allow earlier identification of problems, but will also increase the number of children labeled as "delayed" who are in fact normal.

Variation in Patterns of Development

The pattern by which children reach key milestones varies. There are well-established variants in motor development.

Normal infants progress from immobility to walking, but not all do so in the same way. The most common sequence is the child who sits/crawls-on-all-fours/stands/cruises-around-furniture/walks. The average age of walking in this group of children is 13 months. However, some children do not follow this pattern. "Bottom shufflers" are the most common variant. They do not crawl but shuffle around on their bottoms or sides: a sit/shuffles/stand/walk sequence. They walk late, on the average by 17 months, and tend to be mildly hypotonic early on. Other children crawl with their abdomen on the floor (so-called "commando crawling" or creeping) and a very few just stand up and walk. The limit age of 18 months for walking applies mainly to children who have had crawling as their early mobility pattern. Children who bottom shuffle or "commando crawl" tend to walk later than crawlers, so that within those not walking at 18 months, there will be some children who are simply reflecting variants of normal locomotor patterns.

There is even more variation in the rate of acquisition of language, social skills and behavior, e.g. when children can dress themselves or are

toilet-trained. Some may follow a familiar pattern. It is reassuring (but no more than that) if other members of the family have followed a similar course of development and turned out well, e.g. late in talking or walking. But the concept of limit ages still applies to determining whether or not a child's developmental progress is normal.

The knowledge of difficulties in developmental assessment and diagnosis can therefore only be acquired, after becoming thoroughly acquainted with the normal, and the normal variations that are so common, and efforts to understand the reasons for these variations.

FOUR FUNDAMENTAL AREAS OF DEVELOPMENT

A prerequisite for appreciating and avoiding the pitfalls in developmental diagnosis is an understanding of the four fundamental areas of development.

1) *Motor* milestones are excellent indicators of motor competence but correlate poorly with intellectual capacity.
2) *Language* development, which is measured in terms of expressive and receptive capability, is the best predictor of intelligence.
3) *Problem-solving* skills is perhaps the second best indicator of intellectual function and represent a measure of non-verbal intelligence. They are sometimes referred to as adaptive or visual-motor milestones. The evolution of both the language and problem-solving milestones is independent of motor competence. They may be obscured by motor disability and, as a result, may be more difficult to demonstrate.
4) *Psychosocial* or *affective* milestones consist of two subgroups founded on verbal and non-verbal intelligence, respectively. Those milestones involving interpersonal interaction have a language foundation and therefore indirectly reflect verbal IQ. Activities of daily living milestones are based in the visual-motor domain and depend highly on environmental exposure and practice. Because of this, the activities of daily living milestones reflect intellectual potential to a variable degree, depending on the exposure a child has received. Psychosocial abilities are critical in understanding the whole child and in making a meaningful statement about behavior, but they lend little additional information

to the assessment of intellectual and motor competence. Thus, each of the four areas of development has its own special relevance to the diagnosis of specific motor or cognitive impairment.

HETEROGENEITY OF DEVELOPMENTAL DISABILITIES

Developmental outcome is the end result of the complex, continuous transaction between constitutional or intrinsic factors in the child, and environmental influences and life events. A child's development will be affected differently by different biological and environmental factors. This will depend on the nature of the insult, the stage of neurological development at the time, and the site of the insult. While those areas of the brain that are undergoing the most rapid maturational changes are the most susceptible, it should also be remembered that the developing brain is more plastic in its ability to recover.

Because of the variety of insults that may occur at various stages of development, there is a wide range of neurological impairments and heterogeneity of developmental disabilities. It is dangerous to assume homogeneity in the developmental delays. For example, a child with cerebral palsy and severe motor handicaps, whilst at increased risk of problems in other areas of development, may in fact have normal cognitive functioning. Communication disorders may severely inhibit a child's ability to display his or her intelligence. It is important therefore to define development according to the above four fundamental areas in its very broadest sense, which includes motor, sensory, cognitive, language, social, behavioral and emotional development. For this reason, a multidisciplinary approach to the assessment and management of developmental disabilities is necessary. The disabilities in the various areas of development are not mutually exclusive, but are found in various combinations of differing complexity.

NATURAL BIASES ON AGE OF DEVELOPMENTAL DIAGNOSIS

The way parents and caregivers typically think about their children's growth and development provides some natural biases in the timing at which certain developmental disabilities may be suspected.

During the first 6 to 10 months of life, growth is the primary concern of most parents. The body weight and the rate of weight gain of the child are always the first questions asked at every well child visit. Motor development begins to take priority after about 6 months of life, and is the highest ranking concern by 10 to 12 months. It is only much later, roughly 18 to 24 months of age in most families, that parents begin to pay greater attention to cognitive areas, especially language development.

These natural biases have influenced the age of various developmental diagnoses. Motor disabilities are consistently identified earlier than cognitive deficits, even when severe. The mean age of diagnosis for cerebral palsy is about 12 to 14 months compared to 36 to 42 months for mental retardation. Thus, the pitfalls in infancy are the tendency to miss or downplay motor developmental delay until after about 8 or 10 months of age and to disregard language developmental delay until 24 months or later.

PITFALLS IN ASSESSING MOTOR DEVELOPMENT

Gross motor development is an easy field of development to assess and normal milestones are well-established. Yet this is also a comparatively less important field in developmental assessment. The pitfall is that the child who demonstrates normal gross motor development can lull one into a false sense of security. Although we can be reassured about motor function, gross motor milestones are not indicative and the least predictive of intellectual competence. There is known discrepancy between motor and mental development in the mentally retarded. About one-third to one-half of the severely to profoundly retarded children walk at or before 15 months, the upper limit of normal. On the other hand, children with motor deficits because of cerebral palsy are not necessarily mentally deficient. They may never achieve ambulation or the ability to speak intelligibly, yet they may have average or above average intellects.

The tendency to think only of gross motor skills and omit attention to fine motor milestones is another pitfall in assessing motor development. Manipulative development is far more important than gross motor development, and fine motor delays can be a better and earlier indicator of motor disability. The earliest fine motor milestone is “un-fisting” of the hands (more than 50% of the time) by 3 months of age. The child with moderate or severe cerebral palsy may get his head up well in prone position by

4 months of age because of extensor hypertonus, and roll over occasionally because of abnormal primitive reflexes. Technically this child meets gross motor milestones but his hands are still fisted 100% of the time.

Attention to fine motor skills also forces one to assess hand use that involves problem-solving tasks, bringing one's attention to some aspects of early cognitive assessment. Important developmental features include ambidextrous (average 20 weeks) or unidextrous (average 28 weeks) approach to an object, the rapidity with which he accidentally drops a cube, the question of whether he drops one cube when offered another, tremor or ataxia or other abnormal hand movement when reaching for an object, the transfer of a cube from one hand to another (about 6 months), and the quality of the grasp (evolving from the crude palmar grasp of the cube at about 5 months to the mature pincer grasp between the tip of the thumb and the tip of the forefinger at about 10 to 11 months. In assessing fine motor skills, it is not enough to record just whether a child has acquired a certain skill; one has to note the maturity and the rapidity with which the child performs the task.

PHYSICAL EXAMINATION AND PITFALLS OF "SPOT" DIAGNOSIS

For developmental diagnosis, a full physical examination is essential if serious errors are to be avoided. This will include examination for neurological and physical handicap, congenital anomalies, and defects of vision and hearing. All these conditions may profoundly affect development and developmental tests, and any significant abnormality carries some increased risk of associated mental retardation.

An essential part of the physical examination is the head circumference in relation to the child's weight. Serial measurements are essential to show a falling off from the centile distribution, which is an important indication of a brain defect. Other causes of unusual head measurements, such as familial traits, must also be known. Failure to diagnose subluxation or dislocation of the hip as a cause of limited abduction of the hip might well lead to a diagnosis of cerebral palsy, the limited abduction having been wrongly ascribed to adductor spasm. The presence of obesity may be a factor in delaying gross motor function. Irrelevant physical features include the age of closure of the anterior fontanelle, age of appearance of teeth, an epicanthus, or simian crease and its variations.

During the physical examination, general features are noted and more attention should be paid to the child's positive achievement and those skills he does best, rather than the negative ones. The most important features for the assessment are features that cannot be readily scored: the quality of vocalization, the child's interest in surroundings, alertness, responsiveness, understanding, his concentration and memory, the glint in the eyes, and signs suggestive of mental subnormality such as bruxism when awake, persistence of hand regard after 20 weeks, or mouthing, slobbering, and casting at an age when the child should have grown out of it. It should also be noted that a child might be able to suppress his behavior in a structured environment. This should be appreciated in the assessment of a child with attention deficit hyperactivity disorder.

Developmental diagnosis must never be made on clinical impression. Neither should a "spot" diagnosis be made in a casual manner. It may be easy at a glance from a distance to diagnose Down syndrome and some other syndromic disorders associated with mental subnormality, it is difficult for parents to accept such instant diagnosis from a doctor, followed by his confident statement on the outlook.

The concept that retarded individuals look retarded is prevalent among the general public and even among the medical professionals. Many children with odd facies often may be just taking after one of his parents. Nevertheless, there remains a bias against the intellectual competence of the unusual-appearing child. The converse situation is also a pitfall. Mental superiority may be wrongly diagnosed in a two- or three-year-old child because of charm of manner, absence of shyness and good looks. The retarded child with good looks or good motor skills is typically identified late. Children with autism are most often described as attractive children. These superficial physical attributes, especially if present in a happy or uncomplaining infant, obscure other clear indicators of developmental deficits. Attractive children may manifest any degree of cognitive deficit from learning disability through profound retardation. On the other hand, the child deemed to be facially dysmorphic is not necessarily cognitively deficient.

PITFALLS IN ASSESSING LANGUAGE DEVELOPMENT

There is often a failure to appreciate the importance of early language development to the early recognition of mental subnormality. Infants and

young children by their nature do not express themselves in the examination room. They are usually shy, subdued or failed to cooperate in certain tests. Some children may be genuinely apprehensive in unfamiliar environment, having crossed the stage of stranger anxiety. Therefore, obtaining a precise history of the infant's language milestones becomes very important. Unfortunately, neither the milestones themselves nor the phrasing of the questions to elicit an accurate history are intuitively obvious. In addition, there is a tendency among parents and professionals to excuse language delay until the child reaches about two years of age. Even children who are congenitally deaf with no speech at all are characteristically identified around this age.

The pitfall to be aware of is the concept that language development does not begin until one year of age and is not worthy of concern until two years of age. Avoiding the pitfall hinges on the ability to obtain an accurate history and the knowledge of early language development, both require further professional training and education. Having a table of linguistic and auditory milestones that is easily accessible and that can be routinely consulted upon during well child visits is of great help (Table 2).

PROBLEMS IN EARLY IDENTIFICATION OF AUTISTIC SPECTRUM DISORDER

There is strong evidence that children with autistic spectrum disorder can be reliably identified between 18 to 39 months of age. Systematic review of early videotapes demonstrates prelinguistic communication abnormalities in infancy. Parent questionnaires and other structured screening instruments often retrospectively reveal impairments in pretend play, proto-declarative pointing, joint attention, social interest, and social play as early as 18 months.

A wait-and-see or reassuring attitude towards parents' concern over the child's delayed language acquisition, especially in boys, is the most important factor for late diagnosis. There are problems in identifying developmental delays in socialization and other non-verbal communication behaviors because most health professionals are inadequately or poorly trained in screening these areas of child development. The Denver Developmental Screening Test is not sensitive and specific for autistic spectrum disorder. The nature of impairments is usually subtle. They often represent the absence of normative behaviors (such as not imitating, not

Table 2 Clinical Linguistic and Auditory Milestones

Alert	1 week	Dada/mama (appropriate)	10–11 months
Social smile	1–1.5 months	One word	11 months
Cooing	2 months	One-step command (gesture)	11–12 months
Orient to voice	4 months	Two words	12 months
Orient to bell (I)	5 months	Three words	14 months
"Ah-goo"	4 months	Immature jargoning	12–14 months
Laugh	4–5 months	One-step command (no gesture)	15 months
Razzing	5 months	4–6 words	15 months
Babbling	6–7 months	One body part	15 months
Orient to bell (II)	7 months	Mature jargoning	17–18 months
"Mama/dada"	8 months	5 body parts	17–18 months
Gesture/games	9–10 months	2-word combination	19–21 months
Orient to bell (III)	9–10 months	2-word sentence	21–24 months

Orienting to sound of bell (I): At 5 months, when a bell is rung at one side of the infant's head, the infant turns horizontally to the correct side.
Orienting to bell (II): At 7 months, when a bell is rung at one side of the head, the infant localizes the sound by a visual maneuver consisting of a horizontal followed by a vertical component.
Orienting to bell (III): At 9–10 months, when a bell is rung at one side of the head, the infant localizes the sound by one single visual movement.

gesturing) rather than the presence of noticeably unusual behaviors (such as peculiar use of language).

Differentiation between autistic spectrum disorder and mental retardation may be difficult. Given a child's developmental level, there may not be an opportunity to manifest the behaviors needed for diagnosis. Differentiating milder autistic spectrum disorder from developmental disability may be difficult if the child has not been involved in some preschool program to evaluate their response to structured environments and exposure to other children.

PITFALLS IN DIAGNOSING HEARING IMPAIRMENT

Hearing impairment in children may have substantial long-term consequences if untreated or treated late. Deafness in infancy is not only diagnosed late but routinely goes undiagnosed for more than a year from the

appearance of symptoms. The average age of diagnosis of congenital deafness is 2 to 2.5 years. Lesser degrees of hearing impairment are delayed proportionately longer in their recognition. Therefore, until hearing screening in neonatal and early infancy period becomes routine and universal using sensitive technologies such as oto-acoustic emissions and brainstem auditory evoked response, effort must be made to improve the acumen in diagnosing hearing impairment.

Most congenital deafness is genetic and most of the time it is an isolated deficit. There is every possibility that if a hearing impaired infant is evaluated on the four areas of development, he can virtually pass all developmental milestones. The typical developmental profile of a congenitally deaf child is that the early motor development is normal. Problem-solving skills are also within normal because they have normal intellectual and fine motor capacity. Likewise, early psychosocial milestones such as recognizing parents, stranger anxiety, eating with a spoon, can all be normal. Even expressive language is said to be normal in the first six to eight months of life. They can coo and laugh and even begin to babble before expressive language milestones become noticeably delayed in the form of deviant babbling. Receptive language depends on the infant's failure to respond to noise and early diagnosis of deafness hinges on demonstrating this simple deficit. However, this is exactly where the inexperienced can be most readily fooled. By twelve months of age, babies respond mainly to gesture commands. When a 12-month-old infant is given two cubes and then asked to give one back to the tester who simultaneously reaches out to receive the object, the child will comply. A deaf child responds just as readily as a hearing infant. An infant will respond to simple commands without the reinforcement of an accompanying gesture by only 14 or 15 months of age. It may also be true that infants with deafness are more vigilant than normal-hearing infants, making such a mistake all the more likely. Although the parents of hearing impaired children report suspicions about hearing *per se* less than half the time, they do consistently report the symptoms of hearing impairment, such as inattention, delayed speech, disinterest in musical toys, but their concerns most often may go unheeded by medical professionals.

Another common pitfall with respect to hearing impairment is the inclination to ascribe delayed language development to recurrent episodes of otitis media. It is not unusual to see referrals of children with delayed or absent speech after one year or more of temporizing because the delay

has been dismissed on the basis of middle ear disease. It is true that some language dysfunction can result from the mild and transient hearing impairment associated with acute or chronic middle ear disease. There is considerable controversy over the extent of and the long-term consequences of middle ear disease on language and academic performance.

PRECISION IN TAKING DEVELOPMENTAL HISTORY

Precision in history taking is necessary if errors are to be avoided. It is essential to cover, as far as possible, all fields of development: locomotion, manipulation, play and social behavior, memory, the mode of display of pleasure and displeasure, feeding behavior, sphincter control and speech.

Parents who express concern must not be dismissed. The reliability of parent's story must also be checked against the doctor's own objective findings. Parents of a retarded child are often unwilling to allow themselves to believe what they know is the truth. At the same time, parents may have forgotten or may not even know when their child acquired various skills. They may then fabricate their replies on their child's skills, basing their answers on what they know these skills are usually acquired instead of on what they can remember. In history taking, it is not enough merely to know whether a child does certain thing, but when he began to do it, and how often, with what degree of maturity, and the rate of achievement in general.

The accuracy of the history can be improved by asking about the child's development in relation to that of the siblings or with neighbors' children of about the same age. A simple question about the progress of these children in school will give an idea whether they are likely to be reasonably normal. Asking parents to remember whether specific milestones had been reached by the child's first birthday, checking on family photographs or videotapes, and asking parents to check milestones recorded in the child's baby book are other ways of improving recall. Another useful source is the infant's health booklet, which will contain past measurements of weight, length and head circumference so that percentiles can be plotted to give a picture of the child's growth pattern.

Imprecise histories are of limited value. The information obtained will only be as good as the questions asked. The following are some examples:

- *Beginning to smile* refers to smiling in response to the mother's overtures and not to grimace or twitch when she tickles the face or the baby's facial movement in sleep.
- *Rolling over* refers to rolling completely over, rolling from prone to supine has to be distinguished from rolling from supine to prone, which comes later.
- *Grasping objects* means going for an object without it being put into the hand and has to be distinguished from grasp reflex, or the ability to hold an object placed in the hand.
- *Sitting* means sitting on the floor or another hard surface for seconds, either with the hands forward with support or without, not in a stroller with pillow support.
- *Holding the head up*: the average child can lift his head off the couch when lying supine by about 7 months old. Beware of spasticity.
- *Creeping* (on hands and knees) has to be distinguished from the earlier crawling on the abdomen.
- *Chewing* has to be distinguished from sucking a biscuit.
- *Self-feeding* is not equivalent to an infant of six months holding a biscuit and feeding himself with it. It is not until 15 to 18 months that an average child can feed himself with a spoon and manage a cup without help. The child has to go through numerous stages of trying to load the spoon, of succeeding in loading the spoon but not getting it near his mouth, and later of getting it near the mouth but rotating it and therefore spilling the contents before it enters the mouth. With the cup, the child has to go through the stages of helping to hold it, of suddenly letting go when he is drinking or has had what he wants, of spilling most of the contents, until finally he can pick it up, drink and replace it with only occasional accidents.
- *Walking* means walking a few steps without support.
- *Toilet-trained* means that the child is mainly dry and clean day and night, and has to be distinguished from earlier conditioning when the child voids as the buttocks feel the rim of the potty, whether awake or asleep.
- *Talking* means saying words with meaning, not just "mummum, dadada".

When asking about language development, remember to distinguish between expressive and receptive language. Expressive language is what

the child can say. Ask whether the child's words have consistent meaning. The clarity of pronunciation in the first year following speech development is less important than the extent of the vocabulary and the consistent use of the same sounds for the same object or person. Receptive language is how many words the child can understand, and is usually well in advance of expressive language. A child of 18 months who only says four or five distinct words with meaning may have a receptive language of 200 words or more.

CORRECTION FOR PREMATURITY

The issue of correction for prematurity during developmental assessment is still controversial. By convention, infants born before term are allowed extra credit for the duration of their prematurity when evaluating their developmental progress. Therefore, if birth was at 32 weeks' gestation, eight weeks should be subtracted from the chronological age when assessing development, so that milestones would be expected to be reached two months later than a baby born at full term. Similarly, when an average full term infant begins to smile in response to mother's overtures at 4 to 6 weeks, the infant born three months early would be expected to smile at 4 to 6 weeks plus three months. If the infant is of low birth weight, the duration of gestation must be ascertained to distinguish intrauterine growth restriction.

Current data, however, suggest that full correction for motor milestones should be allowed, but partial or no correction seems to be appropriate for language and problem-solving milestones. The tendency to over-correct because the infant has been sick as a neonate represents a pitfall in the interpretation of assessments. That a child is or has been sick ill represents a factor that adds uncertainty to the evaluation, but it should be dealt with separately and not injected into the sense of correction.

IMPORTANCE OF ENVIRONMENTAL FACTORS

Many adverse environmental factors may profoundly affect development but they may not be directly related to the child's mental endowment. These factors should be actively explored and not be ignored in developmental assessment.

Children who are emotionally deprived, who receive little verbal stimulation, or who are lacking in experiential learning opportunities

such as the chance to practise walking or crawling, to learn to feed and dress himself, or to attend to toilet needs, when they are ready to achieve these milestones will be delayed in these skills. One must also know about the cultural oddities in parental management, such as keeping the baby off his feet for fear of bowlegs and knock-knees. It is also important to know if the mother is out working all day and who the main caregivers of the child are. Maternal depression has been recognized as a major factor adversely affecting the child's growth and development. Besides socio-economic factors, we must also consider poor nutrition, chronic medical conditions requiring prolonged hospitalization, exposure to toxic substances, and accidents.

RELEVANCE OF RISK FACTORS

To avoid mistakes by the omission of relevant and important factors that affect development, a detailed history is essential and it will include all prenatal, perinatal, and postnatal factors. However, the importance of these risk factors must not be exaggerated. When there is doubt about the level of development, the history of a risk factor, such as mental subnormality in a parent, serves only to increase that doubt. A risk factor itself should never lead to a hasty diagnosis that a child is mentally subnormal. A mentally subnormal parent can also have a normal child. Similarly, we must be aware of the pitfall of relating backwardness in certain areas to family pattern of development.

The concept of "at risk" registers is to ensure the close monitoring and assessment of those most likely to be developmentally delayed. The limitations of such registers must be acknowledged. Many developmentally delayed children do not have any identifiable etiological or risk factors, and many children who have suffered clear identifiable insults, and are very much at risk, develop quite normally. Furthermore, the inclusion of all possible risk factors necessitates large numbers of children being placed on the at-risk registers, making the concept logistically difficult and practically impossible most of the time. Nevertheless, there is still an emphasis on the early diagnosis of delay, and certain groups of children such as the very low birth weight infants are followed very closely and their development monitored. The clear advantages of early diagnosis must be counterbalanced by the dangers of inappropriate labeling.

AVOIDING PITFALLS IN DEVELOPMENTAL EXAMINATION

The child is not a passive part of the developmental examination. He should be given the opportunity to perform optimally. He should not be tested when he is ill, sleepy or just woken, miserable for some reasons, hungry, bored, or on medications. For example, it is not appropriate to conduct a developmental examination on a child with epilepsy when he is still in a state of confusion after a major seizure or when he is under the influence of sedative drugs. Head control cannot be properly tested when the child is sleepy or crying. One may also be misled into thinking that there is significant head lag when one pulls a child into the sitting position from lying on his back. A child may refuse to sit without support during the examination if he has a wet or soiled diaper, while he is able to do so normally. Similarly, it would be foolish to try to do developmental assessment on a 22-month-old admitted to hospital with asthma, who is in respiratory distress as well as being "high" from salbutamol medication.

When the child does badly in a test, a decision has to be made about whether he has really been trying. A child may refuse to take part in a test because he is shy. The older child may also refuse or reluctant to cooperate because he regards the tasks as too easy or just silly.

The presence of the parents may occasionally help the child to achieve his best in strange surroundings, but the parents should be told to resist the urge to help the child or to interfere with the tests.

Standardization of test materials is important. The materials used must be the same as those on which the norms are established. Similarly, the tests should be adapted and validated for the local population. Rigid adherence to a method of testing is undesirable. The tester should be able to adjust the order of the tests as soon as he sees that the child is not interested or is becoming bored. The child needs encouragement in his tasks. This is best done by relating to the child in a friendly, enthusiastic way, praising child for good performance, giving new challenges and supporting the child with encouragement during the assessment. One should never try to correct a child or say "no" when he performs a test incorrectly. Therefore, an important part of the training in assessment is to develop skills in playing with and relating to children.

COMMON MISTAKES IN DEVELOPMENTAL DIAGNOSIS

There are important but common pitfalls to avoid in the interpretation of signs that may sometimes indicate cerebral palsy, mental subnormality, and visual or hearing defects.

Cerebral palsy cannot be diagnosed on the basis of isolated retardation in motor development nor on the basis of exaggerated tendon jerks in young baby or even persistent ankle clonus, for these signs may disappear as the baby gets older. The plantar responses in the first year or so are flexor unless there is pyramidal tracts disease.

Excessive extensor tone is a highly important pointer to cerebral palsy. The infant may have good head control in ventral suspension and in the prone position, but gross head lag is obvious when pulled up from the supine to the sitting position. Resistance is also felt when pulling the child up to the sitting position and he may also tend to rise onto his legs.

Adductor spasm in cerebral palsy may be wrongly diagnosed because of limited hip abduction due to hip dislocation or subluxation, or to muscle contracture due to the hypotonic child consistently lying in one position. Spastic cerebral palsy may be wrongly diagnosed in punctate epiphyseal dysplasia on account of limited joint extension. Duchenne muscular dystrophy may be missed in a child with delayed motor development.

Minimal involvement of the upper limbs, as seen when the child is going for cubes or building a tower, can be easily missed, and spastic paraplegia is diagnosed when the correct diagnosis is spastic diplegia. Spastic paraplegia points to the possibility of a spinal lesion rather than a cerebral cause.

Toe walking is often attributed to cerebral palsy. However, it can also be a normal variant, or can be found in congenital shortening of the Archilles tendon, unilateral dislocation of the hip, Duchenne muscular dystrophy, and has been described in children with autism. Spastic children are often wrongly thought to have the mixed form of cerebral palsy because the characteristic awkwardness and splaying out of the hands is thought to be athetosis. A late walker may be unsteady on his feet shortly after first walking without support and may be misdiagnosed as having the ataxic form of cerebral palsy.

Infants may be thought to be blind because of delayed eye following, when the delay is due only to mental subnormality. Similarly, a child with

mental retardation may be thought to be deaf because of delayed response to sound. Isolated delay in speech is rarely due to mental subnormality and tongue tie.

LIMITATIONS IN PREDICTING CEREBRAL PALSY IN PREMATURITY

In premature infants, long-term neuromuscular abnormalities usually cannot be predicted with certainty in the first year of life. There is known variability in the progression of muscular status. Some preterm infants demonstrate muscle tone abnormalities during infancy that resolve subsequently. In contrast, abnormalities can develop in an infant with initially normal tone.

There are also imperfect correlations between brain damage documented on specialized studies, such as cranial ultrasound, CT scan, or even MRI, and later functional outcome. The absence of lesion does not preclude development of disability later, and presence of abnormality does not also preclude normal development. The fact that the neonatal brain has significant plasticity for recovery may account for the difficulties in prediction.

"IQ" TESTS

An important source of error is the over-reliance on purely objective tests to provide unitary score in developmental assessment and misinterpretation of its accuracy and predictive value. Serious mistakes are made in developmental diagnosis if the various factors that affect the course of development are not properly considered, and if the normal variations are not borne in mind. It is rarely desirable to assess the "intelligence quotient" ("IQ") in terms of a single figure in a preschool child, because such a figure cannot take into account the many influential factors that are of importance. Some of the tests used on the two- to three-year-old depend on the acquisition of speech; but a child may be of normal or superior intelligence and yet be late in learning to speak. This will lead to the child being given an unduly low score simply because he cannot speak. Observation might show that the child's understanding of words and his ability to identify objects and to carry out simple acts on request is advanced. A child might also be given a low score in other fields of

development, such as sphincter control, when his lateness in acquiring sphincter control is due to parental mismanagement and bears no relationship to his intelligence. Serious fallacies would result if one were to attempt to calculate the IQ merely by converting each observation in the developmental examination into a single figure, adding all the figures up, and take the average.

There are confusion concerning the terms "developmental quotient" or "DQ" and "intelligence quotient" or "IQ". The DQ indicates how far a child has progressed in all aspects of development, especially behavior, in relation to the average for his age. The IQ relates the child's age to his performance, mainly in verbal and problem-solving tests, on the basis of pass or failure. The DQ is affected by the environment and many other factors unrelated to the child's genetic endowment.

"Intelligence" has not been satisfactorily defined. It cannot be simply described as a single score or figure, and there are many different types of intelligence and of human abilities. It is more sensible to describe areas of strengths and weaknesses in a child according to his age after the assessment. Furthermore, most of the commonly used "IQ" tests are established on a narrowly selected population, instead of on the population as a whole. It follows that there is no valid norms with which to compare the development of children not satisfying these criteria and accuracy is impossible.

PREDICTING OUTCOME

Some conditions causing developmental delay, such as Down syndrome or severe asphyxia, are evident at or shortly after birth. On the other hand, mild mental retardation may not be apparent until the child starts school. We can never diagnose mental retardation on backwardness in just one or two fields of development. High intelligence also cannot be predicted on advanced developmental milestones. Therefore, great caution needs to be exercised in giving a prognosis about a child's eventual outcome. It has been well-established that, except for severe developmental delay, prediction of developmental outcome from test scores in infancy is often inaccurate and fraught with hazard. The younger the child, and the longer the prediction period, the more inaccurate it is likely to be, especially for cognitive development.

It is important to consider the child at a particular point in time, and avoid making long-term predictions about a child's ultimate potential

and achievement, because there is a wide variation in the attainment of milestones, and it is impossible to accurately assess or predict all the constitutional and environmental variables that might affect outcome.

CONCLUSION

Development is a complex, continuous process whereby a child acquires an increasingly sophisticated repertoire of skills across multiple domains. The remarkable consistency of this process allows us to monitor an individual child's progress and to detect any deviation from normality. Developmental delay and deviation is not homogenous, but varies widely according to the etiology. Informal developmental screening and surveillance can be useful in reassuring normality, but care must be taken in drawing any firm conclusions about developmental status without a more formal procedure.

Developmental assessment can be difficult and follow-up of unusual developmental features is important. What is appropriate where milestones are either delayed or advanced is to say where the child's development is in relation to average children of that age and then help the parents with techniques of responding to the child's strengths and weaknesses most appropriately. There is also the role of anticipatory guidance to the parents. Knowing the child's level of development and what the developmental progress over the next few months is likely to be is valuable for informing parents what sort of behavior to expect. This will enable guidance about child management and accident prevention to be given.

Furthermore, even when it may be possible to say something about a child's developmental potential and talents, it is not possible to say what he will do with them. That will depend actually on many factors in the future — the quality of his home, friends and school, his personality, health, and nutrition, and the opportunities that he will have.

FURTHER READING

1. Egan DF, *Developmental Examination of Infants and Preschool Children*, MacKeith Press, Oxford, 1999.

2. Levene MD, Carey WB, Crocker AC, *Developmental-Behavioral Pediatrics*, WB Saunders Company, Sydney, 1999.

3. Illingworth RS, *The Normal Child: Some Problems of the Early Years and Their Treatment*, Churchill Livingstone, London, 1987.

4. Illingworth RS, *The Development of the Infant and Young Child: Normal and Abnormal*, Churchill Livingstone, London, 1987.

5. Pollack M, *Textbook of Developmental Paediatrics*, Churchill Livingstone, Melbourne, 1993.

6. Capute AJ, Accardi PJ, Linguistic and auditory milestones during the first two years of life, *Clin Pediatr* **17**:847–856, 1978.

7. Blasco PA, Preterm birth: to correct or not to correct, *Dev Med Child Neurol* **31**:816–822, 1989.

8. Palfrey JS, Singer JD, Walker DK, Butler JA, Early identification of children's special needs: a study in five metropolitan communities, *J Pediatr* **111**:651–659, 1987.

9. First LR, Palfrey JS, The infant or young child with developmental delay, *N Engl J Med* **330**:478–483, 1994.

This page intentionally left blank

68

Feeding Problems in Young Children: A Developmental Perspective

Ho Lai Yun

INTRODUCTION

Feeding and eating serve a range of biological, psychological and social functions in the life of the developing child. The most basic function is clearly biological, as the child requires adequate and appropriate nutrition to survive and to thrive physically and mentally. However, nurturing is not only nourishing, and feeding is but also one of many ways parents promote children's growth and development. Infant feeding is important in developing and maintaining the emotional relationship between the mother and infant since much of the early interaction between the baby and the mother centers on this activity. It has been shown that breast-feeding mothers have a closer and more expressive relationship with their infants during feeding than do mothers who bottle-feed. Feeding and eating also play an important role in the social life of the child, and are of fundamental significance to the child's experience of the world and his place in it. Meals are usually times when the family is together and provide a focus for family life. Furthermore, a great deal of social interaction with non-family members occurs during mealtimes and this context therefore

serves a wider socializing function. Given the central role feeding and eating play in the life of the developing child, it is not surprising that deviations or disturbances in them are not only a problem for the individual child but can cause great anxiety within the family.

Feeding problems cover a wide spectrum of phenomena of variable significance, ranging from variations of normal behavior such as mild faddiness and pickiness, to conditions of major developmental significance such as failure to thrive. Many of these disorders result from a combination of both the child's own inherent difficulties and parental handling problems. In other cases, feeding difficulties may be the expression of general difficulties within the family. Thus, when a child presents to the clinician with a feeding or eating problem, adequate management depends on viewing this problem in its wider familial and social context. This applies not only to consideration of etiological factors, but also to factors that may be the sequelae of the feeding difficulties, which serve to maintain them (Fig. 1).

Parents enter the feeding relationship with their children with a set of experiences and expectations, and they make decisions about nutrition and feeding practices for their families. Infants and children have their own set of nutritional and caloric needs, as well as certain emotional and developmental requirements. These individual characteristics of the infant and caregiver may be in direct conflict with one another; and if

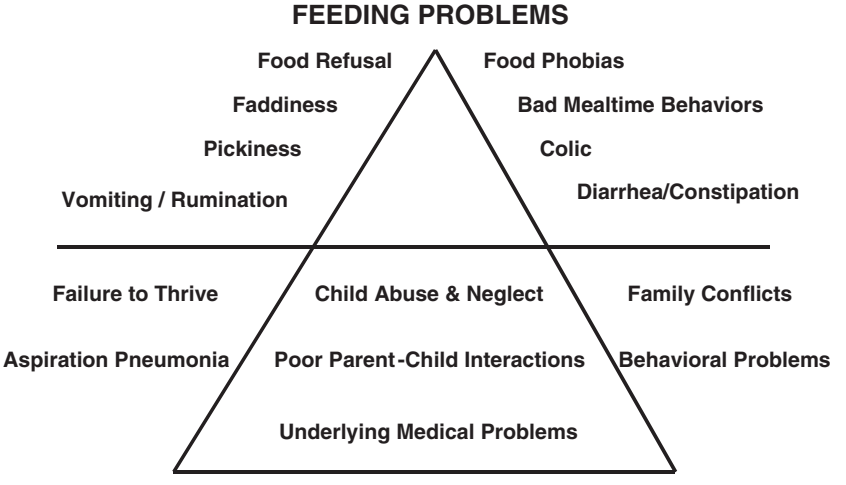


Fig. 1 Feeding problems.

problems develop during the course of this feeding interaction, there may be implications for the style and quality of their relationship in other non-feeding contexts, as well as for subsequent feeding behaviors. Mealtimes may therefore be a pleasure or battleground (Fig. 2).

PARENTAL EXPERIENCES AND EXPECTATIONS

Several complex factors interact to influence the attitude of parents towards feeding their children. They include the parents' upbringing and childhood experiences, social and cultural influences, traditional beliefs, peer and family pressure or support, educational and financial status, and their physical and mental health. Most of the common problems are related to: lack of parental understanding, parental over-concern or overprotection; parental tension; parental attitudes towards food that cause anxiety for the child; parental projection of their wishes to over-eat or under-eat onto their child; and parental ambivalence towards their children.

Many parents are inexperienced in parenting small babies. They may not have an extended family to support them in such apparently unimportant crises. They may also be confused by conflicting and misleading advice from several sources: friends, media, childcare magazines, or even professionals. Some parents consider feeding to be the primary means of nurturing their children. When they place too great an emphasis on

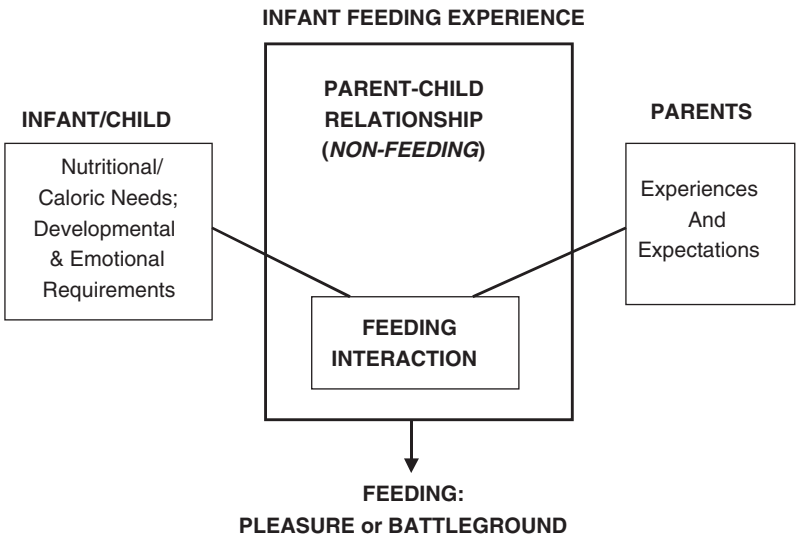


Fig. 2 Infant feeding experience.

feeding, they become overly concerned about the quantity, quality, and variety of foods consumed; the regularity of the child's feeding schedule; and the child's behavior during feeds. They may consciously or unconsciously equate both their ability to feed their child with their own success as parents, and their child's willingness to eat with his or her love for them.

Some parents are not responsive to their infant's feeding signals. For example, they may misinterpret infant cues and misread crying after eating as hunger or lack of satisfaction with a meal. The parent's ability to accept the infant's individual feeding pattern, to use consoling methods other than feeding for non-hunger cries, to wait for the infant to resume sucking after pauses without interfering, and to avoid pushing more on the baby when he or she is satisfied are signs of sensitivity to infant cues. Parents may be less sensitive because they are anxious, depressed, or suffering from other mental problems or psychosocial stresses. Also, they may have a sense that their baby is vulnerable, be under pressure from a spouse or relatives, or even an intense desire to succeed. The frustration and concern that often develop can seriously complicate feedings and result in negative feelings towards the child. From the simple beginning of a child's throwing food, gagging, or refusing to eat, a complex problem between parent and child can develop during feeds. Furthermore, these simple beginnings can result in reduced nutritional intake, poor parent-child interactions, and feeding problems that can be exacerbated as the child grows.

Generally, the most important part of the management of feeding problems is eliciting the parent's expectations of the child and of themselves as caregivers, identifying the psychosocial stresses they are experiencing, assuring them in their roles as parents, and ensuring adequate supports for them. If parents were provided with anticipatory guidance on early feeding scenarios, feeding times would become less challenging to their parenting skills. If parents were to relax and enjoy the mealtime interaction and the child's exploration of foods, acquisition of new behaviors, the common and usually trivial problems that occur during early feeding experiences might disappear so that a reasonable diet and mealtime pattern could evolve naturally for most children.

However, recommendations can only be followed through by parents who do not have significant ambivalence towards their child, that is, both strong feelings of love and anger. The concept of parental ambivalence may be rather contentious but it seems to exist. The more emotionally healthy parent uses repression to keep the angry feelings out of conscious

awareness. This unconscious anger will often influence and interfere with the parent's motivation to follow through recommendations to change his or her behavior concerning food and feeding. Many requests for child evaluation for food refusal problems are dropped by parents when they are suddenly confronted with their angry feelings towards their children. These troubled parents need help to become aware of their unconscious anger, and to accept and cope with their feelings of guilt and shame when this anger is exposed.

MATERNAL-INFANT RELATIONSHIP PROBLEMS

The following types of maternal-infant relationship problems may both cause and present as infant feeding disorders:

1) *Impoverished relationship*: This is characterized by low verbal contact, little play, little eye contact, and lack of pleasure, warmth and affection.
2) *Inconsistent relationship*: The mother's responses are irregular and inconsistent both physically and emotionally. Due to the lack of consistency, the infant is uncertain of what to expect.
3) *Indiscriminate relationship*: The parent gives high intensity reactions but with poor direction. The infant is over-stimulated and his/her real needs are neglected. This occasionally occurs in "colicky" infants.
4) *Interactive mismatch*: Typical cases include a burdensome infant with an unresilient mother, an impossible infant with a competent mother, and a normal infant whose parents hold unrealistic expectations of what constitutes normal infant behavior.
5) *Attachment disorder*: In those mother-infant pairs where the relationship will be tenuous, separation or illness at birth may mean that a proper attachment between the mother and infant does not occur. Premature infants are about three times as likely to be abused as are term infants. This can also happen where the infant is unwanted, where there are overwhelming pressures on the mother, or where the mother is physically or psychiatrically ill.

NUTRITIONAL AND CALORIC NEEDS OF THE CHILD

The most frequently voiced concern about feeding is that children do not eat enough or that they eat too limited a variety of foods. Each child has

his own set of nutritional and caloric needs. A constitutionally or genetically small child will need less food than the parent expects the child to consume for optimal growth. These expectations may be based on the usual daily intake of other siblings, the usual daily intake of the parents themselves, on the expectations of the health workers, or even booklets circulated to new mothers by commercial baby food manufacturers. These booklets often give examples of daily diets that are based on what most infants will eat at any specific age. There are individual variations in caloric requirements of a child. Children who have identical growth and health can have vastly different caloric intakes varying by a factor of two. For example, while growing at the same velocity and with the same quality of health, one child may require twice the calories of another.

The intrinsic growth rate of a child will determine his demand for caloric intake. Many children appear to have a "poor appetite" towards the end of the first year of life that often continues until school age. This can become apparent quite suddenly and often coincides with the onset of the negativism in the infant. Parents will report that their child does not seem to eat enough to sustain his activity. At this time, the child requires less food per unit of body surface or weight because the rate of physical growth has decreased, with a concomitant decrease in caloric needs and appetite, and slimming down of his profile. If the child is happy, active and growing normally, one can be totally confident about reassuring the parents. Unaware of this change in the child's nutritional requirements, many parents become upset when the child's appetite tails off. They may resort to harsh threats or ineffective cajoling in an attempt to encourage intake, which tends to increase the child's resistance to authority and to emphasize to the child the parental interest in his food intake. The child soon learns to use feeding as an effective tool to gain attention, to retaliate, or to spite his parents.

DEVELOPMENT OF FEEDING PROBLEMS

Three major inter-related developmental areas in young children have significant impact on eating behavior and feeding problems. They are: children's temperamental styles, psychosocial and emotional development of the child, and the development of feeding skills.

Children's Temperament

Historically, it was thought that parents shaped their children's personalities; that children were born as "lumps of clay" for parents to mold through daily interactions over the course of childhood. However, this theory does not explain why some children who are the product of fantastic families have problems, and others, who were born into troubled families, do just fine.

Innate temperamental characteristics probably play an important role in personality development. Children are born with a unique combination of temperament characteristics that define an individual's particular behavioral style. These characteristics are derived from constitutional, intrauterine, central nervous system, and postnatal environmental factors. Although they are not fixed, an individual's temperament characteristics are likely to be consistent over time, especially when the interplay between the individual and the environment becomes relatively stable.

There are nine measurable characteristics of infant temperament: activity level, rhythmicity, approach or withdrawal, adaptability, threshold of responsiveness, intensity of reaction, quality of mood, distractibility, and attention span or persistence. The nine temperamental characteristics can be configured into a number of combinations. The three most common constellations are seen in children with easy, slow-to-warm-up, and difficult temperament styles.

The "easy" temperament style is characterized by children with regular biological functioning, a positive approach to new stimuli, a high adaptability to change, mild to moderately intense responses, and a predominantly positive mood. These children quickly develop regular sleep and eating schedules, adapt easily to new foods, schools, or strangers, accept most frustration with little distress, engage easily with other children, and adjust quickly to new game rules. They are indeed a "joy" to parents.

The "slow-to-warm-up" children show mildly negative responses to new stimuli, slow adaptability after repeated contact, and they tend to need a good bit of time before their various responses shift from negative to positive, before they become comfortable in new situations. Although their first encounters with food, people, schools, etc. produce distress reactions, these experiences become better tolerated upon repeated exposure.

Children with a "difficult" temperament style have irregular biological functions and negative withdrawal responses and mood expressions. They have irregular sleep and feeding patterns, slow acceptance of new foods, and prolonged adjustment periods to new routines, situations, and people. They react intensely to everything in the environment and are often described as "colicky". They tend to have both loud and intense crying and laughing. Any kind of minimum frustration leads to tantrums and many behavioral difficulties. They will stay this way for anywhere from six weeks to six months; some difficult traits will remain for years.

These temperament characteristics do not, in and of themselves, cause problems. Rather, it is the "fit" between the child's temperament and the demands and expectations of the parents and other caregivers that can cause a struggle. The question is whether the parents and child complement or antagonize one another, that is the "goodness of fit" between parent and child. If there is a good fit, optimal development is likely. On the other hand, if parental expectations are not consistent with the child's temperament, there is a poor fit; the ensuing dissonance results in stress and potential problem behavior. When parents understand their children's unique strengths and weaknesses in their temperamental styles, they can develop strategies to avoid or resolve conflict.

Given the focused nature of feeding, mealtimes are often times to see children's temperamental characteristics expressed and to evaluate the "goodness of fit" between parent and the child. Feeding struggles can be resolved once incongruities in temperament "fit" between the feeder and the child are mediated. For example, parents of infants with irregular biological functions can be counseled that their babies cannot be expected to conform to strict feeding or sleeping schedules. Similarly, slow-to-warm-up babies may need consistent, gradual and repeated exposure to new foods before they will accept them. These babies also do not elicit parental interaction and are at risk of failure to thrive. When they are hungry, these infants whine a little bit, instead of crying. If their parents are preoccupied or are depressed, this minimal response may not be enough to get their attention.

An understanding of the temperamental characteristics allows the childcare professional to tailor anticipatory guidance based on an individual infant's unique temperamental style. The child's inherent temperament may not change, but the parent's expectations and behavioral

response to the child can be changed to create a better fit and to avoid or lessen the impact of challenges before they arise.

Psychosocial and Emotional Development of the Child

Understanding some of the salient features in the psychosocial development of a child will help parents in adopting a proper approach to feeding their children. Feeding provides many opportunities for parents and their child to learn about each other and to test each other's limits. With each spurt of development, tensions are likely to increase between them, and will influence feeding climate.

During the first 12 months of life, self-regulation is an important determinant of mechanisms of appetite and satiety. As common experience shows, most infants are highly skilled letting their parents know when they are hungry, usually by crying lustily, and also letting them know when they are satiated, by falling contentedly to sleep. Feeding in the early months, be it breast- or bottle-feeding, should be on a demand basis, and the infant will soon follow a reasonably regular schedule. Parents are rarely successful in "training" their baby to follow a strict feeding schedule. However, newborn infants who are slow to make the adjustment to oral feedings, especially those born prematurely, may have an inadequate suck that keeps them from taking in enough calories to grow. Such infants often require transient tube feedings. Immaturity of gastrointestinal motility patterns also may result in excessive spitting. Careful attention to feeding cues, setting a regular schedule for erratic feeders, and gradual advancement of feeding times and volumes, will help them to resolve such problems of homeostasis.

Around 4 months of age, infants become more visually attentive and socially interactive. Sights, sounds, objects take on new meaning and they often turn away from the breast or bottle. Some parents interpret these exploratory moments as a personal rejection and either terminate feedings inappropriately or begin to wean the child prematurely. Parents need reassurance that such a period of refusal is normal and temporary, the baby will eat better in a quiet or darkened room where stimuli and hence distractions are minimal.

Midline play skills also emerge at about the same time, making it possible for infants to hold their own bottles. Some parents take this opportunity to leave the baby alone with feedings. However, infants still need the

comfort, control, and interaction of being held for feedings. Bottle-feeding in a supine position should be avoided to decrease the incidence of otitis media, aspiration, or the bottle becoming a comfort object required for sleep. The practice of propping the bottle up so that an infant can feed unassisted is also more common in neglected or disturbed parent-child relationships.

One of the most important issues in the development of a child is his emerging autonomy and independence from his parents. Normal children express this individuality by learning how to feed themselves. Negativistic or oppositional behavior is a frequent and normal indication of a child's move from dependence to independence and occurs when they are in the process of discovering how to make their own decisions. If this discovery process is continuously thwarted by parents insisting on their own way, or if the negativistic behavior is rewarded by parental attention or other reinforcement, it will be more likely to continue. Feeding time can be protracted and tense and preceded by a sense of anxiety in both parents and child. If the child's behavior spills over into sleeping and toileting with tantrums as well, the parents rapidly become exhausted, recriminatory and out of control, and the child more powerful and fractious. The situation rapidly becomes intolerable.

At 7 or 8 months, children begin to be better able to manipulate objects with their fingers. They become completely involved in exploring the world with their hands. This, of course, becomes the time to introduce finger foods. Unless children can participate actively in a feeding, they will feel a real conflict of interest. If they can hold onto a piece of food, they can be a participant. While they are working at this important new task, they can be fed a whole meal.

Over the next few months, the child begins to express even more individuality in feeding and their desire to feed themselves will increase. As solids are introduced, the child may show strong preferences regarding texture and flavor. Their response will depend on their adaptability, the threshold of their sensitivity to new foods, and their initial responses to newness, which are all a reflection of their temperament. Refusal of food can be dramatic and messy, and parents may resort to feeding games, bribes, or restraining and force-feeding. Meticulous or controlling parents are often upset by the mess of allowing children to try finger-feeding and instead insist on spooning it all in themselves, which can cause the child to refuse solids. It is not surprising to see infant attempts

to block the spoon during feeding in an effort to take control. Parents should understand that allowing finger-feeding practice speeds up development of skillful and neater feeding. One strategy is to offer the infant finger foods and avoid spoon-feeding entirely. If plastic bibs are not enough, allowing the child to eat naked, or over newspaper or plastic, with a bath afterwards, can make the messiness more tolerable.

The burst in independence that occurs by the end of first year is likely to set the scene for more feeding difficulties. The child's new-found freedom in walking brings such choices as whether to walk away from his mother or to walk towards her, whether to sit in a highchair or to refuse, whether to cooperate with her or to resist. This new sense of autonomy invades all the important events of his day. If the parent presses him to eat, he will become all the more determined to have his own way.

Besides wanting to be more independent during feeds, the child begins to accept food and other caregiving only from the mother as part of a new cognitive awareness of differences. Fathers and grandparents may feel hurt or think that the child is getting spoiled and mothers can become exhausted. Mothers should insist on at least one other caregiver participating in feedings regularly, if possible, so they become gradually accepted.

At the same time, the infant also begins to develop the concept of object permanence, partly learned by repetitive games such as peek-a-boo. The child is learning all about disappearing objects and how to hide and retrieve them, and can be charmed by the game for quite some time. The game is fun for parents when played with toys or faces, but it may become less fun when played with food or feeding utensils thrown from the high chair. Sometimes, the strong and predictable reactions from the parents make it even more exciting for the child who is rewarded with a response to his normal negativity.

The most powerful influences on a child's food preferences are cultural traditions, modeling by family members, especially siblings, and the emotions expressed around foods and mealtimes. Parents are advised never to reward a child for eating or to use food as a reward. The appropriate strategy to improve eating includes establishing regular mealtime routines, ensuring a pleasant eating atmosphere, providing models of eating a variety of nutritious foods, facilitating multiple exposures to new foods, offering small servings to avoid overwhelming a child, reducing between-meal calories to enhance appetite, and avoiding pressure related to either the type or volume of intake.

By recognizing the importance of allowing the child the freedom of learning and exploring independence, parents should see their child's resistance less as a personal insult. Parents will find it difficult to enforce their own will and should not try to do so. Even if they do succeed, sometimes by pure might, subsequent mealtimes will become emotionally charged and future problems are more likely to arise.

However, it is equally important for parents to set reasonable limits and rules around feeding in order to help the child to learn to set their own limits and give them a kind of security. Mealtime misbehavior, such as getting up and down from the table, playing with food, and fighting with siblings, occur when parents have not established control over the child's behavior in general. A child in the toddler age must learn to separate the rituals of mealtime from the rest of the day. Children are likely to tease parents to break their rituals, but observing them is important to the whole family. And if children are expected to sit and eat, even for a short time, they will learn the importance of the ritual. Getting a child to eat more is not important enough to justify chasing games on the floor, or allowing a toddler to walk around while eating, with a bottle hanging out of his mouth, or holding pieces of food in one hand as he plays with a toy with the other. An effective management technique is to serve the child in a high chair or at the table and to terminate the meal calmly when playing with food exceeds eating or there is significant misbehavior, regardless of how much food the child has consumed. The quantity of food is not the issue and most children are well-enough established after the first year to survive nutritionally through many bouts of negativism towards food in the second or third year. Alternatively, the child may be offered a "second chance" one hour later to eat the same food. This will make it more likely that all involved adults will be consistent in setting mealtime limits, because they need not feel they are depriving the child of food. The child will soon learn that food is not a plaything and he will become more serious and respectful of mealtimes.

Development of Feeding Skills

The act of feeding is a complex physiological process that depends on two closely inter-related factors: structure and function. In infants and young children, the entire process is dynamic because of ongoing growth and development.

Structural integrity of the mouth, pharynx, and the larynx is essential to the development of competent feeding and swallowing skills. A child with structural defects of the oropharynx such as cleft palate may have considerable difficulty sucking efficiently. Infant anatomy differs from that of adult. Anatomical structures change their physical relationship to one another during growth and consequently their function is affected. Most growth related changes begin in the third and fourth year, after the development of most feeding skills. The maturation of feeding skills, although influenced by anatomical changes associated with growth, is accomplished largely by developmental changes in the central nervous system, coupled with experiential learning. Individual temperament, interpersonal relationships, environmental influences, and culture further compound the basic physiological complexity of feeding.

Oral feeding for the newborn infant is entirely reflexive, under brainstem control. Rooting, nipple latching, sucking, and swallowing are some of the primitive feeding reflexes. Feeding development begins with the reflexive suck, swallow, and breathe pattern of the newborn. Through feeding experience and the process termed encephalization, different sensory inputs are extended past the brainstem to the midbrain, cerebellum, thalamus, and the cerebral cortex. These suprabulbar areas interpret the sensory input and exert their higher level of control on the brainstem motor centers. As a result, the older infant and young child acquire the ability to evaluate the physical character of the food, manipulate it appropriately, and voluntarily ingest it. The sensory information received and the motor output elicited are closely integrated and influence one another. In this way, feeding and swallowing gradually change from a reflexive to a volitional process.

Feeding development is a learned progression of behaviors. This learning is heavily influenced by oral sensation, related gross and fine motor development, and experiential opportunities. It has also been postulated that a sensitive or critical period exists for optimal learning of feeding skills.

Oral sensation involves proprioception, touch, pressure, temperature, and taste. Mouthing one's hands, feet, toys, and other inanimate objects provides needed experience for later feeding and facilitates development of mouth and hand function. For example, the tongue thrusting (or extrusion) primitive reflex and gag reflex are quite powerful in the first 4 months of life, causing an infant to expel solids placed in his mouth.

These reflexes are modified and diminished as baby brings fingers or toys into the mouth and starts to tolerate more solid foods. This may explain why problems arise with the motor-disabled who are incapable of getting limbs and inanimate objects to their mouths. These children often develop oral hypersensitivity or defensiveness. They may gag when a spoon is placed on their tongue; cry during feeding; pull away as the spoon approaches because of unpleasant past experiences; and refuse to try new taste, texture, or methods of feeding.

Although feeding depends on control of the mouth and pharynx, related motor development plays a role. Stability is needed before the infant can learn mobility. Head control and trunk stability provide the necessary gross motor foundation for the fine motor function seen in hands and mouth. The precise oral-motor movements needed in feeding occur after the head and trunk have achieved stability, symmetry, and alignment. The ability to self-feeding necessitates increasing hand-eye coordination and a fine pincer grasp. Until infants develop these abilities from 9 months onwards, they will be unable to feed themselves properly.

The concept of a "critical" or "sensitive period", although hypothetical, is relevant to feeding development. A critical period refers to a fairly well-delineated period of time during which a specific stimulus must be applied in order to produce a particular action. After such a critical period a particular behavior pattern can no longer be learned so readily. The "sensitive period" refers to the optimal time for the application of a stimulus. After the sensitive period, it is more difficult to learn a specific pattern of behavior. This concept offers an explanation for some feeding problems in children. Infants with severe gastrointestinal disease, central nervous system dysfunction, or prematurity, may require enteral or parenteral feeding and may be deprived of oral stimulation for a prolonged period. For these children, many of the pleasurable sensations usually associated with oral stimulation, such as feeding, mouthing of objects, and thumb-sucking, may be replaced by noxious sensations and experiences such as suctioning, placement of nasogastric tubes, and endotracheal intubation. These children may become orally defensive and resistant to oral feedings; food refusal is a frequent outcome.

Young children usually show a preference for familiar foods over novel ones. This "fear" of new tastes and textures is known as "neophobia". Neophobia is a perfectly normal part of any child's developing relationship to the environment, and may well be essential as it helps avoid

the ingestion of potentially harmful foods. Typically, resistance develops following an episode of choking or nausea after eating, or in association with a traumatic incident. It takes more than 10 to 15 tasting exposures to increase the likelihood that a child will eat previously rejected food. While innate preferences and aversions clearly exist, food acceptance patterns can be altered or learned. They are shaped by three factors: (1) opportunities for repeated exposure to new foods; (2) the social context of meals; and (3) associative learning, either conditioned food preferences or conditioned aversions.

Parents often do not appreciate that their children's initial rejection of new foods is a normal and transitory phenomenon, which can be reduced by means of repeated exposures. Rejection should not be interpreted as dislike of the new food, and hence it should not be removed altogether from the child's diet, depriving them of the opportunity to learn to like it. The problem can easily be overcome if parents repeatedly offer new foods in small quantities in a positive, patient manner, without undue coercion, followed by plenty of praise if the child ingests some.

Routine family meals teach children about their culture's rules of cuisine, that is which foods or food combinations their culture finds acceptable and which it does not. By watching what others eat and do not eat, children learn what foods are "dangerous", "disgusting", and which are not. The rules of cuisine also dictate which meals are eaten at which times of day and what foods are typically eaten at these meals. Importantly, parents' predisposed beliefs about the anticipated acceptance or rejection of a food vary from culture to culture and can affect the parents' feeding behavior with regard to that food. These expectations shape the feeding interaction between the parent and child and, ultimately, influence the child's acceptance or rejection of a particular food.

Children also learn food acceptance patterns through associative conditioning, which is simply the pairing of something in the environment with something else, resulting in a new response. When a food is repeatedly associated with a distinctive social context, the child's preferences can be systematically changed in either a positive or negative direction, depending on the social context employed. Social context is the implicit value placed upon that food. An example of this is when we eat something and get sick afterwards. We invariably assume, rightly or wrongly, that it was that particular food that was responsible. What is so remarkable is that people retain these aversions decades later, even when the

aversions are formed after only one pairing. This has been called "one trial learning". For young children with a medical condition, such as gastroesophageal reflux, which leads to vomiting after meals regardless of the type of foods ingested, it is likely that the intense avoiding response will be transmuted into a distinctly maladaptive response to any feeding occasion.

Similarly, to encourage a child to eat certain food about which he has no strong feelings one way or the other, a non-food reward may be promised, e.g. offering to read a favorite story. This can also result in a significant negative shift in the child's attitude to that food. The coercive nature of the feeding strategy generates a negative emotion that becomes associated with the food. The child may think that if he has to be bribed to eat that food, it cannot be very nice. The medical relevance is that if we try to encourage a child to take medicines by offering him a sweet afterwards, we may be virtually guaranteeing that the child will be less likely to accept the medicine without a fuss next time. When foods themselves are the reward component for children's performance of desired behaviors, the pairing may have a positive effect on children's behavior. To reward a child for playing quietly for about one hour without disturbing a busy parent, he is offered an apple at the end of that period. The food, about which the child has only a neutral attitude initially, becomes highly valued by the child. Again, the social context suggests that food must be valuable or it will not be worth giving as a reward.

CONCLUSION

Feeding problems can dominate the lives of those who feel helpless when denied one of their important parenting roles by a young child who is unable or unwilling to eat. Unfortunately, management of these problems is not usually based on an appreciation of the developmental course of normal feeding behavior and appetite regulation. Consequently, many young children, whose problem is relatively minor to begin with, go on to develop intractable patterns of behavior that not only cause anxiety and distress to their parents, but also may well have adverse consequences for their growth and health.

A proper analysis of feeding problems must not only take into account of the overt behavior, and the psychological causes and correlates of that behavior, but also certain developmental and biological character-

istics of the child. Infant feeding problems must be viewed in a developmental perspective, and the individual characteristics of the child and caregiver should be taken into account when assessing possible precipitating and perpetuating factors. Comprehensive preventive and treatment programs involving a team of diverse specialists should be developed to bring a range of expertise to bear on the subject.

FURTHER READING

1. Carey WB, *Understanding Your Child's Temperament*, Macmillan, A Simon and Schuster Macmillan Company, New York, 1997.
2. Skuse D, Identification and management of problem eaters, *Arch Dis Child* **69**:604–608, 1993.
3. Harris G, Booth IW, The nature and management of eating problems in pre-school children, in Cooper PJ, Stein A (eds.), *Feeding Problems and Eating Disorders in Children and Adolescents*, Harwood Academic Publishers, Melbourne, pp. 61–84, 1992.
4. Wolke D, Skuse D, The management of infant feeding problems, in Cooper PJ, Stein A (eds.), *Feeding Problems and Eating Disorders in Children and Adolescents*, Harwood Academic Publishers, Melbourne, pp. 27–59, 1992.
5. Illingworth RS, Lister J, The critical or sensitive period, with special reference to certain feeding problems in infants and children, *J Pediatr* **65**:839–849, 1964.
6. Birch LL, Development of food acceptance patterns, *Dev Psychol* **26**: 515–519, 1990.
7. Brazelton TB, Feeding: Pleasure or battleground?, in *To Listen to a Child*, Addison–Wesley Publishing Company, pp. 93–105, 1984.
8. Rudolph CD, Diagnosis and management of children with feeding disorders, in Hyman P, DiLorenzo C (eds.), *Gastrointestinal Motility Disorders in Children*, Academy Professional Information Services, New York, pp. 33–53, 1994.
9. Stevenson RD, Allaire JH, The development of normal feeding and swallowing, *Pediatr Clin North Am* **38**:1439–1453, 1991.
10. Finney JW, Preventing common feeding problems in infants and young children, *Pediatr Clin North Am* **33**:775–789, 1986.
11. Hammer LD, The development of eating behavior in childhood, *Pediatr Clin North Am* **39**:379–394, 1992.
12. Stein MT, Common issues in feeding, in Levine MD, Carey WB, Crocker AC (eds.), *Developmental-Behavioral Pediatrics*, 3rd ed., WB Saunders, Philadelphia, pp. 392–396, 1999.
13. Satter E, The feeding relationship: problems and interventions, *J Pediatr* **117**:S181–189, 1990.
14. Watt M, Behavioral problems with feeding children, *J Pediatr Obstet Gynaecol* **12**:16–23, 1988.

This page intentionally left blank

Renal Medicine

This page intentionally left blank

69

A Clinical Approach to Glomerulonephritis

Woo Keng Thye

INTRODUCTION

The term glomerulonephritis refers to an inflammation of the kidneys. Every year, about 650 new patients are diagnosed with end stage renal failure (ESRF) in Singapore. The most common cause for ESRF in Singapore is diabetic nephropathy, which accounts for 42% of all patients. Glomerulonephritis (GN) accounts for 29% and is the second most common cause. Fifteen years ago, GN was the most common but nowadays with early detection and measures to retard progression to ESRF, it has fallen to second place.

In many Western countries, Membranous GN is the most prevalent form of GN, accounting for about 35%. In Asian countries, Mesangial Proliferative GN is the most common (30%), as was the case in Singapore in the 1970's and early 1980's.[1] The high incidence of Mesangial Proliferative GN was attributed to infection in these countries. IgA nephritis which used to be an uncommon cause of Nephrotic

Syndrome (9%) is now becoming more common too (15%). However we are now seeing more patients with Minimal Change (30%) and Focal Global Sclerosis (FGS) (19%) not responding to steroids or prednisolone, meaning that they probably have Focal and Segmental Glomerulosclerosis (FSGS), which is less responsive to conventional therapy. What this implies is that the true incidence of FSGS is not at 9% but is probably more. Our low incidence of FSGS is not in keeping with the incidence in Western countries. The incidence of Mesangio-Capillary GN and Crescenteric GN is low (2%). About 20 years ago, Mesangio-Capillary GN was common in the West and it was related to infection; now it is uncommon. The epidemiology of GN follows a racial and geographical distribution which is influenced by the environment (infection) and the foodstuff we ingest (allergens).[1]

The major clinical syndromes are:

1) asymptomatic hematuria and proteinuria;
2) acute nephritic syndrome;
3) nephrotic syndrome; and
4) rapidly progressive glomerulonephritis.

ASYMPTOMATIC HEMATURIA AND PROTEINURIA

Asymptomatic hematuria and proteinuria is the most common presenting sign for a wide variety of GN. In the Singapore context, this is the usual presentation for IgA nephritis, the most common form of GN occurring in Singapore.[2] Patients with such urinary abnormalities are often referred by their general practitioners following a routine investigation for some unrelated complaints. Such cases are also detected on community health surveys or in the course of screening of national service registrants, for example at the Central Manpower Base as in the case of Singapore. Surgeons, too, often refer patients with asymptomatic hematuria after they have been shown to have normal intravenous pyelogram and cystoscopic examination.

History

One should ascertain that the patient is truly asymptomatic. Enquire for a history of gross or macroscopic hematuria, dysuria or frequency of micturition, which may point to a diagnosis of hemorrhagic cystitis. A history of

nocturia, backache, passage of stones, edema or recurrent sore throat may provide useful clues to the underlying basis of the urinary abnormalities.

If the patient has episodes of gross hematuria, determine if there is a relationship to upper respiratory tract infection, fever or exercise as IgA nephritis is associated with synpharyngitic hematuria, i.e. gross hematuria occurring simultaneously with sore throat.

Always ask for a history of systemic illness. Tuberculosis, Systemic Lupus Erythematosus and Henoch–Schonlein purpura may present with urinary abnormalities. A family history of nephritis, hypertension and diabetes mellitus may be important. In a married woman, enquire into a past history of pre-eclampsia.

Physical Examination

In the general examination look for pallor, sallowness, presence of edema and the rash of Henoch–Schonlein purpura or Systemic Lupus Erythematosus. Examine the abdomen for ballotable kidneys or renal masses, which may suggest polycystic kidneys or a renal tumor. Always check the blood pressure; examine the fundi and listen for a renal bruit.

In most cases with asymptomatic hematuria and proteinuria the physical examination is usually normal; nevertheless a complete physical examination is mandatory to exclude any obvious underlying cause for the urinary abnormality.

Investigations

1) An abnormal full blood count may be the first clue to SLE and tuberculosis.
2) Urine Microscopy
    a) RBC count is usually variable and may be anything from 5–10 to 100–300 per high power field.
    b) WBC count: if pyuria is present, exclude urinary tract infection by doing a urine culture. For sterile pyuria, tuberculosis has to be excluded.
    c) Casts: RBC casts point to a glomerulonephritis. Granular casts associated with more than a gram of proteinuria per day denotes a more severe lesion.
    d) Albumin may vary from 1+ to 3+.

3) Quantitation of total urinary protein (TUP) in 24 hours. Normally this should not exceed 0.15 gm. In our experience, a TUP of more than 1 gm generally denotes a more severe glomerular lesion on renal biopsy, i.e. the presence of glomerular scarring (glomerulosclerosis).
4) Blood urea, serum creatinine and creatinine clearance should be documented.
5) Anti-nuclear factor (ANF) should be done when one suspects SLE together with Anti-DNA and serum complement.

Mild microscopic hematuria (< 10 RBC/high-power field (hpf)) in the absence of significant proteinuria is of little prognostic significance. In our experience, renal biopsies of this group of patients usually reveal only mild glomerulonephritis which generally has a good prognosis.

Sometimes on follow-up, patients may develop gross hematuria which may be precipitated by respiratory tract infections or exercise. Such patients on biopsy usually have IgA nephritis.

Proteinuria of 1 gm or more is an indication for a renal biopsy. Biopsy is performed under ultrasound guidance.

Clinical Course

Most patients with asymptomatic hematuria and proteinuria have a benign course as they are likely to have mesangial IgA nephritis, which has a favourable prognosis in most cases. No treatment is required for most of these patients and all they require is reassurance. They could be followed up by their general practitioners and have their blood pressure, urine microscopy, serum creatinine and urinary protein checked once a year. It is important to treat any existing hypertension as uncontrolled hypertension often leads to renal impairment in patients with IgA nephritis.

In those patients with IgA nephritis with significant glomerulosclerosis especially in the presence of severe proteinuria, the prognosis is guarded. Patients with crescents on biopsy have a poorer long-term prognosis.

On long-term follow-up, some patients may develop gross hematuria precipitated by upper respiratory tract infections or exercise. They may have colicky loin pain due to clot colic. These patients require reassurance, rest and plenty of fluids as well as antibiotics for the respiratory tract infections.

Those who develop edema or the nephrotic syndrome will require diuretic therapy. In those with mild diffuse mesangial proliferative GN

with the nephrotic syndrome, a 12-week course of prednisolone therapy starting at 60 mg/day or 1 mg/kg body weight and tailing off by 12 weeks may induce a remission in about 50% of cases. These are patients with selective proteinuria. Hypertension, when it occurs, must be treated aggressively as uncontrolled hypertension can lead to rapid deterioration of renal function, culminating in ESRF.

Currently, the indications for combination therapy with dipyridamole and low dose warfarin in our Department of Renal Medicine are the presence of any one of the following parameters: proteinuria >1 gm/day; hypertension; renal impairment; glomerulosclerosis $>20\%$ on renal biopsy; presence of even a single crescent; medial hyperplasia of blood vessels on biopsy.

Our department's guidelines for treatment of patients with IgA nephritis with asymptomatic hematuria and proteinuria consist of control of systemic hypertension, use of ACE inhibitor and or Angiotensin II Receptor Antagonist (ATRA) for glomerular hyperfiltration, dipyridamole and low dose Warfarin, low protein diet (0.8 gm/kg bw/day) to reduce macromolecular flux of high protein diet as well as to reduce afferent vasodilation in glomerular hyperfiltration and control of serum cholesterol to prevent lipid induced glomerulosclerosis.

On the whole, most patients with asymptomatic hematuria and proteinuria due to IgA nephritis run a benign course, except for about 30% (usually associated with glomerulosclerosis and heavy proteinuria) who develop renal failure over a period of 10 years. These patients would ultimately require renal transplantation or dialysis. In other words, IgA nephritis is not always a benign disease, especially in Singapore, where we have large numbers of people with the disease. It is therefore an important cause of ESRF.

Clinical Trials in IgA Nephritis

The use of Persantin and low dose Warfarin in IgA Nephritis is now accepted as Level I evidence and support Grade A recommendation.[3]

Since ACE inhibitors decrease ACE activity, ACE inhibitor therapy in patients with IgA nephritis would be expected to reduce renal injury in these patients. We have shown that ACEI therapy does lead to decreased proteinuria and retardation of the progression of renal failure in IgA nephritis. In our study we also showed that Angiotensin II Receptor

Antagonist (ATRA) is as effective as ACEI in decreasing proteinuria and preserving renal function.

It has been postulated that ACEI/ATRA (ACE inhibitor/Angiotensin receptor antagonist) may decrease proteinuria in patients with glomerulonephritis by its action on the Glomerular Basement Membrane. We performed a study to examine the relationship between the response of patients with IgA Nephritis (IgA Nx) to ACEI (Enalapril)/ATRA (Losartan) therapy by decreasing proteinuria and its effect on the Selectivity Index (SI) in these patients. Forty one patients with biopsy proven IgA Nx entered a control trial with 21 in the treatment group and 20 in the control group. The entry criteria included proteinuria of 1 gm or more and or renal impairment. Patients in the treatment group received ACEI (5 mg)/ATRA (50 mg) or both with 3 monthly increase in dosage. In the control group, hypertension was treated with atenolol, hydrallazine or methyldopa. After a mean duration of therapy of 13 ± 5 months, in the treatment group there was no significant change in serum creatinine, proteinuria or SI but in the control group, serum creatinine deteriorated from 1.8 ± 0.8 to 2.3 ± 1.1 mg/dL ($p < 0.05$). Among the 21 patients in the treatment group, 10 responded to ACEI/ATRA therapy determined as a decrease in proteinuria by 30% (responders) and the other 11 did not (non-responders). Among the responders, SI improved from a mean of 0.26 ± 0.07 to 0.18 ± 0.07 ($p < 0.001$) indicating a tendency towards selective proteinuria. This was associated with improvement in serum creatinine from mean 1.7 ± 0.6 to 1.5 ± 0.6 mg/dL ($p < 0.02$) and decrease in proteinuria from mean of 2.3 ± 1.1 g/day to 0.7 ± 0.5 g/day ($p < 0.001$).

After treatment, proteinuria in the treatment group (1.8 ± 1.6 g/day) was significantly less than in the control group (2.9 ± 1.8 g/day) ($p < 0.05$). The post-treatment SI in the responder group (0.18 ± 0.07) was better than that of the non-responder group (0.33 ± 0.11) ($p < 0.002$). Eight out of 21 patients in the treatment group who had documented renal impairment had improvement in their renal function compared to 2 in the control group ($\chi^2 = 4.4$, $p < 0.05$). Of the 8 patients in the treatment group who improved their renal function, 3 normalized their renal function.

Our study suggest that ACEI/ATRA therapy may be beneficial in patients with IgA Nx with renal impairment and non-selective proteinuria as such patients may respond to therapy with improvement in protein selectivity, decrease in proteinuria and improvement in renal function. ACEI/ATRA therapy probably modifies pore size distribution by reducing the

radius of large nonselective pores, causing the shunt pathway to become less pronounced, resulting in less leakage of protein into the urine.

Individual antiproteinuric response to ACEI/ATRA therapy varies depending on ACE gene polymorphism as those with the DD genotype respond better to the antiproteinuric effect of ACEI/ATRA therapy. Yoshida in a study examining the role of the deletion polymorphism of the ACE gene in the progression and therapeutic responsiveness of IgA nephropathy using the ACEI lisinopril reported that the ACEI lisinopril significantly decreased proteinuria in the DD genotype patients but not in the II or ID genotype. Similar findings have also been reported by Moriyama.

ACUTE NEPHRITIC SYNDROME

The features are edema with gross hematuria (smoky urine) and frequent association with hypertension. Sometimes the symptoms are complicated by encephalopathy and congestive heart failure. This condition may be caused by bacteria, parasites, viruses, systemic lupus erythematosus, Henoch–Schonlein purpura and Guillain–Barre syndrome.

Post streptococcal glomerulonephritis which classically presents as the nephritic syndrome is better referred to as Post Infectious GN because apart from streptococci, other bacteria and viruses can be the causative agent. It affects children principally, but no age is exempt. There is usually a latent period of 10 to 21 days. The urine characteristically shows a rusty or smoky hue. Mild renal impairment is common. Serum Complements (CH50 and C3) are usually low but normalize after 6 to 8 weeks.

Treatment is usually symptomatic: generally bed rest is recommended during the acute phase. Restrict fluids and salt if edema is present. Treat accompanying hypertension and heart failure. A course of Penicillin is given if throat swab grows streptococci. Dialysis may be required in some instances to tide the patient over the acute renal failure which may complicate the course of a few patients.

For the majority of patients, this is a benign disease, with children faring better than adults. Glassock[5] reported a 99% 5-year and a 97% 10-year survival for children whereas adults have 95% 5-year and a 90% 10-year survival. Potter[6] in a series of 534 patients from Trinidad followed up for 12 to 17 years reported only 2 deaths from chronic renal failure. Both those patients had persistent urinary abnormalities. Of the surviving

patients, 3.6% had urinary abnormalities and another 3.6% had hypertension. All had normal serum creatinine.

The bad prognostic features in this disease are persistent nephrotic syndrome, hypertension, renal impairment and crescents. These bad prognostic features however, as we shall see later, are true for most forms of GN.

NEPHROTIC SYNDROME

Nephrotic Syndrome is a clinical entity of multiple causes characterised by increased glomerular permeability manifested by massive proteinuria of more than 3 gm per day associated with edema and hypoalbuminemia of less than 30 gm. Very often, there is also associated hypercholesterolemia and hypertriglyceridemia and lipiduria. Any glomerular lesion may be associated, at least temporarily, with heavy proteinuria in the nephrotic range.

Causes of Nephrotic Syndrome

- 80% due to GN
- 20% due to miscellaneous causes

1) Diabetes mellitus
2) Amyloidosis
3) Precipitating causes of renal vein thrombosis (RVT)
   a) nephrotic syndrome (hypercoagulable state)
   b) renal amyloid gives rise to thrombosis of intrarenal veins
   c) hypernephroma gives rise to obstruction and RVT
   d) trauma of renal veins
   e) severe dehydration especially in infants suffering from gastroenteritis. RVT is a complication of Nephrotic Syndrome and not a cause of Nephrotic Syndrome
4) Malignancy: Hodgkin's Disease, bronchogenic carcinoma, cancer of the breast, bowel, leukemias, myeloma
5) Infections: Hepatitis B and C, Malaria, syphilis, leprosy
6) Drugs: Trimethadione, penicillamine, phenindione, gold, mercury, bismuth, captopril, NSAIDS
7) Autoimmune Disease, SLE, Cryoglobulinemia, Thyrotoxicosis
8) Congenital Nephrotic Syndrome

Causes: Congenital syphilis, cytomegalovirus infection, mercury poisoning, maternal tuberculosis

9) Miscellaneous: Prophylactic inoculation (smallpox, polio, tetanus), bee stings, pollen allergy

MINIMAL CHANGE DISEASE

Young children are especially affected with a peak from 2–4 years. Minimal Change accounts for 60–70% of all idiopathic nephrotic syndrome in children and 10–30% in adults. In Singapore this is also the commonest lesion in adults (30%). Hypertension and renal impairment are uncommon complications and microscopic hematuria is rare. Shaloub[7] considers this disease a disorder in T cell function with abnormal lymphokine production. Focal Global Sclerosis (FGS) accounts for 20% of nephrotics in adults in Singapore. Clinically, FGS behaves like Minimal Change, is steroid responsive and has a good prognosis.

Treatment consists of a three months' course of prednisolone. In those who fail to respond or where they have frequent relapses, cyclophosphamide or cyclosporine A may induce long-lasting remissions.

This is a disease with a good prognosis even though the relapse rates are high. Depending on a patient's response to prednisolone and his frequency of relapses various categories have been described. Primary responders, non-relapsers form 38%. If a patient has less than 2 relapses in the first 6 months of the initial response he is a primary responder, infrequent relapser (19%). If he had 2 or more relapses in 6 months, he is a primary responder, frequent relapser (42%). Patients who do not respond to steroids following an initial response (secondary non-responder) form 5%.[8]

In a study among adults, Cameron[9] showed that 18% responded with early loss of proteinuria but 70% of them relapsed, 63% repeatedly. A small number, after repeated relapse and remission, acquired steroid unresponsiveness. They displayed focal glomerulosclerosis on renal biopsy. The use of cytotoxics is recommended in this group.

Of adults with Minimal Lesion, 50–60% will remit for 5 years or more. Those who fail to respond to Cyclophosphamide can be given a course of Cyclosporine A for 6 months to a year, or Mycophenolate Mofetil (MMF) (0.75 gm to 1 gm twice daily) for 6 months.

A patient with Minimal Lesion has an excellent long-term prognosis if he has minimal glomerular lesion on light microscopy, foot process

fusion on electron microscopy, absence of immunoglobulins on immunofluorescence (IMF), and complete remission following a course of steroids. Even so, multiple relapses will still occur.

FOCAL & SEGMENTAL GLOMERULOSCLEROSIS (FSGS)

There is a slight preponderance of males with a peak at 20–30 years accounting for 10–20% of idiopathic nephrotic syndrome. The incidence of FSGS is rising in most countries.[10] In Singapore it is now about 15%.

About 70–90% present as the Nephrotic Syndrome; hypertension may be an associated feature.

The majority of patients with this form of nephritis experience a progressive decline in GFR; hypertension and persistent proteinuria. Initially patients present with asymptomatic proteinuria but often they ultimately become nephrotic. Initial studies reported little benefit with treatment. Nowadays with treatment, response range from 30–50% (include partial response). High dose prednisolone (1 mg/kg bw a day) for first 2 months, then 30 mg for 3rd month, and thereafter reduce gradually to 10 mg and maintain till end of 6 months. Those who do not respond to prednisolone should have cyclosphosphamide at 2 mg/kg bw a day together with prednisolone (30 mg/day) for 3 months, then reduce and maintain for another 3 months. In those who fail cyclophosphamide they could be given cyclosporine A at 5 mg/kg bw a day for 3 months, then 4 mg at 4th month, 3 mg at 5th month, 2 mg at 6th month, and maintain. There may be a potential role for FK 506 (0.1 to 0.2 mg/kg bw a day) or MMF.

MESANGIAL PROLIFERATIVE GLOMERULONEPHRITIS

It occurs in about 25% of idiopathic nephrotic syndrome among adults in Singapore but appears to be a much less common cause of nephrotic syndrome in Western countries. IgA nephritis accounts for about 8% of nephrotic syndrome seen in Singapore and is therefore an uncommon cause of nephrotic syndrome.

The long-term evolution of this type of nephritis is not well understood. This discussion excludes IgA nephritis, which has been dealt with

earlier, but it includes IgM nephropathy, IgG and the IMF negative group.

Patients with this lesion who become nephrotic and develop focal and segmental sclerosis have a higher incidence of developing chronic renal failure (CRF). Only 30% of those with nephrotic syndrome experience complete remissions with steroid. They are usually the ones with mild mesangial proliferation with no focal and segmental sclerosis.

In those patients therefore who have mild proliferation of the mesangium with no evidence of focal and segmental sclerosis and negative immunoglobulin staining on IMF, a trial of steroid therapy should be offered as they have a good chance of achieving remission. Habib[11] in fact considered such patients as part of the spectrum of Minimal Change GN. Waldher[12] reported that more than 50% of patients with Mesangial Proliferative GN were steroid-resistant, 70% with associated focal and segmental sclerosis. There is a need for controlled therapeutic trials in patients with this form of nephritis. In our experience, those with selective proteinuria tend to respond to steroids and failing that cyclophosphamide.

Those who fail cyclophosphamide can be offerred cyclosporine A or MMF.

MEMBRANOUS GLOMERULONEPHRITIS

Hypertension and azotemia are late features of the disease. Microscopic hematuria is common but gross hematuria is a rare feature. Renal vein thrombosis is secondary to the glomerulopathy rather than the cause of it.

This disease runs an indolent and slowly progressive course with remissions and exacerbations of the nephrotic syndrome. Children have a better prognosis, with less than 5% CRF after 5 years and a 90% 10-year survival. Adults however have a less benign course; 25% achieve spontaneous remission with another 25% spontaneous partial remission (less than 2 gm proteinuria). Cameron's series[13] had a 75% 5-year and 50% 10-year survival. Even patients with partial remission have a better outlook than those who have no response at all.

A controlled trial with high dose alternate-day prednisolone in the USA has reported a reduction in proteinuria and progression of CRF.[14] However, if there is already abnormal GFR, steroid therapy is not of much use. We would offer a 3 months course of prednisolone therapy and failing that cyclophosphamide. There is a potential role for other agents

like chlorambucil (0.15–0.2 mg/kg bw/day),[15] cyclosporine A (5 mg/kg bw/day) and FK 506 (0.1–0.2 mg/kg bw/day)[16] and MMF.

MESANGIOCAPILLARY GLOMERULONEPHRITIS (MCGN)

All age groups are involved, especially those aged 5–15 years. It occurs in 5–10% of children with the nephrotic syndrome. Fifty percent have associated upper respiratory tract infection and 40% have high anti-streptolysin O titer (ASOT).

This type of nephritis has a relentless but slowly progressive course. The bad prognostic features are low GFR, hypertension, persistent nephrotic syndrome and the presence of diffuse crescents on renal biopsy.

For Type I MCGN (Subendothelial Deposits) the 5-year and 10-year survival are 80% and 60% respectively. For Type II MCGN (Dense Deposit Disease) the respective survival rates are poorer, 60% and 45% at 5 and 10 years. Other series reported a poorer prognosis in the presence of nephrotic syndrome with 40% survival at 10 years, compared to 85% at 10 years for patients with no nephrotic syndrome.

But even in the patients with nephrotic syndrome, the occasional remission has been reported. In general, however, those with Type II disease have a poorer outlook.

For the moment there is no clearly established form of treatment. McEnery[17] reported the beneficial effects of continuous low dose prednisolone whereas Kincaid-Smith[18] reported 3-year survival of 82% using a combination regimen of cyclophosphamide, persantin and warfarin (Melbourne Cocktail) in an uncontrolled trial. We would advocate a 3 months course of prednisolone and failing that cyclophosphamide. Failing that, try cyclosporine or MMF.

MANAGEMENT OF NEPHROTIC SYNDROME

General Treatment

1) Diuretic Treatment

   These are the major agents in treatment:

   a) Frusemide can be used alone. Increase the dose till diuresis occurs. K^+ supplements are required.

b) Spironolactone should be avoided if serum K^+ is high or patient has renal impairment.
c) Hydrochlorothiazide has a synergistic action with frusemide and spironolactone.

We usually use a combination regimen of all 3 diuretics as they have synergistic actions.

2) Treatment of hypertension.
3) Treatment of infections. Use the appropriate antibiotics.
4) Diet: The patient requires a high protein, low salt diet with fluid restriction.
5) Infusion of Na^+-free albumin induces diuresis but its benefit is evanescent.
6) Use ACE inhibitor or ATRA to reduce intra-glomerular hypertension (hyperfiltration) and limit protein loss in urine.

Specific Treatment

1) Investigate and try to elucidate the cause of Nephrotic Syndrome and if possible remove or treat it.
2) Check through list of causes and investigate accordingly. In all patients, always exclude SLE.

Do anti-nuclear factor (ANF), anti-DNA, and serum complement. A patient with membranous GN should be screened for hepatitis B antigen and antibody.

Treatment with Steroids and Cytotoxic Agents

Primary Treatment to induce remission:

1) Minimal change GN and lupus nephritis respond well to a course of prednisolone starting at 60 mg or 1 m/kg bw/day and reducing gradually over a period of 3 months. For those who fail to respond to prednisolone or who are frequent relapsers, cyclophosphamide (2 mg/kg bw) for 3 months is advocated. Those who fail cyclophosphamide can be given a course of cyclosporine A at 5 mg/kg bw for 3 months with reduction over the next 6 months and maintain up to 1 year.
2) Mild diffuse mesangial proliferative GN may respond to prednisolone, failing that, try cyclophosphamide, cyclosporine or MMF.

3) Membranous GN may respond to a course of prednisolone for a period of 3 months, failing that, try cyclophosphamide, cyclosporine or MMF.
4) Focal and segmental glomerulosclerosis may respond to prednisolone and cyclophosphamide for 6 months. If no response, try cyclosporine or MMF.
5) Newer agents include FK 506 (0.1–0.2 mg/kg bw/day)[16] and mycophenolate mofetil (MMF) (0.75–1 gm twice daily)[19] for 6 months.

Persantin and Warfarin Plus Regimen (P and W + Regimen)

All patients who fail to respond to steroids and cytotoxics should be offered P and W + regimen, which would help to retard progression to ESRF.[20]

1) Persantin (dipyridamole) — anti-platelet and anti-PDGF, 75–100 mg tds with low dose warfarin (anti-thrombotic), 1 to 3 mg (INR < 1.6) to retard progression of renal failure.
2) Treat hypertension.
3) ACE Inhibitor to reduce intra-glomerular hypertension and retard progression of renal failure.
4) Angiotensin Receptor Antagonist (ATRA) to reduce intra-glomerular hypertension (Losartan) and retard progression of renal failure.
5) Restricted protein diet (0.8 gm/kg bw) to decrease afferent arteriolar vasodilation and hyperfiltration.
6) Treat high lipids and hypercholesterolemia as cholesterol is toxic to mesangial cells.

Rapidly Progressive Glomerulonephritis (Crescenteric Glomerulonephritis)

This is a clinical syndrome of rapid and progressive decline in renal function, usually resulting in end stage renal failure in weeks to months, where there is extensive and exuberant proliferation of epithelial cells of Bowman's space.

In clinical practice, this condition is diagnosed when a patient has a rapid decline in renal function (acute renal failure), usually with oliguria, hematuria, hypertension in the presence of normal sized or enlarged kidneys. A renal biopsy will show extensive crescents of 50% or more.

Preceding flu-like illness is found in 50% of patients. Hypertension is often mild. Urine microscopy shows many RBC with RBC casts. Serum complements (CH50, C3 and Clq) are often normal. Fibrin degradation products are often present and anti-streptolysin 0 titer is increased in 30% of patients.

Treatment consists of plasmapheresis with steroids and cyclophosphamide ideally administered prior to the onset of oliguria. A quadruple regimen with heparin, prednisolone, cyclophosphamide and anti-platelet agents (dipyridamole) has been used with success, but caution should be exercised when using heparin. A low dose continuous heparin regimen during the acute phase to avoid hemorrhage and then switch over to warfarin therapy is safer. A better alternative is to use pulse therapy with methyl prednisolone (0.5 gm IV daily for 6 days) followed by plasmapheresis. Other measures include restriction of salt and water, treatment of hypertension and supportive dialysis.

Patients with 50–80% crescents on biopsy have less than 30% 5-year and less than 10% 10-year survival. Those with 80% crescents have an 8% 5-year and less than 5% 10-year survival.

REFERENCES

1. Woo KT, Chiang GSC, Glomerulonephritis in Singapore (1987–1997), *Medicine of the Americas* **1**(I), 30–341, 2000.
2. Woo KT, Lau YK, Choong HL, Zhao Y, Tan HB, Cheung W, Yap HK, Chiang GSC, IgA nephropathy: Effects of clinical indices, ACE/ ATRA therapy and ACE gene polymorphism on disease progression, *Nephrology*, **7**: S166–172, 2002.
3. Woo KT, Lee GSL, Pall AA, Dipyridamole and low-dose warfarin without cyclophosphamide in the management of IgA nephropathy, *Kidney Int* **57**(1), pp. 348–349, 2000.
4. Woo KT, Lau YK, Wong KS, Chiang GSC, ACEI/ATRA therapy decreases proteinuria by improving glomerular permselectivity in IgA nephritis, *Kidney Int* **58**:2485–2491, 2000.
5. Glassock RJ, Adler SG, Ward HJ, Cohen AH, Primary Glomerular Diseases, in: Brenner BM, Rector FC (eds.), *The Kidney*, WB Saunders Company, Philadelphia, pp. 1182–1279, 1991.
6. Potter EV, Lipschultz SA, Abidh S *et al.*, Twelve to seventeen year follow up of patients with poststreptococcal acute glomerulonephritis in Trinidad, *N Engl J Med* **307**:725–729, 1982.
7. Shalhoub RJ, Pathogenesis of lipoid nephrosis: A disorder of T cell function, *Lancet* **2**:556–560, 1974.

8. Cameron JS, The long-term outcome of glomerular diseases, in Schrier RN and Gottschalk CW (eds.), *Diseases of the kidney*, 2nd ed., Little Brown, Boston, pp. 1895–1958, 1992.

9. Cameron JS, The problem of focal segmental glomerulosclerosis, in: Kincaid–Smith P, d'Apice AJF, Atkins R (eds.), *Progress in Glomerulonephritis*, New York, John Wiley and Sons, pp. 209–228, 1979.

10. Haas M, Spargo BH, Conventry S, Increasing incidence of focal-segmental glomerulosclerosis among adult nephropathies: A 20-year renal biopsy study, *Am J Kidney Dis* **26**:740–750, 1995.

11. Habib R, Levy M, Gubler MC, Clinicopathological correlation in the nephrotic syndrome, *Paediatrician* **8**:325, 1979.

12. Waldher R, Gubler MC, Levy M, Broyer M, Habib R, The significance of pure diffuse mesangial proliferation in idiopathic nephrotic syndrome, *Clini Nephrol* **10**:171–179, 1978.

13. Cameron JS, The natural history of glomerulonephritis, in: Black D, Jones NF (eds.), *Renal Disease*, Blackwell Scientific Publishers, London, pp. 329–382, 1979.

14. Collaborative Study of the Adult Idiopathic Nephrotic Syndrome, A controlled study of short term prednisolone treatment in adults with membranous nephropathy, *N Engl J Med* **301**:1301–1306, 1979.

15. Ponticelli C, Zucchelli P, Passerini P, Cesana B, Locatelli F, Pasquali S *et al*., A 10 year follow-up of a randomized study with methylprednisolone and chlorambucil in membranous nephropathy. *Kidney Int* **48**:1600–1604, 1995.

16. Mc Cauley J, Shapiro R, Ellis D, Igdal H, Zakis TA, Starzl TE, Pilot trial of FK 506 in the management of steroid resistant nephrotic syndrome, *Nephrology, Dialysis and Transplantation* **8**:1286–1290, 1993.

17. McEnery PT, McAdams AJ, West CD, Membranoproliferative Glomerulonephritis: improved survival with alternate day prednisolone therapy, *Clin Nephrol* **13**:117–124, 1980.

18. Kincaid–Smith P, The natural history and treatment of mesangiocapillary glomerulonephritis, in: Kincaid–Smith P, Mathew TH, Becker EC (eds.), *Glomerulonephritis. Morphology, Natural History and Treatment*, Part I, John Wiley and Sons, New York, pp. 591–609, 1973.

19. Briggs WA, Choi MJ, Scheel PJ Jr. Successful mycophenolate mofetil treatment of glomerular disease, *Amer J Kidney Disease* **31**:213–217, 1998.

20. Woo KT, Current Therapeutic Strategies in Glomerulonephritis, *Ann Acad Med Singapore* **28**:272–278, 1999.

70

Urinary Tract Infection

Choong Hui Lin

Urinary tract infection (UTI) refers to the presence of microorganisms in the urinary tract, causing a host response. It should be differentiated from colonization though colonization precedes infection. Transient introduction of bacteria into the urinary tract probably occurs in the healthy, especially sexually active women.

UTI may present as asymptomatic bacteriuria, urethritis, cystitis or acute pyelonephritis in women. In men, they often present as acute prostatitis, chronic prostatitis and pyelonephritis. In both sexes, predisposing conditions such as obstructive uropathy from congenital anomalies, vesicoureteric reflux, renal calculi, stricture formation and tumors have to be excluded. UTI is very common, second only to respiratory tract infections. Severity may range from mild to severe associated with septicemia, pyonephrosis and abscess formation. Unless there is an underlying anatomic abnormality, urinary tract infection is not likely to cause serious renal disease.

UTI may be classified as complicated or uncomplicated. A UTI is described as uncomplicated when it occurs without the presence of physiologic or anatomical abnormalities, not due to instrumentation and is

community acquired. These occur usually in women as acute cystitis and acute pyelonephritis. At least 25–30% of women between the age of 20–40 years have had at least one infection. An infection is labeled complicated when there are structural or functional abnormalities present in the urogenital tract. Infections in men have been often labeled as complicated as infection rate in men is much lower (as much as 50 times less) than women and is usually accompanied by some predisposing factor. A search for a cause is initiated after the first infection in the man while in woman presenting for the first time with cystitis or easily treatable acute pyelonephritis, investigations into the cause is not necessary. Young healthy men may suffer from an uncomplicated UTI due to acute prostatitis. Factors which suggest a complicated infection include: the male sex, elderly, nosocomial infection, pregnancy, indwelling catheter, recent instrumentation, anatomic abnormality of the urinary tract, previous childhood infection, recent antibiotic use, symptoms persisting for more than 7 days at the time of presentation, diabetes mellitus or immunosuppression.

SIGNS AND SYMPTOMS

Symptoms of frequency, urgency, dysuria, strangury and suprapubic pain suggest a lower tract infection. There may also be cloudy or foul-smelling urine or gross hematuria. In lower tract involvement, there is usually no fever. Upper tract symptoms usually include loin pain associated with a positive renal punch and the patient is febrile. Lower tract symptoms may be present.

In males, symptoms suggesting prostatitis include a sensation of rectal irritation, scrotal pain and pain referred to the front of the thighs. Examination might reveal a tender prostate.

MICROBIOLOGICAL SPECTRUM AND ANTIBIOTIC SUSCEPTIBILITY

Most UTI's are caused by gram-negative organisms. In uncomplicated infections from the community, *E. coli* is the major causative pathogen in 70–95% and *Staph saprophyticus* in 5–20%.[1] From urinary isolates in 1998 in Singapore General Hospital (SGH), *E. coli* accounted for 34%, *Klebsiella sp.* 16%, Enterococcus 15% and Pseudomonas 9%.[2] In a hospital setting,

Table 1 Antibiotic Susceptibility of *E. coli*[2]

Ampicillin	33%
Cephalexin	57%
Ciprofloxacin	48%
Ceftriaxone	94%
Gentamicin	90%
Amikacin	96%
Co-trimoxazole	49%
Nitrofurantoin	82%
Nalidixic acid	65%

nosocomial and more complicated infections are seen. A study of admissions for urinary tract infection in SGH showed positive cultures in only 41.4% yielding *E. coli* (56.4%), Klebsiella (14.5%), Candida (12.7%) and Enterococcus (7.3%) reflecting the tertiary nature of the institution.[3] Eight out of 150 'O' *E. coli* serotypes responsible for 80% of infections are considered uropathogenic. They have certain characteristics including increased adherence to uroepithelium by increased expression of P, F and S fimbriae, hemolysin and aerobactin production.

Certain relationships have been identified. *Staph saphrophyticus* is commonly found in infections involving sexually active young women, Proteus infections in boys aged 1–12 years and in recurrent stone disease, and *Strep fecalis* in infections from elderly men with prostate problems.

Antibiotic use should take into consideration the high resistance shown in *E. coli* isolates. Table 1 shows the antibiotic susceptibility in *E. coli* isolates in 1998.

It is important to remember that nitrofurantoin and nalidixic acid are not useful agents in impaired renal function as they rely on urinary concentration to deliver adequate drug to the urine. However, nitrofurantoin is extremely useful in prophylaxis in normal renal function in women as there does not seem to be in increase in resistance. This drug can also be used in pregnancy and for prophylaxis.

DIAGNOSIS

The diagnosis is usually made on history and examination alone. Treatment of simple cystitis is usually started without the benefit of cultures though it would be ideal to have one performed prior to starting

treatment. In the case of isolated cystitis in women, cultures are probably not necessary unless there is no response to treatment or when relapse occurs. If pyelonephritis is suspected, a specimen for urine culture should be obtained before starting treatment.

In the collection of a urine sample, the importance of proper midstream collection cannot be over-emphasized. Contamination is common if patients are not properly instructed. The patient should be able to pass at least 150–200 ml of urine during the procedure. In women, it is necessary to clean the periurethral area by washing with water or wiping from front to back with moist gauze wet with saline or tap water. Antiseptics should not be used. In men, the foreskin should be retracted and the periurethral area cleaned. In women, the labia should be parted during collection and in men the foreskin remain retracted.[4] A midstream sample is collected after the urine flow is started. The initial part of the stream is allowed to drain for a few seconds after which the collection container is passed under the stream *without stopping* the urine flow. Urine may also be collected via catheter or suprapubic puncture, which is more laborious and less pleasant to perform. The presence of epithelial cells on microscopy suggests contamination.

The following are recommended guidelines for diagnosis based on urinary cultures:[5]

a) Asymptomatic females (supervised collections):
   - Two consecutive positive clean-catch midstream urine (MSU) specimens revealing 10^5/ml or more colonies in the urine with the same organism in all three cultures.
   - Two consecutive positive clean-catch midstream urine (MSU) specimens revealing 5×10^4/ml or more colonies in the urine with the same organism in all three cultures if the organism is of the same serotype (*E. coli*), species (Proteus), phage type (Staphylococci).
   - Single urethral catheter specimen revealing 10^5/ml or more colonies in the urine. Lower counts require repeat collection.
   - Single suprapubic bladder puncture revealing *any number of colonies*. Small numbers with organisms suggestive of skin contaminants such as *Staph epidermidis* or diphtheroids should be repeated.

b) Symptomatic males and females
   - Preferably two consecutive clean-voided specimens or presence of pyuria and bacteria in urinary sediment.

c) Asymptomatic males:
   - Two consecutive clean-voided specimens revealing 10^5/ml or more colonies in the urine with the same organism in both.
   - Three consecutive positive clean-catch midstream urine (MSU) specimens revealing less than 10^5/ml or more colonies in the urine with the same organism in all three.
   - Single urethral catheter specimen or suprapubic bladder puncture, as in asymptomatic females.

With a positive gram-negative culture of more than 10^5 cfu/ml, there was true infection in 92% of cases compared with 74% if there were only 10^3 cfu/ml to 10^5 cfu/ml in bacteremic pyelonephritis. For males, 2 consecutive specimens with 10^5/ml or more colonies of the same organism would suffice. If the counts are lower, three consecutive specimens should be taken. The above takes into consideration the work of Kass, who in 1956 introduced the method of quantitating bacteria to distinguish contamination from true infection.[6] A lower threshold of significance is used when: 1) fastidious organisms such as chlamydia, mycobacterium are cultured; 2) collection was by urinary catheter or suprapubic aspiration; or 3) the patient is symptomatic with a single organism cultured. Negative cultures are common with recent use of antibiotics. In practice, suprapubic bladder puncture is seldom performed.

If allowed to stand for more than 2 hours at room temperature, growth of contaminants may make diagnosis difficult. It is recommended that the specimen be refrigerated at 4–10°C if specimens cannot reach the culture laboratory in 2 hours. Dipslides containing agar media are now available and are convenient to use when collection is done after office hours.

Pyuria is supporting evidence and can be readily seen on urine microscopy. Most symptomatic bacteriuric patients had 10/ml or more in unspun urine.[7,8] Urine specimens from patients with UTIs usually contain more than 1 to 2 WBCs per high power field. However, up to 50% of patients with significant bacteriuria do not have pyuria. To ensure consistency, it has been recommended that variabilities in results for urine microscopy be avoided by standardizing the initital urine volume, resuspension volume, method of centrifuging and use of a hemocytometer or a slide with grid lines for counting.[9]

A test for leukocyte esterase is now commonly incorporated into diagnostic sticks. Leukocyte esterase is an enzyme found in neutrophil

granules. This reacts with an agent on the dipstick pad to give a color change. The test has a positive predictive value of 50% and a negative predictive value of 92%. In this method, the presence of pyuria from contamination cannot be differentiated from pyuria originating from the urinary tract. However, with microscopy, epithelial cells if seen, highly suggest contamination. There may be no obvious pyuria if urinary flow has been increased with increased fluid intake.

The nitrite test is the most common chemical test for bacteria. This tests for the presence of bacteria that can convert nitrate to nitrite. A positive test confirms the presence of bacteria while a negative one does not.

CLINICAL SYNDROMES

Acute Cystitis/Urethritis

Uncomplicated cystitis occurs frequently in women. Urinary frequency, dysuria and suprapubic discomfort are characteristic of cystitis although vaginitis has to be excluded. Gross hematuria occurs in up to one-third of patients with cystitis. This is related to the short urethra, proximity of the anus to the urethral opening and sexual activity. Nearly 90% of community acquired UTI are due to *E. coli* and other Gram-negative bacteria such as Proteus and Klebsiella. *Staphylococcus saprophyticus* account for about 10%.

It is thought that colonization of the vaginal introitus with fecal flora (predominantly *E. coli*) is the initiating step leading to colonization of the urethra and ascending infection to cause urethritis and cystitis and subsequent pyelonephritis. Sexual intercourse may assist migration.

Chlamydia, Neisseria, or herpes simplex infection in young sexually active women may cause the symptoms of acute cystitis. History of a new or multiple sex partners may point to a sexually transmitted disease. There may be accompanying vaginitis and vaginal discharge. Dysuria without frequency and urgency suggests vaginitis rather than cystitis in the presence of vaginal symptoms. Trichomonas and Candida are common causes of vaginal infection. These have to be treated accordingly.

Treatment of acute cystitis starts with advising the patient to increase her fluid intake. Ideally, a urine culture should be taken. This is especially indicated in repeated or relapsing infections or those unresponsive to treatment.

Usual first line therapies include: a 3 day course of bactrim 2 tabs (160 mg trimethoprim/800 mg sulphamethoxazole) bd or ciprofloxacin 250 mg bd, 7 day course of other antibiotics such as nitrofurantoin 50–100 mg tds, amoxycillin 500 mg tds, cephalexin 500 mg qds, cefuroxime 250 mg bd.[10] In principle, one should give the cheapest and least toxic drug first. Single dose therapy is acceptable treatment though less effective than multi-day regimens. The following have been tested: bactrim 4 tabs (320 mg trimethoprim/1.6 g sulphamethoxazole), trimethoprim 400 mg, amoxycillin 3 g, nitrofurantoin 200 mg, cephalexin 3 g, kanamycin 500 mg, ciprofloxacin 500 mg and cefuroxime 500 mg. It appears more effective in young women with cure rate of 90% rather than in women older than 40 years with cure rate of only 46%.[11] The disadvantage is that this dose will not be able to eradicate the vaginal reservoir that a 3–5 day course can. Patients who have recurrent attacks should probably be treated for at least 3 days and single dose therapy reserved for the isolated incidents. These are not suitable as treatment for complicated infections or where there is a relapse, neurogenic bladder, presence of calculi or obstruction, in males or in children. In women, it is usually not necessary to investigate the occasional cystitis.

Women who are still symptomatic after the initial course should have a urine culture performed. If positive, the treatment should be guided by antibiotic sensitivities. If negative and pyuria is present, chlamydia infection is suspected especially if the woman is sexually active and should be treated with doxycycline. If culture is negative and there is no pyuria, treatment with urinary analgesic would suffice.

Uncomplicated cystitis may occur in men, though much more uncommon than in women. Risk factors predisposing are thought to include intercourse with an infected female partner, homosexuality and lack of circumcision. In men, a minimum of a seven-day course is recommended with a fluoroquinolone, bactrim or trimethoprim so as to cover for occult prostatitis.[12] Urine culture is advisable before treatment. Nitrofurantoin should not be used as there is poor tissue penetration into the prostatic bed. Radiological assessment by intravenous urogram should be performed in men even after one infection so to exclude a structural lesion.

Recurrent infections

In women with recurrent cystitis, uropathogenic *E. coli* with increased adherence to the uroepithelium may be the cause. Non-secretors of blood

group H antigen produce other glycosylated lipids, which act as receptors for fimbria of uropathogenic *E. coli*. Patients with a positive P antigen status may be more prone to recurrent disease. Diaphragm and spermicide use alter vaginal flora by increasing the vaginal pH, leading to colonization with more pathogenic flora.

Women should be advised on perineal hygiene such as wiping from front to back to avoid fecal contamination of the periurethral area after micturation or passing motion. Sexually active couples should be advised to wash before intercourse as well as after and the lady to pass urine within 15 min after intercourse. There should be adequate lubrication either through the use of surgical lubricants or prolonged foreplay to avoid small tears and lacerations which encourage bacterial adherence. Post-coital prophylactic antibiotics (bactrim half a tablet, nitrofurantoin 50 mg, cephalexin 250 mg) appear most useful to patients who can associate the onset of symptoms with sexual intercourse. Prophylaxis is given if there are 2 or more infections per year. If there are no structural abnormalities, this should be given for at least 6 months. An intravenous urogram (IVU) should be performed with recurrent infections or a single episode of pyelonephritis to exclude a structural problem such vesicoureteric reflux where treatment of infection is crucial in preventing ongoing renal damage and obstruction. If present, prophylaxis should continue for 2 years from the last infection. Vaccination for prevention of recurrent infection has been studied but is not in widespread clinical use.[13,14] Cranberry products have been shown to reduce the recurrence of lower urinary tract infections by approximately 20%.[15] Recent research has demonstrated that proanthocyanidins found in cranberries have potent antiadhesion properties.[16]

In young men, recurrence suggests prostatitis or an upper tract abnormality and should be treated with a 4- to 6-week regimen of fluoroquinolone or trimethoprim-sulphamethoxazole combination.

Acute Pyelonephritis

The patient with acute pyelonephritis has systemic complaints such as fever, chills and rigors, nausea, vomiting, leukocytosis with pain localizing to the costovertebral angle. This is a more serious infection often associated with septicemia and shock. The infection is usually an ascending one. The bacterial spectrum is similar to that in acute cystitis. Unless there is significant vomiting, many patients can be treated as an

outpatient with oral antibiotics. Oral fluoroquinolones, trimethoprim-sulphamethoxazole combination and cephalosporins such as cefuroxime may be used. Hospitalization is usually required if there is significant systemic symptoms which require intravenous therapy.

Recommended first line intravenous antibiotics [10,17] include:

a) Ampicillin 500 mg 6-hourly with gentamicin 1 mg/kg body weight 8-hourly. This provides coverage for enterocci as well although ampicillin is no longer useful in *E. coli* infections due to resistance in $>45\%$ of isolates. Aminoglycosides continue to be useful agents because of good concentration in renal tissue and there are now recommendations for daily dosing;[18]
b) First or second generation cephalosporin with aminoglycoside;
c) Ceftriaxone 1 g/day is commonly used because of the convenience of once a day dosing; and
d) Ciprofloxacin 200 mg bd also achieve high concentrations in the kidney.

It is good practice to obtain blood cultures before commencing therapy as 12% of patients hospitalized for acute pyelonephritis may have bacteremia and it confirms the offending organism where urine cultures may be contaminated or negative. One expects to see improvement in symptoms within 48 hours. Otherwise, renal abscess and obstruction must be excluded by ultrasound scanning. A switch can be made to oral antibiotics when the patient has been afebrile for 24 hours if he/she has been started on intravenous therapy. The antibiotic chosen then can be guided by urine cultures. Treatment should be for 14 days though it is permissible to stop after 10 days in a mild illness. Some studies have advocated duration as short as 5 days with combinations such as gentamicin and ciprofloxacin. A follow-up intravenous urogram is advised.

Relapses

This implies continuing infection with the same organism within 6 weeks. They should be retreated. If there is another relapse then a 6-week course. Follow-up cultures should be performed to ensure eradication.

Prostatitis

In men, acute prostatitis or acute flares of chronic prostatitis is common and the patient may complain of perianal or rectal pain, pain referred to

Table 2 Evaluation of Prostatic Fluid

VB1	The first voided 10 ml of urine represents the urethral flora.
VB2	Midstream sample represents infection in bladder or above.
EPS	Specimen is collected after massage represents prostatic infection.
VB3	Urine collected after massage flushes out prostatic secretions in the urethra. Useful especially if there are no EPS.
To localize the infection to the prostate, the EPS or VB3 must have a colony count 10 times higher than VB1 and 2.	

VB = voided bladder; EPS = expressed prostatic secretions

the scrotum and upper thighs with accompanying dysuria and frequency. There may also be hemospermia.

E. coli is the most common organism. One should exclude prostatic calculi, benign prostatomegaly and cancer of the prostate and obstruction. In chronic prostatitis, there may be bacteriuria without symptoms. Prostatic massage is not done in acute prostatitis because of acute pain. It is recommended for confirmation if there is little prostatic symptoms. The patient should not have had antibiotic therapy for one month and have not ejaculated for 2 days.[19] The method of Meares and Stamey is used.[20] The significance of each of the specimens collected is listed in Table 2.

Acute prostatitis should be treated for 4 weeks to prevent chronicity. In the inflammed state, antibiotics can cross the prostatic bed easily. Ampicillin and gentamicin can be used initially before cultures are back. This should be followed up with bactrim, ciprofloxacin or erythromycin (or other drugs of the same group) depending on the culture result.

Chronic prostatitis should be treated for 4–12 weeks. If there is a relapse, then he should be retreated and prophylaxis instituted for 2 years. The Prostate Specific Antigen is often raised in prostatic infection and may be used for monitoring of successful treatment, especially if there are no symptoms.

Asymptomatic Bacteriuria

The significance of covert (asymptomatic) bacteriuria would depend on the circumstances in which it occurs. The prevalence is shown in Table 3.

The decision to treat or not to treat depends on the consequences or natural history.

Table 3 Prevalence of Covert Bacteriuria

Population	Prevalence (%)
Infants	1, Boys > Girls
School children (6–18 yrs)	
girls	1.2
boys	0.03, higher in the uncircumsised.
Sexually active women	3–10
Nuns (< 45 yrs)	0.6
Pregnant women	5–6
Elderly (> 65 yrs)	
men	10
women	20

The truly asymptomatic child has been the subject of much debate and children often cannot complain and present with non-specific symptoms such as fever, poor feeding and lethargy. In the presence of an abnormal urinary tract such as vesicoureteric reflux (VUR), therapy should be given as there is a risk of scarring and subsequent renal deterioration. Children less than 4 years of age should be treated as there is a high incidence of VUR. In the older child with a normal urinary tract, there is doubtful therapeutic benefit. For schoolgirls with covert bacteriuria, 30% clear spontaneously. Reinfections are common whether treated or not. The parents and child should be questioned closely for symptoms and if symptomatic, should be treated. Follow-up of growth, blood pressure, urine culture, serum creatinine and radiological investigation has been recommended for recurrent bacteriuria.

The truly asymptomatic non-pregnant woman is generally thought not to require treatment. This is either self-limiting or eradicated by antibiotics given for other conditions. These women should be questioned closely as to whether they really have symptoms. The pregnant patient will be discussed below.

COMPLICATED UTI

This refers to those categories of patients or associated conditions where failure of therapy because of relapse or persistence of infection is common. They need more attentive follow-up. The categories are listed in Table 4. Although *E. coli* is still the predominant, the percentage drops to around 20–54%.[21]

Table 4 Classification of Complicated UTI[21]

1) Structural abnormalities
   - Obstruction, diversion procedures, external drainage
   - Calculi
   - Infected cysts, Medullary sponge kidney
   - Vesicoureteric reflux, neurogenic bladder
   - Fistula
   - Renal abscess
2) Metabolic/Hormonal abnormalities
   - Diabetes mellitis
   - Chronic renal failure
   - Pregnancy
3) Impaired host responses
   - Neutropenia
   - Steroid therapy
   - Congenital or acquired immunodeficiency syndromes
4) Unusual pathogens
   - Yeast
   - Mycoplasma
   - Resistant bacteria

SPECIFIC PATIENT POPULATIONS

The Elderly

Bacteriuria without symptoms is common in the geriatric population. Recurrence rate of bacteriuria is high even when eradicated. Reinfection was common with increasingly resistent organisms. As the elderly are more susceptible to adverse effects of antibiotic therapy, treatment should be withheld unless symptomatic. In a study on ambulatory elderly women, mortality was not shown to be increased with treatment. However, if a patient is to undergo urinary tract instrumentation or surgery, the bacteriuria should be eradicated.

Symptomatic infections should always be treated. Elderly patients are more likely to have bacteremia and hypotension. Choice of drug therapy is as discussed above in younger populations. Care should be exercised in the use of aminoglycosides because of its associated nephro- and ototoxicity. Obstruction such as by prostatomegaly in men and malignant disorders in women should be excluded. Post menopausal women who have recurrent cystitis should consider the use of oral hormone replacement therapy or intravaginal estrogens.

Catheter-Related

Indwelling catheters are implicated in up to 70% of hospital-acquired UTI's. Prophylactic antimicrobial agents reduce risk of bacteriuria but resistant organisms appear. They can be used for short-term catheterization. Intermittant self-catheterization is now recommended for neurogenic bladders. In a local study in spinal cord injury, 78% were admitted to the rehabilitation unit with indwelling catheters. With adequate bladder training using voiding and intermittant self catheterization, indwelling catheters could be avoided in 97%.[22] Wherever possible, especially in the younger patient, indwelling catheters should not be used. If not possible, strict adherence to practices for proper bladder catheter care is necessary.[23] Bacteriuria is not treated unless the patient is symptomatic. Instead, a high fluid intake to promote urine flow, regular and complete bladder emptying should be encouraged.

Diabetics

The majority of UTI's in diabetics are asymptomatic. There is a higher incidence of bacteriuria in diabetic women. Once established, urinary tract infections are more severe with a high incidence of upper tract involvement and with fungal organisms. Poor control, the autonomic bladder and instrumentation are predisposing factors. When asymptomatic bacteriuria is found, it should be treated.

Renal Transplant Recipients

Early (less than 3 months post-transplant) cases of bacteriuria are usually related to the surgery and catheterization of the urinary tract. This is reduced with the use of prophylactic antibiotics such as half a tablet of bactrim, which is also used for prophylaxis of *Pneumocystis carinii* pneumonia. Urinary tract infection in the transplant patient in the early post transplant period is frequently asymptomatic and associated with overt pyelonephritis and bacteremia. Asymptomatic bacteriuria should be treated.

In the late post-transplant period, the incidence does not differ much from the general population. A search should be instituted for reflux in the graft urinary tract or other urological abnormality. Bacteriuria should be treated and prophylactic therapy instituted if reflux is present.

Treatment for symptomatic infection follows the same principles as previously discussed though the physician must be aware of the higher risks of fungal infections.

Pregnancy

Pregnancy itself does not lead to an increase in bacteriuria but it permits urinary colonization established prior to pregnancy to lead to symptomatic infection. Physiological changes of pregnancy leads to dilatation of the renal tract, reduction in urinary peristalsis and urinary stasis. The studies of Kass and others demonstrate that 20–40% of women with bacteriuria detected early in pregnancy will develop acute symptomatic infection later in pregnancy, and only 1–2% without bacteriuria does so. The rate of prematurity increases to 20–50% in symptomatic urinary tract infection (UTI). Other increased risks reported are that of intrauterine growth retardation, low birth weight and fetal mortality. Treating pregnant bacteriuric patients lowers the incidence of symptomatic UTI by 80–90%.[24] Routine screening has therefore been recommended for bacteriuria in pregnancy. It has been recommended that the best time to start screening for bacteriuria is in the 16th gestational week. This is because the highest rates were detected between the 9th to 17th weeks and the best time to obtain the highest number of bacteriuria-free gestational weeks best yield appears to be at 16 weeks.[25] Antibiotics safe to use in pregnancy include amoxycillin, nitrofurantoin, cephalexin. Co-trimoxazole is to be used with caution because of possible terogenecity in early pregnancy and hyperbilirubinemia in the neonate when given in the third trimester. Many still recommend 7-day course with the above drugs with the usual doses as first-line therapy. Failures to single dose therapy must be treated with a conventional course. When the infection is eradicated, monthly follow-up cultures are recommended till term.

Fungal Cystitis

This is common with catheterized patients, diabetic or immunosuppressed. If confined to the lower tract, they usually respond to intravesical instilation of Amphoteracin 15 mg in distilled water after emptying the bladder. If the patient is able to cooperate, the catheter may be removed and the patient instructed not to pass urine for 2 hours.

Otherwise, keep the catheter clamped and released only after 2 hours. A repeat culture should be taken a week later.

SUMMARY

The treatment of urinary tract infection varies depending on the clinical scenario. Not all cases of symptomatic bacteriuria need to be treated. However where the natural history leads to upper tract involvement and bacteremia or in the presence of an abnormal urinary tract, treatment should be started. Antibiotic susceptibility patterns change with time and first-line treatment must take into consideration this factor as well as cost of hospitalization, administration of drug therapy as well as subsequent relapse and retreatment.

REFERENCES

1. Hooton T, Stamm W, Diagnosis and treatment of uncomplicated urinary tract infection, *Infect Dis Clin North Am* **11**(3):551–581, 1997.
2. Antibiogram of urinary isolates, in: *Microbiology Lab Singapore General Hospital*; 1998.
3. Tan HK, Choong HL, Woo KT, Admissions for urinary tract infection to Singapore General Hospital, in: *Chapter of Physicians' 16th Annual Combined Scientific Meeting*; 1996.
4. Whitworth JA, Current concepts in the management of urinary tract infection, *Ann Acad Med Singapore* **11**(1):64–68, 1982.
5. Byungse S, Narins RG, Kunin C, Urinary tract infection — important diagnostic procedures, in: *Diagnostic Techniques in Renal Disease,* Churchill Livingstone, New York, pp. 346–347, 1992.
6. Kass EH, Finland M, Asymptomatic infections of the urinary tract, *Trans Assoc Am Physicians* **69**:56–63, 1956.
7. Little P, Urinary white cell excretion, *Lancet* **1**:1149, 1962.
8. Stamm WE, Measurement of pyuria and its relation to bacteriuria, *Am J Med* **75**(Suppl 1B):53–58, 1983.
9. Pappas PG, Laboratory in the diagnosis and management of urinary tract infections, *Med Clin North Am* **75**(2):313–325, 1991.
10. Use of Antibiotics in Urinary Tract Infection, in: *Use of Antibiotics in Adults, Singapore,* pp. 46–64, 2000.
11. Cardenas J, Quinn EL, Rooker G, Bavinger J, Pohlod D, Single-dose cephalexin therapy for acute bacterial urinary tract infections and acute urethral syndrome with bladder bacteriuria, *Antimicrob Agents Chemother* **29**(3):383–385, 1986.

12. Stamm WE, Hooton TM, Management of urinary tract infections in adults, *N Engl J Med* **329**(18):1328–1334, 1993.

13. Nayir A, Emre S, Sirin A, Bulut A, Alpay H, Tanman F, The effects of vaccination with inactivated uropathogenic bacteria in recurrent urinary tract infections of children, *Vaccine* **13**(11):987–990, 1995.

14. Schmidhammer S, Ramoner R, Holtl L, Bartsch G, Thurnher M, Zelle–Rieser C, An *Escherichia coli*-based oral vaccine against urinary tract infections potently activates human dendritic cells, *Urology* **60**(3):521–526, 2002.

15. Stothers L, A randomized trial to evaluate effectiveness and cost effectiveness of naturopathic cranberry products as prophylaxis against urinary tract infection in women, *Can J Urol* **9**(3): 1558–1562, 2002.

16. Howell AB, Cranberry proanthocyanidins and the maintenance of urinary tract health, *Crit Rev Food Sci Nutr* **42**(3 Suppl):273–278, 2002.

17. Gilbert DN, Moellering RC, Sande MA (eds.), *The Sanford Guide to Antimicrobial Therapy*, Antimicrobial Therapy Inc., Hyde Park, Vermont, USA, 1999.

18. Freeman CD, Nicolau DP, Belliveau PP, Nightingale CH, Once-daily dosing of aminoglycosides: review and recommendations for clinical practice, *J Antimicrob Chemother* **39**(6):677–686, 1997.

19. National Guideline for the Management of Prostatitis, in: *London: Association for Genitourinary Medicine (AGUM), Medical Society for the Study of Venereal Disease (MSSVD)*, 2002.

20. Meares EM, Stamey TA, Bacteriologic localization patterns in bacterial prostatitis and urethritis, *Invest Urol* **5**(5):492–518, 1968.

21. Nicolle L, A practical guide to the management of complicated urinary tract infection, *Drugs* **53**(4):583–592, 1997.

22. Menon E, Tan E, Urinary tract infection in acute spinal cord injury, *Singapore Med J* **33**(4):359–361, 1992.

23. Rubin RH, Tolkof–Rubin NE, Cotran RS, Urinary Tract Infection, pyelonephritis and reflux nephropathy, in: Brenner BM, Rector FC (eds.), *The Kidney: Pathogenesis of Renal Disease*, 4th ed. WB Saunders Co., Philadelphia, pp. 1369–1429, 1991.

24. Andriole VT, Patterson TF, Epidemiology, natural history, and management of urinary tract infections in pregnancy, *Med Clin North Am* **75**(2):359–373, 1991.

25. Patterson TF, Andriole VT, Detection, significance, and therapy of bacteriuria in pregnancy. Update in the managed health care era, *Infect Dis Clin North Am* **11**(3): 593–608, 1997.

71

Renal Hypertension

Wong Kok Seng

INTRODUCTION

Renal hypertension refers to hypertension that is secondary to an underlying renal disease. This may be due to a renal parenchymal disease or renal vascular disease. However it is sometimes difficult to discern if renal parenchymal disease is the cause of the hypertension or if it co-exists with essential hypertension. It is also important to realize that renal artery stenosis does not equate renal vascular hypertension as one may have renal artery stenosis without hypertension. Therefore the co-existence of hypertension and renal disease does not prove that the hypertension is caused by renal disease.

Primary renal disease is uncommon in patients with hypertension (probably around 4%) but hypertension is common in all types of renal disease and becomes more prevalent as the renal function deteriorates. About 60% of patients with glomerular disease have hypertension and about 33% of patients with chronic pyelonephritis have hypertension at presentation.

RENAL HYPERTENSION — DETECTION AND ASSESSMENT

The presence of proteinuria in a patient with hypertension provides an early clue to the possibility of an underlying renal disease. In these patients, adequate investigation and effective treatment would retard the progression to end stage renal failure. Onset of hypertension in the elderly would suggest the possibility of renal vascular hypertension and this is part of the atherosclerotic disease that also affects the peripheral, coronary and cerebral vasculature.

Clinical History

Table 1 lists the clues in the clinical history that suggest an underlying renal disease. These may be useful in selecting patients for further investigation. Onset of hypertension before age 20 years suggests the possibility of chronic pyelonephritis as it is the most common cause of hypertension in such young patients. A family history of renal disease may indicate the need to exclude the possibility of adult polycystic kidney disease, Alport's syndrome or other hereditary renal disease. Accelerated or resistant hypertension should rouse the possibility of renal vascular hypertension. Worsening of renal function after therapy with angiotensin converting enzyme inhibitors would suggest that severe renal vascular disease is present. In fact, one should consider the possibility of occult renal vascular disease when therapy with angiotensin converting enzyme

Table 1 Clues in the Clinical History Suggesting Renal Hypertension

1) Onset of hypertension before age 30 years but without a family history of hypertension
2) Recent onset of significant hypertension after age 50 years
3) Known history or family history of renal disease
4) Symptoms of underlying renal disease, e.g. polyuria, nocturia, dysuria, gross hematuria
5) Accelerated hypertension
6) Resistant hypertension
7) Symptoms of vascular disease (cardiac, cerebral, peripheral)
8) Worsening of renal function after therapy with angiotensin converting enzyme inhibitors
9) Recurrent (flash) pulmonary edema

inhibitors gives rise to a significant rise in serum creatinine (greater than 50 μmol/L) within 7 to 14 days of commencing treatment.

Physical Examination

A full physical examination is essential and the fundi, heart and peripheral pulses need to be examined in order to assess the severity of cardiovascular complications. The cardiovascular system is examined, particularly for heart size and for evidence of arterial disease in the carotid, renal and peripheral vasculature. Examination of the optic fundi will revealed the degree of microvascular disease.

Table 2 shows the clinical clues that suggest the presence of renal hypertension. Large palpable kidneys suggest the presence of polycystic kidneys disease. Features of systemic disease, such as diabetes or systemic lupus erythematosus, may be obvious.

An abdominal or epigastric systolic bruit is relatively common in elderly patients and has a low predictive value for renal vascular disease. On the other hand, the presence of the continuous or systolic-diastolic abdominal bruit would strongly indicate renal vascular hypertension.

Initial Investigations

Routine investigations in a patient with hypertension include urinalysis for blood or protein as well as a microscopic examination of the urine. Blood chemistry will include measurements of potassium, urea, creatinine, fasting glucose and total cholesterol. If urinary abnormalities or elevation of urea and creatinine are noted, a 24-hour urine collection for measurement of proteinuria and creatinine clearance should be performed. An ECG should also be performed.

Table 2 Findings on Physical Examination Suggesting Renal Hypertension

1) Edema (peripheral or pulmonary)
2) Palpable kidneys (e.g. polycystic kidney disease)
3) Signs of uremia
4) Abdominal bruits
5) Features of vascular disease (e.g. carotid or femoral bruits, reduced or absent peripheral pulses)
6) Severe hypertensive retinopathy (grade 3 or 4)

When the clinical history, physical examination and initial investigations suggest the possibility of renal hypertension, further investigations must be carried out to confirm or rule out this possibility as well as to reach a definitive diagnosis.

Specialized Investigations

Renal ultrasound

This is non-invasive and readily performed and is the best method for the diagnosis of polycystic kidney disease. It will indicate if the kidneys are reduced in size or have increased echogenicity and thereby suggesting chronic renal parenchymal disease. It will show if the kidneys are of unequal length. If the inequality is greater than 1.5 cm and is accompanied by other indicators of renal vascular disease, further investigations for renal vascular disease are indicated.

Duplex ultrasound

Color-coded duplex ultrasonography is another non-invasive method to obtain blood flow profiles in the main renal arteries and intrarenal vessels. It allows a direct visualization of the renal vasculature while assessing blood flow velocity and pressure waveforms. Technical improvements in the current machines have practically overcome problems due to obesity and bowel gas. This technique can be used to screen for renal artery stenosis and allows for non-invasive follow-up of patients after interventions, even after placement of stents.

Intravenous urogram

The rapid sequence intravenous urogram has been the traditional test to confirm or exclude renal vascular disease. However, it carries a 15% false-positive and 25% false-negative rates, and its use is now increasingly replaced by isotope renography or intravenous digital subtraction angiography. Nevertheless the intravenous urogram is very useful in the diagnosis of chronic pyelonephritis with renal scarring and this is the most common cause of hypertension in patients aged below 20 years.

Table 3 Criteria for the Captopril Challenge Test

Method:
1) The patient should maintain a normal salt intake and receive no diuretics.
2) If possible, all antihypertensive medications should be withdrawn 3 weeks prior to the test.
3) The patient should be seated for at least 30 minutes; a venous blood sample is then drawn for measurement of baseline plasma renin activity.
4) Captopril (50 mg diluted in 10 mL of water immediately before the test) is administered orally.
5) At 60 minutes, a venous blood sample is drawn for measurement of stimulated plasma renin activity.

Interpretation. A positive test requires:
1) Stimulated plasma renin activity of 12 ng/mL/hr or more and;
2) Absolute increase in plasma renin activity of 10 ng/mL/hr or more and;
3) Increase in plasma renin activity of 150% or more, or 400% or more if baseline plasma renin activity is less than 3 ng/mL/hr.

From Muller FB, Sealey JE, Case DB *et al. Am J Med* **80**:633, 1986.

Captopril challenge test

This is based on the fact that peripheral plasma renin activity is elevated in most patients with functionally significant renal vascular disease and is further augmented by blocking of the renin-angiotensin system. The test is simple to administer and involved the measurements of plasma renin activity before and sixty minutes after oral administration of 50 mg captopril. The criteria for the performance and interpretation of the captopril challenge test is as listed in Table 3.

Isotope renogram

This is an alternative and widely used method for the diagnosis or renovascular hypertension. The original technique of using ^{131}I-hippuran to compare renal blood flow in affected and unaffected kidneys gave a high incidence of false positives and has fallen into disfavor. The captopril-enhanced renogram is preferred for identification of functionally significant renal artery stenosis. Measuring 99Tcm-diethylenetriaminepenta acetic acid (DTPA) uptake and/or ^{131}I-hippuran uptake before and 60 minutes perform this after a 25 mg oral dose of captopril. When functionally significant renal artery stenosis is present, the reduced DTPA

uptake indicates a fall in glomerular filtration rate while the delayed hippurate secretion indicates a prolongation of the mean parenchymal transit time. Using a variety of diagnostic criteria, the captopril-enhanced renogram is used to identify functionally significant renovascular hypertension with a high degree of specificity and sensitivity.

Renal angiography

This is the "gold standard" for the identification and localization of renal arterial lesions. It is invasive and carries the risk of radiocontrast-induced acute renal failure. Increasingly, this technique is used for percutaneous transluminal angioplasty and should not be used only to confirm a suspected renal artery stenosis.

Spiral CT angiography

This technique utilizes the continuously overlapping transaxial images obtained by means of a rotating X-Ray tube, with the scanning time within a single-breath hold of the patient. A three-dimensional reconstruction of the renal vasculature is obtained. This allows for visualization of the whole vascular tree with sensitivity and specificity varying from 90–99%. However, the amount of contrast medium used is high and thereby limits its use in patients with pre-existing renal insufficiency.

Magnetic resonance angiography

This technique is non-invasive and allows for direct visualization of proximal renal artery lesions. It can be used in conjunction with a gadolinium-based contrast and does not carries the risk of nephrotoxicity. Two different imaging methods can be used to diagnose renal artery stenosis with a sensitivity and specificity of 90–100%.

Renal vein renin ratio

This is an invasive technique involving bilateral catherization and sampling from each renal vein as well as the adjacent vena cava above and below the origin of the renal vein. With the advent of non-invasive tests such as those using captopril stimulation of the renin system and improved methods of imaging the renal vasculature, sampling of renal vein renin has fallen into disfavor.

Renal biopsy

This is an invasive procedure and is undertaken in patients with hypertension and suspected renal disease. These patients will usually have proteinuria and the renal function may be impaired. The renal biopsy is also useful in determining whether renal impairment in patients with accelerated hypertension is a consequence of renal damage from hypertension itself, or whether an underlying primary renal disease gives rise to the accelerated hypertension.

HYPERTENSION IN RENAL PARENCHYMAL DISEASE

Mechanisms of Hypertension

Several mechanisms are believed to account for hypertension in renal parenchymal disease. Decreased renal sodium and water excretion leading to extracellular volume expansion is of major importance. Several lines of evidence support this. For example, administration of sodium chloride expands the intravascular space and aggravates the blood pressure in patients with renal failure; and diuretics are effective in the blood pressure in patients with renal failure. At the other end of the spectrum, end stage renal failure patients who receive aggressive ultrafiltration at dialysis will lower their blood pressure towards normal in most instances.

The renin-angiotensin system is also activated especially in patients with renal insufficiency. Increased sympathetic tone also plays a role as the norepinephrine levels and the rate of sympathetic nerve discharge is raised in such patients.

Hyperfiltration Mechanism of Renal Injury

Glomerular hypertension (Fig. 1) in the most likely mechanism by which hypertension leads to progressive renal damage. Systemic hypertension is one of these factors that increase glomerular capillary plasma flow rate or hydraulic pressure. The high systemic pressure is transmitted into the glomerulus and damages glomerular cells and leads to progressive sclerosis. This in turn sets off a vicious cycle by aggravating systemic hypertension, which leads to further glomerular sclerosis.

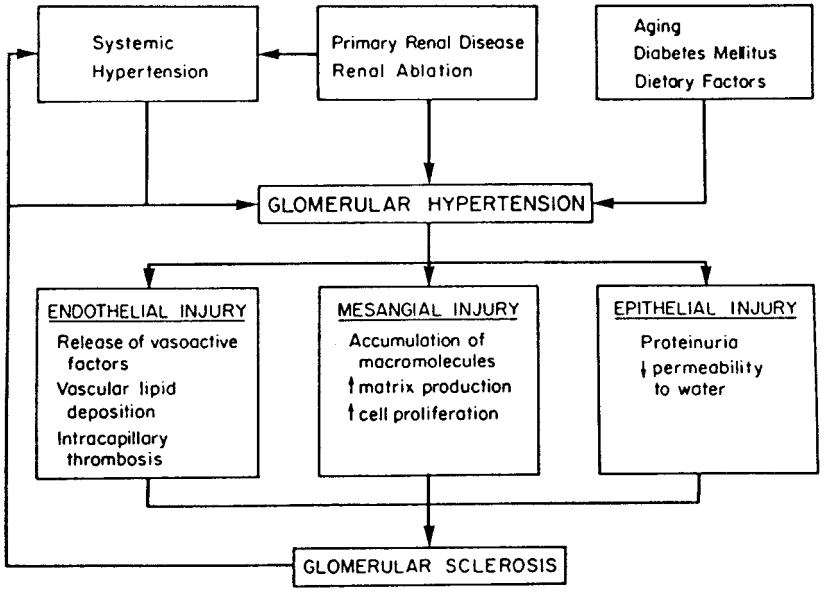


Fig. 1 Pivotal role of glomerular hypertension in the initiation and progression of structural injury.
(Anderson S, Brenner BM; *Q J Med* **70**:185, 1989)

Treatment

Lifestyle measures

These measures include smoking cessation, weight reduction, exercise, and dietary salt and saturated fat restriction. While there is not direct evidence to demonstrate that these measures reduce the overall risk of cardiovascular disease, they are likely to be beneficial.

The importance of dietary salt restriction in patients with hypertension and renal disease cannot be overemphasised. Although in itself unlikely to obviate the need for anti-hypertensive therapy, salt restriction potentiates the blood pressure-lowering effects of many anti-hypertensive agents, especially the ACE-inhibitors.

Drug therapy

The authors of the sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure recommended that therapy to reduce blood pressure be instituted in all

those with diastolic blood pressure above 90 mm Hg diastolic or systolic blood pressure above 140 mm Hg systolic. This recommendation is applicable to patients with renal disease.

Blood pressure should be controlled to 130/85 mm Hg with whatever anti-hypertensive therapy that is necessary. In patients with proteinuria greater than 1 g/day, the target blood pressure should be lower than 125/75 mm Hg (Fig. 2).

ACE-inhibitors have produced impressive results in patients with in patients with renal insufficiency as well as in patients with proteinuria greater than 1 g/day. Unless contraindicated, they should receive an ACE-inhibitor to control the hypertension as well as to slow the progression of renal failure. In most cases, a diuretic is also administered for synergistic effects. However, ACE-inhibitors should be used with caution in patients with serum creatinine above 265 μmol/L or 3 mg/dL.

In the first 3 months of therapy with an ACE-inhibitor, there may be a transient decrease in glomerular filtration rate. If the serum creatinine rises 88 μmol/L or 1 mg/dL above baseline levels, the creatinine and potassium levels should be re-checked after a few days. If the patient is

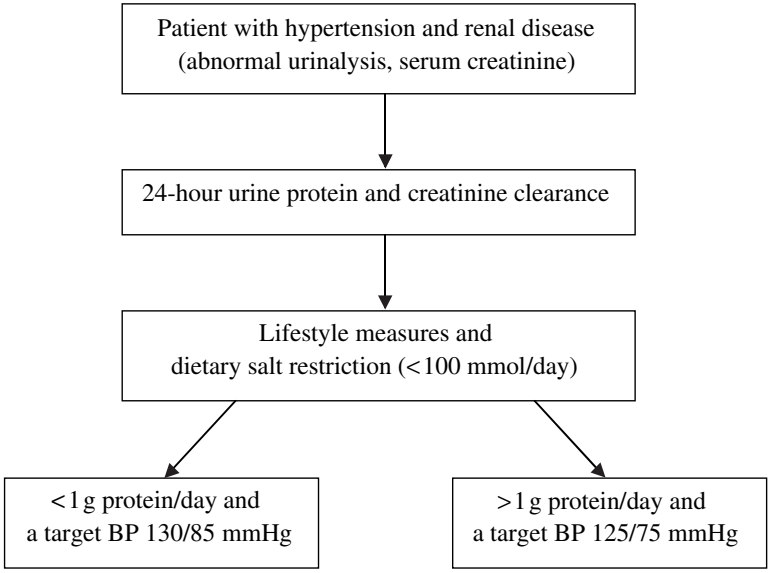


Fig. 2 Treatment of patients with hypertension and renal disease.

euvolemic and the serum creatinine remained persistently high, we need to consider the possibility of renal artery stenosis and the ACE-inhibitors need to be discontinued.

The most important action is to slow progression of renal failure is to lower the blood pressure to target level. All classes of anti-hypertensive drugs are effective and in many cases, several anti-hypertensive drugs may be needed to lower blood pressure to target level.

RENOVASCULAR HYPERTENSION

Renovascular hypertension can be treated in one of 3 ways — medical, surgical, and angioplasty. In the younger patient with renal artery stenosis due to fibromuscular dysplasia, this is amenable to angioplasty with or without surgery. In the older patient with stenosis that is atheromatous in origin, angioplasty and/or surgery may improve but not cure the hypertension, and medical treatment needs to be continued. Some patients are deemed to be poor candidates for revascularization and have to remain on medical treatment.

Medical Treatment

The main indications for medical treatment are listed below:

1) advanced age;
2) poor surgical risk;
3) technical difficulties in performing renal revascularization or angioplasty;
4) irreversible atrophy of kidney distal to the stenosis;
5) doubtful significance of the lesion of renal artery stenosis;
6) hypertension of long duration suggesting pre-existing essential hypertension; or
7) patient's choice.

The aim is to lower blood pressure so as to delay the progression of renal failure. Sodium restriction to less than 100 mmol per day is recommended as sodium retention and concomitant volume expansion is an important component in the hypertension of such patients.

Calcium antagonists are quite effective in the treatment of patients with renovascular hypertension as they have direct arteriolar vasodilatory effects. Another benefit is the favorable effect on renal function in

comparison to ACE-inhibitors. Beta-blockers are also useful, due in part to their ability to inhibit renin secretion — most patients with renovascular hypertension have high renin levels.

Invasive Treatment

The potential advantages of invasive treatment of renovascular hypertension include cure for the hypertension, or improved control of high blood pressure; improvement and preservation of renal function. However, it carries the potential disadvantage of rare but immediate and life-threatening complications.

Percutaneous transluminal renal angioplasty

PTRA has become a common invasive therapy for renovascular hypertension. It has become popular for the following reasons — low cost, short hospitalization, low mortality and morbidity, readily available in most hospitals, and can be used in patients who are poor surgical risks. The success of PTRA is judged by its ability to normalize blood pressure or at least improve control of hypertension, improve renal function, and to widen the stenosis.

Renal artery stenting

Poor technical results from PTRA have led to the current interest in renal artery stent placement. Stents are composed of metallic wire with the unexpanded stent mounted on the uninflated balloon of an angioplasty catheter. Once positioned across the lesion, the balloon is inflated and the stent is expanded and deployed. Over several weeks, endothelization occurs with the renal stent covered with intima.

Surgical treatment

Surgical treatment for renovascular hypertension includes nephrectomy, aortorenal bypass and other techniques for the surgically difficult aorta. Nephrectomy or partial nephrectomy is rarely required as the preservation of renal function is an important goal in the treatment of such patients. The benefits of surgery need to be appraised with regards to its effect on blood pressure control, improved renal function and better patient survival.

Recently, more centers are performing surgical revascularization to preserve renal function in patients with severe atherosclerotic arterial occlusive disease. These patients are at high risk because of age and associated coronary, cerebrovascular, or peripheral vascular disease.

CHOICE OF THERAPY

It has become more common to treat patients with atherosclerotic renal artery stenosis by invasive therapy. Various randomized trials that compared invasive with conservative therapy uniformly demonstrated that blood pressure control and renal function are not very different whether an invasive or a conservative approach is taken as the first step in the management of a patient with renal artery stenosis. Nevertheless, it would be prudent to restrict balloon angioplasty to those whose hypertension persists despite treatment with three or more drugs, or who have progressive occlusive renovascular disease as indicated by an increase in the serum creatinine.

HYPERTENSION IN DIALYSIS PATIENTS

The prevalence of hypertension increases as the renal function declines. About 80–90% of patients have significant hypertension by the time they reach end-stage renal failure. The pathogenesis of hypertension in dialysis patients is multifactorial. Many of these patients have volume-dependent hypertension, especially in those with large interdialytic weight gain. Salt ingestion plays an important role as it gives rise to thirst and increase fluid intake. The renin-angiotensin system is also important in the pathogenesis of hypertension in some patients with end-stage renal failure. In these patients, the blood pressure does not decrease but actually increases during dialysis.

The management is directed towards correcting volume overload in patients with volume-dependent hypertension. Body weight should be reduced by ultrafiltration during dialysis to dry weight. Dry weight is defined as the weight below which further fluid removal would produce hypotension. This approach controls hypertension in a large number of patients. If antihypertensive medication is required, it should be tailored to the patient's need and associated medical conditions. The ACE-inhibitors

are useful as non-volume dependent hypertension in these patients is renin-mediated.

HYPERTENSION IN RENAL TRANSPLANT PATIENTS

Post-transplant hypertension is common and is a significant risk factor for accelerated atherosclerosis and premature coronary disease. It increases the risk of allograft failure and aggravates the deterioration of allograft function. The etiology of post-transplant hypertension is multifactorial. In the immediate post-transplant period, it may be due to delayed graft function or the onset of acute rejection. In the late post-transplant period, other causes of hypertension such as cyclosporine toxicity, recurrent disease, chronic rejection or transplant renal artery stenosis should be considered.

Apart from transplant renal artery stenosis, the management of post-transplant hypertension is largely medical. Calcium antagonists are the most commonly used antihypertensive drugs in these patients. They reduce blood pressure effectively and attenuate both cyclosporine- and endothelin-induced vasoconstriction. Most clinicians prefer to use calcium antagonists of the dihydropyridine class as they are least likely to affect the hepatic P_{450} microsomal system. Therefore adjustment of the dose to control blood pressure is least likely to affect cyclosporine blood levels. Other clinicians, however, utilize calcium antagonists of the non-dihydropyridine class to reduce cyclosporine dosage needed to achieve immunosuppressive blood levels while controlling blood pressure. This approach mandates the monitoring of cyclosporine levels whenever the dose of these non-dihydropyridine blockers is altered.

ACE-inhibitors have been shown to slow the progression of chronic renal failure but the value of ACE-inhibitors for the treatment of hypertension in renal allograft recipients has not been established. The administration of ACE-inhibitors after renal transplant requires caution and close monitoring is required for adverse effects such as acute renal failure, hyperkalemia and anemia. Nevertheless, most transplant recipients tolerate ACE-inhibitors rather well and such therapy may reduce proteinuria and preserve renal allograft function.

Angiotensin II receptor blockers with their potent antihypertensive effect, good safety and tolerance profile and beneficial glomerular effects, could also be potentially useful in transplant patients.

REFERENCES

1. Raine AEG, Ledingham JGG, Secondary hypertension, in: Weatherall DJ, Ledingham JGG, Warrell DA (eds.), *Oxford Textbook of Medicine*, Oxford University Press, pp. 2544–2553, 1996.

2. Mann J, Allenberg JR, Christine Reisch, Dietz R, Weber M, Luft FC, Renovascular hypertension, in: Cameron S, Davison AM, Grunfeld J, Kerr D, Ritz E (eds.), *Oxford Textbook of Clinical Nephrology*, Oxford University Press, pp. 2096–2117, 1992.

3. Smith MC, Dunn MJ, Hypertension associated with Renal Parenchymal Disease, in: Schrier RW, Gottschalk CW (eds.), *Diseases of the Kidney*, Little, Brown and Company, pp. 1333–1365, 1997.

4. van Jaarsfeld BC, Krijnen P, Pieterman H *et al*., The effect of balloon angioplasty on hypertension in atherosclerotic renal artery stenosis, *N Engl J Med* **342**:1007–1014, 2000.

72

Diabetic Nephropathy

Grace Lee and Woo Keng Thye

Diabetic nephropathy is the leading cause of end-stage renal disease (ESRD) in many countries in the world including the United States, most of Europe, Japan, New Zealand, Malaysia and Singapore. In Singapore, diabetic nephropathy accounted for 50% of the new patients starting on dialysis in 2001, with the majority having Type 2 diabetes (91%). This represented a 44.7% increase over the year 1997, where 34.2% of the new ESRD patients had diabetic nephropathy.

The observed increase in the incidence of diabetic nephropathy is related to several factors. Firstly, there has been an increase in the incidence of diabetes mellitus and in Singapore, the National Health Surveys conducted in 1992 and again in 1998 showed the prevalence of diabetes increased from 8.6% to 9.0%. The increased prevalence is probably a result of changes in lifestyle with a greater tendency towards more obesity and also aging of the population. Secondly, treatment of hypertension and coronary artery disease has improved allowing the diabetic patients (especially the Type 2 diabetics) to live long enough to develop nephropathy. Hence, as our population ages, we can expect to see more elderly patients with diabetic nephropathy and cardiovascular disease requiring dialysis.

Once initiated on dialysis, the survival of the diabetic patient is dismal, with less than 20% surviving for more than 5 years. Therefore, it becomes imperative that we find treatment that can effectively delay or prevent the progression of renal disease in diabetes.

DIAGNOSIS

Natural History

The epidemiology for Type 1 diabetes has been more clearly studied that Type 2 diabetes since the time of clinical onset is usually known. Between 25–45% of patients will develop diabetic nephropathy and the peak onset is between 10–15 years after the development of the disease. Patients who do not have proteinuria after 20–25 years have a reduced risk of developing overt renal disease of about 1% per year. Patients with Type 2 diabetes were traditionally thought to have a lower risk of developing diabetic nephropathy but recent data suggest that the renal risk is similar in both Type 1 and 2 disease.

Clinical Features

The mode of presentation depends on the stage of the renal disease. Mogensen described five stages of progressive diabetic nephropathy in Type 1 diabetes (Table 1) and patients with Type 2 diabetes have been

Table 1 Stages of Nephropathy in Type 1 Diabetes

Stage	Clinical Term (Onset)	Clinical Features
I	Initial stage (at diagnosis)	Transient albuminuria related to hyperglycemia
II	Early renal involvement (1.5–5 years)	Normoalbuminuria
III	Incipient nephropathy (5–15 years)	Persistent microalbuminuria Hypertension (±)
IV	Overt nephropathy (10–20 years)	1) Macroalbuminuria 2) Hypertension 3) Declining renal function Edema, ↓ serum albumin, associated complications e.g. retinopathy, neuropathy, cardiac and vascular disease
V	End-stage renal disease (20+ years)	Dialysis required

shown to follow a similar course. *Stage I* occurs at clinical onset of the diabetes and is associated with glomerular hyperfiltration, renal hypertrophy and albuminuria due to the uncontrolled hyperglycemia. When the diabetes comes under metabolic control, the patient enters *Stage II* where the transient albuminuria disappears and the hyperfiltration and hypertrophy resolve. After 5 to 7 years, 25–45% of the patients will enter *Stage III* or *incipient diabetic nephropathy* with the reappearance of persistent microalbuminuria (Table 2). Without intervention, the patient progresses to *Stage IV* or *overt diabetic nephropathy* with the classical diagnostic triad of macroalbuminuria (urine dipstick positive for protein), hypertension and declining renal function. *Stage V* defines end-stage renal failure.

In most instances, patients with Type 1 diabetes are usually asymptomatic and the presence of renal disease is detected through routine monitoring of the urine for albumin. In contrast, patients with Type 2 diabetes frequently have renal disease in the form of macroalbuminuria and/or renal impairment or chronic renal failure at the time of initial diagnosis. Most patients with overt diabetic nephropathy have the nephrotic syndrome with odema, hypoalbuminemia and heavy proteinuria and this is accompanied with hypertension that is often difficult to control.

There are several differences between Type 1 and 2 diabetes. While microalbuminuria is predictive of progressive renal disease in 80% of patients with Type 1 diabetes, only 20% of patients with microalbuminuria develop progressive renal disease in Type 2 diabetes. This is probably because the microalbuminuria in Type 2 diabetes is more reflective of a generalized vascular disease rather than specific renal disease, and

Table 2 Definition of Albuminuria

Type	Albumin/Creatinine Ratio (μg/mg)	24-hour Collection (mg/24 h)
Normoalbuminuria	<30	<30
Microalbuminuria (Dipstick negative)	30–300	30–300
Macroalbuminuria (Dipstick positive)	>300	>300

Measurements should not be performed in the following situations which may temporarily increase urinary albumin excretion: after exercise or excessive protein intake, congestive cardiac failure, hematuria, uncontrolled hypertension or diabetes or urinary tract infection.

explains the observed increased cardiovascular risk in patients with microalbuminuria. Hypertension is also more prevalent in Type 2 diabetes where 70% of the patients will have hypertension during the incipient stage compared to only 25% of patients with Type 1 diabetes. The presence of retinopathy is more strongly associated with nephropathy in Type 1 diabetes with about 90% of patients having retinopathy compared to 55–65% of patients with Type 2 diabetes.

Investigations

The investigations are targeted at: 1) confirming the presence of albuminuria (with a 24-hour urine albumin excretion), 2) the assessment of renal function (serum creatinine and creatinine clearance); and 3) the exclusion of other non-diabetic renal disease.

Even if the diagnosis of diabetic nephropathy is clinically obvious, it is prudent to exclude the following non-diabetic renal diseases: 1) multiple myeloma (urine for Bence–Jones protein, Erythrocyte sedimentation rate, total protein and albumin); 2) collagen vascular disease (Antinuclear antibody), and 3) renal stone disease and other obstructive uropathy (ultrasound examination of the kidneys). In most cases, the diagnosis of diabetic nephropathy is made on clinical features without the need for a renal biopsy. However, when features suggestive of a non-diabetic renal disease are present (Table 3) the patient would require referral to the nephrologist for further evaluation, which could include a renal biopsy.

Table 3 Features Suggesting Non-diabetic Renal Disease

Proteinuria without retinopathy
Nephropathy within 5 years of diagnosis in type 1 diabetes
Renal failure without significant proteinuria
Gross or microscopic hematuria
Rapid decline in renal function
Sudden onset of nephrotic syndrome

SCREENING FOR RENAL DISEASE

Diabetic patients with microalbuminuria are at risk of developing progressive renal disease and screening (Fig. 1) identifies those who will benefit from early intervention. Screening of urine albumin excretion (UAE) should be performed annually in all Type 2 diabetics and also

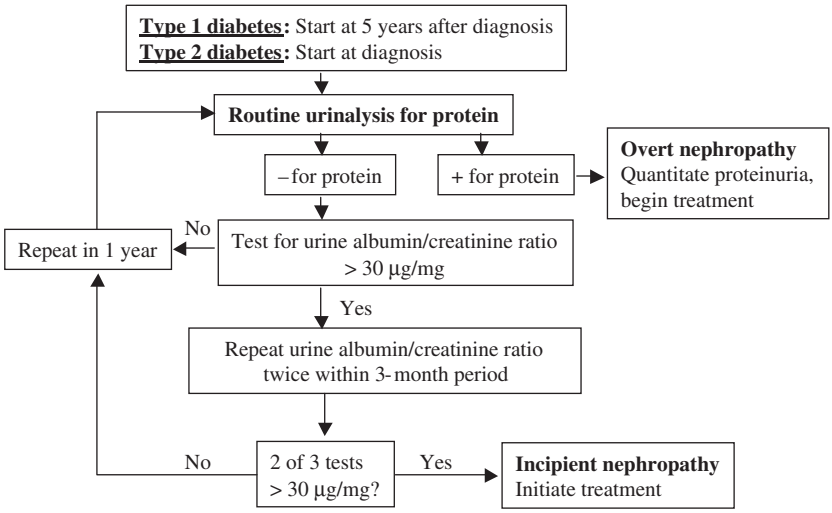


Fig. 1 Annual screening for microalbuminuria in diabetes mellitus.

annually after 5 years of diagnosis and in all Type 1 diabetics. It is important when measuring the UAE to ensure that the patients do not have the conditions listed in Table 2 that may invalidate the test.

TREATMENT

Retarding the progression of renal disease in diabetes requires a multipronged approach comprising of: 1) excellent glycemic control; 2) the use of angiotensin converting enzyme inhibitors (ACE inhibitors) or angiotensin receptor blockers (ARBs); 3) excellent blood pressure control; 4) dietary protein restriction; and 5) control of hyperlipidemia. There are several recent reviews on the treatment of diabetic nephropathy.[1–3]

Glycemic Control

Studies have clearly demonstrated that excellent glycemic control can prevent the development of incipient diabetic nephropathy (microalbuminuria) and progression to overt nephropathy (macroalbuminuria) in type 1 diabetes. The evidence is less conclusive for Type 2 diabetes.

In 1993, the Diabetes Control and Complications Trial (DCCT) convincingly demonstrated that intensive insulin therapy effectively delays

the onset of diabetic microvascular disease, including nephropathy, in Type 1 diabetes.[4] Patients in the intensive therapy arm received insulin in the form of a external insulin pump or by three or more daily insulin injections aiming for a target HbA1c of 6.05%. As the primary end-point of the study was retinopathy, most of the enrolled patients (1441 patients) had mild renal disease with either no albuminuria or microalbuminuria (incipient nephropathy). Although less than 5% of the patients managed to maintain the target HbA1c, intensive therapy reduced the development of incipient nephropathy by 39%, overt nephropathy (macroalbuminuria) by 54% and renal impairment by 60%. There was a two- to three-fold increase in severe hypoglycemia in the intensive therapy group.

For Type 2 diabetes, the United Kingdom Prospective Diabetes Study (UKPDS) enrolled 3867 patients who were assigned to intensive therapy of either a sulphonylurea or insulin, or conventional therapy with diet alone.[5] The HbA1c was maintained at 7.0% in the intensive therapy group and the follow-up period was 10 years. There was a 25% risk reduction in the aggregate microvascular end-points in the intensive therapy group, with a 67% risk reduction of a 2-fold increase in plasma creatinine. There was no difference for any of the aggregate end-points between the sulphonylureas and insulin therapy suggesting that both therapies were equally effective. This was also true for metformin that was used in a subgroup of obese patients.

While excellent glycemic control is important in the early stages of nephropathy, its effectiveness on preservation of renal function is less clear in patients with established overt diabetic nephropathy with macroalbuminuria. In fact, with worsening renal function, patients with overt nephropathy are at risk of developing hypoglycemia because insulin is metabolized in the proximal renal tubules of the kidney and reduced metabolism prolongs the half-life of insulin. The dose of insulin may have to be reduced while those on oral hypoglycemic agents should avoid long-acting sulphonylureas (such as chlorpropamide and glibenclamide) and biguanides such as metformin (because of the risk of lactic acidosis).

Practical recommendations

1) Intensive insulin therapy should be administered in all Type 1 diabetics with normo- and microalbuminuria (incipient nephropathy). In Type 2 diabetics, sulphonulyureas, metformin and insulin therapy have been

found to be equally effective. 2) The target HbA1c level is below 7.0%. 3) Patients with overt nephropathy and impaired renal function are at increased risk of hypoglycemia and should use short-acting sulphonylureas and avoid metformin.

ACE Inhibitors and Angiotensin Receptor Blockers (ARB)

Blockade of the renin-angiotensin system with either an ACE inhibitor or ARB has been shown to delay the progression of diabetic nephropathy and this effect appears independent of its blood pressure lowering properties. In addition to being a potent vasoconstrictor, Angiotensin II has also many effects at the cellular level, including the stimulation of growth factors and cytokines that can lead to cell proliferation and collagen synthesis and the ability to increase glomerular permselectivity. Hence the ability to block the effects of angiotensin II by both the ACE inhibitors and ARBs helps explain the superiority of renal protection provided by these agents compared to other anti-hypertensive medications.

ACE inhibitors have been shown to delay the progression from microalbuminuria (incipient nephropathy) to macroalbuminuria (overt nephropathy) and also retard the progression of renal disease in overt nephropathy in Type 1 diabetes. The evidence for ACE inhibitors in Type 2 diabetes is less conclusive as most of the trials concentrated on the easily definable Type 1 diabetes. The Collaborative Study Group Trial was a large placebo-controlled trial using captopril in patients with Type 1 diabetes and overt nephropathy (urinary albumin levels $>$ 500 mg/day).[6] There were 409 patients who were followed up for 2.7 years and there was a 50% reduction in the composite end-point of doubling of the serum creatinine, ESRD or death for the captopril group. There have also been studies to show ACE inhibitors can decrease progression to overt nephropathy in patients with Type 1 diabetes and microalbuminuria with or without hypertension.

The ARBs have been more extensively investigated in Type 2 diabetes and are similarly able to delay progression of renal disease in hypertensive patients with either incipient or overt nephropathy. Three recent large placebo-controlled trials examined the renoprotective effects of an ARB on the hypertensive Type 2 diabetic with either micro- (incipient nephropathy) or macroalbuminuria (overt nephropathy).[7–9] One trial enrolled patients with incipient nephropathy and found that after 2.7 years of follow-up patients treated with the ARB were less likely to develop overt nephropathy.

The two other trials studied patients with overt nephropathy and similarly found that patients treated with an ARB had a reduced risk of developing the primary composite end-point which was a doubling of the serum creatinine, development of ESRD or death. More indirect support for the use of ARBs in Type 2 diabetes comes from the recently concluded Losartan Intervention For Endpoint reduction in hypertension study (LIFE) where the ARB, losartan significantly reduced cardiovascular morbidity and mortality when compared to the control group using atenolol.[10]

Should normoalbuminuric patients receive prophylactic ACE inhibitors or ARBs? Although there is a study that demonstrated a reduced risk of developing incipient nephropathy when enalapril was used in Type 2 diabetics with no hypertension or albuminuria, there is currently no recommendation to treat all diabetics prophylactically with either an ACE inhibitor or ARB.

Reducing proteinuria

Apart from achieving the target blood pressure (discussed in the following section on Blood Pressure), the reduction of proteinuria should be a goal in itself. Proteinuria has long been regarded as just a marker of renal disease but more recently, there has been an increasing body of evidence to suggest that proteinuria may cause renal injury through the nephrotoxic effects of the proteins. Studies have shown that patients with the greatest reduction in proteinuria have the best preservation of renal function over time. A practical target would be to reduce the proteinuria by at least 50% from the baseline value. Combination therapy in the form of an ACE inhibitor and ARB have shown greater reduction in blood pressure and proteinuria than the individual agents used singly in patients with Type 2 diabetes. Sodium restriction enhances the antiproteinuric effects of the ACE inhibitors and ARBs and sodium intake should be limited to less than 90 mEq/day.

ACE gene polymorphism

The ACE gene comprises of a deletion (D) allele and an insertion (I) allele. The D allele is characterized by higher circulating and tissue ACE levels and as this may result in higher levels of Angiotensin II, it is a potential risk factor. The DD genotype has been associated with declining renal

function and greater risk of progression to ESRD in Type 2 diabetic nephropathy. There is also a higher prevalence of the DD genotype in patients with Type 2 diabetes on dialysis. While these findings suggest that the DD ACE genotype may be a risk factor for progressive renal disease in Type 2 diabetes, there is firm evidence that it predicts the development of nephropathy.

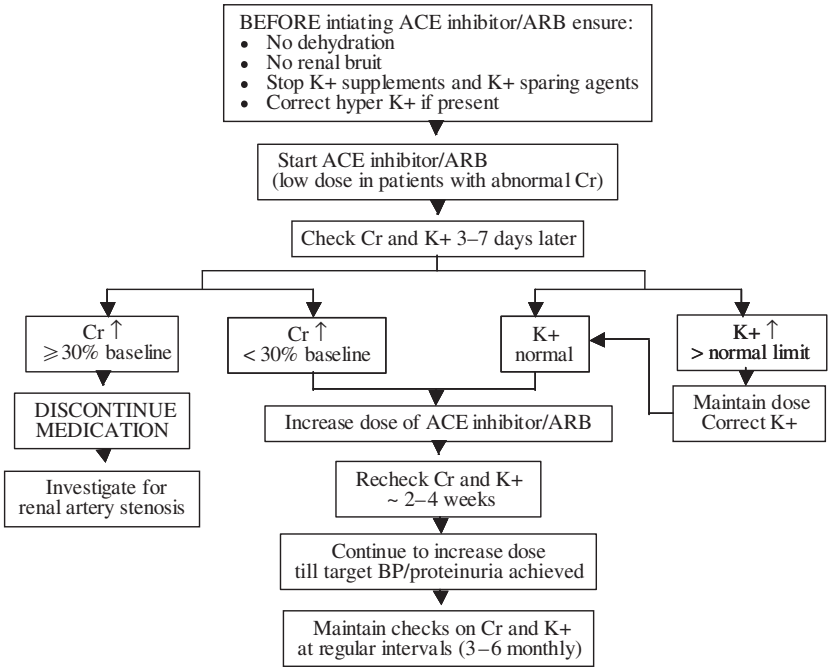
Side effects

The ACE inhibitors and ARBs have two potential side effects: (1) the development of acute, reversible (if detected early) renal failure in patients with bilateral renal artery stenosis or those who are dehydrated; and (2) the development of severe hyperkalemia. Close monitoring of the serum creatinine and potassium should be performed on initiation and when increasing the dose of the ACE inhibitor or ARB, especially in patients with impaired renal function. Before initiating therapy it must be ensured that the patient is not dehydrated, does not have a renal bruit and potassium supplements or potassium sparing agents should be discontinued.

Practical recommendations

1) All patients with diabetes (Type 1 or 2) with
   a) hypertension,
   b) microalbuminuria regardless of blood pressure, or
   c) overt nephropathy should be started on therapy.

   An ACE inhibitor is the treatment of choice in Type 1 diabetes and an ARB in Type 2 diabetes. If the ACE inhibitor cannot be tolerated because of cough, an ARB can be substituted.
2) The goal is to reduce proteinuria by at least 50% from the baseline level.
3) Combination therapy in the form of an ACE inhibitor and ARB can be considered to achieve the goal for proteinuria reduction and/or blood pressure (see below).
4) An abnormal serum creatinine is NOT a contraindication to initiating either an ACE inhibitor or ARB. However, close monitoring of the serum creatinine and potassium should be performed at initiation, with each dose increase and at regular intervals (Fig. 2).



Cr: creatinine, K+ : potassium

Fig. 2 Initiating a patient on an ACE inhibitor or ARB.

Blood Pressure Control

Blood pressure control is the cornerstone of preserving renal function in virtually all forms of renal disease and diabetes is no exception. There have been two recent excellent reviews on the treatment of hypertension in patients with diabetes.[11,12]

Target blood pressure

The glomerulus of the kidney functions optimally at a blood pressure of 80 mmHg and it is therefore not difficult to understand how a high systemic blood pressure can damage the delicate structures of the kidney. In fact, the current blood pressure goals are much lower than the previously accepted target of a blood pressure below 140/90 mmHg. Some studies using ACE inhibitors and/or ARBs in normotensive patients with renal disease have demonstrated that blood pressure can be further reduced

resulting in significant reduction in proteinuria. This suggests that our definition of normotension may actually be too high in patients with renal disease and target blood pressures may need to come down further. In 2003, the JNC VII (Joint National Committee on Detection, Evaluation, and treatment of High Blood Pressure) recommended that all patients with diabetes or chronic kidney disease should have a target blood pressure of <130/80 mmHg.

Antihypertensive agents

It has been observed from clinical trials that an average of 3.2 different antihypertensive agents is required to achieve the recommended target blood pressure. In patients with blood pressures of more than 15/10 mmHg above goal, at least two antihypertensive agents would be required. A step-care approach to achieve the goal blood pressure and reduction in proteinuria is detailed in Fig. 3. However, it must be

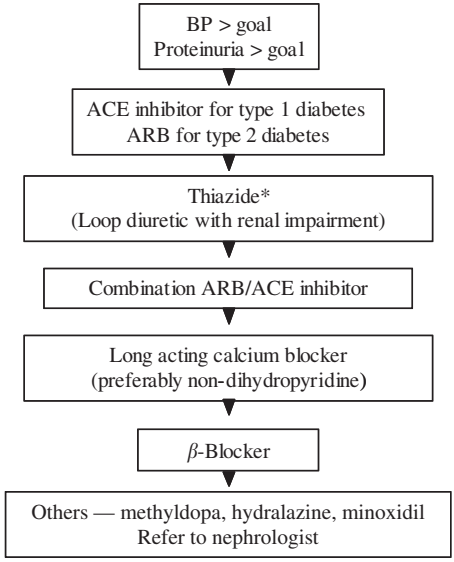


Fig. 3 Step-care approach to achieve blood pressure and proteinuria goals.

remembered that lowering of the blood pressure is more important that the type of antihypertensive agent used.

Initial therapy should be an ACE inhibitor in Type 1 diabetes and an ARB in Type 2 diabetes. The precautions when initiating therapy discussed in the previous section should be observed. If the blood pressure goal is not achieved, the second drug should be a diuretic in the form of a thiazide (hydrochlorothiazide up to a maximum of 25 mg/day) in patients with normal renal function or a loop diuretic in patients with renal impairment. As the blood pressure in diabetic patients is frequently salt sensitive, the diuretics assist through increased urinary sodium excretion. Furthermore, the use of an ACE inhibitor or ARB in combination with a diuretic has also been shown to result in a further reduction in proteinuria (an additional desirable effect). The next step would be to use a combination of ACE inhibitor and ARB (with diuretic) as studies in both Type 2 diabetes and non-diabetic renal disease have shown that a combination of the two agents can more effectively reduce blood pressure and proteinuria than either agent used singly.

The third-line drug would be a long-acting calcium channel blocker (CCB), e.g. diltiazem, amlodipine. Although the dihydropyridine CCBs are powerful antihypertensive agents, there has been concern about their cardiovascular safety in patients with Type 2 diabetes. Furthermore, studies have also shown that the short-acting dihydropyridine CCBs may actually increase proteinuria in both Type 1 and 2 diabetes. They should, therefore, not be used as monotherapy. However, several authors have suggested that a combination of a CCB and ACE inhibitor may be superior to either alone and in practical terms, a CCB is often necessary to achieve the target blood pressure.

Subsequent add-on therapy would include a β-blocker followed by other agents such as methyldopa, hydralazine or minoxidil. A β-blocker should not be used in combination with a non-dihydropyridine CCB in elderly patients and those with conduction abnormalities.

While much emphasis has been placed on achieving the blood pressure goals, care has to be observed in the elderly patients as many will have cardiovascular disease associated with reduction in both cerebral and renal autoregulation. Blood pressure should be slowly lowered in this group of patients and monitored both in the sitting and upright positions to account for orthostatic hypotension.

Dietary Protein Restriction

Although there have been many studies examining the effect of reduced dietary protein intake on the progression of renal disease, there is no conclusive evidence of the benefit of a low protein diet in patients with diabetes. Moreover, with the additional fat and carbohydrate restrictions imposed on the diabetic patient, both compliance and malnutrition may become problems. However, most authors would recommend a moderate protein restriction of 0.8 g/kg body weight/day in diabetic patients with nephropathy.[1]

Treatment of Hyperlipidemia

Hyperlipidemia is frequently associated with diabetes, especially Type 2 diabetes. Most commonly, the lipid abnormalities include an elevated total cholesterol and triglycerides and low HDL (high-density-lipoprotein) cholesterol. The hyperlipidemia is worsened by renal failure and responsible for the excessive cardiovascular disease in diabetic patients with ESRD. Hyperlipidemia may directly contribute to the development of glomerulosclerosis and studies have shown that oxidized LDL (low-density-lipoprotein) cholesterol causes mesangial cell proliferation at low doses and is cytotoxic at high doses.

It is therefore prudent in diabetic patients with renal disease who are all at high risk of cardiovascular disease to adopt the NCEP ATP III (National Cholesterol Education Programme Adult Treatment Panel III) guideline of a target LDL-cholesterol of less than 100 mg/dL. Treatment with an HMG-CoA reductase inhibitor has been shown to reduce proteinuria in patients with Type 2 diabetes.

HMG Co-A reductase inhibitors can effectively reduce cholesterol but in patients with renal disease, there is an increased risk of drug-induced myopathy. This risk is increased if the HMG Co-A reductase inhibitor is combined with cyclosporine, gemfibrozil or other lipid-lowering agents. Hence, when initiating a patient with renal disease on an HMG Co-A reductase inhibitor, the patient should receive a reduced dose and be monitored closely for myopathy after initiation.

Summary

A summary of the therapeutic recommendations is presented in Table 4.

Table 4 Therapeutic Strategies in Diabetic Nephropathy (Types 1 and 2)

Treatment	Stages of Diabetic Nephropathy				
	I	**II**	**III**	**IV**	**V**
Glycemic control	Maintain excellent glycemic control Target HbA1c $<7.0\%$			Watch for hypoglycemia with worsening renal function	
ACE inhibitor and/or ARB	Only if hypertension is present		STRONGLY RECOMMENDED Type 1 diabetes — ACEi Type 2 diabetes — ARB		For control of BP
BP Control	Excellent BP control; $<130/80$mmHg				
Sodium restriction			Sodium intake of <90mEq/day to achieve BP and proteinuria goals		
Protein restriction				0.8g/kg BW/day	
Hyperlipidemia control	Target LDL cholesterol level <100 mg/dL				

ACEi — ACE inhibitor, BP — blood pressure, BW — body weight

INDICATIONS FOR REFERRAL TO NEPHROLOGIST

The diabetic patient should be referred to the nephrologist in the following circumstances:

1) when non-diabetic renal disease is suspected (Table 3);
2) when target blood pressure or reduction in proteinuria cannot be achieved; and
3) when the patient develops progressive chronic renal failure (creatinine ~ 350 μmol/L) requiring preparation of renal replacement therapy.

REFERENCES

1. Ismail N, Becker B, Strzelczyk P, Ritz E, Renal disease and hypertension in non-insulin-dependent diabetes mellitus, *Kidney Int* **55**:1–28, 1999.

2. Mogensen CE, Cooper ME, Diabetic renal disease: from recent studies to improved clinical practice, *Diabet Med* **21**:4–7, 2004.

3. Remuzzi G, Schieppati A, Ruggenenti P, Nephropathy in patients with type 2 diabetes, *N Engl J Med* **346**:1145–1151, 2002.

4. The Diabetes Control and Complication Trial Research Group, The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, *N Engl J Med* **329**:977–986, 1993.

5. UK Prospective Diabetes Study (UKPDS) Group, Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes, *(UKPDS 33) Lancet* **352**:837–853, 1998.

6. Lewis EJ, Hunsicker LG, Bain RP, Rohde RE, The effect of angiotensin-converting enzyme inhibition on diabetic nephropathy, *N Engl J Med* **329**:1456–1462, 1993.

7. Parving HH, Lehnert H, Brochner–Mortensen J, Gomis R, Anderswn S, Arner P, The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes, *N Engl J Med* **345**:870–878, 2001.

8. Lewis EJ, Hunsicker LG, Clarke WR, Berl T, Pohl MA, Lewis LB, Ritz E, Atkins RC, Rohde R, Raz I, Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes, *N Engl J Med* **345**:851–860, 2001.

9. Brenner BM, Cooper ME, de Zeeuw D, Keane WF, Mitch WE, Parving HH, Remuzzi G, Snapinn SM, Zhang Z, Shahinfar S, Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy, *N Engl J Med* **345**:861–869, 2001.

10. Lindholm LH *et al.* for the LIFE study group, Cardiovascular morbidity and mortality in patients with diabetes in the Losartan Intervention for Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol, *Lancet* **359**:1004–1010, 2002.

11. Arauz–Pacheco C, Parrott MA, Raskin P, The treatment of hypertension in adult patients with diabetes, *Diabetes Care* **25**:134–147, 2002.

12. Bakris GL, Williams M, Dworkin L, Elliott J, Epstein M, Toto R, Tuttle K, Douglas J, Hseuh W, Sowers J for the National Kidney Foundation Hypertension and Diabetes Executive Committees Working Group, Preserving renal function in adults with hypertension and diabetes: a consensus approach, *Am J Kidney Dis* **36**:646–661, 2000.

This page intentionally left blank

73

Contemporary Management of Renal Failure

A. Vathsala

INTRODUCTION

Renal failure is characterized by progressive deterioration in the physiological functions of the kidney leading to accumulation of toxic substances in the blood and to alterations in functions of various hormones. These effects lead to a clinical syndrome associated with fatigue, anorexia, weight loss, nausea, vomiting, pruritus, pericarditis, hypertension, fluid overload and neurological disturbances, which if untreated, culminates in coma and death due to end-stage renal failure (ESRF).

Though historically, the term renal failure referred primarily to the later stages in its development including chronic renal failure (CRF) and ESRF, it has become increasingly apparent that these later stages can be prevented or delayed with appropriate interventions initiated at the earlier stages. Thus, recently, the scope of management of chronic renal disease has been extended to include chronic kidney damage from its earliest stages in addition to CRF and ESRF. This article proposes a schema for the holistic management of chronic renal disease that begins with identifying patients who have chronic renal disease, evaluating the cause,

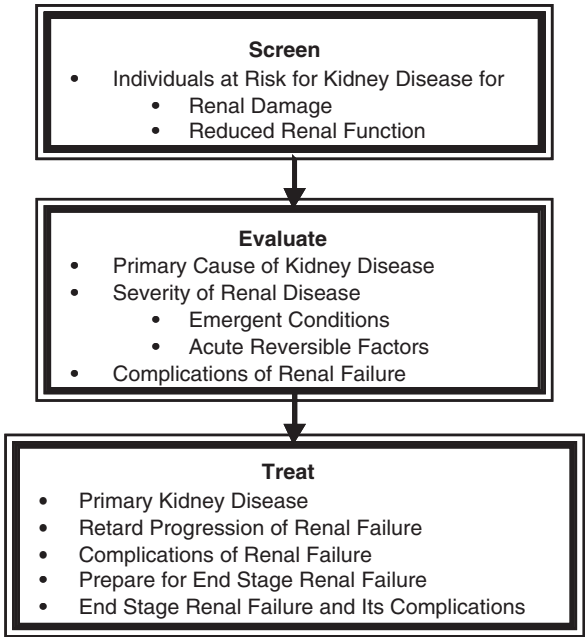


Fig. 1 Schema for management of chronic kidney disease.

severity and complications of renal failure, retarding the progression of renal failure, optimizing preparation for and timely initiation of ESRF therapy and its management in adults.

DEFINITION

A definition of chronic renal disease has been recently devised by the National Kidney Foundation (NKF, USA), and is based on the presence of chronic renal damage and on the level of kidney function.[1] The presence of structural, functional or pathological abnormalities of the kidney or abnormalities in the composition of the blood, urine or in imaging tests of the kidney, persisting for more than 3 months is deemed as evidence for chronic renal damage. Glomerular filtration rate less than 60 mL/min/1.73 m^2, with or without other evidence for renal damage for more than 3 months is also deemed evidence for chronic renal disease. Given that renal damage is progressive in nature, the NKF has also established a framework for defining the various stages of chronic renal disease, based on the degree of renal functional impairment (Table 1).

Table 1 Stages of Chronic Renal Disease

Stage	Description	GFR[a] mL/min/1.73 m^2	High Blood Pressure or Laboratory Abnormality	Symptoms
1	Kidney damage with normal or ↑ GFR	⩾90	Sometimes[b]	Sometimes
2	Kidney damage with mild ↓ GFR	60–89	Possible	Sometimes
3	Moderate ↓ GFR	30–59	Mild	Possible
4	Severe ↓ GFR	15–29	Moderate	Mild
5	Kidney failure	<15 or dialysis	Severe	Moderate

[a]GFR — Glomerular Filtration Rate.
[b]Refers to instances where patients may have symptoms and signs due to nephrotic syndrome, urinary tract symptoms, tubular syndromes etc.
Adapted from K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification.[1]

Based on these definitions and stages, it is readily apparent that, the vast majority of patients become symptomatic from chronic renal disease when they have already lost more than 50% of their renal function. This reiterates the need for screening of at-risk individuals as the first step in approaching the problem of chronic renal disease.

SCOPE OF CHRONIC RENAL DISEASE

In the USA, it has been estimated that nearly 21% of the population may have some form of renal disease or are at risk for renal disease: 4.6% have moderate to severe reduction in renal function or are on dialysis, 6.3% have evidence of early kidney damage with normal to mildly decreased renal function and at least 10% of the population are at risk for renal disease.[1] Among the various causes of ESRF, glomerular diseases such as diabetic nephropathy (Range: 22–44% in Asia) and glomerulonephritis (Range: 10–42% in Asia) are the leading causes worldwide. Vascular disease including hypertensive nephrosclerosis is the third leading cause (Range: 5–19% in Asia), while tubulo-interstitial diseases such as "chronic pyelonephritis" and vesicoureteric reflux, toxic nephropathy, analgesic nephropathy, urinary tract obstructive disorders such as

obstructive uropathy as well as cystic diseases are less common causes. Of particular concern is the increasing incidence of ESRF in Singapore and in the rest of Asia; whereas in 1995, there were 403 new cases of ESRF starting dialysis in Singapore, in 2000, there were 513 new cases of ESRF starting dialysis, a nearly 20% increase (Singapore Renal Registry, 2001, unpublished).

SCREENING FOR CHRONIC RENAL DISEASE

Screening for chronic renal disease in adults is contingent on firstly identifying patients at risk for chronic renal disease and secondly performing the necessary screening tests on these patients to identify those with chronic renal damage and/or decreased renal function.

At Risk Populations

The following categories of patients are at risk for renal disease and should be screened for renal damage:

- diabetes
- hypertension
- older age ($\geqslant 60$ years)
- family history of chronic kidney disease
- autoimmune diseases
- systemic infections
- urinary tract disorders
- neoplasia
- those with exposure to drugs toxic to the kidney
- those with recovery from acute renal failure
- those with reduced kidney mass

Screening Tests

Patients at risk for chronic renal disease should be screened for markers of renal damage and level of renal function. Proteinuria is a hallmark of renal damage; as normal individuals excrete minimal quantities of protein in the urine, persistent proteinuria ($>1+$ on dipstick), if detected on two consecutive tests, is suggestive of renal damage and indicates the need for further evaluation. Once detected, a positive urine dipstick for

protein should be confirmed with a quantitative protein assessment. Spot urinary measurements of protein/creatinine ratios are an accurate estimate of urinary protein excretion rate. This test is performed by measuring protein and creatinine concentration (in mg/dL) in a spot urine specimen; the protein concentration is divided by the creatinine concentration and the unit-less number (the protein/creatinine ratio) equals the 24 hour urine protein excretion rate (normal values are < 0.15, equal to < 0.15 g protein/day).

While a serum creatinine value above the normal reference range for that laboratory is indicative of reduced renal function, its measurement alone has been deemed inadequate. The usual method of estimating renal function or glomerular filtration rate (GFR) in clinical practice has been timed urine collection for creatinine clearance. However, due to the tendency of timed urine collections to overestimate GFR and their cumbersome nature, various prediction equations have been used to assess GFR instead. As these equations have acceptable accuracy among western populations, the NKF (USA) guidelines accept these prediction equations as surrogate markers for GFR, making timed urine collections for creatinine clearance redundant. Nevertheless, as these prediction equations were largely obtained from larger-sized western populations, their accuracy in estimating GFR in the smaller-sized Asian population is unestablished. Timed creatinine clearance will estimate GFR more accurately in at risk Asian populations.

Other laboratory tests that are markers of renal damage include:

- microhematuria;
- other abnormalities on examination of urine sediment, e.g. pyuria, red blood cell casts etc; and
- imaging studies such as Ultrasonography, Intravenous Pyelography, CT scan, Magnetic Resonance Imaging, Nuclear Medicine scans.

Thus patient groups at risk for renal disease as listed above should undergo urine protein, urine sediment and blood creatinine measurements to confirm the presence of chronic renal disease. In those with abnormalities detected on screening, proteinuria should be confirmed and quantitated by urine protein/creatinine ratios and renal function estimated with creatinine clearance and the stage of renal disease classified as in Table 1. The speed of further assessment and management obviously is based on the presence of symptoms and severity of renal disease — more

symptomatic and more severe disease requires more rapid assessment and management. In the absence of evidence of renal disease on initial screening of at-risk individuals, the screening should be repeated annually.

EVALUATION OF CHRONIC RENAL DISEASE

After identifying patients with chronic renal disease, the next steps in the evaluation are to firstly establish the cause and to secondly, to establish its chronicity and severity. A full history, physical examination and various diagnostic tests are useful in the evaluation (Table 2). As laboratory tests to evaluate cause of renal disease can be quite extensive, investigations should be tailored to the individual patient.

Importantly, during the course of evaluation of patients with renal disease, the chronicity and severity of renal disease should be established. Chronicity is favored by the presence of symptoms of renal failure, anemia, elevated parathyroid hormone and small kidneys on ultrasound/or increased renal parenchymal echogenicity. The evaluation should assess the severity of renal failure; emergent conditions requiring urgent dialysis or other treatment as well presence of acute reversible factors, which, upon correction, would result in reversal of acute renal dysfunction (Table 3) should be identified and appropriately managed.

Part of the evaluation of chronic renal disease includes an assessment of the manifestations and complications of renal failure. These manifestations are more prominent with increasing severity of chronic renal disease and contribute significantly to its morbidity (Figure 2).

Chronic renal disease, once diagnosed, invariably progresses over time to CRF and eventually to ESRF. The interval of progression to CRF and ESRF is variable and is affected by the cause of renal disease and by other factors. Diabetic nephropathy, glomerulonephritis and polycystic kidney disease are generally associated with a faster rate of decline in renal function than tubulo-interstitial diseases and hypertensive nephrosclerosis. In diabetics, worse glycemic control is associated with a faster decline in GFR; furthermore, regardless of cause of renal disease, those with more proteinuria, lower serum albumin, higher blood pressure and smoking also experience a faster decline. Finally, patients with chronic renal disease progress rapidly to ESRF if they experience renal insults that cause acute renal dysfunction. Thus, optimal treatment of

Table 2 Evaluation of Primary Cause of Renal Disease

Medical history	• Diabetes
	• Hypertension
	• Previous renal problems including urinary abnormalities, stones, difficulty micturition, dysuria, urgency, gross hematuria, urinary tract obstruction
	• Previous auto-immune disease
	• Previous serious systemic infections or renal failure
	• Previous failed/normal urinary examinations:
	  • Insurance examination
	  • Pre-employment checks etc.
	  • Other screening for urine abnormalities
	• Pregnancy: hypertension, pre-eclampsia, recurrent abortions
	• Previous pelvic surgery
	• Hepatitis

Drug history	• Non Steroidal Anti-Inflammatory Agents (NSAIDS) • COX2 Inhibitors • Traditional herbs or medications
Family history	• Kidney failure, or urinary abnormalities • Diabetes • Hypertension • Polycystic kidney disease • Urinary abnormalities

Physical examination	• Blood pressure (postural drop?)
	• Uremic fetor, asterexis
	• Skin: rash, scratch marks, vasculitis, stigmata of embolic disease, gouty tophi
	• Assess volume status:
	  • Jugular venous pressure
	  • Crepitations in lungs
	  • Edema
	• Cardiomegaly, gallop rhythm
	• Abdomen:
	  • Organomegaly
	  • Renal bruits
	  • Flank tenderness
	  • Palpable masses
	• Peripheral vascular disease
	• Peripheral neuropathy
	• Joints: Arthritis (auto-immune disease, gout)
	• Digital rectal examination (prostate) in men

Laboratory confirmation of renal disease	• Urine protein/creatinine ratio (or Urine albumin/creatinine ratio) or 24-hour urinary protein • Urine microscopy • Serum creatinine • 24-hour creatinine clearance

Table 2 (*Continued*)

Laboratory evaluation of cause of renal disease	• Urine phase contrast microscopy • Ultrasound kidneys/Intravenous pyelography/ CT scan/MRI/Nuclear medicine studies • Fasting glucose/HbA1C • Anti nuclear factor/Double stranded DNA etc. • Serum complement levels (CH50, C3, C4) • Anti neutrophil cytoplasmic autoantibody • Anti glomerular basement membrane antibody • Serum/Urine immunoelectropheresis for myeloma • Cryoglobulins • Hepatitis B Surface Antigen, Anti-HCV, HIV • Eosinophiluria • Renal biopsy
Other tests	• Electrocardiogram • Chest radiograph • Fasting lipids

chronic renal disease in its early stages includes treatment of the primary disease, retardation of progression of renal failure by ameliorating risk factors for progression and preventing acute declines in renal function, and most importantly educating and counseling patients on these measures (Table 4).

While the treatment of primary renal disease *per se*, is beyond the scope of this chapter, excellent glycemic control (HbA1C target $<7\%$) clearly reduces the risk of progression of diabetic nephropathy in both Type I and Type II diabetes.[2,3] Two other measures, namely excellent blood pressure control and renoprotection with Angiotensin Converting Enzyme Inhibitors (ACEI) or Angiotensin Receptor Blockers (ARB) have been shown to retard progression of renal disease and should be implemented in all patients as part of the holistic management of chronic renal disease.

Blood pressure control

The majority of patients with chronic renal disease have hypertension and the level of blood pressure is directly proportional to progression; thus strict blood pressure control would be expected to retard progression

Table 3 Assessment for Severity of Renal Disease

Symptoms of renal failure	• Nausea, vomiting • Oliguria, nocturia, polyuria • Dyspnea, orthopnea, edema
Laboratory evaluation of severity of renal disease	• Urea/Electrolytes/Glucose • Full Blood Count • Calcium/Phosphate • Total Protein/Albumin • Parathyroid Hormone
Renal emergencies	• Rapidly rising serum creatinine • Fluid overload, especially pulmonary edema • Severe hyperkalemia • Severe metabolic acidosis • Encephalopathy: Coma, seizures etc • Pericarditis

Acute reversible factors	• Pre-renal causes:
	  • Volume depletion/Hypotension
	  • Severe congestive heart failure
	• Renal causes:
	  • Acute glomerulonephritis (Crescenteric change)
	  • Acute tubular necrosis (Vasomotor nephropathy) due to:
	  • Sepsis
	  • Nephrotoxic agents
	  • Acute interstitial nephritis
	  • Angiotensin-converting enzyme inhibitors/ Angiotensin II Receptor blockers/ Radiocontrast agents
	  • Rhabdomyolysis
	  • Cholesterol emboli
	• Post-renal causes: Urinary tract obstruction
	• Renovascular:
	  • Accelerated hypertension
	  • Stenosis/Thrombosis

of renal failure. The Modification of Diet in Renal Disease (MDRD) study evaluated the impact of different levels of blood pressure control on rate of GFR decline in non-diabetic renal disease.[4] As patients with higher levels of proteinuria experienced the greatest benefits from the lower blood pressure (BP $<$ 125/75 mm Hg), these lower target blood pressure levels are recommended for patients with proteinuria $>$ 1 g/day. These lower target blood pressure levels are also recommended for all diabetics regardless of level of proteinuria.[5] For patients with non-diabetic renal

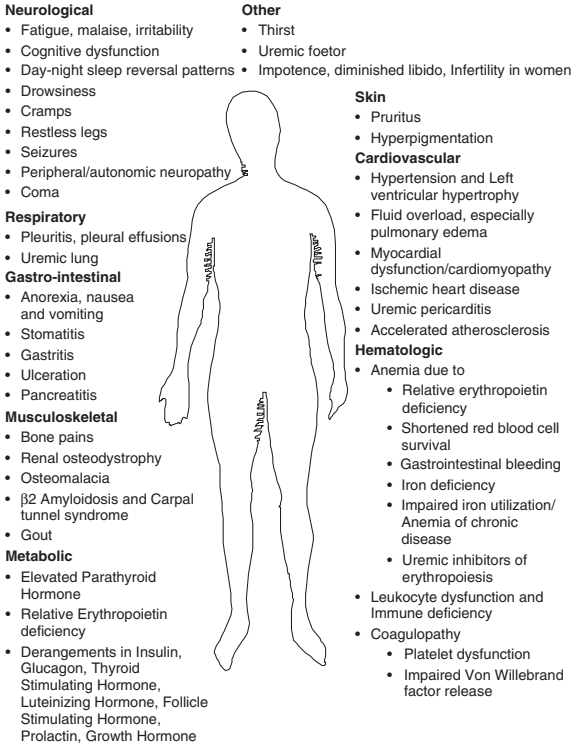


Fig. 2 Manifestations of chronic renal disease.

disease and proteinuria >1g/day, guidelines for target blood pressures have been less stringent at 130/85 mm Hg.[6] However more recent guidelines by various bodies recommend an even lower BP target of 130/80 mm Hg in patients with chronic renal disease with proteinuria <1 g/day.[7]

All anti-hypertensives can be used in patients with chronic kidney disease, though dihydropyridine calcium channel blockers (e.g. Amlodipine) should not be used alone in patients with chronic kidney disease due to their association with faster decline in renal function.[8] On the average, 2 to 3 anti-hypertensives may be needed to control blood pressure to the desired levels suggested above.

Reno-protection

While any anti-hypertensive drug appears to be equally effective in reducing blood pressure, ACEI and ARB theoretically confer an

Table 4 Early Treatment of Chronic Renal Disease

Treat underlying disease	• Control blood glucose strictly in diabetics • Control blood pressure • Primary disease (glomerulonephritis etc)
Monitor renal function	• Urea, electrolytes and creatinine • Urine protein/Creatinine ratio • Creatinine clearance

Retard progression of renal failure	• Control blood pressure
	  • $<130/80$ mm Hg for those with proteinuria <1 g/day
	  • $<125/75$ mm Hg for those with proteinuria $\geqslant 1$ g/day and diabetics
	• Angiotensin converting enzyme inhibitors/ Angiotensin receptor blockers
	• Control blood glucose strictly in diabetics

Treat and prevent reversible factors	• Correct:
	  • Volume depletion/hypotension
	  • Heart failure
	• Treat:
	  • Sepsis
	  • Urinary tract obstruction
	  • Vascular stenosis/thrombosis
	  • Acute interstitial nephritis
	  • Uncontrolled hypertension
	• Stop/avoid:
	  • Nephrotoxic agents
	  • Radiocontrast agents
	  • Agents causing rhabdomyolysis (Statins, Colchicine etc.)

Educate and counsel	• Regarding disease, severity, treatments and prognosis • Refer Dietician, Pharmacist, Psychologist, Social worker

additional reno-protective effect in patients with renal disease by ameliorating intra-glomerular hypertension and reducing proteinuria; they may also be beneficial in reducing angiotensin II mediated cell proliferation and fibrosis. Several large studies have demonstrated the effects of ACEI to lower blood pressure, decrease proteinuria and slow the progression of renal failure to a greater extent than other anti-hypertensive agents in patients with non-diabetic and diabetic renal disease, making these drugs the preferred choice for treatment of high blood pressure in these patients.[9,10] Recent studies have also demonstrated the efficacy of ARB to reduce proteinuria and delay progression of renal failure in Type II diabetics with various stages of nephropathy.[11] Though similar evidence for ARB's in non-diabetic renal disease are currently lacking,

ARB's may have similar effects to that of ACEI and may be particularly useful for patients unable to tolerate the latter because of cough. Combined use of ACEI and ARB has been advocated by some to reduce proteinuria and control blood pressure, however, they should be used cautiously in combination in patients on non-cardio-selective β blockers.

When used in patients with chronic renal disease, both ACEI and ARB's are associated with transient increases in serum creatinine (due to their hemodynamic effects) and hyperkalemia and should be used with caution in patients with serum creatinine greater than 265 μmol/L or 3 mg/dL. A rise in serum creatinine >88 μmol/L following ACEI/ARB therapy has been suggested as an indication to exclude renal artery stenosis and discontinuation of ACEI/ARBs. Patients with hyperkalemia (serum $K+>5.5$ mmol/L) respond to dietary potassium restriction or addition of a diuretic; failing this ACEI/ARB doses may need to be reduced. In addition to pharmacologic therapy, sodium intake should be restricted to less than 100 mmol/day so as to improve blood pressure control as well as ameliorate fluid overload. As hypertension in renal failure occurs frequently in relation to fluid overload, fluid restriction and treatment with high dose, loop diuretics is often necessary to optimize therapy.

Other measures that are often instituted with the diagnosis of chronic renal disease are dietary protein restriction and control of hyperlipidemia. Though protein restriction has been shown to retard progression of renal failure, this requires intensive nutritional counseling to prevent malnutrition.[4] Thus modest protein restriction diet consisting of 0.8 g protein/kg/ body weight/day of high biologic value protein can be instituted when the creatinine clearance is less than 25 mL/min. As patients with renal failure have a high incidence of hyperlipidemia with elevations in total and low density cholesterol and triglycerides, treatment with dietary intervention and pharmacologic therapy is likely to reduce risk for cardiovascular disease as in the general population, though this may not have any renoprotective effects.[12]

LATE COMPLICATIONS OF CHRONIC RENAL DISEASE

Appropriate management of the later stages of chronic renal disease including treatment of complications of renal failure such as anemia and secondary hyperparathyroidism; timely initiation of dialysis can

Table 5 Treatment of Late Stages of Chronic Renal Disease

Monitor complications of renal failure	• Hemoglobin, Iron status • Serum calcium, Phosphate, Alkaline phosphatase • Serum parathyroid hormone • Serum albumin

Treat complications of renal failure	• Exclude other causes of anemia
	  • Gastrointestinal bleeding
	  • Iron/vitamin deficiency
	  • Thalassemia
	• Correct anemia
	  • Iron/vitamin deficiencies
	  • Recombinant erythropoietin
	• Prevent and treat hyperparathyroidism
	  • Phosphate binders
	  • Calcium carbonate or calcium acetate
	  • Active vitamin D analogues
	• Fluid overload with diuretics
	• Acidosis with bicarbonate supplements
	• Hyperlipidemia with statins or fibrates

Prepare for end-stage renal failure	• Evaluate suitability for
	  • Renal transplantation
	  • Hemodialysis
	  • Peritoneal dialysis
	• Refer transplant or dialysis coordinator
	• Refer psychologist, social worker
	• Do serology for Hepatitis B, Anti HCV, HIV
	• Consider tissue typing, vascular mapping
	• Refer surgeon
	• Consider timely initiation of dialysis

significantly reduce the morbidity associated with CRF (Table 5). Recombinant Erythropoietin (rEPO) has significantly facilitated the management of anemia of renal failure. At doses of 80 to 120 units/kg/week in divided doses and in the presence of adequate iron stores, rEPO administration reverses the anemia of renal failure, improves appetite and reduces hospitalization rates in patients with renal failure. Current recommendations for treatment of anemia aim for target hemoglobin levels of 11 and 12 g/dL in pre-menopausal females and males/post menopausal females respectively.[13] However, financial and logistical considerations frequently limit the use of rEPO in pre-dialysis patients and patients often receive only enough rEPO to remain asymptomatic and transfusion-independent.

Nevertheless, with progression of renal disease, accumulation of uremic toxins and symptoms of fluid overload worsen; conservative measures become inadequate and the patient requires renal replacement therapy to treat ESRF. As early dialysis has been suggested to reduce hospitalization and morbidity, current recommendations suggest a strategy of "timely" initiation of dialysis when the creatinine clearance is between 9–14 mL/min/1.73 m^2 with incremental dialysis being administered as residual renal function diminishes.[13] However, in Asia, socioeconomic considerations often prevent early initiation of dialysis, and the majority of patients start dialysis when serum creatinine is greater than 900 μmol/L (or 10 mg/dL) and earlier only if symptoms of uremia and fluid overload supervene.

RENAL REPLACEMENT THERAPIES

An integrated and individualized management is the ideal approach for treatment of ESRF and each patient must be educated regarding the options for renal transplantation, hemodialysis or peritoneal dialysis. The choice of one modality over another for any individual patient is based on medical suitability, the availability of these options as well as socioeconomic and cultural circumstances. The presence of multiple underlying co-morbidities may prompt some patients to opt for conservative management without renal replacement therapy and these patients will require the necessary social and psychological support. In the holistic management of renal failure, every patient needs to be evaluated medically, financially and psychosocially and prepared appropriately with a view towards the timely initiation of renal replacement therapy of ESRF.

Renal Transplantation

Generally, renal transplantation (RTX), with kidneys obtained from live or cadaveric donors, is the ideal form of treatment for many categories of patients with ESRF as it is the most successful in reversing the metabolic consequences of uremia. Patients with underlying systemic malignancy, severe cerebrovascular or cardiovascular disease, significant liver disease, and serious infections are excluded from consideration for RTX. Additional criteria such as age or minor coronary artery disease or cerebrovascular disease are occasionally imposed in some countries due to

lack of adequate donor kidneys for RTX. For patients suitable for RTX, family members (live-related and emotionally related) are counseled and those willing to be considered for renal donation are evaluated to exclude systemic disease as well as diseases that may pose a risk for renal deterioration after uninephrectomy (example: baseline renal dysfunction, hypertension and diabetes). In the immunological evaluation of a live-related donor, the donor with the best tissue match is given preference as RTX between Human Leukocyte Antigen (HLA) identical siblings still yields the best long-term results. However, with newer immunosuppressive drugs, live donor RTX even across HLA mismatched donor-recipient pairs also yields excellent results and matching alone need not be a consideration for live donor RTX. After ensuring compatibility between donor and recipient, and screening for infections that may be transmitted to the recipient with RTX, the most suitable donor undergoes evaluation of renal vascular anatomy to confirm suitability. If a suitable donor is available, the pre-ESRF patient can undergo pre-emptive RTX, prior to the need for dialysis. Those without a suitable live-donor are started on dialysis and placed on a waiting list for a cadaveric RTX.

Live donor uninephrectomy was in the past performed through a flank incision; currently, the surgery can be performed via a laparoscopic approach, a method which does away with the flank incision and reduces post-operative hospitalization. Cadaveric kidneys are harvested from patients declared brain dead following cerebrovascular accident or trauma. Criteria for selection of recipients for cadaveric RTX vary between different countries but are generally based on HLA matching, waiting period on dialysis, pre-sensitization and other medical considerations. Surgical preparation for the potential recipient for RTX is straightforward; native nephrectomy is unnecessary unless in the presence of an infected/obstructed native kidney. Live donor kidneys are transplanted immediately after harvest, while cadaveric kidneys are preserved either with cold or machine perfusion for up to 48 hours; transplantation requires anastamoses of donor renal vessels to recipient iliac vessels with implantation of the donor ureter into the bladder with a neocystostomy.

Life-long immunosuppression is required to prevent rejection of the allograft in the majority of patients. Standard immunosuppression in many centers is with cyclosporine (CsA), azathioprine and corticosteroids, with CsA doses adjusted to levels. Newer agents include tacrolimus (FK506) as a substitute for CsA, mycophenolate as a substitute for azathioprine,

interleukin 2 receptor antagonists and sirolimus, the latter, a calcineurin inhibitor-sparing agent. With the introduction of these newer immunosuppressants into clinical practice in recent years, the choice of therapy is varied and can be tailored to the individual, based on immunological risk and other clinical factors. Induction protocols using anti-lymphocyte preparations may be especially useful in high-risk candidates such as re-transplants and those with sensitization. Acute rejection episodes which occur in 10–40% of transplants (based on risk and regimen used) can be treated with high dose corticosteroid pulse therapy or anti-lymphocyte preparations such as anti-CD3 monoclonal antibody (OKT3) with successful reversal in over 90% of cases.

With current immunosuppressive protocols, short and long-term results of renal transplantation are excellent: one-year graft survival rates for recipients of live donor and cadaveric transplants are 96.4% and 87.9% respectively at our center under CsA immunosuppression and compare favorably with the 93% and 87% one-year graft survivals respectively reported from the Registry of the United Network of Organ Sharing, UNOS.[14] The importance of HLA match is still evident with 93% three-year graft survival for HLA identical sibling transplants versus 87% and 86% for those receiving one haplotype and live unrelated transplants respectively. Nevertheless, results of live donor transplants are still superior to that of cadaveric transplants, supporting the continued use of living donors as the preferred modality of renal replacement therapy. With improving immunosuppressive protocols, graft loss due to acute rejection is becoming less frequent and patient death with graft function is of increasing importance.[14] As cardiovascular mortality and sepsis remain the leading causes of death in RTX, the challenge for the future is to further reduce mortality in RTX with optimal immunosuppressive and disease management strategies.

Hemodialysis

Though RTX is the ideal form of renal replacement therapy, most ESRF patients do not have a suitable live donor and need dialysis to sustain life. Dialysis is a process by which the solutes diffuse between the blood and dialysate across a semi-permeable membrane down a concentration gradient; waste products such as urea, creatinine, and phosphates diffuse from the blood into the dialysate while bicarbonate and calcium from the

dialysate diffuse back into the blood. In hemodialysis (HD), diffusion occurs between the blood and the dialysate through the hollow fibers of an artificial semi-permeable membrane comprising the hemodialyzer.

A permanent vascular access is needed to provide delivery of blood at flow rates of 200 to 350 mL/min to the hemodialyzer. The preferred access is a radiocephalic or brachiocephalic arteriovenous fistula in the non-dominant hand, a procedure in which the radial or brachial artery is anastamosed to the cephalic veins. After surgical creation of a fistula, the venous limbs of the fistula dilate and the vessel walls becomes arterialized under the influence of systemic arterial pressures thereby permitting repeated needling and the high blood flow rates required for HD. In patients with inadequately developed superficial veins, an arterio-venous access can be created using a graft made of synthetic material such as polytetrafluoroethylene. The graft can then be needled directly and repeatedly to provide the necessary blood flow for HD. Access infection and failure due to the occurrence of stenosis or thrombosis are major causes of morbidity and hospitalization in HD patients. In patients without a permanent vascular access, temporary access can be secured using a catheter placed in the internal jugular, subclavian or rarely the femoral vein. Their use is also complicated by infections and vascular stenosis, such that timely creation of a permanent vascular access in anticipation of its future use is crucial in the management of HD patients.

The three components of the HD apparatus are the hemodialyzer, the HD machine and the dialysate. Standard hemodialyzers are made from hollow fibers comprised of cellulose, substituted cellulose or other synthetic materials such as polysulfone, polyacrylonitrile and polymethylmethacrylate and provide surface areas between 0.8 m^2 to 2.0 m^2 for dialysis. After the first use, a hemodialyzer can be rinsed free of blood, cleansed and sterilized for reuse for up to 7 uses for the same patient. The HD machine itself consists of a blood pump, a dialysate mixing and delivery system and appropriate safety monitors to detect pressures, blood and air leaks and ion conductivities within the extracorporeal circuit. In a typical circuit, blood is moved from the arterial limb of the patient's vascular access to the blood compartment of the hemodialyzer with the aid of a blood pump; a heparin pump permits the continuous infusion of heparin into the patient for anticoagulation of the circulating blood. The dialysate delivery system mixes the bicarbonate-based dialysate concentrate with treated water to its final dilution, heats it to body

temperature and delivers it to the dialysate compartment of the hemodialyzer circuit at rates of 500 mL/min. As the flow of blood and dialysate are in countercurrent directions, concentration gradients between the blood and dialysate are maximized, thereby permitting maximal cross diffusion of solutes across the dialyzer membrane. In addition to solute exchange, fluid removal is an important component of the dialysis procedure and is achieved by application of transmembrane pressure to the blood in the dialysis circuit, thereby permitting ultrafiltration of blood via hydrostatic pressure. Following transit through the hemodialyzer, the purified blood is returned to the patient through the venous limb of the vascular access.

The HD procedure is generally performed over a 3 1/2 to 4 1/2 hour period 3 times a week. Goals of dialysis are to normalize some electrolytes and ions such as potassium and hydrogen ions, remove uremic toxins and make the patient euvolemic by the end of the dialysis. Adequate dialysis when delivered to a patient with adequate nutrition is associated with lower hospitalization rates and lower morbidity and mortality. Thus for each patient, adequacy of HD is periodically assessed during a single dialysis by monitoring the delivered dose of dialysis with urea kinetic modeling; based on this, and other parameters such as lean body weight, residual renal function and dietary protein intake, dialysis can be individualized by prescribing the appropriate duration of dialysis and type of dialyzer to be used.

Though HD is a safe procedure, and is the most common modality of renal replacement therapy for ESRF in Asia, complications unique to the procedure do occur as a result of the intensive nature of the procedure. Intradialytic hypotension may occur due to absolute volume depletion, fluid shifts between extracellular and intracellular compartments and changes in osmolality or cardiac arrythmias. Anaphylactoid reactions may occur due to complement activation by dialyzer membranes upon exposure to blood usually with the first use of dialyzer or due to exposure to ethylene oxide-altered proteins, when ethylene oxide has been used to sterilize the dialyzer. Dialysis associated neutropenia and hypoxaemia can be related to the first use syndrome. The need for anticoagulation during HD often increases the risk for bleeding.

Overall mortality for HD patients is dependent on age and presence of co-morbid risk factors such as diabetes and cardiovascular disease. For 20- to 29-year-olds prevalent on hemodialysis in the USA in 1999, annual

death rate was 50.7 per 1000 patient years at risk in comparison to the death rate of 224.8 per 1000 patient years for those between age 60–64 years (proportionately higher with age). These results are in contrast to the death rates of 8.3 and 61.4 per 1000 patient years, respectively, in the USA, for transplanted patients in the same age groups. Five year survival for the entire cohort of patients entering a HD program between 1991 and 1995 in Singapore are comparable at 85.2% (Choong HL, personal communication). As with RTX, the leading causes of death in dialysis patients are cardiac or infectious in origin with cerebrovascular disease, malignancy and hemorrhage comprising the remainder.[15]

Peritoneal dialysis

Peritoneal dialysis (PD) is an alternative form of dialysis in which exchange of toxins and solutes occurs between the blood and the dialysate across the peritoneal membrane lining the peritoneal cavity. During the procedure, 2 to 2.5 L of sterile peritoneal dialysate is instilled into the peritoneal cavity by gravity and allowed to dwell for a period of time. Toxins diffuse from plasma to the dialysate during the dwell period and are removed with the dialysate when the dialysate is drained out of the peritoneal cavity, after which fresh dialysate is instilled into the cavity. In Continuous Ambulatory Peritoneal Dialysis (CAPD), the patient manually performs the exchanges 4 to 5 times a day and is fully ambulant during the dwell periods. In Automated Peritoneal Dialysis (APD), a machine called a cycler performs a scheduled number of exchanges during the night, freeing the patient from having to do multiple exchanges during the daytime.

The preference for PD over HD is often based on medical and non-medical factors. Patients who prefer the convenience of a home-based dialysis procedure or those who stay far from a HD facility may opt for PD for non-medical reasons. Others who are unable to tolerate HD for various reasons such as significant cardiac disease, extensive vascular disease or those in whom vascular access is problematic are, in fact, better medically suited for PD. Nevertheless, there are some contraindications to PD, including the presence of either intra-abdominal adhesions that could limit dialysate flow, uncorrectable mechanical defects such as a diaphragmatic or other irreparable hernia, severe inflammatory bowel or diverticular disease.

The physiology of PD is not significantly different from that of HD. Toxins, solutes and fluid move between the blood and dialysate through

the processes of diffusion and ultrafiltration and transport of smaller molecules is more efficient than that of larger molecules. However, unlike in HD, the peritoneal membrane permits the transport of larger molecules and proteins to diffuse and lead to protein losses through the dialysate. Fluid removal is achieved in PD through osmotic ultrafiltration. The dialysate in PD contains the osmotically active substance, glucose, in high concentrations (1.25%, 2.5%, 4.25%); the osmotic pressure generated by the glucose then draws water from the blood across the peritoneal membrane into the dialysate, thereby effecting fluid removal. Glucose absorption from the dialysate does occur and may predispose to obesity and hypertriglyceridemia in PD patients in general and increase insulin requirements in diabetic patients in particular. Newer dialysates such as those containing amino acids or Icodextrins may ameliorate problems of malnutrition and inadequate ultrafiltration occurring with PD using the conventional glucose-based dialysates.

The apparatus for PD is simple and includes dialysate solutions, the peritoneal access catheter and a transfer set that is used to connect the access device to the peritoneal dialysate bags. Peritoneal access catheters are inserted surgically into the abdomen through a subcutaneous tunnel with the catheter tip placed in the pelvis. Dialysate bags are connected to the peritoneal access catheter through a special transfer set that minimizes the risk for peritonitis. The transfer set currently used in Singapore uses a twin bag system, which has yielded a peritonitis rate of one episode every 38 months (Chew STH, personal communication), comparable to the peritonitis rate of 1 in 45 months reported in other centers. Adequacy of PD can be measured using urea and creatinine clearances and studies have demonstrated an association between greater urea clearance and a decreased relative risk of death. Thus after initiating PD, peritoneal membrane transport characteristics and adequacy of dialysis are measured and dialysis schedules modified to individualize and optimize PD prescription.

Peritoneal dialysis is a simple technique that is easy to perform and has the advantages of portability, fewer dialysis-related symptoms and does not require anticoagulation. However, and despite the simplicity of PD techniques, the major problem with this modality is peritonitis. Peritonitis occurs generally as a result of transmigration of bacteria from the skin into the abdominal cavity and is defined as a cloudy dialysate effluent associated with a dialysate white blood cell count greater than 100/mm.[3] Concomitant symptoms of abdominal pain are usually present,

though fever is less common. Though infections with the usual organisms such as *Staphylococcus epidermidis* or *aureus* usually respond to intraperitoneal antibiotics, those due to other organisms such as *Pseudomonas aeruginosa*, while less common, are less responsive to antibiotic therapy and may require catheter removal and even lead to peritoneal fibrosis and failure. Other complications of PD are pericatheter leaks, catheter malfunction due to catheter migration or omental wrapping, hernia development and exit site infections. Transfer from PD to HD may become necessary in those receiving inadequate dialysis or ultrafiltration and in those developing technical problems due to the procedure. One year death rate on PD for 20–29 year olds and 60–64 year olds, prevalent on PD in USA in 1999, is reported as 56.1 and 284.0 per 1000 patient years respectively, the excess risk in comparison to HD being attributed to the higher incidence of cardiovascular deaths in the former.[15]

CONCLUSIONS

With the increasing numbers of patients with chronic renal disease, its contemporary management has shifted from that of treatment of CRF and ESRF to its early diagnosis and retardation of progression.

REFERENCES

1. K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification and Stratification, *Am J Kidney Dis* **39** (Suppl 1):S14–S266, 2002.
2. The Diabetes Control and Complications Trial Research Group, The effects of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, *NEJM* **329**:977–986, 1993.
3. UK Prospective Diabetes Study Group, Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes, *BMJ* **317**:703–713, 1998.
4. Peterson JC, Adler S, Burkart JM *et al*., Blood pressure control, proteinuria, and the progression of renal disease: The Modification of Diet in Renal Disease Study, *Ann Intern Med* **123**:754–762, 1995.
5. Hansson L, Zanchett A, Carruthers SG, Bahlof B *et al*., Effects of intensive blood pressure lowering and low-dose aspirin in patientswith hypertension: Principal results of the Hypertension Optimal Treatment (HOT) randomized trial. The HOT Study Group, *Lancet* **351**:1755–1762, 1998.
6. The Sixth Report of the Joint National Committee on prevention, detection, evaluation and treatment of high blood pressure, *Arch Int Med* **157**:2413–2416, 1997.

7. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. Chobanian AV, Bakris GL, Black HR, Cushman WC *et al.*, National Heart, Lung, and Blood Institute Joint Naitonal Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. *JAMA*. 2003 May 21;**289**(19):2560–2572.

8. Agodoa LY, Appel L, Bakris GL, Beck GJ *et al.*, African American Study of Kidney Disease and Hypertension (AASK) Study Group. Effect of ramipril vs amlodipine on renal outcomes in hypertensive nephrosclerosis: A randomized controlled trial, *JAMA* **285**:2719–2728, 2001.

9. Jafar TH, Scmid CH, Landa M *et al.*, Angiotensin-converting enzyme inhibitors and the progression of nondiabetic renal disease: a meta-analysis of patient-level data, *Ann Intern Med* **135**:73–87, 2001.

10. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD, The effect of angiotensin-converting enzyme inhibition on diabetic nephropathy, *NEJM* **329**:1456–1466, 1993.

11. Brenner BM, Cooper ME, de Zeeuw D, Keane WF *et al.*, The RENAAL Study Investigators. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy, *NEJM* **345**:861–869, 2001.

12. Levey AS, Controlling the epidemic of cardiovascular disease in chronic renal disease: Where do we start? *Am J Kidney Dis* **32**(5):853–906, 1998.

13. NKF DOQI Clinical Practice Guidelines, *Am J Kidney Dis* **30** (Suppl 2): S15–S240, 1997.

14. Cecka JM, The UNOS scientific renal transplant registry, in: Cecka JM, Terasaki PI (eds.), *Clinical Transplants 1999*, UCLA Tissue Typing Laboratory Los Angeles, California, USA, pp. 1–21, 2000.

15. US Renal Data System, USRDS 2001 Annual Data Report, Atlas of End-Stage Renal Disease in the United States. National Institutes of Health, *National Institute of Diabetes and Diabetes and Kidney Diseases*, Bethesda MD 483–507, 2001.

Respiratory Medicine

This page intentionally left blank

74

Approach to the Patient with Respiratory Disease

Philip Eng

Respiratory diseases are a major cause of morbidity and mortality all over the world. In the primary care setting, upper respiratory tract infections are the most common reason why patients consult their doctors. Lung cancer is the second most common type of cancer in Singapore, with dismal 5-year survival rates of 14%. Pneumonia is the third most common cause of death in Singapore and is often the end result of many chronic debilitating diseases like cerebrovascular disease and cancer.

Principal causes of death in Singapore:

1) cancer;
2) ischaemic heart disease;
3) pneumonia;
4) cerebrovascular disease; and
5) injuries.

Most common Cancers in Singapore:

	Male	Female
1)	Lung	1) Breast
2)	Colorectal	2) Colorectal
3)	Stomach	3) Lung

RAPID EVALUATION

In the initial evaluation of a patient with suspected respiratory disease, a detailed history is the cornerstone. The only exception is when one is dealing with a critically ill patient where rapid assessment and control of the ABCs, i.e. airway, breathing and circulation, must take precedence. Examples are those with problems of:

- Airway, e.g. apnea, stridor;
- Breathing, e.g. respiratory distress due to severe pneumonia; and
- Circulation, e.g. hypovolemic shock from bleeding, septic shock.

Once this is done, history taking usually involves relatives, colleagues and bystanders in this setting. Clinical examination is also more complicated as there is a need to verify that all the devices e.g. central venous catheters, endotracheal tube … etc. are in the right place and functioning properly.

In this modern era where many patients seek to undergo routine health screening, one must also be mindful of the fact that there are patients who are totally asymptomatic, yet have abnormal CXR findings leading to more sinister diagnoses like lung cancer.

HISTORY

In the history taking, certain symptoms may point to a respiratory disorder but this is not invariable. The most common symptom of respiratory disease is shortness of breath or dyspnea. Dyspnea that is unrelated to exertion is commonly due to asthma. This symptom can be brought on by cold, emotion, stress or upper respiratory tract infections. In asthma, the dyspnea is usually but not invariably accompanied by wheezing. Also, the symptoms are rapidly relieved by inhaled short acting beta 2 agonists like Salbutamol. Symptoms of mild asthma may also resolve spontaneously.

Dyspnea can also be due to other organ dysfunction. Severe anemia, sepsis and metabolic acidosis are common non-cardio-respiratory causes of shortness of breath. The more common situation, however, is to try to differentiate it from cardiac disease. Cardiac dyspnea is usually due to congestive heart failure and is effort related and relieved by rest. Orthopnea, paroxysmal nocturnal dyspnea and bilateral ankle edema, if present helps clinch the diagnosis. Orthopnea is the sensation of shortness of breath when lying down and is a classic symptom of left ventricular failure as the venous return increases in the supine position, aggravating pulmonary congestion. Paroxysmal nocturnal dyspnea is explained by the same mechanism. This is not invariable as there are some reasons why respiratory patients may get short of breath in the supine position. Some examples are morbid obesity, diaphragmatic paralysis and patients with severe chest wall disease, e.g. kyphoscoliosis.

In contrast, some patients experience dyspnea when adopting the standing position (platypnea). If desaturation is documented, this is called orthodeoxia. The classic cause of this is anatomical pulmonary arteriovenous malformation (+/− Hereditary Hemorraghic Telangiectasia), or functional ones due to the hepatopulmonary syndrome related to liver cirrhosis.

Wheezing is another common symptom or sign seen in patients with respiratory disease e.g. Chronic Obstructive Pulmonary Disease (COPD) and asthma. It is best described as a high pitched sound best heard during expiration caused by turbulent airflow through narrowed airways. Features in the history suggestive of asthma are young age (<50), non-smoking status, past history or family history of atopy (asthma, allergic rhinitis, eczema). The patient is usually totally well between episodes. As described above, the wheezing and shortness of breath in asthma is also unrelated to exertion unlike in COPD. Tumors, notably carcinoid, mucoepidermoid cancer, anaphylactic reactions, tracheobronchitis, e.g. due to Sarcoidosis, Wegener's, Tuberculosis, Aspergillus, can also cause wheezing. Rarer causes of wheezing are acute pulmonary embolism and herpes simplex tracheobronchitis in those with the Acute Respiratory Distress Syndrome (ARDS). In unilateral wheezing in a child, the diagnosis of an inhaled foreign body must always be entertained.

Chronic cough (>3 months) is another classic symptom found in respiratory patients. Chronic cough in patients with normal CXR is usually

due to one of the following: smoking, ACE inhibitors, cough variant asthma, post nasal drip syndrome and gastroesophageal reflux disease. About 15% of patients on ACE inhibitors do complain of chronic cough. Chronic cough is much rarer in patients using angiotensin 2 receptor antagonists.

Hemoptysis is another symptom suggestive of respiratory disease. The character of hemoptysis may help differentiate the cause. Pink frothy hemoptysis is suggestive of acute pulmonary edema. Hemoptysis going on for years is suggestive of a long-standing process like bronchiectasis or arteriovenous malformations. Massive hemoptysis is usually due to active tuberculosis, bronchiectasis or mycetoma. Streaky hemoptysis of recent onset for weeks is always worrying for lung cancer, especially if it happens in a middle-aged male smoker. It is important to rule out "pseudohemoptysis", i.e. bleeding that is not from the lower respiratory tract. Gastrointestinal bleeding is characterized by blood mixed with food particles. The acidic nature of the fluid, confirmed on litmus, points towards a gastric origin. Another source for confusion is bleeding from the post-nasal space and upper airway. In Singapore, one has to consider nasopharyngeal cancer, especially if there is unilateral epistaxis, decreased hearing or rhinorrhea.

Chest pain due to respiratory disease is extremely varied. It should be noted that the lung parenchyma and the visceral pleura do not have pain fibers. Chest wall pain due to a fractured rib is exquisitely tender whereas that due to pleurisy varies with respiration. Pleural effusion and lung cancer involving the chest wall gives a dull ache. Sometimes, the chest tightness due to asthma can be mistaken for chest pain. It is important also not to miss the searing pain characteristic of the initial presentation of herpes zoster of the intercostal nerves.

Hoarseness of voice can be due to left recurrent laryngeal nerve palsy, which is usually due to compression by aortopulmonary mediastinal lymph nodes from metastases from primary lung cancer. Other causes include tuberculous lymph nodes, sarcoidosis and lymphoma. Infrequently, tuberculous laryngitis can present like an acute febrile illness with prolonged hoarseness of voice.

Snoring is another symptom of respiratory disease, and is usually reported by the person's spouse. Accompanying daytime somnolence, headaches and nocturnal choking all help point towards obstructive sleep apnea.

SMOKING HISTORY

A smoking history is extremely important in the evaluation of a patient not just confined to those with respiratory disease. Smokers are predisposed to develop cancers of the lung, larynx, esophagus, pancreas. In addition, they are at risk for coronary artery disease, peripheral vascular disease, COPD and cerebrovascular accidents. Cigar smokers are also at similar risk. It is also important to realize that passive smokers are also predisposed to developing lung cancer and the data is best established in nonsmoking spouses of smokers. The risk is also dose dependent, i.e. it depends on the number of sticks smoked per day and the number of years smoked. The new marketing strategy of the tobacco industry of "light cigarettes" with less tobacco content per stick is targeted at those who are less informed. The data shows that those who use such cigarettes tend to take deeper inhalations to achieve the same "kick". The smoking prevalence in Singapore has somewhat stabilized over the past 5 years at 15% but what is alarming is that there has been a 20-fold rise in female smokers aged 18–24 over the past 16 years from 0.4% to 8% in 2001. It is notable that in many developed countries in the world today including the US, UK, Sweden and Finland, female smoking rates are about even with male smokers and hovering at about 28%. Oblivious to such is the alarming increase in smoking prevalence in Asian countries like Japan, Korea and China, which tobacco industries have now turned their attention to.

HIV

The most common opportunistic infection in patients with the Human Immunodeficiency Virus (HIV) is oral candidiasis. Indeed, in patients who are not known to be immunosuppressed by cancer or chemotherapy or prolonged antibiotics, it would be unusual to find oral candidiasis unless he is harboring the HIV virus. The most common life threatening opportunistic infection in patients with the Human Immunodeficiency virus (HIV) is the Pneumocystis Carinii Pneumonia (PCP). PCP remains a relatively common index presentation in patients subsequently found to have the HIV. The more common risk factors for HIV in Singapore are sexual promiscuity in heterosexuals, followed by homosexuals and intravenous drug abusers. This history must be sought for in all cases who present with a community acquired pneumonia where there are risk factors for acquiring the HIV.

OCCUPATIONAL HISTORY

An awareness of common occupation-related disease is important for diagnosis. A history of working in a sand quarry should immediately alert one to the possibility of silicosis. Occupation on board a ship as an electrician should make one think of asbestos as this was a popular material for electrical insulation more than 30 years ago. However, the most common cause of occupation-related lung disease today is occupational asthma. The list is endless, but common ones include the use of glutareldehyde in health care workers, flour amongst bakers, soldering flux, and isocyanate.

A continuously updated list on the Asmanet website (http://asmanet.com) currently includes more than 361 occupational agents shown to be involved in occupational asthma.

CLINICAL EXAMINATION

There are few clinical signs that are specific to respiratory disease. In the examination, one must be observant to the fact that many clues lie outside the chest.

Clubbing can be due to respiratory disease like idiopathic pulmonary fibrosis, lung cancer, bronchiectasis and other suppurative lung disease. All patients with clubbing must be examined for features of hypertrophic pulmonary osteoarthropathy (HPOA). This is usually evident as tenderness over the distal end of long bones like the radius/ulna or the tibia/fibula and is usually seen on X-rays as periostitis. In such patients it is mandatory to do a CXR to rule out lung cancer, e.g. squamous cell carcinoma. Even if lung cancer is proven, evidence of HPOA is not a criteria for inoperability. Other causes of clubbing include cyanotic congenital heart disease, atrial myxoma, infective endocarditis, liver cirrhosis, thyrotoxicosis, inflammatory bowel disease and congenital.

Cervical lymphadenopathy is another important sign as it can be a sign of metastatic lung cancer. Other causes include tuberculosis, sarcoidosis, chronic lymphocyctic leukemia, lymphoma and metastatic cancer from the gastrointestinal tract (especially if the deep cervical lymph node is involved).

In patients with chronic respiratory disease, one must examine and look for features of complications, i.e. chronic respiratory failure, cor pulmonale, polycythemia due to chronic hypoxia and cyanosis due to

increased concentrations of reduced hemoglobin. Cor pulmonale can be demonstrated by evidence of loud P2, right parasternal heave and ankle edema. Evidence of chronic hypoxia include irritability and tachycardia. Ventilatory failure as manifested by severe hypercapnia may result in vasodilation, bounding pulse, asterexis, and sweating.

Stridor is a very important sign that should never be missed. It is best described as the death rattle because death is imminent unless efforts are taken to correct it immediately. The best way to pick up stridor is to auscultate over the neck area and compare it with the breath sounds. In true stridor, the breath sounds are decreased and there are added sounds especially during inspiration best heard over the throat. Common causes are upper airway obstruction from tumors like laryngeal cancer, acute epiglottitis, laryngospasm, foreign bodies, post extubation glottic edema, tracheal strictures, bilateral vocal cord paralysis and neuromuscular dysfunction.

Breath sounds are important as they are a clue to the function of the respiratory system. Expiratory phase can be prolonged in those with obstructive airway disease. Diminished breath sounds are a clue to very severe obstruction. Crepitations or rales or crackles on auscultation indicate that there is fluid in the alveoli, usually due to acute pulmonary edema or pneumonia. Occasionally these disappear when the maneuver is repeated after coughing, indicative of minor atelectasis. Basal crepitations, which sound like "velcroe", are indicative of idiopathic pulmonary fibrosis. It is important to differentiate crepitations from pleural rub as the latter is heard both during inspiration and expiration. Pleural rub is a sign of pleural inflammation, often seen in parapneumonic effusions.

Percussion is an important part of the physical examination of the respiratory system. Abnormal unilateral resonance is commonly found in pneumothorax whereas bilateral hyperresonance is found in COPD. Dullness is found in underlying consolidation, atelectasis and pleural effusion although it is classically stony dull in the latter.

RADIOGRAPHIC DIAGNOSIS

In the approach to the patient with respiratory disease, a standard PA CXR is the most important investigation. Clinical diagnostic algorithms emanate from pattern recognition of the most obvious CXR abnormality.

In the evaluation of patients with abnormal mediastinal masses, the clue lies in the localization of the mass except in the case of lymphomas

and aortic aneurysm, which can occur anywhere in the mediastinum. Superior mediastinal masses are usually due to retrosternal thyroid goiters and thymomas whereas anterior mediastinal masses are usually due to teratomas, thyroid goitres and thymomas. Middle mediastinal masses are usually either bronchogenic cysts, pericardial cysts, pericardial fat and hiatus hernia. Posterior mediastinal masses are usually either diaphragmatic hernia and neurogenic tumors.

In the evaluation of patients with lung infiltrates , the most important step is to determine the distribution as upper lobe infiltrates are usually due to silicosis, tuberculosis, ankylosing spondylitis, histiocytosis X and aspergillosis. Lower lobe infiltrates are commonly due to idiopathic pulmonary fibrosis, asbestosis, bronchiectasis and aspiration pneumonia.

Causes of diffuse miliary nodules (< 2 mm) include miliary tuberculosis, disseminated histoplasmosis, previous Varicella pneumonia, silicosis, and pulmonary alveolar microlithiasis. If the pattern is that of diffuse nodules (up to 4 cm) , then others like disseminated adenocarcinoma, septic infarcts due to bacteremia and sarcoidosis enter into the differential diagnoses. The pattern of metastatic carcinoma is usually that of diffuse nodules, especially in the peripheries and bases of varying sizes compared to septic infarcts where the nodules tend to be more similar sized.

Bilateral hilar adenopathy is usually due to sarcoidosis, silicosis and lymphoma. In silicosis, the pattern is classically that of eggshell calcification with upper lobe fibrosis.

In the approach to a slowly resolving/non-resolving pneumonia, one must always think of aspiration, bronchoalveolar cell carcinoma, obstructed bronchus (due to tumor or foreign body), mechanical complications e.g. empyema or lung abscess, inappropriate antibiotics, e.g. tuberculosis or multi-resistant organisms, e.g. Burkholderia pseudomallei, penicillin resistant streptococcal pneumonia.

REFERENCES

Global Initiative for Asthma, *Global strategy for Asthma Management and Prevention*, National Institutes of Health, 2002.

Cancer Incidence in Singapore 1993–1997, Singapore Cancer Registry, 2000.

Hospital Statistics, Ministry of Health 2001, www.moh.gov.sg.

75

Asthma

Tan Keng Leong

EPIDEMIOLOGY

Asthma is a common disorder that is encountered in clinical practice. Worldwide, epidemiological studies in both children and adults suggest that asthma is largely under-diagnosed and consequently, under-treated. In Singapore in the year 2000, bronchitis, emphysema and asthma as a group accounted for 0.7% of total deaths and was the 9th principal cause of death. Although rising trends in mortality from asthma in the 1970's and/or 1980's have been reported in a number of countries, including Hong Kong, New Zealand, England and Wales, France, Italy and the United States, there was no evidence of a temporal increase in asthma mortality from 1976 to 1995 among adults in Singapore. The local prevalence rate of adult asthma based on a cross-sectional population-based sample study in 1992 was 2.4% in men and 2.0% in women. Marked ethnic differences exist, with Malays and Indians having higher asthma mortality and morbidity rates than Chinese in Singapore. Data from local prevalence studies indicate that the proportion of children under 14 years diagnosed with asthma have increased from 5% in 1967 to 20% in 1994.

DEFINITION

Asthma is defined as a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role, in particular, mast cells, eosinophils, T lymphocytes, neutrophils, and epithelial cells. The chronic inflammation causes an associated increase in airway hyperresponsiveness that leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night and in the early morning. These episodes are usually associated with widespread but variable airflow obstruction that is often reversible either spontaneously or with treatment.

Airflow inflammation produces four forms of airflow limitation: acute bronchoconstriction, swelling of the airway wall, chronic mucus plug formation, and airway remodeling.

CLINICO-PATHOLOGIC FEATURES OF ASTHMA

The classic triad of symptoms associated with asthma consists of cough, dyspnea and wheezing. However, it is not unusual for one or more of these typical symptoms to be absent, or for asthmatics to present with other symptoms such as chest tightness, chest discomfort, phlegm production or hyperventilation syndrome. Similar respiratory symptoms may also be seen in various other disorders including left ventricular failure, bronchiectasis and chronic obstructive pulmonary disease. Asthma may present as chronic cough, especially at night, with exercise or with viral illness.

Dyspnea and/or wheezing are the common symptoms of airflow obstruction. Eliciting the nature of the dyspnea may aid in differential diagnosis. In non-asthmatic chronic airflow limitation, the dyspnea is usually chronic and progressive. Episodic dyspnea or cough, on the other hand, is more typical of asthma. Variable airflow obstruction and airway hyperreactivity are typical of asthma.

Asthma is a chronic inflammatory disorder of the airways with recurrent exacerbations. Symptoms may be precipitated or aggravated by upper respiratory tract infections, inhalant allergens, cigarette smoke, exercise, occupational exposure to triggers, drugs (e.g. aspirin,

non-steroidal anti-inflammatory drugs) and pets. Exposure to allergens and respiratory (viral) infections are the main factors responsible for causing exacerbations of asthma and/or the persistence of symptoms. Inflammation, airway remodeling and altered neural control of the airways are responsible for both recurrent exacerbations of asthma and more permanent airflow obstruction. During an exacerbation of asthma, contraction of airway smooth muscle, edema, and hypersecretion tend to close the smaller (noncartilaginous) airways. To compensate, the patient breathes at a higher lung volume to increase outward retraction of the airways, thereby helping to maintain their patency. The more severe the airflow limitation, the higher the lung volume must be to keep the airways open. The combination of hyperinflation and advanced airflow limitation in an asthma exacerbation also increases the work of breathing.

Bronchial biopsy specimens in patients with mild asthma reveal significant collagen deposition beneath the epithelial basement membrane, in addition to the inflammation. The presence of the subepithelial basement membrane fibrosis suggest that progressive airway fibrosis may ensue in long-standing asthma, especially when anti-inflammatory therapy is inadequate. Long-standing asthma may progress to chronic airflow limitation.

The physical examination of the respiratory system may be normal as the airway obstruction in asthma is variable and often reversible either spontaneously or with treatment. The clinical findings of widespread wheezing are characteristic of asthma although they are not specific for the diagnosis. Wheezing may be absent in severe asthma exacerbations. Physical signs, which may be present during an exacerbation of asthma, include tachypnea, difficulty speaking, reduced ribcage expansion and hyperresonance, activity of accessory muscles, pulsus paradoxus, drowsiness and central cyanosis. With worsening airflow obstruction, hyperinflation results in the loss of cardiac dullness and increase in the anteroposterior diameter of the chest.

Among the common obstructive pulmonary disorders, there exist some similarities in their pathophysiologic features and clinical presentations. Overlap among these disorders is common. The importance of differentiating asthma, emphysema and chronic bronchitis lies in the different approaches to treatment and prevention, particularly in light of the major advances in the last decade and differences in prognosis.

DIAGNOSIS

The diagnosis of asthma is based on appropriate clinical history and evidence of reversible airflow obstruction. To establish a diagnosis of asthma, the clinician should determine that episodic symptoms of airflow obstruction are present, airflow obstruction is at least partially reversible and alternative diagnoses are excluded.

A clinical diagnosis of asthma is often prompted by compatible symptoms such as episodic dyspnea, wheezing and chest tightness. Favorable symptomatic response to bronchodilator therapy, seasonal variability of symptoms, nocturnal symptoms, exacerbation of symptoms on exposure to stimuli such as aeroallergens, exercise, cold air, air pollutants, upper respiratory tract infection, or strong odours, a positive family history of asthma and a personal or family history of atopic diseases aid in supporting the diagnosis.

A detailed past and present occupational history is essential in the evaluation of individuals exposed to occupational agents known to cause asthma.

The diagnosis of asthma is established by demonstrating reversible airway obstruction. Variability greater than 20% in serial FEV1 or peak expiratory flow (PEF) suggests the diagnosis of asthma.

DIFFERENTIAL DIAGNOSIS

In the evaluation of a patient with new-onset asthma, particular attention should be paid to atypical presenting features such as hemoptysis, weight loss, clubbing or airflow obstruction that does not reverse with bronchodilators. Atypical presenting features should prompt consideration of alternative diagnoses.

In children, foreign body aspiration and viral bronchiolitis may mimic asthma. Upper airway obstruction by tumour or laryngeal edema, bronchiectasis and pulmonary embolism can occasionally be confused with asthma. Transmission of expiratory wheezing from the upper airway can mimic asthma on auscultation of the chest. Clinical differentiation between upper and lower airway obstruction may sometimes be difficult. In upper airway obstruction, the clinical clue is maximal wheezing over the laryngeal (neck) area. Persistent wheezing localized to one area of the chest raises the suspicion of endobronchial disease such as

tumor, foreign body aspiration or stricture. Vocal cord dysfunction, a condition due to functional (non-organic) causes can also masquerade as asthma. The condition is characterized by paradoxical adduction of the vocal cords during inspiration and definitive diagnosis is established by visualization of the vocal cords by laryngoscopy during an attack.

Late-onset asthma occurring in association with vasculitis and marked eosinophilia should raise the possibility of the diagnosis of Churg–Strauss syndrome. In travelers to and foreign workers from filarial endemic areas who present with paroxysmal cough and wheezing worse at night, low-grade fever and eosinophilia, the diagnosis of tropical pulmonary eosinophilia should be considered.

Asthma is prevalent in all age-groups while in general, chronic obstructive pulmonary disease (COPD) is a disease of older patients. COPD (that is not secondary to congenital alpha1-antitrypsin deficiency) commonly presents in the 5th decade of life. In the older-aged patients, COPD and heart failure should be considered in the differential diagnosis. Smoking history is valuable in distinguishing non-asthmatic from predominantly asthmatic obstruction. In symptomatic COPD, there is usually a history of at least 20 pack-years of cigarette smoking.

PULMONARY FUNCTION TESTS

In patients with suspected airflow obstruction, pulmonary function tests are essential for the diagnosis, assessment of severity of disease and monitoring of response to treatment. Although asthma can often be diagnosed on the basis of symptoms, measurements of lung function, especially the reversibility of lung function abnormalities, greatly enhance the diagnostic confidence.

Spirometry is helpful in the diagnosis of asthma, the assessment of its severity and in the monitoring of the progression of asthma and response to therapeutic intervention. Spirometry, unlike peak flow measurement, differentiates the physiologic pattern of airflow obstruction from pulmonary restriction. A reduction in the ratio of forced expiratory volume in one second (FEV1) to forced vital capacity (VC) from the predicted range is diagnostic of airflow obstruction. The severity of airflow obstruction is classified on the basis of FEV1, expressed as a percent of predicted values. An increase in FEV1 of at least 200 mL and 12% after bronchodilator administration indicate bronchial reversibility. At least a 12% improvement in FEV1

either spontaneously, after inhalation of bronchodilator, or in response to a trial of glucocorticosteroid therapy, favors a diagnosis of asthma.

Bronchoprovocation testing is another strategy for diagnosing asthma in patients with normal lung function. The provocative stimuli used include inhaled methacholine, exercise, inhaled histamine and hyperventilation of cold and dry air. While a positive test result (indicating bronchial hyper-responsiveness) is not entirely specific for asthma, a negative bronchoprovocation challenge has a very high negative predictive value in excluding asthma.

In primary health care settings, peak expiratory flow (PEF) meters are relatively inexpensive and may be helpful in the diagnosis of asthma. A 15% or greater improvement in PEF following inhalation of bronchodilator or in response to a trial of corticosteroid therapy is supportive of the diagnosis of asthma. A diurnal variation in PEF of more than 20% is diagnostic of asthma. However, in mild intermittent asthma or in severe asthma, the variability in PEF may not be present.

CHEST RADIOGRAPH

The chest radiograph is usually normal in patients with asthma. The chest radiographic features of airway obstruction include hyperinflation manifesting as depression and flattening of the diaphragm on the posteroanterior film and increase in the retrosternal airspace on the lateral chest radiograph. In patients presenting with clinical features suggestive of new-onset asthma, chest radiograph is valuable in excluding alternative diagnosis.

OTHER TESTS

An elevated eosinophil count and serum immunoglobulin E (Ig E) may indicate the presence of atopy. However, markedly elevated eosinophil percentages ($>15\%$) should prompt consideration of alternate diagnosis such as drug reactions, parasitic infections, and syndromes of pulmonary infiltrates with eosinophilia. High Ig E levels (>1000 ng/mL) in association with asthma and central bronchiectasis suggests the diagnosis of allergic bronchopulmonary aspergillosis (ABPA).

Measurements of the allergic status (e.g. skin prick test or specific Ig E in serum) aid in the identification of risk factors so that appropriate environmental control measures and allergen avoidance measures can be recommended.

TREATMENT

Under diagnosis and under treatment are major contributors to asthma morbidity and mortality. Long-term preventive treatment is the cornerstone of good asthma control. Patient education is an essential part of the overall management of asthma.

The goals of management in asthma are to achieve and maintain control of symptoms, prevent asthma exacerbations, maintain pulmonary function as near to normal levels as possible, maintain normal activity levels (including exercise), provide optimal pharmacotherapy with minimal or no adverse effects, prevent the development of irreversible airflow limitation and prevent asthma mortality.

Asthma severity is classified by the presence of clinical features prior to commencing treatment and/or by the amount of daily medication needed for optimal control (Table 1). The severity of asthma is subdivided into four steps: Intermittent, Mild persistent, Moderate persistent and Severe persistent.

Short-acting inhaled beta2-agonists taken as needed to treat symptoms (reliever medication) are usually sufficient therapy for intermittent asthma. Persistent asthma, either mild, moderate, or severe, is controlled with daily anti-inflammatory therapy (preventer medication). Inhaled corticosteroids are the most potent inhaled anti-inflammatory agent currently available. Inhaled corticosteroids should be given to patients with persistent asthma (i.e. needing reliever medication one or more times a week). In patients with persistent asthma, inhaled corticosteroids have been shown to improve asthma control and reduce mortality from asthma. Early intervention with inhaled corticosteroids can improve asthma control and normalize lung function, and may prevent irreversible airway injury.

A stepwise approach to pharmacologic therapy is recommended. Pharmacologic therapy is usually initiated at a higher level at the onset to establish prompt control and then stepping down. Response to therapy is monitored clinically and with serial measurements of lung function (spirometry or peak expiratory flow).

If control is sustained for at least 3 months, a gradual stepwise reduction in treatment may be attempted to identify the minimum therapy required to maintain control. Discontinuation of long-term preventive treatment with inhaled steroids should be attempted with great caution. After stopping inhaled steroids, patients are at an increased risk of severe

Table 1 Classification of Asthma Severity by Clinical Features and Current Daily Medication

	Current Treatment Step		
	Step 1: Intermittent	**Step 2: Mild Persistent**	**Step 3: Moderate Persistent**
Clinical features on current therapy		Level of severity	
Step 1: Intermittent Symptoms < once a week Brief exacerbations Nocturnal symptoms not > 2x a month Normal lung function	Intermittent	Mild persistent	Moderate persistent
Step 2: Mild persistent Symptoms > once a week but < once a day Nocturnal symptoms > twice a month but < once a week Normal lung function	Mild persistent	Moderate persistent	Severe persistent
Step 3: Moderate persistent Symptoms daily Exacerbations affect activity and sleep Nocturnal symptoms > once a week 60% < FEV1 < 80% predicted 60% < PEF < 80% of personal best	Moderate persistent	Severe persistent	Severe persistent
Step 4: Severe persistent Symptoms daily Frequent exacerbations Frequent nocturnal symptoms FEV1 < 60% predicted PEF < 60% of personal best	Severe persistent	Severe persistent	Severe persistent

asthma and fatal asthma relapse. If control is not achieved, prior to stepping up drug therapy, it is important to review patient's inhaler technique, compliance and environmental control such as avoidance of allergens or other trigger factors. Short-acting inhaled beta2-agonists taken as needed are used in the treatment of acute asthma symptoms and exacerbations and in the prevention of exercise-induced bronchospasm. Increasing use of short-acting beta2-agonists for acute asthma symptoms indicates inadequate control of asthma and the need for increasing anti-inflammatory therapy. Regularly scheduled, daily use of short-acting beta2-agonist is not recommended. Increasing use of beta2-agonist has been associated with increased risk for death or near-death in patients with asthma. Nocturnal symptoms may be controlled with long-acting inhaled beta2-agonist or sustained-release theophylline. In persistent asthma, the addition of a long-acting inhaled beta2-agonist to inhaled steroids results in better asthma control and reduction in severe exacerbations when compared with doubling the dose of inhaled steroids. In asthmatics with symptoms not controlled with 400–800 mcg of inhaled budesonide or equivalent, long-acting inhaled beta2-agonist may be added. Alternatives to long-acting inhaled beta2-agonist for add-on treatment to inhaled steroids in persistent asthma include oral sustained release-theophylline, leukotriene modifier, or long-acting oral beta2-agonist.

Management of acute asthma exacerbations includes inhaled beta2-agonist to provide prompt relief of airflow obstruction, increasing the dose of inhaled corticosteroids, initiating systemic corticosteroids for moderate-to-severe exacerbations to suppress and reverse airway inflammation and oxygen to relieve hypoxemia. The management of acute asthma in the hospital setting is outlined in Fig. 1. Addition of ipratropium to an aerosolized solution of beta2-agonist in adults has been shown to cause additional bronchodilatation, particularly in those with severe airflow obstruction and to reduce hospitalizations. Asthmatics, particularly those with moderate-to-severe persistent asthma and those with a history of severe exacerbation, should be provided with a written action plan to guide self-management during acute asthma excerbations. Asthma self-management plans are cost-effective and have been shown to lead to significant reductions in morbidity and patients' need for medical services. Self-management plans based on either peak flow or symptoms are of similar efficacy.

Patient education should aim to improve understanding, compliance, skills and self-management. Review of patient technique in using

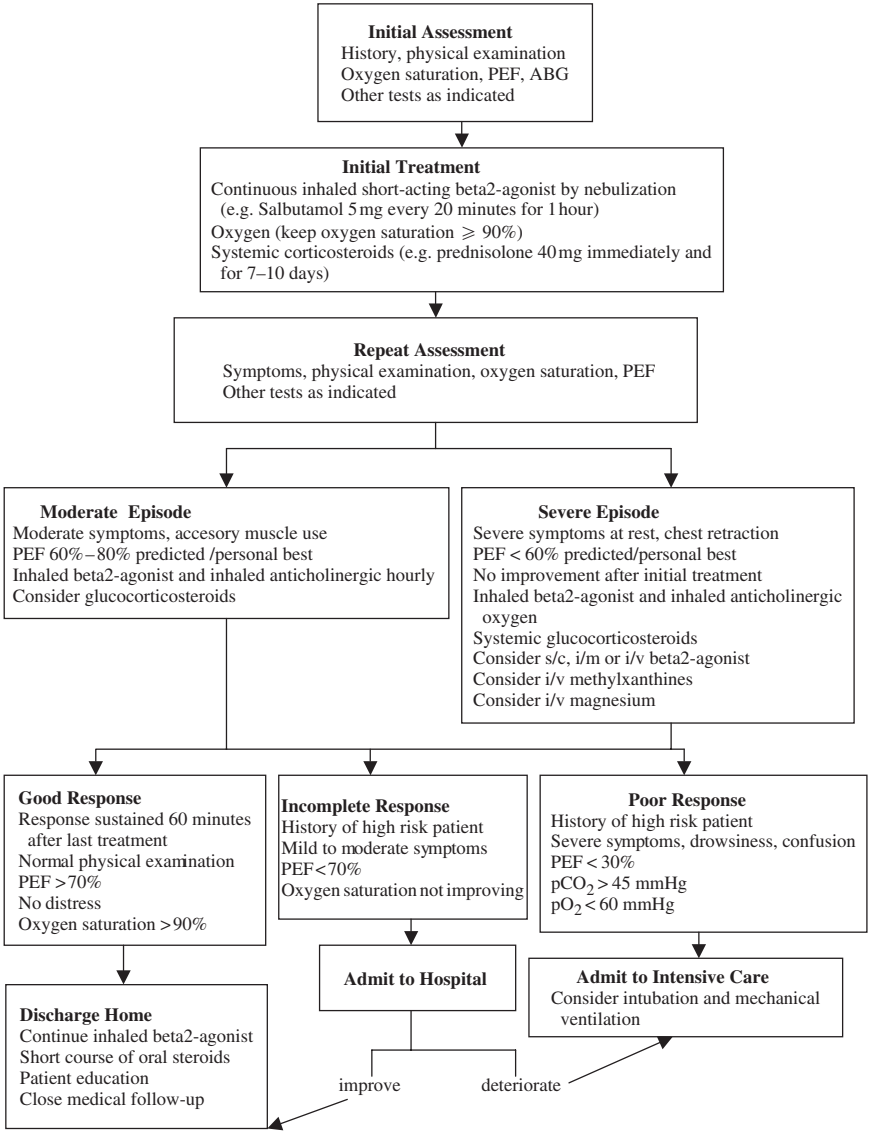


Fig. 1 Management of acute asthma in hospital.

medications and drug delivery devices is mandatory. Patients should also be advised to avoid or control allergens, irritants or other factors that worsen their asthma. Rhinitis, sinusitis and gastroesophageal reflux should be treated when present. Annual influenza vaccinations are recommended for patients with persistent asthma.

OCCUPATIONAL ASTHMA

Occupational asthma is defined as a disease characterized by variable airflow limitation and/or airway hyper-responsiveness due to causes and conditions attributable to a particular occupational environment and not to stimuli encountered outside the workplace. The term "work-related asthma" has been used for cases of pre-existing asthma, which are aggravated by the working environment.

Two types of occupational asthma are distinguished by whether they appear after a latency period. Sensitizer-induced asthma is characterized by a variable time during which "sensitization" to an agent present in the worksite takes place. Irritant-induced asthma occurs without a latent period after substantial exposure to an irritating dust, mist, vapour or fume. Reactive airways dysfunction syndrome (RADS) is a term used by some to describe irritant-induced asthma caused by short-term, high-intensity exposure.

Occupational asthma is the most common occupational respiratory disease in Singapore. In Singapore, occupational asthma is a legally notifiable industrial disease under the Factories Act and a compensable occupational disease under the Workmen's Compensation Act. It is a condition associated with disability in the workplace and may still be largely under-reported.

More than 250 agents capable of causing occupational asthma have been reported. Substances that cause occupational asthma are classified either as high molecular weight allergens or low molecular weight compounds. High molecular weight agents that cause occupational asthma include laboratory animal allergens, flour, detergent enzymes and fish and seafood protein. Examples of low molecular weight agents include acid anhydrides, metals, isocyanates, western red cedar, amines, colophony and antibiotics. The most common causative agent in Singapore is isocyanates. Other common causative agents include solder flux, welding fumes and wood dust.

A detailed occupational history and a knowledge of the common causative agents and association with various occupations and industries are helpful in identifying possible causative agents. Work-relatedness may be suspected based on the history demonstrating improvement when away from work and onset of symptoms during working periods.

The diagnosis of occupational asthma is established by demonstrating the presence of asthma, relationship between the asthma symptoms

and work, and exposure to a specific causative agent. Serial peak expiratory flow rate monitoring during periods at work and away from work is a useful tool in documenting work-relatedness and has been found to be both sensitive (86%) and specific (89%) for the diagnosis of occupational asthma (Fig. 2). The gold standard for the diagnosis of occupational asthma is a positive specific bronchial provocation test to the causative agent. It may not be necessary to require a specific bronchial provocation test to diagnose or confirm every case of occupational asthma. Besides the time and expense that is required from the patient and the investigating doctor, there are risks involved as well. Indications for specific bronchial provocation testing include: 1) to document new causative agents; 2) to distinguish between multiple known agents; and 3) to provide objective evidence in a difficult case or where it is not possible to do a serial peak expiratory flow rate monitoring (e.g. in a patient who has already left employment).

It has been recommended that all workers confirmed to have occupational asthma be permanently transferred to a job with totally no exposure to the causative agent. It may also be possible to substitute the offending agent with a safer one. Improved local exhaust ventilation and enclosure of specific processes may also be helpful. With irritant-induced asthma, the use of personal protective equipment may lower exposures to levels that do not induce bronchospasm. With sensitizer-induced asthma, however, the worker should be precluded from further exposure to the sensitizing agent. It may be necessary to completely remove the worker from

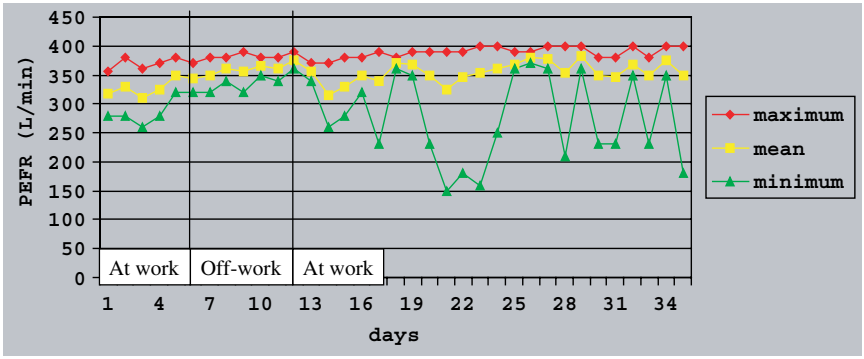


Fig. 2 Serial peak expiratory flow rate (PEFR) monitoring in a worker with occupational asthma.
Note the PEFR improvement at home and deterioration at work.

the workplace because even exposure to minute quantities of the offending agent may induce bronchospasm. Prevention of additional cases of occupational asthma should be considered in all work places in which cases are diagnosed. Protection of workers by the use of appropriate ventilation systems, respiratory protective equipment, and education about appropriate procedures should be recommended.

REFERENCES

Ng TP, Adult asthma prevalence, morbidity and mortality and their relationships with environmental and medical care factors in Singapore, *Asian Pac J Allergy I* **17**:127–135, 1999.

National Asthma Education Program. *Expert Panel Report 2. Guidelines for the diagnosis and management of asthma*, Office of Prevention, Education and Control, National Heart, Lung and Blood Institute, National Institutes of Health, US Dept of Health and Human Services, Public Health Service, Bethesda, NIH Publication No. 97–4051, 1997.

Global initiative for asthma, *Global strategy for asthma management and prevention*. National Institutes of Health. National Heart, Lung and Blood Institute, 2000.

Bernstein IL, Chan-Yeung M, Malo JL, Bernstein DI, *Asthma in the workplace*, Marcel Dekker Inc., New York, 1993.

Chan-Yeung M, Assessment of asthma in the workplace, *Chest* **108**:1084–1117, 1995.

Lee HS, Chee CB, Clinical update on occupational asthma, *Ann Acad Med Singapore* **30**:546–550, 2001.

Kor AC, Lee HS, Chee CB, Wang YT, Occupational asthma in Singapore, *Singapore Med J* **42**:373–377, 2001.

This page intentionally left blank

76

Chronic Obstructive Pulmonary Disease

Loo Chian Min

INTRODUCTION

Chronic bronchitis and emphysema are two distinct diseases characterized by chronic airflow limitation. They are often present in combination and are collectively termed as chronic obstructive pulmonary disease (COPD). COPD is one of the leading causes of death worldwide, ranking fourth or fifth in developed countries. It has been increasing in prevalence and mortality, and is projected to be the third leading cause of death worldwide by 2020. In Singapore, bronchitis, emphysema and asthma as a group account for 0.8% of total deaths and is the eighth principal cause of death. Although there are similarities between COPD and asthma, it is important to differentiate them as treatment and prognosis differ.

DEFINITION

COPD is a disease state characterized by airflow limitation that is not fully reversible. Airflow limitation is usually progressive and associated

with an abnormal inflammatory response of the lungs to noxious particles or gases. Chronic bronchitis is defined clinically as the presence of a chronic productive cough on most days, for a minimum of 3 months, for at least 2 consecutive years, in a patient in whom other causes of chronic cough have been excluded. Emphysema is defined anatomically as abnormal, permanent enlargement of the airspaces distal to the terminal bronchiole, with destruction of their walls and without obvious fibrosis. Airway obstruction due to other specific etiologies, like bronchiectasis and obliterative bronchiolitis, is not classified as COPD. Figure 1 shows the relationship between emphysema, chronic bronchitis and asthma. It may sometimes be difficult to differentiate asthmatics with partially reversible airway obstruction from COPD with airway hyperreactivity. This group of patients is sometimes classified as COPD.

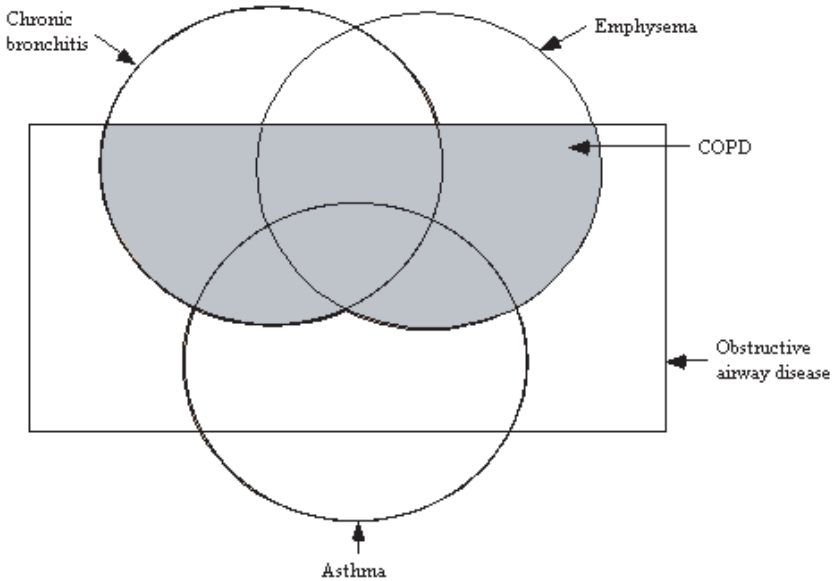


Fig. 1 Venn diagram showing the relationship between emphysema, chronic bronchitis and asthma. It is possible for asthma, chronic bronchitis or emphysema to have no airway obstruction. There is a subset of asthma that is difficult to differentiate from COPD and is sometimes classified as COPD (see text).

Table 1 Risk Factors for COPD

Host factors	Genes e.g. alpha-1 anti-trypsin deficiency Airway hyper-responsiveness Lung growth
Environmental factors	Tobacco smoke Occupational dust and chemicals e.g. coal Air pollution Socioeconomic status Infections

RISK FACTORS

Evidence concerning risk factors for COPD came from epidemiological studies that identified associations rather than cause-effect relationship. As such, much is still unknown regarding risk factors and development of COPD. Table 1 shows the risk factors for developing COPD.

Tobacco smoke is by far the most important risk factor for developing COPD. In fact, diagnosis needs to be carefully considered in a non-smoker who has been labelled with COPD. However, not every smoker will develop COPD. It is believed that genetic predisposition is important for the development of COPD. Alpha1 anti-trypsin deficiency a well studied gene abnormality that causes emphysema. Other candidate genes like gluthatione S-transferase P1, micosomal epoxide hydrolase, tumor necrosis factor α are still under investigation. Occupations like coal and gold miners and farmers have been shown to be associated with higher risk of developing COPD. The exact role of air pollution, airway hyper-responsiveness, socioeconomic status and infection is still unclear.

PATHOGENESIS

The pathogenesis of COPD is still not fully understood. Data are mainly obtained from animal models and *in vitro* studies. Three mechanisms are thought to be important in the pathogenesis of COPD:

1) airway inflammation;
2) imbalance of proteinases and anti-proteinases in the lung; and
3) oxidative stress.

Noxious gases and particles are thought to cause lung inflammation. There is increased production of proteinases as well as oxidative injury.

These are normally counteracted by anti-oxidants and antiproteinases. However, there seems to be an imbalance in these counter mechanisms in some patients. If repeated damage cannot be adequately repaired by the body, chronic bronchitis or emphysema may develop. This is indeed simplified as the actual mechanism is complex and there is a lot of interplay between the different mechanisms. How other environmental and host factors affect the pathogenesis is unclear.

The main inflammatory cells that are found in airways of COPD are CD8 T lymphocytes, macrophages and neutrophils, producing mediators like interleukin 8, tumor necrosis factor-α, leukotriene B4, etc. In contrast, the inflammatory cells (eosinophils, CD4 T lymphocytes, activation of mast cells) found in asthma are different from that of COPD. It is because of different inflammatory response that occurs in these conditions. The best example of proteinase-anti-proteinase imbalance is that of α-1 anti-trypsin deficiency. In this genetic disorder (rare in Singapore), α-1 anti-trypsin, which inhibits a number of proteinases like neutrophil elastase, is deficient. As a result, these individuals are at higher risk of developing emphysema. There is data to suggest that imbalance in these enzymes may either be due to over production of proteinases or under production of anti-proteinases. There is now evidence to suggest that oxidative stress is important in the development of COPD. There is overproduction of oxidants with respect to anti-oxidants. Markers of oxidative stress like nitric oxide (NO) and hydrogen peroxide are found to be increased in breaths of COPD patients.

PATHOPHYSIOLOGY AND CLINICAL CORRELATION

The pathological changes in COPD are correlated with the physiological abnormalities and clinical features. Physiological abnormalities usually develop in the following order:

- mucous hypersecretion;
- ciliary dysfunction;
- airflow limitation;
- lung hyperinflation;
- gas exchange abnormalities;
- pulmonary hypertension; and
- cor pulmonale.

Chronic inflammation leads to goblet cell hyperplasia, which causes increased mucous production. Chronic inflammation also causes squamous metaplasia that leads to ciliary dysfunction. At this early stage of disease, patient may only have cough and increased sputum production. Irreversible narrowing of peripheral airways (small bronchi and bronchioles) occurs in COPD due to remodeling (fibrosis and narrowing). This is the main cause of airflow limitation in COPD. Two other causes are loss of elastic recoil of the lung and small airway collapse. The former is due to parenchymal destruction; the reduced elastic recoil reduces the driving pressure in the alveoli during expiration. Small airways collapse when their alveolar attachment and support are lost, causing early airway closure during expiration. Hyperinflation of the lungs occurs due to inadequate emptying. As airway obstruction worsens, patient will begin to experience increasing dyspnea. Reversible airway obstruction may be contributed by smooth muscle contraction and airway inflammation. As not every part of the lung is equally affected, the distribution of ventilation and perfusion will be heterogeneous. Ventilation-perfusion mismatch will then occur, contributing to gas exchange abnormalities. As more alveoli and interstitium are destroyed, there will be a loss of vascular bed and reduced surface area for gas exchange. These also contribute to gas exchange abnormalities in later stages of COPD. Eventually pulmonary hypertension develops secondary to chronic hypoxia and loss of vascular bed. Cor pulmonale occurs when there is right heart failure secondary to pulmonary hypertension, with the patient experiencing fluid retention.

CLINICAL FEATURES

The cardinal feature of COPD is dyspnea. This is the first presentation in most COPD patients. In contrast to asthma where dyspnea is episodic, COPD tends to have progressive dyspnea with periods of exacerbation. Exertional dyspnea is a common complaint and usually begins when a patient is in his sixties. This may be present for many years before the patient seeks medical attention. In the early stage, patient often notices dyspnea only on strenuous activities. It may be the only symptom in early COPD. This gradually progresses with time and eventually dyspnea is noticeable even with minimal exertions. In late stages, dyspnea may be present at rest, with severe limitation of activities. Wheeze may be

noticed by patients and may be present in the absence of acute exacerbation. This can be confused with asthma. Cough productive of mucoid sputum are frequent associations. However, purulent sputum should alert the doctor to possible concomitant respiratory infection. Orthopnea, paroxysmal nocturnal dyspnea and leg swelling are non-specific symptoms. The first 2 symptoms could occur in later stages of COPD. All the 3 symptoms could also be present in cor pulmonale secondary to pulmonary hypertension or concomitant coronary heart disease with congestive cardiac failure. Weight loss and poor appetite are also common in the late stages of COPD. A history of tobacco abuse is present in most patients with COPD. One has to be careful in diagnosing COPD in a person without a smoking history. Other causes of COPD, e.g. occupational exposure, should also be sought for.

In earlier stages, physical examination may be unremarkable. Nicotine stains of the finger nails may be present as clues to tobacco abuse. In later stages, hyperinflation of the chest wall may be present with use of accessory muscles of respiration. Purse lip breathing, retraction of intercostal muscles and paradoxical abdominal movement may be present during respiratory distress. Cyanosis may be seen in severe hypoxemia. Asterixis can be present in those with respiratory failure. Bounding pulse and conjunctival hyperemia are signs of carbon dioxide retention. There will be raised jugular venous pressure and ankle edema in patients with cor pulmonale. However, ascites and anarsarca are relatively uncommon. Percussion of the lungs will show increased resonance due to air trapping and hyperinflation. Lung auscultation frequently just reveals decreased breath sounds with prolonged expiratory phase. Rhonchi may be heard in some patients, especially during an acute exacerbation. Loud pulmonary second sound may be present on cardiac auscultation in patients with pulmonary hypertension and heart sounds are frequently distant.

INVESTIGATIONS

Pulmonary Function Studies

Spirometry is central in the diagnosis, severity grading and follow-up of COPD. All suspected COPD patients should have spirometry performed to confirm airflow limitation. An obstructive pattern as demonstrated by reduced FEV_1 / FVC with varying decrease of FEV_1 is expected. FEV_1 is easily performed, has less variability and has more accurately predictable

Table 2 NHLBI/WHO Classification of Severity of COPD

Stage	Characteristics
0: At risk	Normal spirometry Chronic symptoms: cough, sputum
1: Mild COPD	FEV_1 / FVC < 70% $FEV_1 \geq 80\%$ predicted ±Chronic symptoms
2: Moderate COPD	FEV_1 / FVC < 70% $30\% \leq FEV_1 < 80\%$ predicted 2a: $50\% \leq FEV_1 < 80\%$ predicted 2b: $30\% \leq FEV_1 < 50\%$ predicted ±Chronic symptoms
3: Severe COPD	FEV_1 / FVC < 70% $FEV_1 < 30\%$ predicted, or, $FEV_1 < 50\%$ predicted + respiratory failure or cor pulmonale

normal values that other parameters of airway dynamics like peak flow or airway resistance. Bronchodilator challenge should also be performed to assess for airflow reversibility which occurs in asthma but not in COPD. A change in FEV_1 of 12% and 200 ml is significant by American Thoracic Society (ATS) criteria. FEV_1 has been shown to correlate with mortality from COPD in previous studies. Table 2 shows a classification of severity of COPD adopted by NHLBI/WHO.

Total lung capacity (TLC) is increased in COPD consistent with hyperinflation. Residual capacity (RV) and RV/TLC ratio are both increased, demonstrating air-trapping in this disease. Diffusing capacity is reduced in more severe emphysema due to interstitial destruction and loss of alveolar capillaries.

Chest Radiograph

A chest radiograph should be performed in the initial assessment of a new undiagnosed COPD patient or in an existing COPD patient with acute respiratory symptoms. In frontal and lateral chest radiographs, features of hyperinflation like low flattened diaphragm, increased retrosternal airspace and a long narrow heart can be seen. Hyperlucent lungs with reduced vascular shadowing and hyperinflation are features of emphysema. Bullae may be seen and represents local disease. In pulmonary hypertension, prominent hilar vascular shadows may be present.

CT Thorax

CT thorax has no role in the routine management of COPD.

Arterial Blood Gas Measurements

Arterial blood gas (ABG) results may be normal in early stages of COPD. As the disease progresses, worsening hypoxemia occurs. This is later followed by hypercapnia and compensatory metabolic alkalosis. Blood gas abnormalities can also occur during acute exacerbations.

NATURAL HISTORY AND PROGNOSIS OF COPD

The natural history of COPD is well-studied. It is generally slowly progressive, with many patients remaining asymptomatic for many years. Symptoms usually begin when patients are in their fifties or sixties. The exception to this is α_1 antitrypsin deficiency, where emphysema occurs 10–20 years earlier. The following factors have been shown to be associated with poorer prognosis:

- advanced age;
- low FEV_1;
- tobacco smoking;
- hypoxemia;
- hypercapnia; and
- pulmonary hypertension or cor pulmonale

Studies have differing survival rates and this is related to patient selection criteria.

MANAGEMENT

Acute Exacerbation

Acute exacerbation is suspected when symptoms like dyspnea, cough and sputum are increased. In the assessment of COPD exacerbation, it is important to try to identify and correct precipitating factors:

- infection;
- cigarette smoking;
- air pollution; and
- allergen exposure.

However, a precipitating factor may not be found in up to one-third of patients. Also, one must not forget that other concomitant medical conditions like heart failure, pneumonia, pneumothorax, pulmonary embolism, etc. may mimic a COPD exacerbation.

In the assessment of an acute exacerbation, it is very important to ascertain the baseline functional status, arterial blood gas and spirometry measurements. They form the basis for comparison and are important for subsequent management of the patient. Table 3 gives a summary of the history, physical findings and investigations of patient with acute COPD exacerbation.

There is increase in breathlessness that is frequently accompanied by more cough and sputum. Fever and purulent sputum indicate an infective exacerbation. Wheeze may not always be present. Other associated symptoms include poor appetite, reduced exercise tolerance and malaise. Drowsiness may indicate acute carbon dioxide retention. Physical examination frequently shows tachypnea and use of accessory muscles of respiration. Fever and wheeze may be present. Raised jugular venous pressure and ankle edema are present in heart failure.

Investigations will depend on the severity of exacerbation. None may be necessary during a mild exacerbation. In a more severe exacerbation, relevant investigations will help in the assessment as well as aid treatment of these patients. Spirometry during an acute exacerbation is usually not required and often does not add information that affects management.

Indications for hospital admission include:

- marked worsening of symptoms;
- severe baseline COPD;
- poor response to initial treatment;
- old age;
- significant comorbidities;
- new arrhythmias;
- hemodynamic instability;
- respiratory failure; and
- poor home support.

Indications for ICU admission include:

- worsening hypoxemia while on high FiO_2;
- worsening respiratory acidosis and hypercapnia despite treatment;

Table 3 Assessment of COPD Exacerbation

History	**Findings**
1) Baseline status	It is important to ascertain baseline functional status, arterial blood gas and spirometry as a basis for comparison
2) Severity of new or worsening symptom	This has to be compared to the baseline of each patient. A big increase indicates a more severe exacerbation
3) Response to any prescribed treatment	Patients who are not responding to treatment tend to have a more severe exacerbation
4) Number of previous hospitalizations	Patients requiring frequent hospitalization for COPD exacerbation are at higher risks
5) Fever and purulent sputum	Fever and purulent sputum indicate infection
6) Wheeze	Wheeze is not related to severity or cause of exacerbation, and may not be present
7) Home oxygen therapy	This indicates a more severe baseline COPD and allows comparison of severity of hypoxemia
8) Usual medication	Gives an indication of baseline and present starting bronchodilator dose and frequency
Sign	**Findings**
1) Mental status	Drowsiness indicates a severe exacerbation and possible acute hypercapnia
2) Tachypnea/ability to speak & converse	Gives an indication of severity of dyspnea. Respiratory rate > 30/min indicates severe exacerbation
3) Fever	Suggests the presence of infection
4) Wheeze	Wheeze may be absent
5) Raised JVP and edema	Indicates right heart failure
6) Cyanosis	Indicates severe hypoxemia
7) Use of accessory muscles of respiration	Frequently present in an exacerbation
8) Paradoxical chest wall or abdominal movement	Present in severe exacerbation
9) Hemodynamic instability	Needs admission to intensive care unit
Investigations	**Findings**
• Full blood count	Raised white cell suggests infection. Polycythemia may be present in chronic hypoxemia.
• Urea, creatinine, electrolytes	Useful baseline. Important to assess for hypokalemia in frequent β_2-agonist or diuretic use
• Arterial blood gas	Assess both oxygen and ventilation. Respiratory acidosis indicates a severe exacerbation.

Table 3 (*Continued*)

Investigations	Findings
• Pulse oximetry	It is a useful non-invasive means of checking oxygenation but it does not give information on the level of carbon dioxide
• Serum theophylline level	It is useful for patients treated with theophylline as it has low therapeutic index
• Sputum gram stain and culture	It is useful if a respiratory tract infection is suspected
• Chest radiograph	It helps to assess for alternative diagnosis
• Electrocardiogram	It helps to assess for cardiac ischemia, arrhythmia and right heart hypertrophy

- severe respiratory distress that is not responding to initial therapy;
- hemodynamic instability;
- life-threatening arrhythmia; and
- impaired mental status.

Mild exacerbations may be managed as an outpatient but more severe ones require hospitalization. Table 4 summarizes the management of acute exacerbation of COPD.

Bronchodilator therapy

Inhaled bronchodilator therapy remains the cornerstone of treatment. Combined short-acting β_2-agonist and anti-cholinergic are frequently used although β_2-agonist alone may be used initially. There is presently no strong evidence that combining both drugs improve outcome. Inhalation via spacer or nebuliser may be used although dyspneic patient may prefer nebuliser therapy. Frequency and/or dose need to be increased from baseline.

Methylxanthines

Routine use of methylxanthines is not warranted in acute exacerbation of COPD. Aminophylline or theophylline may be considered in severe exacerbation that is not responding to first line treatment.

Steroids

Glucocorticoid therapy has been shown to be beneficial in the treatment of an acute COPD exacerbation. A course of up to 14 days therapy with 30–40 mg daily of prednisolone may be used. There is no role for inhaled steroids in the treatment of acute exacerbation of COPD.

Antibiotics

Antibiotics are only indicated in the presence of bacterial infection. Purulent sputum alone is sufficient for commencement of antibiotic therapy. Sputum culture may be useful to guide antibiotic therapy if positive. *Streptococcus pneumoniae*, *Hemophilus influenzae* and *Moraxella catarrhalis* are the most common bacteria involved. It may also be considered in patients with severe exacerbations.

Adjunct therapy

Oxygen therapy is useful to correct hypoxemia. However, oxygen therapy needs to be controlled as some patients may develop hypercapnia and acute respiratory acidosis. Reassessment of arterial blood gas after 30 minutes of oxygen therapy is useful. Oxygen therapy may be monitored using pulse oximetry or arterial blood gas. The later is indicated when hypercapnia needs to be monitored. Adequate nutrition and hydration should be maintained. Tube feeding and intravenous hydration may be necessary in a dyspneic patient. There is no study to show that mucolytics or routine chest physiotherapy is useful in acute exacerbation of COPD. Chest physiotherapy may be considered in patients who produce >25 ml of sputum a day or when atelectasis is present.

Mechanical ventilation

Invasive or non-invasive mechanical ventilation may be indicated in severe exacerbation. Non-invasive positive pressure ventilation (NIPPV) has been shown to improve patient outcomes in acute exacerbation of COPD. Selection of patient is important for a successful outcome. Invasive mechanical ventilation is necessary for those who fail NIPPV, or in those who are too ill for NIPPV. However, before this is instituted, it is important to consider the risk and benefit of mechanical ventilation.

Table 4 Management of Acute Exacerbation of COPD

Bronchodilator therapy
- Inhaled short-acting β_2-agonist, e.g. salbutamol
- Inhaled anti-cholinergic, e.g. ipratropium bromide
- Delivery either through spacer or nebulizer
- They may be combined
- Routine use of methylxanthines is not warranted

Glucocorticoid
- Intravenous hydrocortisone or methylprednisolone

Antibiotics
- Antibiotic is indicated only when there is evidence of bacterial infection, clinical or laboratory
- Antibiotic choice: β-lactam, β-lactam + β-lactamase inhibitor, new generation macrolides, doxycycline, newer generation fluoroquinolones

Adjunct therapy
- Controlled supplemental oxygen
- Chest physiotherapy only if there is atelectasis or sputum production > 25 ml
- Nutrition
- Hydration

Mechanical ventilation
- Indicated when ventilation cannot be sustained by patient, worsening gas exchange despite treatment

The main concern in a severe COPD patient is whether he will become ventilator dependent. Unfortunately, there is no foolproof way of predicting such an outcome. Degree of airway obstruction does not accurately predict ventilator dependence and should not be used alone for making a decision on whether to institute mechanical ventilation. Other considerations include social and cultural factors, comorbid conditions, baseline performance status, patient and family wishes or expectations, and quality of life.

Stable COPD

The aims of treatment of stable COPD are to:

1) control symptoms;
2) improve quality of life;
3) improve exercise tolerance;
4) prevent exacerbations;
5) prevent and treat complications;

6) prevent progression of disease; and
7) reduce mortality.

Most of the treatment options address the first 5 points, but only smoking cessation and oxygen therapy had been able to prevent progression of disease and reduce mortality from COPD. Treatment should be stepwise and gradually increased depending on the disease severity. Both pharmacological and non-pharmacological components of treatment should be addressed.

Bronchodilators

Bronchodilator therapy is the main pharmacological treatment of COPD. Short-acting inhaled β_2-agonists are given as necessary in patients who have mild COPD with few or intermittent symptoms. Those patients with moderate COPD with regular symptoms should be given regular inhaled anticholinergics and/or short-acting β_2-agonists. Long-acting β_2-agonists may be added in patients who do not respond to the above treatment and theophylline. However, studies have not consistently shown these agents to be beneficial in stable COPD.

Methylxanthines

Long-acting theophylline may be added if symptoms are not well controlled on regular inhaled bronchodilators. Its mode of action in COPD is unknown. It only has a mild bronchodilator effect. Its main effect could be improvement in diaphramatic function.

Glucocorticoids

Large studies have not shown that chronic inhaled glucocorticoids improve lung function or quality of life in stable COPD. However, inhaled glucocorticoids may be considered in these subsets of COPD patients:

- those who exhibit positive bronchodilator response in spirometry, i.e. 200 cc and 12% improvement in FEV_1 (ATS criteria) or 15% increase in FEV_1 alone (European Respiratory Society criteria);
- those who had significant improvement in spirometry (200 cc and 15% increase in FEV_1 by ATS criteria) after a trial of 2 weeks of 0.5–1 mg daily of prednisolone or 2–3 months of moderate to high dose of inhaled glucocorticoids; and

- those with $FEV_1 < 50\%$ with frequent exacerbations requiring antibiotics or glucocorticoids. In this group, treatment should be discontinued if no improvement is noted after an adequate trial of treatment.

Vaccination

Influenza vaccination has been shown to reduce mortality and serious illness in COPD patients by 50%. It is recommended every autumn in temperate countries. However, it is uncertain whether influenza vaccination should be routinely advised in the Singapore context as there is no local seasonal change and the epidemiology of influenza in Singapore is unknown. Pneumococcal vaccination using 23-valent vaccine may reduce mortality and morbidity in COPD patients.

Others

$Alpha_1$ antitrypsin augmentation may be prescribed for patients with proven α_1 antitrypsin deficiency. There is presently no evidence that respiratory stimulant, mucolytics, antioxidants or prophylactic antibiotics is beneficial in COPD. Vasodilator therapy has not been shown to be effective in pulmonary hypertension secondary to COPD. Chronic use of antitussives is also not recommended in COPD.

The non-pharmacological aspect of COPD management needs to go hand-in-hand with pharmacological treatment. To date, only oxygen therapy and smoking cessation have been shown to retard disease progression and improve mortality in indicated patients with COPD.

Oxygen

Long-term home oxygen therapy is indicated in chronically hypoxemic patients with stable COPD. Although supplemental oxygen may be prescribed during an acute exacerbation, its need should be reassessed when patient has become stable and reached his baseline. Only when the patient continues to be hypoxemic should long-term oxygen therapy be advised. Table 5 shows the criteria for long-term home oxygen therapy.

Oxygen should be used for more than 15 hours a day and titrated to $PaO_2 \geq 60$ mmHg or $SaO_2 \geq 90\%$. Long-term use improves survival, pulmonary hemodynamics, exercise capacity, mental alertness and lung

Table 5 Criteria for Long-Term Home Oxygen Therapy

Assessment must be made when patient is clinically stable and on appropriate therapy
$PaO_2 \leq 55$ mmHg or $SaO_2 \leq 88\%$ (with or without hypercapnia)
$PaO_2 \leq 56$–59 mmHg or $SaO_2 \leq 89\%$ with either of the following: 1) Pulmonary hypertension 2) Edema due to heart failure 3) Hematocrit $> 55\%$

PaO_2 — arterial oxygen partial pressure, SaO_2 — arterial oxygen saturation.

mechanics. For these patients, titration can be performed for exercise and sleep although empiric increment by 1–2 litres per minute is also acceptable. There is currently no evidence that oxygen use for hypoxemia during sleep or exercise without daytime resting hypoxemia improves survival.

Smoking cessation

Smoking cessation is recommended for all COPD patients who continue to smoke. Physicians should routinely inquire about smoking habits and provide advice on the effects of tobacco for those who continue to smoke. These patients should be assessed for readiness and encouraged to quit smoking on every clinic visit. Those who are ready should either be referred to dedicated smoking cessation clinics or be counseled on smoking cessation techniques. Patient should be encouraged to set a quit date rather than reducing gradually. Nicotine replacement therapy (NRT) and slow-release bupropion have been shown to be effective in smoking cessation. Formulations of NRT include transdermal patch, intranasal device, inhaler, lozenges and gum. There were concerns regarding the use of NRT in patients with cardiovascular disease but studies have shown it to be safe in these patients as the delivered nicotine dose is much less than that of cigarettes. Second-line drugs in smoking cessation include nortriptyline and clonidine, but they should be used with care and by experienced physicians. The safety of drugs for smoking cessation has not been studied in pregnancy. Pregnant smokers should be given psychosocial intervention and routine use of pharmacotherapy cannot be recommended in this group of patients.

Pulmonary rehabilitation

Pulmonary rehabilitation is a multidisciplinary and multimodality program that aims to achieve and maintain maximum level of independence and functioning in a patient with chronic lung disease. Although it has been widely studied in COPD, its efficacy in other chronic lung diseases has not been well investigated. It has been shown to improve exercise endurance and capacity, quality of life, health-related cost, reduce dyspnea and hospitalization. It has no effect on physiological parameters like lung function and gas exchange. COPD patients often follow a vicious cycle where dyspnea leads to inactivity, which leads to skeletal muscle atrophy, which leads to increased dyspnea and further physical inactivity. Pulmonary rehabilitation is an intervention to break this cycle. All patients who are with dyspnea should be referred for pulmonary rehabilitation. Patients should be assessed prior to entering a program to ensure there are no contraindications like unstable comorbid conditions or dyspnea due to non-COPD causes. A program usually lasts 4–8 weeks. Central to a program will be exercise training, which consists of both strengthening and endurance exercises. The program is individually tailored to the patient's requirements. Exercises are gradually increased and patients are taught to continue with their own exercises even when the program is completed. Other components of pulmonary rehabilitation include education, dietary advice, smoking cessation, psychological counseling and self management skills. Table 6 shows the various team members of a pulmonary rehabilitation program.

Table 6 Members of a Pulmonary Rehabilitation Program

Team member	Component
Physician	Team leader. Provide medical input, interpret exercise test, evaluate outcome
Physiotherapist	Exercise training, teach breathing techniques, chest therapy as needed
Nurse	Coordinate activities, organize program, patient education
Psychologist	Counseling, smoking cessation, stress management, teach relaxation techniques
Pharmacist	Educate drug usage and supervise delivery devices
Dietician	Dietary advice
Occupational therapist	Vocational and activities-of-daily-living (ADL) counseling

Education

Patient education is important in any chronic disease. Although there are no studies to show that it improves outcome, it is certainly helpful in improving the ability to cope with the illness. The scope of education will depend on social and cultural factors. Education should include the following topics:

- nature of the disease and complications;
- medication use, including inhaler techniques;
- how to cope with symptoms;
- recognition and treatment of exacerbations; and
- end of life issues and advanced directives.

Mechanical ventilation

Long-term mechanical ventilation may be necessary in advanced COPD when it becomes impossible for patients to maintain adequate gas exchange. Depending on the degree of compromise, patients may need continuous or partial ventilatory support. The latter is usually achieved with nocturnal NIPPV. Studies on long-term continuous ventilator support have shown good survival. However, quality of life is definitely affected. Whether this is a viable option depends on the cultural and social background, home and family support of the patient.

Bullectomy

Bullae are defined as abnormally dilated airspaces of >1 cm in diameter within the lung parenchyma. The reason for bulla formation is unknown but is thought to be due to ball-valve effect where obstruction of airway leads to progressive air-trapping and dilatation of distal airspaces that are already damaged by emphysema. Giant bulla that causes compression of the adjacent lung tissue may be excised if it is thought to contribute to dyspnea. Careful selection of patient is necessary. CT thorax, radionuclide ventilation-perfusion scan and lung volume measurements by plethysmography and inert gas dilution is necessary for evaluation. Poorer results of surgery occur in patients where bullae occupy less than one-third of the hemithorax, and in those with diffuse emphysema or chronic bronchitis.

Lung volume reduction surgery (LVRS)

Lung volume reduction surgery is a palliative surgical procedure for selected patients with pulmonary emphysema. It has been shown to improve dyspnea, quality of life, exercise tolerance and lung mechanics. Improvement in gas exchange is not consistent among studies. Long-term follow-up studies have not shown any reduction in mortality in COPD. The mechanisms of improvement are hypothesised to be improved lung elastic recoil, improved respiratory muscle function, and correction of ventilation-perfusion mismatch. Patient eligibility must be carefully assessed and pulmonary rehabilitation before surgery is compulsory. Patients should have significant dyspnea despite maximum medical therapy, pulmonary emphysema with minimum airway hyperreactivity, and predominantly upper lobe disease with lower lobe compression. Recent data shows that patients with very low FEV_1 and diffusing capacity have high mortality after LVRS. Lung transplant is indicated in this group of patients instead. Surgery may be performed via median sternotomy, bilateral thoracotomy or video-assisted thoracoscopy. Approximately one-third of both upper lobes are resected in this procedure to allow the lower parts of the lungs to expand and restore normal pulmonary mechanics. Studies show that improvement in lung function may last for 2–3 years before deteriorating again. It may be used as a bridge to lung transplantation.

Lung transplantation

Lung transplantation is appropriate in patients with end-stage COPD that is unresponsive to medical treatment. Indication for lung transplantation in COPD include dyspnea of class 3 by New York Heart Association classification, $FEV_1 < 25\%$ predicted, $PaO_2 < 60$ mmHg, $PaCO_2 > 45$ mmHg, secondary pulmonary hypertension, rapid decline in lung function or repeated life-threatening exacerbations. The inclusion and exclusion criteria are complex and patients need to be carefully evaluated and selected. Patients who are suitable for lung transplant have poorer pulmonary physiology than those who are eligible for LVRS. Data from the United Network of Organ Sharing in USA suggests that lung transplantation after 2 years may not confer survival benefit for end-stage COPD. Either single or bilateral sequential lung transplant may be performed for COPD. Common complications after lung transplantation

are operative mortality, acute and chronic rejection, and opportunistic infections. At present, the main problem in lung transplantation worldwide is the shortage of donor organs.

SUMMARY

Chronic obstructive pulmonary disease (COPD) usually comprises of various degrees of both chronic bronchitis and emphysema. It is characterized by chronic airflow limitation that is not fully reversible. The main cause of COPD is tobacco smoking. It has a progressive course with gradually worsening lung function and symptoms. There are exacerbations that tend to become more frequent as the disease progresses. The cardinal symptom is dyspnea. Other possible associated symptoms are coughing, sputum, reduced exercise tolerance and wheezing. It can sometimes be confused with asthma. Relevant investigations in the assessment of COPD include chest radiograph and spirometry. Arterial blood gas assessment may be necessary in the later stages of COPD. Management should include both pharmacological and non-pharmacological treatment. Inhaled bronchodilators (short-acting β_2-agonist and anticholinergic) are the form of treatment. A step-wise approach should be adopted in the pharmacological treatment. A short course of systemic steroid improves outcome in an acute exacerbation. Antibiotic is only indicated if bacterial infection is suspected. A trial of inhaled steroids in chronic COPD treatment may be given if indicated. Smoking cessation should be advised for current smokers. Pulmonary rehabilitation has been shown to improve quality of life, dyspnea score and exercise tolerance of symptomatic COPD. However, it does not alter the course of the disease. Long-term oxygen therapy, when indicated, reduces mortality in severe COPD. Lung volume reduction surgery and lung transplantation are only useful in selected patients with end-stage COPD on maximal medical therapy.

REFERENCES

1. National Heart, Lung and Blood Institute/World Health Organisation, Global Initiative for Chronic Obstructive Lung Disease, Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease, NHLBI/WHO Workshop Report, NHLBI Publication Number 2701, March 2001.

2. American Thoracic Society, Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease, *Am J Respir Crit Care Med* **152**:S84–S96, 1995.

3. Siafakas NM, Vermeire P, Price NB, *et al*., Optimal assessment and management of chronic obstructive pulmonary disease (COPD): The European Respiratory Society Task Force, *Eur Respir J* **8**:1398–1420, 1995.

4. The COPD Guideline Group of the Standards of Care Committee of the BTS, BTS guidelines for the management of chronic obstructive pulmonary disease, *Thorax* **52**(suppl 5):S1–S28, 1997.

5. Fiore MC, Bailey WC, Cohen SJ, *et al*., Treating Tobacco Use and Dependence, Clinical Practice Guideline. U.S. Department of Health and Human Services, Public Health Service, Rockville, MD June 2000.

6. American Thoracic Society, Pulmonary rehabilitation, *Am J Respir Crit Care Med* **159**(5 Pt 1):1666–1682, 1999.

7. Medical Research Council, Report of the Medical Research Council Working Party, Long term domiciliary oxygen therapy in chronic hypoxic cor pulmonale complicating chronic bronchitis and emphysema, *Lancet* **1**(8222):681–686, 1981.

8. Nocturnal Oxygen Therapy Trial Group, Continuous or nocturnal oxygen therapy in hypoxemic chronic obstructive lung disease: a clinical trial, *Ann Intern Med* **93**:391–398, 1980.

9. Yusen RD, Lefrak SS, Evaluation of patients with emphysema for lung volume reduction surgery, *Semin Thorac Cardiovasc Surg* **8**:83–93, 1996.

10. Sciurba F, Early and long-term functional outcomes following lung volume reduction surgery, *Clin Chest Med* **18**:259–276, 1997.

11. National Emphysema Treatment Trial Research Group. A randomized trial comparing lung volume reduction surgery with medical therapy for severe emphysema. *New Engl J Med* **348**:2059–2073, 2003.

12. Trulock EP, Lung transplantation, *Am J Respir Crit Care Med* **155**:789–818, 1997.

This page intentionally left blank

77

Sleep Disorders

Anne Hsu

This chapter is a major revision of the previous one that the chapters on Sleep-disordered Breathing and Sleep Disorders have been merged onto this chapter. The subject of sleep disorders is too vast to be covered in one chapter, hence the aim here is to provide an overview of sleep disorders. The more common sleep disorders will be reviewed briefly but many disorders are beyond the scope of this overview.

The knowledge of sleep, sleepiness and sleep disorders has advanced dramatically over the last 3 decades. Although sleep medicine has become established as a specialty field, it is important for medical students, the general practitioners, the internists and public health authorities to appreciate the prevalence of sleep disorders, and the consequences of sleepiness and sleep disorders. This knowledge will serve to enhance the lives, both qualitatively and quantitatively, of patients and will also have a significant impact on the socioeconomic burden and the medicolegal aspect. To name a few of the socioeconomic impact of sleepiness and sleep disorders identified by the USA National Commission on Sleep Disorders Research:[1]

- The official final determination of cause of the Exxon Valdez, Bhopal, Challenger, Chernobyl accidents was fatigue-related impairment of judgment and performance in the workplace.

- 40 million Americans suffer from chronic sleep disorders. 95% of these are undiagnosed and untreated.
- In 1990, 200 000 motor vehicle accidents in the USA were due to falling asleep at wheels. The direct cost of sleepiness and sleep disorders to the American public was US$15.9 billion. Indirect costs may have added another US$150 billion.

FUNCTIONAL ANATOMY OF SLEEP AND CLINICAL CORRELATES

We spend our lives in three different states of being: wakefulness; non-REM (rapid eye movement) sleep; and REM sleep. Each state has its own distinct neuroanatomic, neurophysiologic, and neuropharmacologic mechanisms and behavioral features. The neuronal system that controls this sleep-wake cycle is found in the brainstem, hypothalamus and basal forebrain with the relay nuclei nested in the thalamus and the target organ in the cortex. The electrophysiologic expression of wakefulness is desynchrony of the electroencephalogram (EEG) by activation of the midbrain reticular neurons, which directly excite thalamocortical projection. During sleep, distinct stages of REM sleep and non-REM sleep alternate in a well-organised fashion. REM sleep originates in the brainstem[2] and is characterized by rapid eye movements (Fig. 1), the hallmark of REM sleep, dream content and muscle atonia. The EEG is desynchronized. The ponto-geniculo-occipital spike activity, seen in phases in REM, may eventually reach the cortical areas and trigger off fragmentary imagery, which we recognise as dreams. Muscle atonia occurs in all skeletal muscles except for the ocular and diaphragmatic muscles. Hence, significant hypoventilation occurs, and this can occur profoundly in patients with diaphragmatic dysfunction such as those with chronic obstructive pulmonary disease and those with diaphragmatic paralysis.

Neurophysiologic and behavioral features of more than one state may occur simultaneously, resulting in bizarre and important clinical syndromes.[3] REM sleep intrusion into wakefulness accounts for the symptoms of narcolepsy, such as cataplexy, sleep paralysis and hypnogogic hallucinations. In cataplexy, a strong emotion, such as laughter, triggers off REM sleep atonia, thereby causing paralysis. Non-REM sleep intrusion into wake occurs as micro-sleep periods that interrupt the wakefulness

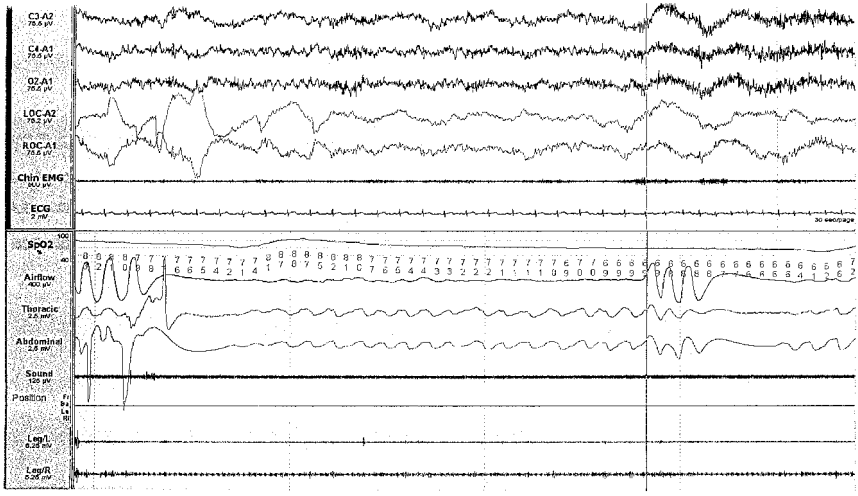


Fig. 1 This 30-second epoch of a standard polysomnogram shows obstructive apnea occurring during REM sleep. The vertical line corresponds to arousal with brief termination of the preceding obstructive event that has lead to oxyhemoglobin desaturation to a nadir of 61%.

Channel abbreviations are as follows:
C3-A2, C4-A1, O2-A1 = EEG; LOC-A2, ROC-A2 = left, right oculogram; EMG = chin EMG; ECG = electrocardiogram; Sp02 = oxyhemoglobin saturation via pulse oximeter; Airflow = via nasal thermistor; Thoracic, Abdominal = chest, abdominal movements; Sound = records snoring in this case; Position = body position; leg/L, leg/R = left, right leg EMG.

state, hence causing automatic behavior. The most prominent dissociated state arising from REM sleep is REM sleep behavior disorder (RBD). The muscle tone of the wakefulness state persists even during REM sleep. The preservation of the motor tone in patients with RBD may result in vigorous and injurious motor activity as they can act out their dreams.

CLASSIFICATIONS OF SLEEP DISORDERS

The International Classification of Sleep Disorders lists 84 known sleep disorders[4] classified under dyssomnias, parasomnias, medical/psychiatric sleep disorders and proposed sleep disorders. Dyssomnias are disorders that produce either difficulty initiating or maintaining sleep or excessive sleepiness. Parasomnias are disorders of arousal, partial arousal, and sleep stage transition. Sleep disorders associated with medical or psychiatric

disorders are not primarily sleep disorders but are medical or psychiatric conditions that have either sleep disturbance or excessive sleepiness as a major feature of the disorder. Proposed sleep disorders are for disorders for which there is insufficient information available to confirm the unequivocal existence of the disorders.

A more useful and practical diagnostic classification is classifying sleep disorders into the following four categories:

1) insomnia;
2) hypersomnia;
3) disorders of sleep/wake cycle; and
4) parasomnias or movement disorders associated with sleep or partial arousal.

INSOMNIA

Insomnia is the most common of all sleep-related complaints. A 1991 study found that 36% of Americans reported having insomnia.[5] Although local data shows that the prevalence of insomnia is lower, there is an increasing trend over the last 1–2 decades. A survey reported in 1986, of clinics in Singapore showed that 8% of patients had insomnia and a subsequent study revealed that 25% of the 612 elderly people in the community had sleep difficulties.[6] The prevalence of insomnia in a Singapore community was 15% reported in 1996[7] and in an unpublished survey, the Insomnia Tracking Study, completed in early 2002 indicated that this has risen significantly. The morbidity of insomnia includes impairment of quality of life and cognitive function; and increased motor vehicle accidents, work absenteeism, healthcare use and risk of depression.[1,5,8] Insomnia is a symptom, and not a disorder in and of itself. It is often the symptom of many underlying medical, psychiatric, and psychological conditions and may be the presenting symptom of other primary sleep disorders. It could be transient, lasting several nights in response to a precipitating factor; short-term if it is less than 3 weeks; and chronic if it persists beyond 3 weeks. Untreated short-term insomnia places the individual at risk for a more chronic psychophysiological or conditioned insomnia, which becomes difficult to treat and hence better managed by a sleep specialist or a psychiatrist. The primary cause of chronic insomnia is rarely immediately apparent and the likelihood of more than 1 cause is significant. The

two most frequent causes of chronic insomnia are anxiety and depressive disorders. It is important to establish the cause as specific therapy such as anxiolytics, antidepressant, treatment of sleep apnea can than be instituted with good outcomes. Sedative-hypnotic treatment has a role, particular in short-term insomnia and may serve to prevent the development of chronic insomnia. Effective agents include the benzodiazepines and the newer non-benzodiazepines. A combination of behavioral and pharmacological therapy is often effective for chronic insomnia.[9] Sleep hygiene (Table 1) should be reviewed. A good sleep history and sleep diary is often diagnostic and few insomniac patients require polysomnographic study for other suspected primary sleep disorders such as sleep apnea, sleep state misperception and periodic leg movements.

Periodic Leg Movements

Periodic leg movements of sleep (PLMS) is a frequent cause of insomnia or excessive daytime sleepiness in elderly patients. Stereotyped movements of the limbs occur periodically, predominantly in the legs during sleep. It is not myoclonic and is identical to the Babinski response where there is extension of the toes and dorsiflexion of the ankle. There could be flexion of the knees or hips. It lasts from 0.5 to 5 seconds and is shown on the

Table 1 Sleep Hygiene Tips

- Maintain regular wake time regardless of the hours of prior sleep.
- Avoid excessive time in bed. Get out of bed if you have not fallen asleep or don't feel sleepy after 30 minutes. Go to another room but go back to bed when you feel sleepy.
- Sleep only as much as you need for refreshment. Avoid getting too much sleep. The quality of our sleep is more important than quantity.
- Use the bed only for sleeping and sex.
- Make your bedroom a pleasant place to sleep, e.g. use a comfortable bed with clean linen, keep your bedroom quiet and dark or dim.
- Develop a bedtime routine. Do things that give you a sense of security, relaxation and comfort. Reserve the hour before bedtime for quiet activities. Read a "light" novel or watch a relaxing TV program; do not finish office work or discuss family finances with your spouse, for example.
- Eat a light snack before bedtime if hungry.
- Do not watch the clock.
- Avoid nicotine, caffeine, and alcohol.
- Exercise regularly and early (before 7 pm) in the day.
- Avoid naps, except if shift worker.

electromyogram of the anterior tibial muscle as a burst of muscle activity causing arousal during non-REM sleep. The average interval between jerks is 20 to 40 seconds. These movements occur in clusters that last several minutes. At least 5 bursts per hour i.e. PLMS index of 5 on the polysomnography, are required for a diagnosis. PLMS occurs in majority of patients with restless leg syndrome (RLS) and forms a component of it. RLS may not occur in all cases of PLMS. It is characterized by irresistible leg movements accompanied by a creeping unpleasant sensation deep in the limbs typically occurring at rest and relieved with movements. It affects at least 5% of the general population but seems to pose less of a problem in Singapore. RLS and PLMS have been observed in pregnancy, chronic renal failure, peripheral neuropathy, iron deficiency anemia and hypothyroidism, although it can occur in healthy people. Both can disrupt sleep and hence treatment with Clonazepam, Sinemet or opioids at bedtime may be necessary.

HYPERSOMNIA

Sleepiness is the normal consequence of sleep deprivation or the use of hypnotics. A very common cause of sleepiness in our modern society is "voluntary" sleep deprivation. This societal sleep deprivation is due to many factors: shift work, air travel allowing rapid time zone shifts, 24-hour TV, stock exchanges and information technology. When sleepiness is excessive or inappropriate especially in the absence of "voluntary" sleep deprivation, it often indicates a clinical disorder termed as hypersomnia or daytime hypersomnolence or excessive daytime sleepiness (EDS), which is the result of many sleep disorders. The well-documented consequence of sleep loss is impairment of performance with respect to judgement, reasoning, learning and reaction time particularly at tasks involving sustained attention.[10] Falling asleep on the job and behind the wheel have resulted in grave consequences and in some fields, e.g. internship, legislation has been developed restricting duty hours.[11]

The following common sleep disorders with hypersomnia as a key feature will be discussed:

1) Obstructive Sleep Apnea
2) Narcolepsy
3) Idiopathic hypersomnia
4) Periodic hypersomnia

Obstructive Sleep Apnea

Obstructive sleep apnea (OSA) is the most common sleep disorders encountered in the Sleep Disorders Unit in Singapore General Hospital and in many Sleep Centres worldwide. Sleep-related disordered breathing consist of breathing disorders that occur only during sleep such as obstructive sleep apnea and central sleep apnea, as well as breathing disorders which can have significant worsening during sleep (such as chronic alveolar hypoventilation, Table 2).

Prevalence of OSA

OSA is very prevalent in the general population. The largest and most often quoted population survey to date found the prevalence of OSA, to be 24% and 9% in middle-aged adult males and females respectively.[12] Of these, not all are associated with symptoms and the investigators concluded that 4% of males and 2% of females in the middle-aged workforce of the USA have clinically important degree of sleep apnea, i.e. OSA syndrome (OSAS) defined as OSA with excessive daytime sleepiness. There is less prevalence data in Asia compared to the Western population. A recent community-based study of sleep apnea among middle-age men in Hong Kong showed similar OSAS prevalence of 4.1%.[13] The prevalence of

Table 2 Etiology of Chronic Ventilatory Failure

"DRIVE" PROBLEMS
• Primary alveolar hypoventilation
• Structural defects: CNS tumor, bulbar poliomyelitis, Arnold–Chiari malformation, cerebrovascular accident
• Metabolic: severe metabolic alkalosis, uremia, hemodialysis
• Cheyne-stokes respiration associated with severe left ventricular dysfunction
"PUMP" PROBLEMS
• Neuromuscular disorders
• Neuropathic processes, such as motor neuron disease
• Neuromuscular junction disorders, such as myastehnia gravis and Eaton–Lambert syndrome
• Myopathies, such as muscular dystrophy and acid maltase deficiency
• Restrictive abnormalities, such as kyphoscoliosis
"MIXED" DISORDERS
• Myxoedema
• Chronic obstructive pulmonary disease
• Obesity-hypoventilation syndrome

OSAS amongst medical inpatients in Singapore General hospital was 20%, 30% and 12% for Chinese, Malays and Indians respectively.[14] Another outpatient hospital-based study conducted in the same hospital reported an extrapolated population prevalence of 15% for OSAS.[15]

Pathogensis and sequaela of OSA (Fig. 2)

OSA is characterized by repetitive upper airway, from the pharynx to the larynx, collapse during sleep. Unlike the trachea and bronchi, whose patency is maintained by cartilaginous support, the patency of the upper airway depends critically on the action of its dilator muscles. Sleep predisposes one to narrowing and, in susceptible persons, collapse of the upper airway by reducing the tone of the upper airway muscles and their reflex response to the sub-atmospheric airway pressures generated during inspiration. The pathogenic factors that lead to instability of the upper airway patency during sleep include:

1) the size of the channel, which is related to the adjacent soft tissue, e.g. tonsillar structures and "preset" skeletal structures; and
2) the compliance of the airway wall, which is related to the neuromuscular function, e.g. in bulbar palsy or myasthenia gravis.

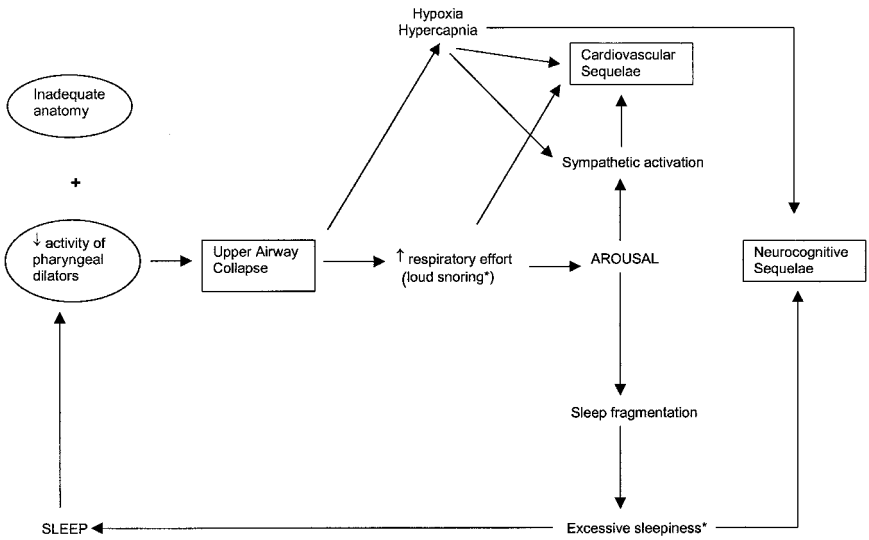


Fig. 2 Pathophysiology of Obstructive Sleep Apnea (OSA)
* Key symptoms of OSA

The risk factors for OSA include:

1) Obesity.
2) Older age.
3) Male gender. The male to female ratio is 3:1.[16] In women, OSA increases markedly after 55–65 years; after menopause.[17]
4) Familial and genetic risk.[18] Children of subjects with OSA tend to inherit similar craniofacial abnormalities, which may subsequently predispose them to the development of OSA.
5) Race. Thus far, most epidemiological studies on OSA have involved mainly Caucasians. The limited data on other racial groups including African-Americans, Hispanics and Asians gave higher prevalence compared to the Caucasian population.[15,19,20]
6) Alcohol and sedatives predisposes one to upper airway closure. Smoking causes upper airway inflammation.
7) Specific medical conditions include hypothyroidism, acromegaly, Down's syndrome and all neuromuscular dysfunction affecting the upper airway muscles, e.g. bulbar poliomyelitis.

Loud habitual snoring and excessive daytime sleepiness are the two key symptoms of OSA. When complete obstruction of the airway develops, however, the snoring is interrupted by periods of silence lasting 10 seconds to a couple of minutes. This coincides with the complete cessation of airflow. During these episodes of apnea, futile respiratory efforts continue and oxyhemoglobin desaturation develops until a brief awakening or arousal terminates the apnea (Fig. 1). Hence the airway patency is restored. These events are usually accompanied by a generalized startle response, snorting and gasping. After a few deep breaths, the patient returns to sleep, only to have the cycle of events repeated as many as few hundreds times through the night.

Clinical features of OSA

Symptoms During Sleep

1) Loud habitual snoring. A local study,[15] similar to other published studies, showed that snoring is the most common presenting symptom. The snoring is interrupted by apnoeic episodes associated with snorting or gasping. Some patients are awaken with snorting or with feeling of choking.

2) Abnormal motor activity during sleep. Patients or their bed partners may complain of excessive tossing and turning during sleep.
3) Sleep disruption. Patients are often unaware of the frequency and intensity of arousals from sleep. Some may complain of insomnia.
4) Oesophageal reflux.
5) Nocturia and enuresis due to increased secretion of atrial natriuretic peptide in response to reflex pulmonary vasoconstriction during apneic episodes; increased venous return and increased intra-abdominal pressure associated with respiratory efforts against a closed upper airway.
6) Heavy sweating at night.

Daytime Symptoms

1) Excessive daytime sleepiness. This is the chief clinical consequence of OSA. This is thought to be related to the fragmentation of sleep by recurrent arousals, the loss of deeper levels of sleep and the effects of hypoxemia on the cerebral function. Many studies have documented the debilitating effects of sleep deprivation in patients with OSA: poor vigilance, cognition, memory, impaired school and work performance, changes in personality and sexual problems, "sleep drunkenness" (disorientation and morning confusion) and, perhaps most worrying of all, a significant increase in driving accidents.[21,22] Sleepiness scale can be used to subjectively quantify sleepiness and one of the common scales used is the Epworth Sleepiness Scale[23] (Table 3). This is a simple, self-administered questionnaire, which can provide a measurement of the subject's general level of daytime sleepiness.
2) Morning headaches due to hypoxemia, with or without hyerpcapnia.
3) Dry throat on waking up.
4) Loss of hearing. Loud habitual snoring may lead to hearing impairment for both the snorer and his or her bed partner after many years.

Physical examination

Evaluation of the following should be done:

1) Upper airway patency. Crowding of the oropharynx due to large adenoids or tonsils, elongated soft palate, edematous reddish uvula or large tongue.

Table 3 Epworth Sleepiness Scale[24]

Situation	Chance of dozing (score 0 to 3)
• Sitting and reading	
• Watching television	
• Sitting, inactive in a public place	
• As a passenger in a car for an hour without a break	
• Lying down to rest in the afternoon when the circumstances permit	
• Sitting and talking to someone	
• Sitting quietly after a lunch without alcohol	
• In a car, while stopped for a few minutes in traffic	

0 = would never doze, 1 = slight chance of dozing, 2 = moderate chance of dozing, 3 = high chance of dozing. A score of > 10 is considered as abnormal.

2) Craniomandibular abnormalities such as micrognathia or retrognathia.
3) Obesity. The body mass index and the neck circumference should be measured. A measurement of $> 40\,cm$[15] for the latter is significant.
4) Associated syndromes or conditions. OSA may coexist with other respiratory disorders (the so-called "overlap" syndrome), namely, obesity hypoventilation syndrome and chronic obstructive pulmonary disease. The Pickwickian syndrome, which describes an obese, hypersomnolent individual with polycythemia and leg edema, has often been considered to be synonymous with the OSA syndrome. It, however, more closely describes a patient with the obesity-hypoventilation or the "overlap" syndrome, rather than obese individuals with OSA alone. Neuromuscular disorders such as myasthenia gravis, myotonia dystrophica and poliomyelitis can be the contributory factor of OSA. Other conditions associated with OSA are acromegaly and hypothyroidism.
5) Complications of OSA. Numerous reports have suggested that OSA are linked to systemic hypertension, pulmonary hypertension, increased risks of arrhythmias, ischemic heart disease, stroke and premature death.[24–27] It is estimated that a third of OSA patients have systemic hypertension and conversely a third of hypertensive patients have OSA.

Objective evaluation of OSA

Attended full night polysomnography (Fig. 1) has been established by consensus[28,29] as the accepted test for the diagnosis and determination of the severity and treatment of OSA. The polysomnography (PSG) consists of the evaluation of a minimum of seven parameters, including electroencephalogram, electro-oculogram, chin electromyogram, electrocardiogram, airflow, respiratory effort and oxygen saturation. Body position during sleep must be documented. Leg movement recording is desirable, but optional. Trained personnel must be in constant attendance and be able to intervene. A minimum of 6 hours sleep recording is required. Unattended portable sleep studies are not recommended for the routine assessment of OSA. Currently, there are no established rules for the assessment of the severity of OSA. The apnea-hyponea index (AHI), which is the number of apneas or hyponeas per hour of sleep, and the minimum oxygen saturation during sleep are often used in the analysis of results of the PSG to grade the severity of the condition.

There are a few tests that can objectively measure excessive daytime sleepiness. Multiple sleep latency test is the preferred study as it is a standardized and well-validated measure of physiologic sleepiness.[30] This test is discussed later.

Management of OSA

Indications for the treatment of OSA are:

1) to alleviate excessive daytime sleepiness (EDS); and
2) to reduce the potential cardiovascular and neurocognitive morbidity and premature mortality associated with OSA.

While the data is still emerging for the latter indication, treatment is warranted for EDS as this can be disabling. There is strong evidence in support of reduction in driving accidents with CPAP therapy for OSA.[22]

Therapeutic measures for OSA are:

1) Continuous positive airway pressure (CPAP).
2) Surgery. As OSA is mainly a problem of multilevel upper airway collapse, this has to be well addressed for a successful surgical outcome. This well explains why tracheostomy is completely effective, though not well accepted, and uvulopalatopharyngoplasty by itself has relatively low efficacy for treatment of OSA.

3) Oral appliances.
4) Behavioral such as attain an ideal body weight, avoid sedatives or alcohol or smoking and sleeping on sides.

The most consistently efficacious, safe therapy and therefore the most commonly recommended, is CPAP therapy. CPAP mechanically splints open the entire upper airway, hence is universally effective regardless of the site of obstruction. However, the major disadvantage is that it is not curative for OSA, i.e. lifelong usage is generally expected.

Narcolepsy

Narcolepsy is not a rare disease.[31] It affects up to 1 in every 1000 individuals in North America, making it relatively as common as Parkinson disease or multiple sclerosis. It is known that there are many instances of symptomatic patients who are not diagnosed as narcolepsy for between 10 to 40 years. Men and women are equally affected and its onset occurs in the second or third decade of life. It is a primary sleep disorder characterized by excessive daytime sleepiness. In addition, patients with narcolepsy can have abnormal manifestations of REM sleep, such as cataplexy, hallucination on falling asleep or waking up and sleep paralysis.[31] A patient can have sleepiness and electrophysiological evidence of narcolepsy, but without the other features. This condition is called monosymptomatic narcolepsy. Some patients can develop cataplexy years later.

Patients have a irresistible tendency to sleep. This has given rise to the term "sleep attacks". Despite many catnaps, patients with narcolepsy do not sleep any more hours than normal people. Abnormal drowsiness (apart from sleep episodes) is the cause of poor job performance, scholastic failure, family disruption and anti-social behavior. Automatic behavior is the result of ongoing drowsiness punctuated by microsleep.

Cataplexy is a symptom unique to narcolepsy. It is episodic weakness without any loss in consciousness and occurs in all voluntary muscles in the body, except the eye muscles and diaphragm, and can last for minutes. Laughter is the most common emotion that brings on the weakness. It could be mild enough to cause buckling of knee, jaw dropping, head dropping or dysarthria. The marked weakness could cause the patient to fall. It is this history that distinguishes cataplexy from transient ischemic attacks, myasthenia gravis, periodic paralysis, syncope and atonic seizure.

Sleep paralysis and hypnagogic hallucinations are terrifying experiences for the patient during the onset of sleep or upon waking up. The hallucinations of narcoleptic patients are dream-like, visual events with auditory and tactile perceptions or a feeling of levitation. They differ from dreams in that there is no theme or story, e.g. a static image (a hallucination of someone standing over the bed) is present.

The etiology of narcolepsy remains unknown although there seems to be a genetic component[32] as over 90% of narcoleptics have HLA-DR2 or HLA-DQ1. Recent studies also indicate the nacroleptics have deficiency in the hypothalamic neuropeptide hypocretin.[33]

Objective evaluation of narcolepsy

Patients should have an overnight polysomnography followed by a multiple sleep latency test. The former excludes other forms of sleep disorders and the latter demonstrates sleep onset REM (the electrophysiological signature of narcolepsy) in 2 or more tests with an average sleep latency of five minutes or less. HLA typing supports, but does not determine, the diagnosis of narcolepsy as up to 30% of the general population without narcolepsy may have the same HLA type.[32] The measurement of cerebrospinal fluid concentrations of hypocretin as a new diagnostic tool may be a promising application.[33]

Management of narcolepsy

There is no cure for narcolepsy. The hypersomnia in these patients responds well to stimulant medication, such as amphetamines and newer drug, Modafinil; with little evidence of tolerance, dependence, or abuse. Naps of between 15 to 20 minutes are refreshing for these patients and they can cut down on the dosage of stimulant medication. The ancillary symptoms of cataplexy, hypnagogic hallucinations, and sleep paralysis respond to antidepressants, a REM suppressant. Many patients with narcolepsy are poorly tolerant of irregular wake/sleep patterns, including shift-work.

Idiopathic Hypersomnia

This condition is characterized by hypersomnia and naps without cataplexy or REM sleep abnormality. Sleepiness dominates the patient's life.

The diagnosis is made based on normal overnight polysomnography and multiple sleep latency test. Treatment is with stimulant medication.

Peroidic Hypersomnia

These are uncommon, but are well-defined conditions of episodic hypersomnia that typically occur weeks or months apart. The Kleine–Levin syndrome is one such example and it is characterized by excessive sleepiness associated with hyperphagia and hypersexuality. The symptoms begins in young, adolescent boys and its etiology is unknown. Menstrual hypersomnia is another form of periodic hypersomnia, which strikes in the days preceding menstruation. The onset is soon after menarche and may disappear after pregnancy.

PARASOMNIAS

Parasomnias are defined as unpleasant or undesirable behavioral phenomena, which occur during sleep.[34] They are initially attributed to psychiatric disorders but recent clinical and ploysomnographic analysis has revealed that they are in fact the result of a large number of completely different conditions, most of which are diagnosable and treatable. These disorders are manifestation of central nervous system activation usually transmitted through skeletal muscle or autonomic nervous system channels. And the behavior may result in injury or violence to self or others. They can be classified as:

1) Arousal disorders, including confusional arousal, sleepwalking and sleep terrors which are common in children. Treatment with medication or hypnotherapy is often effective.
2) Sleep-wake transition disorders, including rhythmic movement disorders, sleep starts, sleep talking and nocturnal leg cramps.
3) Parasomnias associated with REM sleep, including nightmares, sleep paralysis, impaired sleep-related penile erections and REM sleep behavior disorder (RBD). RBD was predicted by animal experiments (bilateral perilocus ceruleus lesions) in 1965 and were subsequently identified in humans.[35] In RBD, there is absence of REM atonia and this permits "acting out of dreams", often with violent, injurious behavior. The patient might dream that he was rescuing his wife from

attackers and, in the process, actually hit and injure her whilst sleeping next to her. RBD is more common in the elderly. Although it can occur in otherwise healthy people, there is a higher prevalence of central nervous disorders such as Parkinson's disease. The administration of Clonazepam at bedtime suppresses these attacks.

4. Other parasomnias such as nocturnal seizures, sleep bruxism, enuresis and sleep-related abnormal swallowing syndrome. Nocturnal seizures may be misdiagnosed because of their tendency for bizarre behaviors, exclusively nocturnal timing, and clustering in time.[36]

DISORDERS OF THE SLEEP/WAKE CYCLE

The sleep-wake pattern of humans follow a circadian rhythm. The impact of circadian disorders is on the alertness, concentration and performance of the individuals during the time when they have to be awake. Circadian disturbances can be triggered off by:

1) intrinsic factors, i.e. malfunction of the biologic clock *per se*. Primary circadian dysrhythmia include: delayed sleep phase syndrome, advanced sleep phase syndrome, non 24-hour sleep phase and the irregular sleep phase syndrome.
2) extrinsic factors, i.e. due to environmental effects upon the underlying clock. The societal revolution in the last century has given rise to steady erosion of the circadian rhythm in shift workers and air/space travelers. Jetlag occurs when the body is transported rapidly to a new time zone while the internal circadian clock remains at the home time and it takes a while for the body clock to adjust to the new time zone.

These secondary circadian disorders such as shift work and jetlag are usually apparent upon simple questioning of the patients. The primary disorders may be much more difficult to diagnose, as they may masquerade as other disorders such as hypersomnia, insomnia, substance abuse, or psychiatric conditions. Identification of these disabling disorders is important, as treatment exist. Several agents can manipulate circadian rhythms. They are phototherapy (using bright light) for shift workers and jetlag; chronotherapy for delayed sleep phase syndrome; melatonin, which synchronizes the circadian rhythm and facilitates sleep, for jetlag and advanced sleep phase syndrome.

EVALUATION OF SLEEP DISORDERS

Clinical

A detailed sleep history and focused physical examination should be performed. The latter include evaluation of the oronasomaxillofacial and neck region, cardiopulmonary, neurologic and psychiatric examinations. Supporting information from roommates, family members or care-givers is usually required for a complete and accurate assessment of sleep behaviors. Sleep/wake diaries completed by the patient or observer over weeks may be helpful.

Laboratory Studies

It is recommended that patients who required sleep studies be referred to the sleep specialist as these tests are complex and are best tailored to the specific clinical situation.

Polysomnography (PSG)

The technology used for the diagnosis of sleep disorders employs standard electrophysiologic recording systems which allows sleep to be accurately quantified, sleep stages to be fully characterized and the analysis of sleep-related events. The basic PSG consists of electroencephalogram, electro-oculogram, chin electromyogram, electrocardiogram, respiratory parameters (at least, airflow) and anterior tibialis electromyogram. Multiple respiratory parameters, including respiratory effort, oxygen saturation and airflow, must be monitored to evaluate sleep disordered breathing. Other parameters may be monitored as clinically indicated, such as multiple channel electroencephalogram for parasomnias, esophageal sensors for upper airway resistance syndrome or reflux, or penile tumescence for erectile function.

Multiple sleep latency test (MSLT)

The MSLT is a standardized and well-validated measure of quantifying daytime sleepiness[30] and in differentiating the subjective complaints of "sleepiness", "fatigue" and "tiredness". The MSLT consists of four to five 20 minute-nap opportunities scheduled at 2-hour intervals in the

day. It quantitates sleepiness by standard electrophysiological test that measures how quickly a subject falls asleep on sequential naps by measuring the latency between 'lights out' and sleep onset. A mean latency of 5 minutes or less indicates severe daytime sleepiness. The MSLT also identify the abnormal occurrence of REM sleep during a nap and the number of naps during which REM sleep appears is noted as well. As many factors can affect sleep latency and REM sleep during the daytime, proper interpretation requires a preceding night PSG to measure the quality and quantity of night sleep prior to the MSLT.

Actigraphy

This is a more objective way of analyzing the sleep/wake pattern compared to sleep diaries completed by patients. An actigraph is a small wrist-mounted device that records the activity plotted against time, usually for 1 to 2 weeks. The data collected is transferred into a computer where the software displays activity versus time demonstrating the rest/activity pattern. This information compliments the subjective sleep log obtained concurrently.

LEGAL ASPECTS OF SLEEP AND ALERTNESS

Sleep disorders and lack of alertness can cause various legal problems. It is a vexing question whether a person who causes harm while sleep-walking is criminally liable. The association between daytime sleepiness and accidents is well-known. It is becoming increasingly apparent that various drugs, including sedatives, tranquilizers and hypnotics, seriously affect a person's ability to drive or operate heavy machinery as much if not more than alcohol does. When a driver falls asleep at the wheel, the court may convict him of a driving offence on the grounds that he should have pulled off the road when he began to feel drowsy. However, such sleep may be totally unpredictable and it occurs in microsleep. Due to sleep deprivation regardless of the cause, short microsleeps occur during periods of wakefulness. They take the form of short 1- to 10-second bursts of Stage 1 or 2 sleep, enough to result in an accident that the person is not aware of. This typically results in frequent automobile accidents especially along highways, driving on the wrong side of the road and ignoring traffic signals. These attacks are difficult to distinguish from automatism, which is associated with partial complex seizures, or absence seizures, transient

global amnesia or simple day-dreaming. The concept of microsleep and its implication is not well-known among doctors and the public.

CONCLUSION

Sleep disorders medicine has exponentially evolved over the past few decades. Although there remains many intriguing issues to be resolved, it is timely that doctors and the public be more aware of this field of medicine and certainly, it should be a core subject well taught in the medical undergraduate school.

REFERENCES

1. National Commission on Sleep Disorders Research, Report of the National Commission on Sleep Disorders Research, Sup. of Docs., U.S. Government Printing Office, DHHS Pub. Washington, DC, 1992.
2. Douglas NJ, White DP, Pickett CK, Respiration during sleep in normal man, *Thorax* **37**:840–844, 1982.
3. Mahowald M, Schenck C, Status dissociates — a perspective on states of being, *Sleep* **14**:67–79, 1991.
4. Diagnostic Classification Steering Committee, Thorpy MJ, Chairman, International Classification of Sleep Disorders : Diagnostic and coding manual, Rochester, MN: American Sleep Disorders Association, 2000.
5. The Gallop Organisation. *Sleep in American*, National Sleep Foundation Los Angeles, California, pp. 1–30.
6. Kua EH, Depressive disorder in elderly Chinese people, *Acta Psychiatr Scand* **81**(4):386–388, 1990.
7. Yeo BK, Perera IS, Kok LP, Tsoi WF, Insomnia in the community, *Singapore Med J* **37**(3):282–284, 1996.
8. Roth T, The relationship between psychiatric diseases and insomnia, *Int J Clin Pract Suppl* **116**:3–8, 2001.
9. Walsh JK, Mahowald MW, Avoiding the blanket approach to insomnia, *Postgraduate Med* **90**:211–224, 1991.
10. Monk TH (ed), *Sleep, Sleepiness and Performance,* John Wiley and Sons, Chichester, 1991.
11. Leung L, Becker CE, Sleep deprivation and house staff performance, Update 1984–1991, *J Occupational Medicine* **12**:1153–1160, 1992.
12. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S, The occurrence of sleep-disordered breathing among middle-aged adults, *N Engl J Med* **328**:1230–1235, 1993.

13. Ip MSM, Lam B, Lauder IJ, Tsang KWT, Chung KF, Mok YW *et al.*, A community study of sleep-disordered breathing in Middle-aged Chinese Men in Hong Kong, *Chest* **119**:62–69, 2001.

14. Tan YK, Khoo KL, Low JAYH, Wong ZW, Theng CTS, Ong YY *et al.*, Ethnicity, Obstructive Sleep Apnea and Ischaemic Heart Disease, *Ann Acad Med*, Singapore, **28**:214–216, 1999.

15. Puvanderan K, Goh KL, From snoring to sleep apnea in a Singapore Population, *Sleep Research Online* **2**(1):11–14, 1999.

16. Strohl KP, Redline S, Recognition of obstructive sleep apnea, *Am J Respir Crit Care Med* **154**:279–289, 1996.

17. Guilleminault C, Quera-Salva MA, Partinen M, Jamieson A, Women and the obstructive sleep apnoea syndrome, *Chest* **3**:983–988, 1988.

18. Guilleminault C, Partinen M, Hollman K, Powell N, Stoohs R, Familial aggregates in obstructive sleep apnea syndrome, *Chest* **107**:1545–1551, 1995.

19. Kripke DF, Ancoli-Israel S, Klauber MR, USA population estimates for sleep-disordered breathing: high rates in minorities, *Sleep Res* **24**:268, 1995.

20. Ong KC, Ckerj AA, Comparison of the severity of sleep-disordered breathing in Asian and Caucasian patients seen at a sleep disorders centre, *Respir Med* **92**:843–848, 1998.

21. Naegele B, Thouvard V, Pepin JL, *et al.*, Deficits of cognitive executive functions in patients with sleep apnea syndrome, *Sleep* **18**:43–52, 1995.

22. Findely L, Smith C, HopperJ, Dinnen M, Suratt P, Treatment with nCPAP decreases automobile accidents in patients with sleep apnea, *AJRCCM* **161**:857–859, 2000.

23. Murry WJ, A new method of measuring daytime sleepiness: the Epworth Sleepiness Scale, *Sleep* **14**:540–545, 1991.

24. Roux F, D'Ambrosio C, Mohsenin V, Sleep-related breathing disorders and cardiovascular disease, *Am J Med* **108**(5):396–402, 2000.

25. Nieto FJ, Young TB, Lind BK, Shahar E, Samet JM, Redline S *et al.*, Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study, *JAMA* **283**(14):1829–1836, 2000.

26. Kessler R, Chaouat A, Weitzenblum E, Oswald M, Ehrhart M, Apprill M, Kreiger J, Pulmonary hypertension in the obstructive sleep apnea syndrome: prevalence, causes and therapeutic consequences, *Eur Respir J* **9**(4):787–794, 1996.

27. Wessendorf TE, Teschler H, Wang YM, Konietzko N, Thilmann AF, Sleep-disordered breathing among patients with first-ever stroke, *J Neurol* **247**(1):41–47, 2000.

28. Standards of Practice Committee of the American Sleep Disorders Association, Practice parameters for the use of portable recording in the assessment of obstructive sleep apnea, *Sleep* **17**:372–377, 1994.

29. Indications for Polysomnography Task Force, American Sleep Disorders Association Standards of Practice Committee, Practice Parameters for the Indications for Polysomnography and Related Procedures, *Sleep* **20**:406–422, 1997.

30. American Sleep Disorders Association, The clinical use of multiple sleep latency test, *Sleep* **15**:268–276, 1992.

31. Aldrich MS, The neurobiology of narcolepsy-cataplexy, *Progr in Neurobiology* **41**:538–541, 1993.

32. Practice parameters for the use of stimulants in the treatment of narcolepsy, Standards of practice committee of the American Sleep Disorders Association, *Sleep* **17**:348–351, 1994.

33. Overeem S, Scanmmell TE, Lammers GJ, Hypocretin/orexin and sleep: implications for the pathophysiology and diagnosis of narcolepsy, *Curr Opin Nuerol* **15**(6):739–745, 2002.

34. Schenck CH, Hurwitz, TD, Mahowald MW, REM sleep behavior disorder: a report on a series of 96 cases and a review of the literature, *J Sleep Res* **2**:224–231, 1993.

35. Mahowald MW, Schenck CH, REM sleep parasomnias, in: Kryger MH, Roth T, Dement WC (eds.), *The Principles and Practice of Sleep Medicine*, 3rd ed., WB Saunders Philadelphia, pp. 724–741, 2000.

36. Mahowald MW, Schenck CH, Parasomnia purgatory — the epileptic/nonepileptic interface, in: Rowan AJ, Gates JR, (eds.), *Non-epileptic Seizures*, Butterworth-Heinemann, pp. 123–139, 1993.

This page intentionally left blank

78

Lung Cancer

Constance Lo and Philip Eng

INTRODUCTION

Lung cancer remains the number 1 cancer in males and number 3 cancer in females (after breast and colorectal cancer) in Singapore to date. This is despite the decreasing prevalence of lung cancer since the mid-1980s. The age-standardized rate of lung cancer is 54 and 17 per 100 000 per year for males and females respectively.[1]

Epidemiology of Lung Cancer

The age-adjusted incidence of lung cancer among men exceeds by two-fold that of women. The peak incidence is at age 70–79. Lung cancer is the leading cause of cancer mortality in most countries.[2] The varying rates of incidence rates correlates well with cigarette smoking habits prevalent at least 10 years prior. Much of the geographical and cultural variations in lung cancer incidence can be at least partially attributed to the variations in smoking habits, including age when the habit was picked up, type of cigarettes smoked and duration of smoking.

Lung cancer is largely felt to be a "preventable" disease. There is a strong etiologic link between lung cancer and cigarette smoking, being implicated in $>85\%$ of lung cancer cases. Tobacco smoke is a potent cocktail of carcinogens, including polynuclear aromatic hydrocarbons, N-nitrosamines, aromatic amines, and other organic (benzene, acrylonitrile) and non-organic compounds (arsenic, acetaldehyde). Increases in lung cancer are seen 20–30 years after a parallel increase in cigarette smoking.[3] The risk of developing lung cancer increases with both the duration of smoking and the number of cigarettes smoked. While manufacturing companies have the touted lower tar content and the use of filters to reduce lung cancer risk among smokers, changes in smoking patterns (for example, deeper inhalations) reduce this theoretical advantage. Other forms of tobacco smoke, in the form of pipe and cigar smoking are also linked to lung cancer. Although smoking cessation reduces the risk of developing subsequent lung cancer among smokers, the risks still remain elevated above non-smokers, even up to 40 years of abstinence. Environmental tobacco smoke also has the same carcinogenic composition of mainstream smoke, and involuntary smoking is associated with an increased risk of lung cancer.

Other environmental agents incriminated in the causation of lung cancer include radon, asbestos, arsenic, nickel chromium and combustion-generated carcinogens. In developing countries, exposure to the fumes from cooking stoves and fires has been associated with elevated lung cancer risk.

Lung cancer is known to complicate diffuse fibrotic lung disease. While lung cancer also frequently complicates chronic obstructive lung disease, this association may arise from a common etiology: cigarette smoke.

Finally, genetic and dietary factors clearly influence the development of lung cancer, as ultimately, only approximately 10–15% of lifetime smokers develops lung cancer, and 10–15% of patients with lung cancer are non-smokers. Studies have suggested a protective association between lung cancer and increased fruit and vegetable consumption. In addition, other micronutrients have been suggested to decrease lung cancer risk, including vitamin C and selenium. Lastly, the tendency for familial aggregation of lung cancer suggests an inheritable tendency. This may be secondary to inherited variations in carcinogen metabolism and DNA repair.

Signs and Symptoms

More than 90% of patients with lung cancer will have symptoms at presentation. The symptoms are diverse, and arise from: 1) the primary tumor; 2) regional/intrathoracic spread; 3) distant metastasis; and 4) paraneoplastic syndromes.

Symptoms

The common symptoms are:

- Cough
- Wheeze or stridor
- Hemoptysis
- Dyspnea
- Chest pain

Symptoms arising from the primary tumor depend on its location. Central tumors (in the mainstem, lobar and proximal segmental bronchi) are associated with cough, hemoptysis, wheezing, and dyspnea secondary to central airway obstruction with post-obstructive pneumonitis. Although cough is the most common presenting symptom in lung cancer, it is often difficult to distinguish the cough due to chronic bronchitis (in the smoker) versus that due to lung cancer. Suspicious features of cough in a chronic smoker are a change in the character of the cough, a change in the quality and quantity of expectoration, and hemoptysis. Although bronchitis is the most common cause of hemoptysis, the incidence of bronchogenic carcinoma varies from 19–29%. Bronchoscopic series also show that lung cancer can occur in up to 9% of patients with hemoptysis and an unsuspecting chest radiograph. The indications for bronchoscopic airway examination in patients with hemoptysis and a normal chest X-Ray include: age over 40 years, a significant smoking history and male sex.[4,5]

Conversely, tumors arising in the distant airways (peripheral tumors) tend to have less cough, and present with chest wall pain.

Intrathoracic extension

Symptoms from intrathoracic extension may be from direct extension of the tumor or lympthatic spread. This may cause symptoms from compression of vascular structures (superior vena caval obstruction),

nerve invasion (recurrent laryngeal, and phrenic nerves, brachial plexus involvement and sympthathetic nerve involvement), visceral (pericardium, diaphragm and esophagus) , and chest wall involvement (pain).

One study showed that lung cancer is one of the commonest causes of metastatic carcinoma of unknown primary. The most common site of distant metastasis from lung cancer the bones, the liver, adrenal glands, intraabdominal lymph nodes, brain, spinal cord and skin.

Paraneoplastic Symptoms

- Loss of appetite and weight
- Hypertrophic osteoarthropathy
- Hypercalcemia
- Ectopic ACTH Secretion
- Syndrome of Inappropriate ADH Secretion
- Lambert–Eaton Syndrome

Other Symptoms

Clearly within any individual tumor stage, and adverse prognosis was seen with concomitant systemic symptoms of anorexia, weight loss and fatigue. This is thought to be a reflection of a subgroup with more extensive disease.

Asymptomatic

It is the occasional individual who has a chest X-Ray done for other reasons, e.g. pre-employment or health screening, where a solitary pulmonary nodule is the only presentation. This presents the best, though rare, scenario for a curative resection.

Diagnostic Tests for Lung Cancer

- Sputum cytology
- Lymph node biopsy (fine needle aspiration cytology and excision biopsy)
- Closed pleural biopsy and thoracentesis
- Percutaneous lung biopsy (peripheral lesions)
- Bronchoscopy (proximal lesions)
- Thoracoscopy
- Thoracotomy (rarely)

A variety of techniques are available to confirm a diagnosis of lung cancer. The diagnostic procedure of choice depends on the patient's disease presentation, and specimens for histology or cytology may be obtained from either the primary tumor or from secondary sites of involvement, as for example, lymph nodes, and skin nodules. Central tumors may be assessed by sputum cytology, (which has a lower yield of 20–30%) or flexible fibreoptic bronchoscopy, which yields as high as 90% when the tumor is visible endoscopically). Peripheral lung lesions may be sampled by bronchoscopic lung biopsy or transthoracic needle aspiration biopsies. Pleural effusions should be sampled by repeated pleural aspirations, with a view to thoracoscopy after 2 consecutive negative samples. Peripheral lung nodules may also be sampled by thoracoscopic lung biopsies. Finally, patients with a nodule that is suspicious for malignancy and limited disease should be considered for excision biopsy and lobectomy should lung cancer be confirmed.

CLASSIFICATION OF MALIGNANT LUNG TUMORS

IN 1999, the World Health Organization (WHO) revised its classification of lung and pleural tumors.[6] A discussion of the changes is beyond the scope of this chapter, and the reader is requested to refer to the article for its details. Bronchogenic carcinoma is derived from epithelial tissue, and histological classification is based on light microscopy. The 4 most common subtypes of bronchogenic carcinoma are squamous cell carcinoma, adenocarcinoma, large cell carcinoma and small cell lung cancer. Currently, adenocarcinoma has surpassed squamous cell carcinoma as the most frequent histologic subtype, representing approximately 31% of lung cancers, with squamous cell at 30%, large cell at 9%, and small cell at 18%.[7]

The important clinical question is in differentiating small cell lung cancer from non-small cell lung cancer, as, traditionally, reasonable results may be obtained with chemotherapy in the former cell type.

STAGING OF NON SMALL CELL LUNG CANCER

The International System for Staging Lung cancer by Mountain[8] is currently the most widely used staging system. This classification regroups the

TNM subsets based on survival in a contemporary time frame. The stage of disease at presentation is thus a strong prognostic factor in survival.

Primary Tumor

Tx	Primary tumor cannot be assessed or tumor proven by the presence of malignant cells in sputum or bronchial washings, but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumor
T1s	Carcinoma *in situ*
T1	Tumor < 3 cm in greatest dimension surrounded by lung or visceral pleura without bronchoscopic evidence of invasion more proximal than lobar bronchus
T2	Tumor with any of the following: > 3 cm in widest dimension; involves main stem bronchus > 2 cm from carina; invades visceral pleura; associated atelectasis or obstructive pneumonitis that extends to hilar but does not involve entire lung
T3	Tumor of any size that directly invades any of the following: chest wall, diaphragm, mediastinal pleura, parietal pericardium, or tumor in main stem bronchus < 2 cm distal to carina but without involvement of carina, or associated atelectasis or obstructive pneumonitis of entire lung
T4	Tumor of any size that invades: mediastinum, heart, great vessels, trachea, esophagus, vertebral body, carina, or tumor with malignant pleural or pericardial effusion, or with satellite tumor nodules within the ipsilateral primary tumor lobe of the lung

Regional Lymph Nodes

Nx	Regional Lymph nodes cannot be assessed
N0	No regional lymph node metastases
N1	Metastases to ipsilateral peri bronchial or hilar lymph nodes, and intrapulmonary nodes involved by direct extension of the primary tumor
N2	Metastases to ipsilateral mediastinal lymph nodes and/or subcarinal lymph nodes
N3	Metastases to contra lateral mediastinal lymph nodes

Distant Metastases

Mx	Presence of metastases cannot be assessed
M0	No distant metastases
M1	Distant metastases

Stage Grouping

0	Carcinoma *in situ*			
IA	T1N0M0			
IB	T2N0M0			
IIA	T1N1M0			
IIB	T2N1M0	T3N0M0		
IIIA	T3N1M0			
	T1N2M0	T2N2M0	T3N2M0	
IIIB	T4N0M0	T4N1M0	T4N2M0	T4N3M0
	T2N3M0	T3N3M0	T4N3M0	
IV	Any T, Any N, M1			

MANAGEMENT OF NON-SMALL CELL LUNG CANCER

Surgical Treatment

Surgery with curative intent remains the cornerstone for early lung cancers (stage IA and IB, IIA and IIB, and selected patients with IIIA). Patients with good performance status and locally advanced disease (Stage IIIA) may benefit from pathologic down staging with neoadjuvent chemo radiotherapy prior to surgical resection.[9,10] Standard resection techniques are lobectomy, bilobectomy or pneumonectomy. Limited resections with wedge resections and segmentectomy appear to have a higher incidence of recurrence when compared with standard lobectomy.[11] Other lung sparing surgical procedures include bronchial and sleeve operations. Preliminary assessment of resectable lung cancers includes a complete history, physical examination, and computerized tomogram of the chest to assess the tumor and mediastinum. The importance of mediastinal lymph node status in overall prognosis cannot be

underscored and is a powerful prognostic indicator. Surgery as a sole modality of treatment in patients with greater than N2 disease is poor.[12,13] Mediastinal nodal enlargement seen on computerized tomogram of the chest has a reported sensitivity and specificity of 64% and 62% respectively, using a short axis diameter of 1 cm.[14] Because of this low specificity, enlarged lymph nodes should be biopsied for accurate staging before curative surgery is precluded.[15] Bone scans should be performed on the finding of elevated serum calcium, or elevated alkaline phosphatase and brain scans in the event of symptoms or localizing signs. Medical evaluation of the patient's respiratory status includes pre-operative spirometry. A patient with a pre-operative FEV_1 (forced expiratory volume in one second) of greater than 2 litres, can be expected to tolerate a pneumonectomy.[16,17] Other techniques to predict post-operative pulmonary complications include quantitative perfusion lung scanning[18] and exercise testing.[19,20]

Metastatic Disease

Radiotherapy

Radiotherapy may be used in radical and palliative settings. Radical radiotherapy aims to provide local control of the tumor in patients with clinical Stage I and II disease who are medically in-operable due to comorbid problems, for example, ischemic heart disease. In contrast, palliative radiotherapy is indicated for metastatic disease to the brain, bones, spinal cord, superior vena caval obstruction and endobronchial obstruction.

Chemotherapy

Cisplatinum based chemotherapy has been showed to improve survival outcome in patients with metastatic disease when compared to best supportive care. At one year, 10% more patients receiving palliative chemotherapy would be alive compared to the supportive care group.[21] However, it is clear that much of this data was gleaned from trials where only patients with good performance status (ECOG < 2) are included. Hence, the decision whether or not palliative chemotherapy is to be useful should only be made following a lengthy discussion with the patient on the relative risks, benefits and costs involved.

Interventional bronchoscopy for malignant airway occlusion

A variety of endoscopic techniques have evolved to relieve malignant major airway occlusion. This includes neodymium:yttrium-aluminium-garnet (Nd:YAG) laser resection for malignant exophytic lesions, implantation of tracheobronchial airway stents for extrinsic airway compression and brachytherapy, which allows local irradiation to a tumor following bronchoscopic placement of an after loading catheter.

Malignant pleural effusion

Management of malignant pleural effusion is entirely palliative. There is no doubt that tube thoracostomy followed by pleurodesis is the best approach in most patients with malignant pleural effusion. Intermittent thoracocentesis results in frequent recurrences and possibility of iatrogenic pleural space infection. Efforts continue in the quest for the best pleurodesing agent. Bleomycin, tetracycline and talc slurry remain the agents to instil via tube thoracostomy. Pleurodesis may also be performed under direct vision using talc poudrage via thoracoscopic approach (video assisted thoracoscopy, medical thoracoscopy), and is frequently reserved for the young and fitter patients, yet its cost-effectiveness remains controversial.

Malignant pericardial effusion

The approach to malignant pericardial effusion is mainly surgical with options of pericardial window or pericardiectomy. In the critically ill patient with cardiac tamponade and hypotensive, bedside pericardiocentesis with ultrasound guidance in the ICU setting is life-saving.

Small Cell Lung Cancer

Small cell lung cancer accounts for 20–25% of all cases of bronchogenic carcinoma. Small cell lung cancer is an aggressive disease with a short symptom period preceding diagnosis. Nevertheless, at the time of diagnosis, 70% of patients already frequently have metastatic disease. Without treatment, survival may be as short as 6–12 weeks for such patients.[22]

Small cell lung cancer is commonly staged according to the 2-stage system developed by the Veterans Administration Lung Cancer Study

Group. The essence of the 2-stage classification is the ability to encompass the disease within 1 radiation portal. Hence, limited stage disease is defined as tumor confined to the ipsilateral hemithorax and its regional lymph nodes (hilar or mediastinal), with or without ipsilateral supraclavicular lymph node involvement. Extensive disease is defined as disease beyond the confines of the ipsilateral hemithorax, and includes metastatic disease and pericardial disease.

As the majority of patients present with metastatic disease, staging includes chest radiographs, computerized tomography of the chest, liver, adrenals and head, and bone scans. Routine bone marrow aspirates and biopsy in staging is not recommended outside research protocols. Common sites of dissemination include the bone, liver, the central nervous system, lymph nodes and pleura.

Treatment

Chemotherapy remains the cornerstone in the management of small cell lung cancer, given 1) at the time of diagnosis, small cell carcinoma is frequently a systemic neoplasm, and 2) the chemosensitive nature of small cell lung cancer. The treatment scenario can be divided into limited-stage, extensive-stage and relapsed disease.

Limited-stage disease

Limited-stage disease is treated with curative intent with chemotherapy and radiation therapy, with approximately 20% of patients achieving a cure.

Combination regimens using chemotherapy and radiotherapy are employed in the treatment of limited stage disease. Popular chemotherapeutic protocols include 1) Etoposide and Cisplatin (EP) 2) Cyclophosphamide, doxorubicin, vincristine (CAV), and 3) Cyclophosphamide, doxorubicin and etoposide (CAE).

The most commonly used regimen is EP, as it has the most favorable toxicity profile. The addition of thoracic radiotherapy (to control local disease has a modest advantage in survival[23,24] and should be considered early, while weighing against the overall treatment toxicity. The overall response rate is 80–90%, with a 2-year survival rate of 20–40%.[25]

Extensive disease

Chemotherapy remains the mainstay of extensive stage disease, with regimens similar to limited stage disease. Routine thoracic irradiation is not prescribed and its role is relegated to palliation of symptomatic sites. The overall response rate is 50–70%, with 2-year survival of 5–10%.

Recurrent disease

Following successful chemotherapy, most patients experience relapse and die within 2 years. The prognosis for recurrent small cell lung cancer is poor, with median survival of 4–5 months when treated. Salvage protocols depend on the previous chemotherapeutic regimen, its response, and the duration of response.

Prophylactic cranial irradiation

Patients with limited disease achieving a complete remission should be considered for prophylatic cranial irradiation, which reduces the occurrences of brain metastases.[26]

Surgery for small cell lung cancer

Surgery in not considered standard treatment in the management of small cell lung cancer, as studies have not shown a survival advantage with the addition of surgery in limited-stage disease. There has been some interest, however, in patients that present with a solitary pulmonary nodule. In historical series of patients who present with a solitary pulmonary nodules that is subsequently found to be small cell following resection, impressive 5-year survival reached 40–53%.[27] Surgery was often combined with post-operative chemotherapy. Hence, this encouraging result has revived the controversial role of surgery in the management of very small cell lung (T 1-2, N0 tumors).

The Solitary Pulmonary Nodule (SPN)

The optimal management of an isolated pulmonary opacity seen on plain chest X-Ray is not an unusual clinical dilemma. On the one hand, the resection of malignant nodules provides the best opportunity for cure, with 5-year cure rates as high as 80% with stage I lesions.[28] This should be

weighed against the risks of an unnecessary thoracotomy for benign lesions. A solitary pulmonary nodule is defined as a well-circumscribed spherical lesion that is completely surrounded by normal lung tissue, and is not associated with adenopathy or atelectasis. The size is variable, but is generally agreed to be less than 3 cm in diameter. The causes of the solitary pulmonary nodule can be divided into malignant and benign causes. Examples of common causes of solitary pulmonary nodules are as follows:

Malignant	**Benign**
Primary bronchogenic carcinoma	Granulomas: tuberculosis, histoplasma, coccidiodomycosis
Solitary pulmonary metastasis	Harmatoma
Bronchial carcinoid	Abscess
Pulmonary sarcoma	Pulmonary arteriovenous malformations
	Pulmonary sequestration
	Pulmonary infarcts
	Bronchogenic cysts
	Wegener's granulomatosis
	Rheumatoid nodule
	Amyloid
	Echinococcal cyst

The majority of solitary pulmonary nodules (> 50%) have a benign etiology.[29,30] Malignant causes account for 20% of cases, and to up to 40% in certain surgical series.[31,32] Factors that predict benignity include:

1) *Radiographic stability*
   A nodule that has remained stable in size over duration of 2 years is likely to be benign, and no further investigations beyond annual chest radiographs are necessary.[33] It cannot be emphasized enough that the first step in the evaluation of SPN is a review of previous chest X-Rays.

2) *Calcification pattern*
   Characteristically, calcification in a difuse, laminated or central pattern favors a benign process. Harmatomas tend to have a popcorn calcification pattern. Eccentric or stippled calcification, is seen in both benign and malignant lesions.

Predictors of malignancy

The presence of **spiculation** and irregularity on the nodule edge seen on both the plain X-Rays and computed tomographic studies is suggestive of

a malignant nodule. In addition, **larger nodules** (e.g. >3 cm), and advanced age tend to favor a malignant nodule.[34] While **cavitation** of a nodule is seen in both benign and malignant nodules, cavitory nodules with a wall thickness >1 mm were more likely to be malignant (84% malignant), compared with <5 mm (95% benign).[35]

Histological diagnosis may be obtained by transthoracic needle aspiration, bronchoscopic biopsy, thoracoscopy and thoracotomy.

Screening for Lung Cancer

The five-year survival for patients with lung cancer, at 7–13%[36,37] is an appalling figure. At the time of presentation, the majority of patients have advanced disease beyond curative surgery.[41] The idea of earlier detection of bronchogenic carcinoma via screening methods is an appealing concept. The most celebrated trials involving lung cancer screening were in the 1970s by the Mayo Clinic,[38] John Hopkins[39] and Memorial Sloan–Kettering Hospitals.[40] In the Memorial Sloan–Kettering Lung project and, and John Hopkins Lung project, subjects were randomized to receive annual chest X-Rays (control group), or annual chest X-Rays with four monthly sputum cytology (screened group). No significant differences were found between the two groups with regards to overall cancers detected, number of late stage cancers and number of resectable cancers. The Mayo Lung project had a slightly different protocol, the subjects first undergoing a prevalence screen comprising a chest radiograph and sputum cytologic examination. Only patients found to be cancer-free were subsequently randomized to a control group, who received standard recommendation comprising annual chest X-Rays and sputum cytology, with no efforts at enforcing compliance, or the experimental group, who received four monthly chest X-Rays and sputum cytology. The Mayo Lung Project found the proportion of early stage and resectable cancers to be higher in the experimental group. Overall mortality between the two groups, however, did not reach statistical significance. Much more recently, the Early Lung Cancer Action Project (ELCAP) looked at the use of low radiation dose computer tomography (low dose CT) in people at high risk of lung cancer. They reported that low dose CT can greatly improve the possibility of detecting non small calcified nodules and thus of lung cancer when compared with CXR.[41] This sensitivity has to be weighed against a high rate of false positive results, and high financial

cost. Moreover, a reduction in mortality has not been demonstrated with low dose CT screening. Other screening tools under evaluation include autoflorescence bronchoscopy, and monoclonal antibodies to detect neoplastic changes in the sputum of high-risk patient groups.

With ongoing enthusiasm into developing the ideal screening method for lung cancer, we should not neglect the role of primary prevention in reducing lung cancer mortality. Efforts should continue in discouraging the smoking habit and facilitating smoking cessation.

REFERENCES

1. Chia KS, Lee HP, Seow A, Shanmugaratnam K, Trends in Cancer Incidence in Singapore 1968–1992, Singapore Cancer Registry Report No. 2, 1996.
2. Miller BA, Ries LAG, Hankey BF *et al.* (eds.), SEER. Cancer statistics review: 1973–1990, Nat Cancer Institute. NIH Pub. No. 93-2789, 1993.
3. Advisory Committee to the Surgeon General of the US Public Health Service, Smoking and Health. Washington DC, Public Health Publication No. 1103, US Government Printing Office, 1964.
4. Adelman M, Haponik EF, Bleeker ER, Britt EJ, Cryptogenic hemoptysis: clinical features, bronchoscopic findings and natural history in 67 patients, *Ann Intern Med* **102**:829–834, 1985.
5. Lederle FA, Nichol KL, Parenti CM, Bronchoscopy to evaluate hemoptysis in older men with non-suspicious chest roentgenograms, *Chest* **95**:1043–1047, 1989.
6. Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E, in collaboration with Sobin LH, and pathologists from 14 countries, *Histologic Typing of Lung and Pleural Tumours*, 3rd ed., World Health Organization, Springer, 1999.
7. Pretreatment evaluation of non-small cell lung cancer, The American Thoracic Society and the European Respiratory Society, *Am J Respir Crit Care Med* **156**:320–332, 1997.
8. Mountain CF, A new international staging system for the staging of lung cancer, *Chest* **111**:1718–1723, 1997.
9. Kirn DH, Lynch TJ, Mentzer SJ, Lee TH, Struss GM, Elias AD, *et al.*, Multimodality therapy of patients with stage IIIA, N2 non small cell lung cancer. Impact of preoperative chemotherapy on resectability and downstaging, *J Thorac Cardiovasc Surg* **106**:696–702, 1993.
10. Martinin N, Kris MG, Flehinger BJ, Gralla RJ, Bains MS, Burt ME, Preoperative chemotherapy for Stage IIIA (N2) lung cancer: the Sloan–Kettering experience with 136 patients, *Ann Thorac Surg* **55**:1365–1373, 1993.
11. Ginsberg RJ, Rubinstein LV, Randomized trial of lobectomy versus limited resection for T1 NO non small cell lung cancer. Lung cancer Study Group, *Ann Thorac Surg* **60**:615–622, 1995.

12. Pearson FG, Delarue NC, Ilves R, Todd TR, Cooper JD, Significance of positive superior mediastinal lymph nodes identified at mediastinoscopy in patients with resectable cancer of the lung, *J Thorac Cardiovasc Surg* **83**:1–11, 1982.

13. Martini N, Flehinger BJ, The role of surgery in N2 lung cancer, *Surg Clin North Am* **67**:1037–1049, 1987.

14. Mcloud TC, Bourgouin PM, Greenberg RW, Kosiuk JP, Templeton PA, Shepard JA, Bronchogenic carcinoma: analysis of staging in the mediastinum by CT and correlative lymph node mapping and sampling, *Radiology* **182**:319–323, 1992.

15. Dales RE, Stark RM, Raman S, Computerized tomography to stage lung cancer. Approaching a controversy using meta-analysis, *Am Rev Respir Dis* **141**:1096–1101, 1990.

16. American College of Physicians, Preoperative pulmonary function testing, *Ann Intern Med* **112**:793–794, 1990.

17. Dunn WF, Scanlon PD, Preoperative pulmonary function testing for patients with lung cancer, *Mayo Clin Proc* **68**:371–377, 1993.

18. Markos J, Mullan BP, Hillma DR, Musk AW, Antico VF, Lovegrove FT, *et al.*, Preoperative assessment as a predictor of morbidity and mortality after lung resection, *Am Rev Resp Dis* **139**:902–910, 1989.

19. Bechard D, Wetstein L, Assessment of exercise oxygen consumption as pre-operative criterion for lung resection, *Ann Thorac Surg* **44**:344–349, 1987.

20. Smith TP, Kinasewitz GT, Tucker WY, Spillers WP, George RB, Exercise capacity as a predictor of post-thoracotomy morbidity, *Am Rev Resp Dis* **129**:730–734, 1982.

21. Non-small Cell Lung Cancer Collaborative Group: Chemotherapy in non-small cell lung cancer: a meta-analysis in patients from 52 randomized clinical trials, *BMJ* **311**:899–909, 1995.

22. Souhami RL, Law K, Longevity in small cell lung cancer: a report to the lung cancer subcommittee of the United Kingdom Co-ordinating Committee for Cancer research, *Br J Cancer* **61**:584–589, 1990.

23. Kotalik J, Yu E, Markman BR, *et al.*, Practice guideline on prophylactic cranial irradiation in small-cell lung cancer, *Int J Radiat Oncol Biol Phys* **50**:309–316, 2001.

24. Pignon JP, Arriagada R, Ihde DC, Johnson DH, Perry MC, Souhami RT, *et al.*, A metaanalysis of thoracic radiotherapy for small cell lung cancer, *New Engl J Med* **327**:1618–1624, 1992.

25. Johnson BE, Bridges JD, Sobczeck M, Patients with limited stage small cell lung cancer treated with concurrent twice daily chest radiotherapy and etoposide/cisplatin followed by cyclophasphamide doxorubicin and vincristine, *J Clin Oncol* **14**:806–813, 1996.

26. Kristjansen PE, Hansen HH, Prophylactic cranial irradiation in small cell lung cancer, *Lung Cancer* **87**:161–162, 1995.

27. Kreisman H, Wolkove N, Quiox E, Small cell lung cancer presenting as a solitary pulmonary nodule, *Chest* **101**:225–231, 1992.

28. Lillington GA, Management of solitary pulmonary nodules, *Dis Mon* **37**:271–318, 1991.

29. Swenson SJ, Jett JR, Payne WS, Viggiano RW, Pairolero PC, Trastek VF, An integrated approach to the evaluation of the solitary pulmonary nodule, *Mayo Clin Proc* **65**:173–186, 1990.

30. Lillington GA, Systemic diagnostic approach to pulmonary nodules, in Fishman AP (ed)., *Pulmonary Diseases and Disorders*, 2nd ed., McGraw-Hill Book Company, New York NY, **3**:1945–1954, 1988.

31. Khouri NF, Meziane MA, Zerhouni EA, Fishman EK, The solitary pulmonary nodule assessment, diagnosis and management, *Chest* **91**:128–133, 1987.

32. Midthun DE, Swensen SJ, Jett JR, Clinical strategies for the solitary pulmonary nodule, *Annu Rev Med* **43**:195–208, 1992.

33. Good CA, Wilson TW, The solitary circumscribed pulmonary nodule: study of seven hundred five cases encountered roengographically in a period of three and one half years, *JAMA* **166**:210–215, 1958.

34. Midthun DE, Swensen SSJ, Jett JR, Approach to the solitary pulmonary nodule, *Mayo Clin Proc* **68**:378–385, 1993.

35. Woodring JH, Fried AM, Significance of wall thickness in solitary cavities of the lung: a follow up study, *Am J Roentgenol* **140**:473–474, 1983.

36. Boring CSS, Squires TS, Tong T, Cancer statistics, *CA Cancer J Clin* **42**:19–38, 1992.

37. Bignall JR, Martin M, Smithers DW, Survival in 6068 cases of bronchial carcinoma, *Lancet* **967**:1067–1070.

38. Fontana R, Sanderson DR, Woolner LB, Lung cancer screening: the Mayo programme, *J Occup Med* **28**:746–750, 1986.

39. Tockman M, Survival and mortality from lung cancers in a screened population: the John Hopkins Study, *Chest* **89** (suppl):325S–326, 1986.

40. Melamed MR, Flehinger RB, Heelan RT, Perchick WA, Martinin N, Screening for early lung cancer: results of the Memorial Sloan– Kettering study in New York, *Chest* **86**:44–53, 1984.

41. Henschke CI, McCauley DI, Yankelentz DF, Naidich DP, McGuiness G, Mietiren OS, *et al*., Early lung cancer action project: overall design and findings from bedside survey, *Lancet* **354**:99–104, 1999.

79

Pulmonary Infections

Loo Chian Min and Leng Poh Hock

The entire respiratory tract extends from the nasal passage to the alveolus and infections can occur anywhere along this route. In addition, infections of the sinuses and pleural space contribute significantly to respiratory dysfunction. The host response to infection, virulence and load of infecting organisms and sensitivity of the microbe to antibiotics, all determine the overall outcome of a host who has pulmonary infections. We shall discuss upper and lower respiratory tract infections, infections of the pleural space and uncommon but life-threatening pulmonary infections in this chapter.

ACUTE UPPER RESPIRATORY TRACT INFECTIONS

Common Cold

Etiology

Viruses, e.g. *adenovirus*, *influenza*, *parainfluenza* and *respiratory syncytial virus*. Transmission is by aerosolized droplets and/or contaminated secretions on fomites and hands.

Clinical features

It is usually an acute illness involving the nasopharynx with fever and cough. After an incubation period of 12 hours to 5 days, infection develops with local inflammation and submucosal edema and subsequent sloughing of the affected epithelial cells of the nose and pharynx. General malaise and myalgia can occur. Complications are rare but sinusitis and otitis media can occur.

Treatment

Management is mainly symptomatic with antipyretics and nasal decongestants. Antibiotics are not routinely indicated unless clinical symptoms persist and complications of pneumonia, bronchitis or sinusitis occur. Randomized controlled clinical trials have shown that neuraminidase inhibitors, e.g. zanamivir and ostelmivir, can reduce symptoms of influenza by 1–1.5 days if used within 48 hours of onset of illness. Cost and side effects limit their widespread use. In addition, amantadine and rimantidine are effective against influenza A. Vaccination with inactivated live virus should be considered in patients who are >65 years old, have cardiopulmonary disease, diabetes, renal failure, immunosuppression or resident of a chronic care facility. Amantadine and rimantadine can be administered simultaneously with the vaccine to prevent influenza.

Sinusitis

Etiology

Bacteria: *Streptococcus pneumoniae, Hemophilus influenzae, Moraxella catarrhalis, anaerobes, Staphylococcus aureus*, Gram-negative bacteria
Viruses: Influenza, parainfluenza, rhinovirus virus.
Others: Actinomyces, Nocardia, Rhizopus and Rhizomucor species.

Sinusitis can be part of a sinopulmonary spectrum of diseases including ciliary dysmotility syndrome, Wegener's granulomatosis, cystic fibrosis and diffuse panbronchiolitis.

Clinical Features

These include low-grade fever, malaise, blocked nose, purulent nasal discharge, facial pain, toothache, headache and pain on mastication. Examination may reveal discomfort and abnormal transillumination over

the appropriate sinus. Purulent secretions through the ostium of the involved sinus may be seen on speculum examination or nasoendoscopy. Sinus radiograph may show fluid levels, complete antral opacity or mucosal thickening, although it is not as sensitive as the CT (computed tomogram) scan, especially for the ethmoidal and sphenoidal sinuses. A diagnostic antral puncture and culture of the secretions is useful for complicated cases.

Treatment

Antibiotics need to cover both *S. pneumoniae and H. influenzae*. Amoxycillin, amoxycillin-clavulanate or macrolides are the preferred first line antibiotics. Duration of therapy is usually extended for 7 days after resolution of symptoms. Nasal decongestants should also be used.

Complications

Rare complications include local and cerebral abscesses, cerebral infarction, meningitis, osteomyelitis, cellulitis, and cavernous sinus thrombosis.

Pharyngitis

Etiology

Bacteria: Group A Streptococcus, sexually transmitted diseases (gonorrhea, syphilis, Herpes Simplex virus), chlamydia, *Corynebacterium diptheria, Arcanobacterium haemolyticum*.
Viruses: Adenoviruses, Coxsackie viruses, Epstein–Barr virus, retrovirus.

Clinical features

The hallmark is a sore throat, often with associated fever and malaise. The presence of membrane suggests bacterial infection and culture for *Corynebacterium diphtheriae* should be obtained. Localized cervical lymphadenopathy is frequent in streptococcal infection and infectious mononucleosis (IMS).

Treatment

Antibiotics should not be used if the cause of the sore throat is viral in origin. However, antibiotics should not be withheld in the severely ill,

suspected or previous rheumatic fever, recurrent tonsillitis, concomitant laryngitis and stridor. A throat swab yielding group A Streptococci is not diagnostic of streptococcal pharyngitis for this may be a normal throat commensal. Bacterial infections contribute to a third of cases of pharyngitis. Therapy with penicillin or erythromycin is justified for persistent and severe symptoms. Ampicillin and its derivatives should be avoided for sore throats, as these are well-recognized to produce a widespread skin rash in IMS.

Epiglottitis

This condition is well-recognized in children. It has been described with increasing frequency in adults over the last few years. *H. influenzae* is often the offending microbe.

Clinical features

These range from sore throat, cough and dysphagia to rapidly increasing life threatening symptoms of upper airway obstruction and respiratory distress over 1–3 days. This can be a cause of acute upper airway obstruction. Findings include fever, neck tenderness, stridor and pooling of saliva in the throat. The epiglottis should not be visualized unless immediate endotracheal intubation or tracheostomy is available. The diagnosis is suggested by lateral radiographs of the neck, which show laryngeal and epiglottic swelling (the "thumb sign").

Treatment

Antibiotics active against the likely pathogens, in particular *H. influenzae* should be started immediately. Intravenous third generation cephalosporins provide adequate coverage in this situation. Facilities for urgent upper airway management should be available at all times.

Laryngotracheitis

In adults larygngotracheitis is usually viral in etiology, although *M. catarrhalis* can be the causative agent. Secondary infection of the trachea by *S. pneumoniae* and *H. influenzae* is a complication of viral upper respiratory tract infection. S*taphylococcus aureus* tracheitis can be a sequelae of

herpes virus infection. Airway intubation and tuberculosis can lead to tracheitis and airway stenosis. Initially there may be a dry cough with retrosternal soreness. Beta-agonist aerosols will help wheeze and cough. Good hydration and physiotherapy facilitates removal of viscid secretions. Appropriate antibiotics should be prescribed.

Acute Bronchitis

Acute bronchitis may occur at the time of or shortly after an upper respiratory infection. The etiological agents include viruses, particularly adenovirus or influenza virus in adults, and respiratory syncytial or parainfluenza virus in children; *Mycoplasma pneumoniae*, *H. influenzae* and *S. pneumoniae* are also involved. If the patient has chronic pulmonary disease, then additional bacterial pathogens are usually implicated (e.g. *Pseudomonas aeruginosa or S. aureus* in bronchiectasis, *H. influenzae* or *M. catarrhalis* in chronic bronchitis).

Clinical features

Cough, chest discomfort and dyspnea are the main symptoms. A cough productive of blood stained mucopurulent sputum with fever may be present, especially when there is superimposed bacterial infection. Wheezes and crepitations may be heard on lung auscultation. A persistent cough should arouse the suspicion of infection with *Bordetella pertussis.*

Treatment

Antibiotics can hasten recovery. In general, macrolides, cephalosporins and ampicillin with beta lactamase inhibitor for 7–10 days are useful agents.

ACUTE INFECTIVE EXACERBATION OF CHRONIC BRONCHITIS

Acute infective exacerbation of chronic bronchitis responds to antibiotic therapy, particularly if increased sputum production, dyspnea and sputum purulence are noted. Macrolides, amoxycillin-clavulanate and fluoroquinolones are effective therapies.

Bronchiectasis

Bronchiectasis refers to destruction of the distal bronchial walls resulting in dilatation and ectasia. There is a predisposition to proximal airway involvement in aspergillosis. Upper lobe involvement may be seen in mycobacterial infections, allergic bronchopulmonary aspergillosis (ABPA), chronic mycotic infections and cystic fibrosis.

Etiology

Infection: Mycobacterium tuberculosis, measles pneumonia, whooping cough, adenoviruses, mycoplasma and pneumococcal pneumonia
Mucociliary clearance defects: Cystic fibrosis, Immotile cilia syndrome, Kartegener's syndrome
Immunoglobulin disorders: Congenital and acquired agammaglobulinemia and hypergammaglobulinemia IgE (Job's syndrome)
Immunologic: Allergic bronchopulmonary aspergillosis (ABPA)
Other causes: Chronic bronchial obstruction from benign tumors, foreign body and stenosis; inhalational injury, recurrent aspiration (such as from gastroesophageal reflux), inflammatory bowel disease, sequestrated lung, relapsing polychondritis, post-lung transplantation, chronic granulomatous disease of childhood, alpha 1-antitrypsin deficiency.

Associated conditions

Felty's syndrome, Yellow nail syndrome (triad of discoloration of nails, leg lymphoedema and pleural effusion), Young's syndrome (obstructive azoospermia) and Macleod–Swyer–James syndrome.

Clinical features

Those with minimal or focal bronchiectasis may be asymptomatic or have mild symptoms. They may present with a cough productive of purulent sputum with or without hemoptysis during infective exacerbations. In the severely affected patients, the cough productive of copious purulent sputum is continuous. Accompanying features include intermittent fever, anorexia, weight loss, clubbing, arthralgia, hypertrophic pulmonary osteoarthropathy, cyanosis and cor pulmonale. Bronchospasm may be present during infective exacerbations, or in cases of ABPA. Otherwise the

only sign present is localized or widespread coarse crepitations. Specific features of associated conditions mentioned above should be sought for.

Investigations

Chest radiography: Changes relevant to bronchiectasis are tramline shadows (suggestive of bronchial wall edema), cystic lesions with or without air fluid levels (suggestive of saccular bronchiectasis), volume loss, crowding of bronchovascular markings (indicative of damaged or infected areas), evidence of past tuberculosis or mycosis, and pulmonary fibrosis (this may have associated traction bronchiectasis).

High resolution CT scan: This has replaced the bronchogram as the diagnostic test because of the ease, availability, and safety factors. The specificity and sensitivity are more than 90%. Findings include signet-ring shadows (dilated bronchus with adjacent bronchial artery), bronchial wall thickening, dilated bronchi extending to the periphery, bronchial obstruction from inspissated secretions, consolidative changes during infective exacerbations, volume loss and cystic changes.

Sputum cultures: Acid-fast bacillus smear and mycobacterial cultures should be obtained to rule out tuberculosis.

Others: Relevant tests for associated conditions or etiologies such as sinus radiograph, serum immunoglobulins, aspergillus precipitins, skin prick test or specific IgE to aspergillus, barium swallow and bronchoscopy (especially for localized bronchiectasis with volume loss).

Complications

Hemoptysis can occur, usually during infective exacerbations in up to 50% of patients with bronchiectasis and can be life-threatening. Those with diffuse and severe bronchiectasis can progress to respiratory failure. Secondary infections with multi-drug resistant bacteria, fungi and atypical mycobacteria can complicate the management. In recent years, *Burkholderia cepacia* has emerged as a respiratory pathogen of increasing importance in patients with cystic fibrosis. Its presence is a marker of poorer survival.

Treatment

The 2 key measures are physiotherapy and antibiotics for symptom control and prevention of complications.

Physiotherapy: Postural drainage should be done twice daily, usually on waking up in the morning and before bedtime. Cough expectorant and mucolytics appear to be ineffective but humidification, nebulized saline and bronchodilators can be beneficial.

Antibiotic therapy: A 10 to 14-day course is indicated for a clearly defined infective exacerbation as manifested by increase sputum purulence with or without fever and increased dyspnea. Common pathogens encountered are *H. influenzae, S. pneumoniae* and *S. aureus.* However in those with severe bronchiectasis with recurrent or persistent infection or colonization, coverage for *P. aeruginosa* may be necessary. Acute infective episodes may sometimes require hospitalization for intensive therapy, including chest physiotherapy and intravenous antibiotics. Nebulized aminoglycosides has also been reported to be useful in patients with cystic fibrosis and severe bronchiectasis with persistent infection.

Surgery: This is reserved for localized, resectable and symptomatic bronchiectasis or in patients with massive hemoptysis. Bronchoscopy is useful in localizing the bleeding pulmonary segment. If surgery is elective, a full pulmonary function test is warranted. Bronchial artery embolization can be considered if the expertise is available.

Specific therapies: These include intravenous immunoglobulin replacement for patients with immunoglobulin deficiency, relief of airway obstruction, control of recurrent aspiration, treatment of upper respiratory tract infections such as chronic sinusitis.

Pneumonia

Pneumonia is defined as infection of the lung parenchyma where there is accumulation of secretions and inflammatory cells in the alveolar spaces. Despite the advent of potent antibiotics, it is still the third principal cause of death in Singapore. The management of pneumonia requires an understanding of the clinical spectrum of pneumonia and an awareness of the current antibiograms in specific settings.

Classification

A useful and practical classification of the types of pneumonia includes reference to the clinical and environmental circumstances under which the pneumonia is acquired and to the prior clinical state of the patient.

Initial choice of empiric antibiotic influences mortality and morbidity outcome and is based on the clinical classification of the type of pneumonia (Table 1) and risk stratification of the patient.

Community Acquired Pneumonia

Community acquired pneumonia (CAP) is defined as an acute infection of the pulmonary parenchyma in a patient not hospitalized or residing in a long-term care facility for more than 14 days. It is usually accompanied by symptoms of acute infection, presence of auscultatory findings consistent with pneumonia and new infiltrates on a chest radiograph.

Etiology

The list of pathogens causing community acquired pneumonia and their frequency can differ geographically. Knowledge of prevalence and sensitivities of the pathogens can greatly aid initial empiric therapy.

Streptococcus pneumoniae is the most common CAP pathogen in Singapore and other parts of the world.

Atypical pneumonia is used to describe pneumonia that have clinical features that are unlike those of the usual bacterial pneumonia. Extrapulmonary manifestions are more common. Organisms causing atypical pneumonia include:

- *Mycoplasma pneumoniae*;
- *Legionella species*;
- *Chlamydia* (TWAR and *psittaci* strains);
- *Coxiella brunetti* (Q fever); and
- Viruses.

Burkholderia pseudomallei: Melioidosis is more prevalent in some parts of Southeast Asia, including Singapore, and Northern Australia. Sporadic cases found elsewhere are related to travel to these endemic areas. Incubation period varies from days to years. Transmission occurs by direct inoculation from soil through small cuts or abrasions, inhalation of contaminated dust and ingestion or aspiration of contaminated water. Human-to-human transmission is extremely rare. Predisposing factors include diabetes, renal disease, alcoholism, and occupations such as rice farming, army and construction. Presentation may be acute with fulminant

Table 1 Clinical Classification of Pneumonia and Common Microbial Causes

CAP*	NP+ (includes VAP++)	Aspiration Pneumonia	Opportunitic Pneumonia
Streptococcus pneumoniae (15–76%)	Gram-negative bacilli (50–70%): *Klebsiella pneumoniae* *Pseudomonas aeruginosa* *Escherichia coli*, *Acinetobacter baumannii* *Hemophilus influenzae* *Enterobacter species*, *Stenotrophomonas maltophilia* *Legionella pneumophilia*	Community acquired: *Peptostreptococci* *Peptococcus* *Bacteroides fragilis* *Melaninogenicus* *Fusobacterium nucleatum*	*Pneumocystis carinii* Fungus: *Candida* species, *Aspergillus* species *Cytomegalovirus* *Herpes simplex* *Nocardia asteroides* *Actinomyces israelii* *Mycobacterium tuberculosis complex* *Mycobacterium avium intracellulare*
Hemophilus influenzae (3–46%)			
Mycoplasma pneumoniae (2–14%) (in those < 40 years old)		Nosocomial: *Staphylococcus aureus* *Escherichia coli* *Klebsiella pneumoniae* *Pseudomonas aeruginosa*	
Chlamydia pneumoniae or *C. psittaci* (3–12%)	Gram-positive: *Staphylococcus aureus* (including MRSA) *Streptococcus pneumoniae* Other streptococci		
Influenza A (may be complicated *by Staphylococcus aureus* infection)			
Legionella pneumophila (0–15%)			
Staphylococcus aureus (3–14%)			
Moxarella catarrhalis (1–2%)			
Klebsiella pneumoniae (3–14%)			
Burkholderia pseudomallei (in endemic areas)			
Mycobacterium tuberculosis complex (in endemic areas)	Anaerobes		
	Fungi		
	More resistant organisms with VAP		

*CAP-community acquired pneumonia, +NP-nosocomial pneumonia, ++VAP-ventilator associated pneumonia

septicemia, subacute with fever and localized pneumonia or lung abscess, or chronic, similar to a tuberculous infection. Mortality is high in patients with advanced age, septicemia, smoking history and renal or heart failure. Relapse is common, even after appropriate antibiotic treatment.

Clinical features

Symptoms and signs (Table 2) are neither sensitive nor specific in defining the etiology of CAP.

Respiratory symptoms vary but classically include cough, dyspnea, sputum production with or without hemoptysis and pleurisy. Clinical signs may be minimal or florid such as cyanosis, tachypnea, confusion and signs of consolidation on chest examination.

In atypical pneumonia, constitutional symptoms such as headaches, malaise, arthralgia and gastrointestinal tract symptoms may predominate and precede the chest symptoms by several days. Marked confusion is seen in patients with any severe pneumonia but is a feature of legionellosis and psittacosis. In the elderly, the classic symptoms and signs of pneumonia may be absent, with the only indication being a raised pulse and respiratory rate associated with deterioration in physical or mental state. Travel history is important, bearing in mind the epidemiology of the endemic pathogens of each geographical region.

Investigations

Although an early etiologic diagnosis is optimal in the management of CAP, the responsible pathogen is not defined in as many as 50% of patients, despite extensive diagnostic investigations. There must be a

Table 2 Common Symptoms and Signs in CAP

Symptom	Sign
Fever	Confusion
Cough	Tachypnoea
Purulent sputum production	Dullness to percussion
Perspiration	Inspiratory crackles
Hemoptysis	Bronchial breathing
Dyspnea	Increased vocal resonance
Chest pain and pleurisy	Pleural rub

balance between reasonable diagnostic studies and empiric treatment. Table 3 shows the common investigations that are performed for CAP.

Chest radiograph: This is indicated for all patients with CAP. Radiographic abnormalities include consolidation, diffuse or peribronchial infiltrates, but these correlate poorly with the causative organisms. Cavitation may occur in tuberculosis, melioidosis, Gram negative bacteria, anaerobic, *Staphylococcal* or *pneumococcal* pneumonia and rarely, *Legionella* infection. Parapneumonic effusions may occur in up to 20% of cases and are uncommon in atypical pneumonia.

Sputum examination: The sputum sample must reflect the bronchial airway secretions and not oropharyngeal secretions for proper interpretation of the specimen. Criteria for acceptability of respiratory specimens for bacteriological studies take into account the relative number of polymorphonuclear cells and squamous epithelial cells. Sputum culture is neither sensitive nor specific and should be interpreted together with the sputum Gram stain. Induced sputum is useful for *pneumocystis carinii* or *Mycobacterium tuberculosis*.

Table 3 Investigations for Community Acquired Pneumonia

Blood
• Full blood count
• Serum urea, creatinine and electrolytes
• Liver function test
• Arterial blood gas
• Blood cultures ($\times$2, pre-treatment)
• Serology (generally not useful in acute diagnosis)
Sputum
• Gram stain and bacterial culture
• Acid-fast bacillus stain and mycobacterial culture
Urine
• Urine legionella antigen
• Urine pneumococcal antigen
Imaging
• Chest radiograph
Others (when indicated)
• Pleural fluid assessment
• Bronchoscopy with bronchoalveolar lavage

Blood and pleural fluid culture: Blood cultures should be taken from 2 separate sites before initiation of antibiotics in hospitalized patients. It has low sensitivity but is highly specific. The yield of blood cultures in studies of CAP ranged from 6.7–27%. The presence of bacteremia has prognostic significance as the risk of complications increase in Gram-negative bacteremia. Pleural effusion with CAP should be aspirated. A positive culture or Gram stain is significant since this is a normally sterile site.

Serological studies: Serological studies are not helpful in the initial evaluation of patients with CAP and should not be routinely performed. It may provide data for epidemiological surveillance. In such cases, paired serum samples 14 days apart may be used to document a 4-fold rise in antibody titres. The presence of cold agglutinins supports the diagnosis of *M. pneumoniae* infection, with a sensitivity of 30–60%, but with poor specificity. A *Legionella* antibody titre of $\geqslant$ 1:256 may be used as a criteria for presumptive diagnosis but a convalescent serology should also be sent. Melioidosis serology may be useful and can be present in people with previous exposure but no clinical disease. A positive culture for *B. pseudomallei* is diagnostic.

Antigen detection: Antigen detection is useful only in selected instances. Direct fluorescent antibody (DFA) staining of respiratory secretions for *L. pneumophila* is difficult, and shows poor results when not performed by experts using only certain antibodies. Urinary antigen test is 70% sensitive and 100% specific. It detects antigen of *L. pneumophila* serogroup 1, which accounts for up to 90% of cases of legionella pneumonia. An immunochromatographic membrane to detect *S. pneumoniae* antigen in the urine is reported to have sensitivity and specificity of 86% and 94% respectively.

Invasive tests: There are no studies to show that the early use of invasive tests improves outcome of CAP compared to empiric therapy. Fibreoptic bronchoscopy is useful in the evaluation of a non-resolving or progressing CAP. It is subject to contamination by upper airway flora. Quantitative culture of bronchoalveolar lavage (BAL) fluid or protected specimen brush (PSB) specimen is superior to conventional technique. It is, however, costly and requires technical expertise. Transtracheal aspiration and percutaneous lung biopsy are rarely performed because of higher risks and relatively low yield. Invasive testing should be considered for

those patients in whom etiologies are unknown with non-invasive studies and are not responding to usual antibiotics.

Complications

The list of possible complications is as follows:

- lung abscess;
- complicated parapneumonic effusion or empyema;
- respiratory failure;
- acute respiratory distress syndrome; or
- septic shock.

Treatment

Supportive measures include adequate hydration and rest, supplemental oxygen or ventilatory support if necessary. Chest physiotherapy is not routinely helpful in patients with pneumonia.

Numerous guidelines have been published for the treatment of CAP. Empiric treatment should take into account the epidemiology and antibiotic susceptibility pattern of causative organisms. Some guidelines for empiric treatment of CAP are given in Table 4. Risk assessment is an important consideration in the management. Patients who are well with no comorbidities may be treated as outpatients, whereas seriously ill, elderly or the very young, those with an underlying respiratory disorder, heart or other systemic disease, immunocompromised hosts and those with unsatisfactory home circumstances should be hospitalized. Oral macrolides covers both pneumococcal and atypical pneumonia. Doxycyline is the antibiotic of choice for *Chlamydia* infection. Combination therapy with beta-lactams and macrolides is indicated in patients with CAP who are hospitalized to broaden coverage to include the usual CAP organisms and atypical bacteria as well. Newer generation fluoroquinolones are useful as single agents as they are active against *S. pneumoniae, M. pneumoniae, Legionella* species and Gram-negative bacteria. For severe CAP in endemic areas, empiric treatment should include high-dose ceftazidime to cover *B. pseudomallei*. Intravenous ceftazidime or imipenem for 2 weeks is the initial treatment of choice for melioidosis. This is followed by maintenance therapy with a combination of doxycycline and co-trimoxazole for 6 months to reduce relapse rate.

Table 4 Guidelines for Empiric Therapy for Community Acquired Pneumonia

Singapore (2000)	IDSA (2000)	ATS (2001)	ERS (1998)	BTS (2001)
Oupatient CAP No modifying factors: Macrolide, or doxycycline Modifying factors: Macrolide, or 2nd G* cephalosporin, or beta-lactam/ beta-lactamase inhibitor	*Outpatient CAP* Macrolide, or doxycycline, or newer G quinolone	*Outpatient CAP* No modifying factors: Newer macrolide, or doxycycline Modifying factors: Beta-lactam (high-dose amoxycillin, cephalosporin, or beta-lactam/beta-lactamase inhibitor) + macrolide or doxycycline, or newer G quinolone	*Outpatient CAP* Aminopenicillin, or tetracycline, or oral cephalosporin, or newer G quinolone, or oral streptogramins, macrolide	*Outpatient CAP* Amoxycillin or erythromycin Clarithromycin for those with GI intolerance to erythormycin
Hospitalized CAP General ward: High dose penicillin ± macrolide, or 3rd G cephalosporin or beta-lactam/ beta-lactamase ± macrolide, or newer macrolide, or newer G quinolone ICU**: Macrolide + ceftazidime*** ±	*Hospitalized CAP* Newer G quinolone, or 2nd, 3rd, 4th G cephalosporin + macrolide	*Hospitalized CAP* No modifying factors: iv azithromycin, or or beta-lactam + doxycycline, or newer G quinolone - Modifying factors: iv beta lactam + macrolide or doxycycline; or iv newer G quinolone ICU: No risk for pseudomonas:	*Hospitalized CAP* 2nd, 3rd G cephalosporin, or beta-lactam/beta-lactamase inhibitor, or benzyl-penicillin/ amoxycillin iv ± macrolide, or 2nd G quinolone (ciprofloxacin, ofloxacin), OR newer G quinolone alone	*Hospitalized CAP* Non-severe CAP: Oral Amoxycillin or iv ampicillin + erythromycin or clarithromycin, or levofloxacin alone Severe CAP: iv amoxycillin/ clavulanate or cefuroxime or cefotaxime or

Table 4 (*Continued*)

Singapore (2000)	IDSA (2000)	ATS (2001)	ERS (1998)	BTS (2001)
cloxacillin or clindamycin or vancomycin, or newer G quinolone + ceftazidime		iv beta-lactam + iv macrolide or iv quinolone Risk for pseudomonas: iv antipseudomonal beta-lactam + anti-pseudomonal quinolone (ciprofloxacin), or iv antipseudomonas beta-lactams + iv aminoglycosides + iv macrolides or iv nonpseudomonas quinolone		ceftriaxone + erythromycin or clarithromycin (±rifampicin), or iv levofloxacin + benzylpenicillin

IDSA-Infectious Disease Society of America, ATS-American Thoracic Society, ERS-European Respiratory Society, BTS-British Thoracic Society

*G-generation, **ICU-intensive care unit, ***ceftazidime-high dose to cover melioidosis

Newer G quinolone: levofloxacin, moxifloxacin, gatifloxacin. Newer macrolide: azithromycin, clarithromycin. Beta-lactam/beta-lactamase inhibitor: ampicillin-sulbactam, piperacillin-tazobactam

No controlled trial exists to specifically address the issue of duration of treatment. The following should be taken into consideration:

- pathogen;
- coexisting illness;
- severity of illness;
- treatment response; and
- complications.

Generally, antibiotics for bacterial infection should be continued for 7–10 days, or for 3 more days after fever has settled. Pneumonia with parenchyma necrosis should probably be treated for at least 2 weeks. Atypical or staphylococcal pneumonia should also be treated for 2–3 weeks. Resolution of fever is the first and best indicator of clinical response. Complete radiologic recovery may take 6 weeks to 6 months (especially elderly patients). The common causes of treatment failure are:

- incorrect diagnosis: pulmonary embolism, pulmonary edema;
- resistant organism: penicillin-resistant *S. pneumoniae*;
- resistant infection: *B. pseudomallei*, *M. tuberculosis*, Legionellosis, *S. aureus*;
- complication: empyema, abscess, fever related to drug therapy; and
- underlying disease: lung cancer, cardiac failure, immunodeficiency.

Nosocomial Pneumonia

Nosocomial or hospital-acquired pneumonia is by definition a lung infection that develops in hospitalized patients and was neither present nor incubating at the time of admission. This infection develops at least 2 or more days after hospital admission or there is a recrudescence of infection in those who had pneumonia on admission. Nosocomial pneumonia is the leading cause of death from hospital acquired infections with an associated mortality of 20–50%. It also contributes significantly to the morbidity of hospitalized patients.

Etiology

Table 1 shows the common causative organisms of nosocomial pneumonia. The list of common and emerging resistant organisms are as follows:

- extended spectrum beta-lactamase (ESBL) Gram-negative bacteria (contributed by prior third-generation cephalosporin usage);

- vancomycin-resistant *Enterococcus* (rare in Singapore);
- multi-drug resistant *Acinetobacter baumanii*;
- *Stenotrophomonas maltophilia* (related to carbepenem usage);
- methicillin-resistant *S. aureus* (MRSA); and
- fluconazole-resistant *Candida* species.

Common ESBL Gram-negative bacteria include *Klebsiella pneumoniae*, *Serratia marascens, Citrobacter diversus, Enterobacter species, E. coli* and *Proteus mirabilis*. The main pathogenesis of nosocomial pneumonia is aspiration of upper respiratory secretions colonized with pathogenic organisms. An effective infection control program is important for reducing the prevalence of nosocomial infections.

Clinical features and investigations

Clinical features and investigations are similar to other types of pneumonia although debilitated or very ill patients may not have the classical features of fever, purulent sputum or leucocytosis.

Treatment

Mortality and morbidity of nosocomial pneumonia is much higher than community acquired pneumonia. Antibiotic choice depends on antibiogram and epidemiology of causative organisms of the institution. Specific antibiotics like anti-pseudomonal penicillins, fluoroquinolones and cephalosporins, aminoglycosides and carbepenems for *P. aeruginosa*; vancomycin for MRSA; ampicillin/sulbactam, fluoroquinolones and carbapenems for *Acinetobacter*; co-trimoxazole for *S. maltophilia*; carbepenems, fluoroquinolones and aminoglycosides for ESBL bacteria; and amphotericin for fungus, may be necessary.

Ventilator Associated Pneumonia (VAP)

This is a nosocomial pneumonia developing in mechanically ventilated patient more than 48 hours after intubation. This is a cause of an increase in ICU morbidity and mortality. Table 5 lists some common factors for VAP. Many studies have shown the lack of diagnostic accuracy using clinical features of fever, leucocytosis and purulent endotracheal aspirates. New or progressive radiographic lung infiltrates should suggest the presence of

VAP. Tracheal aspirates, non-bronchoscopic blind catheter lavage or brushing, and bronchoscopic protected specimen brush (PSB) or bronchoalveolar lavage (BAL) are invasive methods to obtain airway secretions directly. Treatment is the same as for nosocomial pneumonia.

Aspiration Pneumonia

Acute aspiration of gastric contents into the lung can cause severe chemical pneumonitis resulting in acute lung injury. Bacterial infection can set in due to aspiration of anaerobes (especially in those with poor dental hygiene) and Gram-negative colonies (in hospitalized patients) from the upper aereodigestive tract. Necrotizing pneumonia, lung abscess, empyema and bronchiectasis are important sequelae. Predisposing causes of aspiration pneumonia are shown in Table 6. Initial antibiotic of choice is clindamycin or penicillin with beta lactam. For nosocomial

Table 5 Risk Factors for Ventilator-Associated Pneumonia

Mechanical ventilation of more than 7 days*	Reintubation
Prior antibiotic administration*	Thoracic or upper abdominal surgery
Multi-organ dysfunction	Duration of hospitalization prior to mechanical ventilation
Chronic obstructive lung disease	Use of positive end-expiratory pressure
Advanced age	Intracranial pressure monitoring and/or depressed consciousness
Histamine type-2 receptor antagonist	Daily ventilator circuit changes
Nasal intubation and/or sinusitis	Supine head positioning, i.e. head of bed not elevated
Aspiration of gastric content	

*associated with antibiotic resistant pathogen
Modified from Kollef MH, Silver P, Respir Care 1995;40:1130

Table 6 Predisposing Causes of Aspiration Pneumonia

Pregnancy
Gastroesophageal reflux
Esophageal stricture
Tracheoesophageal fistula
Vocal cord palsy
Reduced conscious level due to anesthesia, drugs, alcohol, stroke, epilepsy

aspiration pneumonia, antibiotic coverage should include organisms commonly found in the institution, especially Gram-negative bacteria. Predisposing causes for aspiration should also be treated.

Pneumonia due to Opportunistic Infections

Pneumonia is a common cause of morbidity and mortality in hosts with all types of immune deficiency. The increasing magnitude of this problem is due to the increase in numbers of HIV-infected cases and proliferation of usage of immunosuppressive drugs for organ transplantation, malignancy and autoimmune disorders.

Etiology

The causative organisms involved depend on the type of immunodeficiency (Table 7). Immunodeficiency can be mixed and may vary with the type and stage of the primary disease and treatment course.

Investigations

Early use of invasive procedures such as bronchoscopic sampling, needle aspirates and open lung biopsies may be required. Identification of

Table 7 Relationship between Type of Immunodeficiency and Causative Organism

Type of Deficiency	Cause	Causative Agents
Neutropenia	Chemotherapy (early phase)	Bacterial infection (especially Gram-negative and *S. aureus*), fungi (especially *Aspergillus* and *Candida spp.*), community acquired pneumonia organisms
T lymphocyte dysfunction	HIV Immunosuppressive drugs like steroids, cyclosporin, late phase of chemotherapy	Bacterial infection, fungi, *Pneumocystis carinii*, herpes viruses, mycobacterium, *Nocardia, Toxoplasma gondii, Strongyloides stercoralis, Legionella pneumophila*
Immunoglobulin(Ig) deficiency	Hyposplenia, Ig deficiency	*Streptococcus pneumoniae, Hemophilus influenzae, Mycoplasma pneumoniae*

organisms, e.g. *Cytomegalovirus* in bronchoalveolar lavage, could imply pathogenicity or colonization. Correlation with the clinical findings is required.

Treatment

Empiric antimicrobial therapy depends on the clinical setting and the suspected organisms.

Lung Abscess

This is a pus-containing necrotic lesion of the lung parenchyma. Bacteria, mycobacteria, fungi and parasitic infections can cause lung abscess. Bacterial infection is the most common, in particular the anaerobic bacteria from aspiration in the presence of periodontal disease. Other bacteria such as *S. aureus, K. pneumoniae, P. aeruginosa, Nocardia* and *Actinomyces* species and less commonly, *Legionella sp* are also known etiologic agents. In endemic areas, *Burkholderia pseudomallei and Mycobacterium tuberculosis* need to be excluded. Bacteremia, especially from line sepsis, can result in multiple nodules with cavities from septic emboli. Lung abscesses can occur in malignant lesions, pulmonary infarction and pneumoconiosis.

Predisposing factors

Factors predisposing to the development of lung abscess are malnourishment, poor dental hygiene, recurrent aspiration, diabetes, hematological malignancies, immunocompromised states and renal failure.

Clinical features

Fever, cough productive of purulent or blood-stained sputum which may be foul-smelling (due to the presence of anaerobes) and weight loss are symptoms that may be present for days to weeks. The source of sepsis may be evident such as poor oral hygiene and presence of infected vascular access. Clubbing may be present in chronic cases.

Investigations

The chest radiograph classically shows a cavitating lesion with an air fluid level. Sometimes it can be difficult to differentiate a lung abscess

from fluid in a cyst or bleb. If the lung abscess has ruptured into the pleural space, resulting in an empyema with a bronchopleural fistula, the presentation would then be a hydropneumothorax. Sputum Gram stain and cultures should be obtained. Bronchoscopic and percutaneous lung aspiration are other means of obtaining samples for Gram stain and culture studies. Besides sampling secretions, bronchoscopy may be performed to exclude suspected obstructing lesions. Otherwise, routine bronchoscopy for lung abscess is not necessary.

Complications

Rupture of abscess into the airway or the pleural space may cause sudden respiratory decompensation.

Treatment

The majority of lung abscesses respond to antibiotic therapy. The initial choice of antibiotics should cover Gram-positive, Gram-negative and anaerobes. Empiric choice of a third generation cephalosporin/quinolone with clindamycin or a penicillin/beta-lactam antibiotic is appropriate. Antibiotics should be given for at least 4 weeks. Surgical resection is considered when there is no clinical response to antibiotics after 2 weeks, there is an abscess rupture, or when the patient cannot mount an immune response and the abscess is deemed resectable.

Empyema

The pleural space is a sterile cavity and pleural infections can occur via direct inoculation, hematogenous seeding or an adjacent lung infection with spillover effect. Bacterial pneumonia is the most frequent cause of empyema. Other causes are septicemia, trauma, thoracic surgery, perforation of esophagus or lung abscess. Causative organisms are *S. aureus* (especially trauma related), anaerobes, Gram-negative bacilli (especially nosocomial acquired) and *S. pneumoniae* and other streptococci. *Mycobacterium tuberculosis* is also a known cause.

Clinical features

These include fever, cough, purulent sputum, dyspnea and pleurisy. There is a pleural effusion with decreased chest movement, stony dull percussion note, decreased breath sounds, and decreased vocal fremitus.

Investigations

Chest radiography, especially the lateral decubitus film, should be obtained. The absence of free flowing pleural fluid on the decubitus view indicates loculation and pleural fluid sampling and drainage can be done safely under imaging (ultrasound or CT) guidance. Pleural fluid should be analyzed for Gram stain and culture, acid-fast bacillus smear and TB culture, pH, LDH, glucose, total protein, cytology and fungal smear and culture. CT thorax is helpful in assessing loculations, identification of associated lesions such as lung abscess, tumor or bronchopleural fistula, and verifying the position of the chest tube. Bronchoscopy is indicated if an endobronchial tumor is suspected or a bronchopleural fistula is present.

Treatment

Early diagnosis is important. Aspiration of gross pus necessitates early tube drainage and appropriate antibiotics. Loculated empyema may need tube positioning under imaging guidance. Controlled trials show that instillation of fibrinolytic agents such as urokinase and streptokinase into the pleural space can shorten hospital stay. There is some evidence that early intervention with video-assisted thoracoscopy and drainage can improve outcome and reduce morbidity.

OTHER LIFE THREATENING PULMONARY INFECTIONS

Anthrax

The threat of bioterrorism has made inhalational anthrax a serious threat to the world community. Anthrax, caused by Gram-positive *Bacillus anthracis*, cause infections via the inhalational, cutaneous or gastrointestinal route. The incubation period is several days and can be up to a few weeks. An initial flu-like illness consisting of fever, non-productive cough and myalgia is then followed by hemorrhagic mediastinal lymphadenitis. Lymphatic obstruction in the lungs predispose to pulmonary edema. In addition, the release of exotoxins results in the septic shock syndrome with rapid progression to death. Early identification by Gram's stain and treatment are essential in preventing mortality. Intravenous ciprofloxacin is given empirically if *B. anthracis* resistant to penicillin and doxycycline is suspected.

Plague

This is a disease spread by rodent fleas and was the cause of "Black Death" during the 14th century. Pneumonic plague, caused by inhalation of *Yersinia pestis* or by hematogenous dissemination, has an acute presentation of cough and hemoptysis, with progression to respiratory failure, stridor and cyanosis. This is highly contagious. The incubation period is 1 to 4 days. The diagnosis is suggested by Gram stain of the sputum or lymph node aspirate which reveal a Gram-negative coccobacillus. Intravenous aminoglycoside (gentamicin or streptomycin) is treatment of choice. Alternative antibiotics include doxycycline and chloramphenicol.

Tularemia

Pulmonary tularemia, caused by *Franciscella tularensis*, may spread to humans as a form of zoonosis. Exposure to an aerosol of this bacteria from live domestic animals or dead wildlife or hematogenous dissemination from the ulceroglandular form of tularemia can cause pneumonia. It presents initially as an atypical pneumonia, with progression to bilateral patchy infiltrates, cavitating lesions and empyema. This form of pneumonia is suspected after a careful travel and occupational history is obtained. Treatment is effective with IV streptomycin or gentamicin, but not tobramycin.

MYCOBACTERIAL INFECTIONS

Mycobacterium Tuberculosis

Tuberculosis (TB) has remained a major public health problem in underdeveloped and developing countries. Over the last decade, there has been an increase in incidence of TB worldwide because of the increase in numbers of HIV infected cases.

In parallel to this phenomenon, there is also a rise in the emergence of multi-drug resistant tuberculosis (MDR-TB). Almost all cases of TB are caused by *Mycobacterium tuberculosis*, which is one of the mycobacteria belonging to the Mycobacterium tuberculosis complex (Table 8). Transmission is via droplet nuclei expelled when an infected person coughs or sneezes. Infectiousness is highest amongst close contact of infected persons with laryngitis, sputum smear-positive or cavitating

Table 8 Classification of Mycobacteria

Mycobacterium Tuberculosis Complex (MTC)
M. tuberculosis, *M. bovis*, *M. africanum*, *M. microti*
Nontuberculous Mycobacteria (NTM)
Slowly growing organisms
E.g. *M. avium complex*, *M. kansasii*, *M. xenopi*, *M. malmoense*, *M. scrofulaceum*, *M. ulcerans*, *M. marinum*
Rapidly growing organisms
E.g. *M. abscessus*, *M. chelonei*, *M. fortuitum*

pulmonary TB. Latent TB is generally non-infectious. Incubation period is from weeks to years. Ten percent of infected persons with an intact immune system will develop TB disease at some point of their lives. Certain conditions such as diabetes mellitus, immunocompromised states including HIV infection, chronic renal failure, malignancy, gastrectomy, malnutrition and silicosis increase the risk that a latent TB infection will progress to a disease state.

Clinical features

Pulmonary TB accounts for 90% of TB cases. Mild disease may be asymptomatic and is evident only on a screening chest radiograph. Symptom duration at presentation is usually weeks to months with cough, hemoptysis, loss of weight, loss of appetite, fever, night sweats, chills and fatigue. Hemoptysis is generally minor but can also be life-threatening. Hoarseness usually indicates laryngitis. Patients with miliary TB may be very ill and can deteriorate rapidly. Although lung parenchymal involvement is most common, patients may present with just a pleural effusion (commonest cause of exudative pleural effusion amongst patients <60 years in endemic areas) and less commonly empyema. Airway TB has been found at bronchoscopy or post-mortem examination in 15–42% of patients with active pulmonary TB. These patients may have been misdiagnosed to have asthma or chronic obstructive airway disease or even bronchogenic carcinoma. Airway stenosis may develop many years later despite appropriate antituberculous chemotherapy. The incidence of residual airway stenosis range from 12–57%. Extrapulmonary disease such as lymphadenitis (especially cervical), osteomyelitis, arthritis, genitourinary infection, meningitis or gastrointestinal involvement may occur with or without the presence of pulmonary infection. The possibility of underlying HIV infection should be

considered, especially in those with atypical chest radiographic findings, extrapulmonary or disseminated disease.

Investigations

There are three main investigations that have diagnostic value. They are:

1) *Mantoux skin test*: A positive reaction of 5 mm or more induration indicates exposure of the immune system to tubercle bacilli. In endemic areas, a positive reaction may not be clinically significant for disease occurrence unless the reaction is 15 mm or more with clinical and/or radiological features of TB. A strongly positive reaction of 20 mm or more induration is often associated with blistering and intense erythema. False negative reactions may occur in severely ill, immunocompromised hosts and patients with sarcoidosis.
2) *Chest radiograph* abnormalities may be the first to indicate the presence of tuberculosis. This includes cavitating nodular lesions, fluffy infiltrates or nodules especially in the apical and posterior segments of upper lobe and superior segment of lower lobe; diffuse alveolar infiltrates, miliary shadows, lymph node enlargement, pleural effusion and enlarged cardiac size due to pericardial effusion. Atypical appearances such as lower lobe basal segment infiltrates and prominent lymphadenopathy are seen in the immunocompromised host, in particular HIV-infected person.
3) *Bacteriological examination* for smear and culture can be obtained from the sputum collection, laryngeal swab, bronchoscopic sampling (washings, brushings and biopsy), gastric aspirates, pleural fluid (15% positivity) and tissue (80% positivity for caseous granulomas on histology and $>75\%$ positivity for culture). A series of 3 early morning specimens should be collected on different days for sputum (supervised collection), laryngeal and gastric samples. Bronchoscopic sampling may be obtained if there is no sputum or when bronchoscopy is performed to exclude other causes such as bronchogenic carcinoma. Smear examination results should be available within 24 hours. A positive result in endemic areas should be taken as MTC infection unless proven otherwise. However in low prevalence areas, a positive test may be due to NTM and nucleic acid amplification test may be helpful in distinguishing the 2 types of mycobacterial infection. Culture results may take between 10 days (for BACTEC

radiometric system) to 8 weeks (for conventional method). Follow-up bacteriologic examinations are recommended for assessing the patient's infectiousness and response to therapy. Documentation of conversion to smear and culture negative sputum should be achieved during treatment.

The role of nuclei acid amplification (NAA) tests using polymerase (PCR) and ligase (LCR) chain reaction is still not clearly defined.

Complications

Tuberculosis is a chronic inflammatory infection that heals with tissue destruction and fibrosis. Significant sequelae includes bronchiectasis, fibrothorax, end-stage lung destruction, severe airway stenosis with associated obstructive complications and lung fibrosis with mycetoma formation, recurrent or massive hemopytsis and respiratory failure. Chronic bronchopleural fistula may occur in TB empyema.

Treatment

Tables 9 and 10 show the first- and second-line drugs for tuberculosis therapy respectively. The principles of treatment are:

- use the safest, most effective and shortest course of therapy; and
- ensure patient compliance to treatment.

When adequate treatment is given, almost all patients will recover and remain well. Tuberculosis has to be treated with at least 3 drugs in the initial phase and then at least 2 drugs in the continuation phase for a total of 6 to 24 months depending on the drug regime.

Those at risk of MDR-TB are individuals with a history of non-compliance with TB treatment, reactivation of past TB, contacts of MDR-TB cases, persons from areas where prevalence of MDR-TB is $>4\%$ and persons whose smears or cultures remain positive after 2 months of TB treatment. In Singapore the incidence of MDR-TB is $<4\%$.

Review of drug susceptibility test should always be done within 2 months of initiation of drug therapy. Fixed-dose combination drugs such as *Rifamate* (isoniazid 150 mg and rifampicin 300 mg) may enhance adherence to drug therapy. The beneficial effects of corticosteroids have not been well established for TB pleural effusion and airway TB. It may be

Table 9 Antituberculous Therapy for Adults (First-line Drugs)

First-line Drugs	Main Adverse Effects
Isoniazid 5 mg kg day 15 mg kg thrice wk	Hepatitis* (increases with age, hepatic disorders and alcohol consumption) Peripheral neuropathy (give pyridoxine for individuals at risk, e.g. diabetes mellitus, uremia, malnutrition, malignancy, alcoholic, pregnancy)
Rifampicin 10 mg kg day 15 mg kg thrice wk	Gastrointestinal (GI) intolerance (most common) Hepatitis* Drug interaction** Rash Flu-like symptoms Hematological abnormalities
Pyrazinamide 15–30 mg kg day 50–70 mg kg thrice wk	Hepatitis* Rash Arthralgias Hyperuricemia (treat only if symptomatic) GI intolerance
Ethambutol 15 mg kg day 25 mg kg thrice wk	Optic neuritis (usually does not occur with 15 mg/kg daily dose for 2 months) GI intolerance
Streptomycin (intramuscular) 15 mg kg day 25 mg kg thrice wk	Ototoxicity (avoid or reduce dose in those >60 yrs old) Renal impairment

Intermittent regimens should be used with directly observed therapy
*hepatic dysfunction is relatively common with these drugs which may be continued unless the enzymes are >5 times normal or 3 times above the baseline
**rifampicin is a hepatic enzyme inducer, hence dosage of contraceptive pills, corticosteroids, warfarin, anticonvulsants, cyclosporin, theophylline, ketoconazole, oral hypoglycemics and antiarrhythmic drugs have to be increased

useful in those with pronounced systemic features, meningitis, adrenal insufficiency and tuberculous pericardial effusion.

Referral to centres with experts managing tuberculosis such as the Tuberculosis Control Unit in Singapore, is recommended for:

1) Non-adherence to drug therapy. This is a major problem worldwide because of adverse reactions and prolonged therapy. All patients non-adherent to drug therapy must be placed under directly observed

Table 10 Second-line Antituberculous Therapy in Adults

Second-line Drugs	
Ciprofloxacin	Ethionamide
Ofloxacin	Cycloserin
Amikacin	Clofazimine
Kanamycin	Dapsone
Capreomycin	

therapy (DOT) where a health care worker watches the patient swallow each dose of medication.

2) Multi-drug resistant TB, i.e. TB resistant to at least 2 drugs, e.g. isoniazid and rifampicin. This is extremely difficult to manage with high mortality and morbidity rates and is as infectious as drug-sensitive TB. The second-line anti-TB drugs (Table 10) are less tolerable, costly and a more prolonged course of treatment is required for MDR-TB. All these lead to higher non-compliance rates. Surgery may have a role in recalcitrant cases.
3) Relapse cases because of the possibility of non-adherence to drug therapy and MDR-TB.
4) Persistence of symptoms or smear positive after 3 months of therapy.

Lastly, major research efforts are being made in recent years to develop new vaccines for tuberculosis. This renewed interest was prompted by the increasing incidence of MDR-TB worldwide and also the variable protective efficacy of 10–80% by the only currently approved BCG vaccine.

NONTUBERCULOUS MYCOBACTERIA

Nontuberculous mycobacteria (NTM) is also known as environmental mycobacteria and mycobacteria other than tuberculosis (MOTT), Table 8. Most of them are found in natural waters and soil. Human to human transmission has not been documented. Two group of populations are most likely to be affected: 1) those with structural lung disease; and 2) HIV-infected persons. Presentation tends to be subacute in structural lung disease. *M. avium intracellulare* is the most common NTM in HIV patients and the clinical course could be acute and disseminated.

Diagnosis of NTM requires a positive culture. Treatment is usually prolonged and may require multiple drugs. Surgery is an option in localized and poorly responding lesions. As treatment of NTM is generally difficult, prolonged and ineffective, it is important to ensure the presence of disease rather than colonization before commencing treatment. Also, prophylaxis against NTM (with a macrolide) should be considered in HIV patients with low CD4 counts.

MYCOSES

Table 11 shows the common fungi that infect the lungs. They tend to occur in patients with impaired immunity but may also affect previously well individuals

Cryptococcosis

Cryptococcosis is caused by the encapsulated yeast, *Cryptococcus neoformans*, which is found in soil, food and animals especially the excreta of pigeons. Human infection is acquired through inhalation of aerosolized organisms. Conditions that predispose to cryptococcosis are diabetes mellitus and immunocompromised states such as lymphoma, leukemia and corticosteroid ingestion. Infection in normal hosts is often characterized by granulomatous inflammation. Cryptococcal meningitis has to be excluded in every case. These patients present with fever, headache, nausea, anorexia

Table 11 Pulmonary Fungal Infections

Common (worldwide)
Cryptococcosis
Aspergillosis
Candidiasis
Mucormycosis
Pneumocystis carinii
Uncommon (found in endemic areas)
Histoplasmosis*
Blastomycosis
Coccidioidomycosis
Paracoccidioidomycosis
Sporotrichosis

*In the recent years, sporadic cases of histoplasmosis have been seen here. These patients are either from Indonesia or have travelled to endemic areas.

and incidental lung infiltrates on the chest radiograph. Chest radiographic features include cavitating nodular infiltrates, solitary nodule or mass especially in the lower lobes, occasional hilar adenopathy and pleural effusion. Differential radiographic diagnoses include carcinoma and other granulomatous infections e.g. tuberculosis and histoplasmosis. Organisms can be seen in respiratory secretions, biopsy of lung lesions and cerebrospinal fluid (CSF) by using the Indian ink stain. Cryptococcal antigen titers should be routinely done in the serum and CSF, as this may indicate disseminated infection. Cryptococcus can also infect the skin, bone and prostate gland. Treatment is as follows:

1) Observation only if sputum isolated for *C. neoformans* in a normal host with no chest radiographic and CSF abnormalities.
2) Oral fluconazole in those with pulmonary cryptococcosis with negative serum serology and CSF abnormalities.
3) Amphotericin with or without flucytosine for disseminated infection.
4) Lifelong maintenance with fluconazole is needed for AIDS patients.

Aspergillosis

This can be caused by *Aspergillus fumigatus, A. niger or A. flavus* resulting in a spectrum of human diseases:

1) Allergic response in allergic bronchopulmonary aspergillosis.
2) Mycetoma (aspergilloma) of the respiratory tract in preexisting lung diseases.
3) Chronic necrotizing aspergillosis of lung tissue and tracheobronchial aspergillosis in the immunocompromised host.
4) Acute invasive pulmonary aspergillus with systemic hematogenous spread (angioinvasive form). This form most commonly occurs in neutropenic patients.
5) Bronchocentric granulomatosis. This is a more focal form of aspergillus infection involving the airways. It can be seen in asthmatics.

The mere presence of aspergillus in respiratory secretions is not diagnostic and tissue invasion should be evident. However, in patients with hematologic malignancies who have dry cough, hemoptysis, pleurisy, dyspnea, prolonged neutropenia on broad-spectrum antibiotics and pulmonary infiltrates, the presence of aspergillus in the respiratory secretions

should be treated as a pathogen. Pulmonary abnormalities on chest radiograph are that of nodular infiltrates and consolidation with or without cavitation. Mortality in this group is high despite adequate treatment with amphotericin.

Candidiasis

Candidiasis is usually caused by *Candidia albicans* and is more likely to affect the immunocompromised host (especially in hematologic malignancies with prolonged neutropenia), diabetics, patients with burns, renal failure, cirrhosis, prosthesis, catheters and on prolonged antibiotics. Lung involvement is uncommon in the normal host. Chest radiograph shows patchy or diffuse infiltrates. Pulmonary candidiasis is often part of disseminated candida infection as primary lung candidiasis is rare. As the oropharynx is colonized with *Candida* in ill patients, histologic demonstration of tissue invasion is required to prove *Candida* pneumonia. However, the febrile immunocompromised host with positive blood cultures with or without lung infiltrates (on chest radiograph) must be treated. The 2 common drugs used for treatment of candidiasis are amphotericin and fluconazole. All infected vascular access, catheters and prosthesis must be removed. *Candida krusei* and *Torulopsis glabrata* are resistant against the azole group of antifungals and hence therapy is best initiated with amphotericin.

PARASITIC INFECTIONS

Parasitic lung infections are uncommon in Singapore. Important parasitic infections with pulmonary involvement include falciparum malaria (adult respiratory distress syndrome), amoebiasis (lung abscess), microfilaria *Wuchereria bancrofti* (tropical eosinophilic lung) and paragonimiasis (eosinophilic pleural effusion).

REFERENCES

1. Cifu A, Levinson W, Influenza, *JAMA* **284**:2843–2847, 2000.
2. Bisno AL, Acute pharyngitis, *N Engl J Med* **344**:205–211, 2001.
3. Monto A, Gravenstein S, Elliot M, Colopy M, Schweinle J, Clinical signs and symptoms predicting influenza infection, *Arch Intern Med* **160**:3243–3247, 2000.
4. Tan YK, Khoo KL, Chin SP, Ong YY, Aetiology and outcome of severe community acquired pneumonia in Singapore, *Eur Respir J* **12**:113–115, 1998.

5. Hui KP, Chin NK, Chow K, Brownlee A, Yeo TC, Kumarasinghe G, Chan TB, Tan WC, Prospective study of the aetiology of adult community acquired bacterial pneumonia needing hospitalisation in Singapore, *Singapore Med J* **34**:329–334, 1993.

6. Lee KH, Hui KP, Tan WC, Lim TK, Severe community acquired pneumonia in Singapore, *Singapore Med J* **37**:374–377, 1996.

7. Clinical Practice Guidelines, Use of antibiotics in adults. Ministry of Health, Singapore, 2000.

8. Bartlett JG, Dowell SF, Mandell LA, File TM, Musher DM, Fine MJ, Practice guidelines for the management of community-acquired pneumonia in adults, *Clin Infect Dis* **31**:347–382, 2000.

9. Niederman MS, *et al*., Guidelines for the management of adults with community-acquired pneumonia: diagnosis, assessment of severity, antimicrobial therapy, and prevention, *Am J Respir Crit Care Med* **163**:1730–1754, 2001.

10. ERS Task Force Report, Guidelines for management of adult community-acquired lower respiratory tract infections. European Respiratory Society, *Eur Respir J* **11**:986–991, 1998.

11. British Thoracic Society, British Thoracic Society guidelines for the management of community-acquired pneumonia in adults admitted to hospital, *Thorax* **56**(suppl 4): 1–64, 2001.

12. Clinical guidelines on anthrax, botulism, plague and smallpox, Ministry of Health, Singapore, 2001.

13. Guidelines on the clinical diagnosis, laboratory investigation, Treatment and prevention of melioidosis, Singapore General Hospital, 1999.

14. Colice GL, *et al*., Medical and surgical treatment of parapneumonic effusions. An evidence-based guideline. ACCP Consensus Conference, *Chest* **118**:1158–1171, 2000.

15. Chin NK, Lim TK, Controlled trial of intrapleural streptokinase in the treatment of pleural empyema and complicated parapneumonic effusions, *Chest* **111**:275–279, 1997.

16. Wait M, Sharma S, Hohn J, Dal Nogare A, A randomized trial of empyema therapy, *Chest* **111**:1548–1551, 1997.

This page intentionally left blank

80

Acute Respiratory Distress Syndrome

J. Raghuram and Phillip Eng

INTRODUCTION

The Acute Respiratory Distress Syndrome (ARDS) continues to exert its toll on patients in the intensive care units worldwide. Despite advances, ARDS continues to tax resources in terms of financial and human costs. Asbaugh and colleagues[1] first described this condition in 1967 when 12 patients presented with severe dyspnea, diffuse bilateral infiltrates on chest radiograph, decreased lung compliance and hypoxemia, which was refractory to supplemental oxygen therapy but responded to positive pressure ventilation. Autopsies in 7 of these patients revealed marked edema of both lungs with atelectasis and surfactant deficiency mimicking infant respiratory distress syndrome. Subsequently, it was recognized that ARDS affects patients of all age groups and as such, it would be appropriate to refer to it as acute rather than adult respiratory distress syndrome.[2]

Incidence

The true incidence is still controversial. This is primarily because the reported incidence varies depending on the diagnostic criteria and

population studied. In the USA it is estimated to range from 3.5 to 7 per 100 000 individuals per year.[3,4] A recent survey suggested that patients with ARDS occupy approximately 9% of all ICU beds in the USA.[5]

Definition

Clinical experience indicates that this disease varies in severity from patient to patient. It was agreed at the American-European Consensus Conference on ARDS that the term Acute Lung Injury (ALI) be used to represent this wide clinical spectrum of presentation.[2] ARDS is being reserved to represent the most severe end of this spectrum. The following definitions are widely accepted and utilized by clinicians and researchers currently.[2]

Acute lung injury (ALI)

Represents a constellation of symptoms and signs resulting from acute and persistent inflammation of the lung associated with increased vascular permeability. The characteristic features of this syndrome include:

1) acute onset of symptoms and signs;
2) chest radiographic evidence of bilateral pulmonary infiltrates;
3) the ratio of the partial pressure of oxygen (PaO_2) to the inspired fraction of oxygen (FiO_2), i.e. PaO_2/FiO_2 should be $\leqslant 300$ mmHg regardless of the level of Positive End Expiratory Pressure (PEEP); and
4) no evidence of elevated left atrial pressure, i.e. the measured pulmonary capillary wedge pressure (PCWP) should be $\leqslant 18$ mmHg.

ARDS

All the 3 characteristic features of ALI apply except that the degree of hypoxemia associated with ARDS is worse, i.e. $PaO_2/FiO_2 \leqslant 200$ mmHg.

It is not essential to measure PCWP in all cases of ALI/ARDS unless clinical suspicion of heart failure is high.[6] It must be borne in mind that in both ARDS and ALI the degree of hypoxemia does not correlate well with the underlying pathologic changes occurring in the lung. Furthermore, severity of hypoxemia does not have a predictable clinical impact on the outcome.[7]

PATHOLOGY

Early in the course of this disease, an intense inflammatory response characterized by the presence of large quantities of neutrophils and mesenchymal cells is seen in the interstitium. Associated with this there is accumulation of protein-rich fluid within the parenchyma and alveolar space.[8] As the disease progresses, proliferation of the Type II pneumocytes occurs replacing the Type I pneumocytes that under normal conditions covers 95% of the alveolar surface. This inflammatory process progresses at an unpredictable pace, leading to parenchymal fibrosis and obliteration of the adjacent vascular endothelium by fibrin.[9] Hyaline membrane formation, which represents a mixture of debris from cells lining the alveolar surface and fibrin, occurs within the first two days of the onset of the injury to the lung. Approximately 7 to 10 days later, collagen deposition by proliferating fibroblasts occur along the alveolar lining with progression of the fibrotic process.[10] The extent of involvement of the fibrotic process varies. In some patients, this process progresses at an unrelenting pace and leads to permanent irreversible scarring, while in others complete recovery is seen within days. The exact reason for this difference is unknown. However, the extent of the initial insult, development of nosocomial pneumonia and iatrogenic injury to the lung (e.g. oxygen toxicity and barotrauma) may all have a role to play.

PATHOGENESIS

The pathogenesis of ARDS/ALI can be thought of as arising from injury either direct to the pulmonary epithelium and endothelium or indirect secondary to an extra-pulmonary septic or inflammatory focus. This eventually leads to loss of function of the alveolar-capillary unit. This is histologically described as Diffuse Alveolar Damage (DAD). DAD is defined as endothelial and Type I pneumocyte cell necrosis associated with hyaline membrane and proteinaceous alveolar edema formation. There appears to be a temporal sequence of progression of DAD: an exudative phase (1^{st}–3^{rd} day) followed by the fibroproliferative phase (4^{th}–7^{th} day) and eventually the fibrotic phase (after the 1^{st} week). The inflammatory response leading to DAD includes both cellular and humoral factors.[2] Neutrophils, monocytes, macrophages and lymphocytes

being the key players in the cellular response. The mechanism by which neutrophils attach to endothelial lining has been the focus of interest recently because if adhesions can be prevented, lung injury can be reduced or eliminated altogether.[11] The components of the humoral system include the complement, fibrinolytic and coagulation factors interacting with cytokines, lipid mediators, proteases, nitric oxide and other growth factors.[12] These factors are generally thought of as the main effectors of cell injury leading to defects in surfactant metabolism and ultimately ventilation perfusion mismatch.[2] Neutrophils together with alveolar macrophages are in part responsible for the orderly removal of cellular debris and repair of the damaged alveolar epithelium.[11] This process of repair may become disordered in some patients resulting in an exuberant fibrosis leading to inefficient gas exchange and persistent hypoxemia. The reason why acute lung injury resolves in some patients but progresses to extensive fibrosis in others remains unknown. The factors that may be responsible for the development of fibrosis are the severity of the initial insult to the lung, persistence of the inflammatory process and the development of microthrombi within the pulmonary microvasculature.[10]

RISK FACTORS

There are currently more than 60 different causes of ARDS identified. New therapeutic modalities continue to emerge to treat various disease states. Many of these modalities of treatment may result in adverse pulmonary reactions. As such, additional risk factors for the development of ARDS are becoming apparent. Three prospective studies have shown that ARDS most commonly occurred inpatients with the following conditions.[13–15]

1) *Sepsis syndrome*: Bacteremia with systemic manifestations, e.g. metabolic acidosis, hypotension.
2) *Aspiration of gastric contents*
3) *Near drowning*: Immersion injury associated with loss of consciousness, metabolic acidosis or hypothermia.
4) *Pulmonary contusion*: Development of localized infiltrate on chest radiograph within 6 hours of blunt trauma to the overlying chest wall.

The risk of developing ARDS increases as the number of potential risk factors increase.[13,14] Sepsis is very often suspected as the underlying cause of ARDS/ALI and as such many clinicians would begin treatment

for possible infections immediately when ARDS/ALI is encountered.[17] Gram-negative organisms are responsible for most of the cases of ARDS seen in hospitalized patients. ARDS developing in patients outside the hospital is most likely due to a viral infection. *Pneumocystis carinii* pneumonia (PCP) was not considered as a risk factor in all 3 studies.[13–15] However, the American-European Consensus Conference group decided that pulmonary infections including PCP should be considered as risk factors for the development of ARDS/ALI when the physiologic criteria are met.[2] It was also decided that the risk factors be divided into 2 groups:

1) *Direct lung parenchymal injury*
   a) aspiration of gastric contents;
   b) diffuse pulmonary infections;
   c) near drowning;
   d) toxic gas inhalation; and
   e) lung contusion.
2) *Indirect lung injury*
   a) sepsis syndrome;
   b) multiple trauma;
   c) multiple blood transfusion; and
   d) cardiopulmonary bypass.

MANAGEMENT

In general, managing patients with ARDS involves measures that aid in sustaining cellular physiological functions (i.e. gas exchange, organ perfusion and aerobic metabolism).[16] As a result the therapy involves largely supportive measures such as mechanical ventilation, positioning patient, judicious use of fluids and vasoactive agents to maintain hemodynamic stability and systemic perfusion and various techniques to optimize oxygen transport and utilization. The management of these patients with ARDS/ALI should be aimed at establishing the underlying cause. Diagnosing and treating infections are particularly important in this regard. Patients with diffuse pulmonary infiltrates and ARDS should undergo bronchoscopy, bronchoalveolar lavage and possibly transbronchial lung biopsy to establish or rule out the presence of an infection.

Most patients with ARDS/ALI will require positive pressure ventilation to maintain adequate oxygenation. Some patients breathing

spontaneously may be able to achieve sufficient tissue oxygenation and perfusion without mechanical ventilatory support. The clinical objectives of mechanical ventilation in ARDS are to:

1) reverse hypoxemia and respiratory acidosis;
2) relieve respiratory distress; and
3) prevent and reverse pulmonary atelectasis and respiratory muscle fatigue.

Positive End Expiratory Pressure (PEEP)

The conventional approach to ventilating patients with ARDS was to use a volume-oriented approach to ventilation to achieve these goals. Gattinioni and colleagues have shown that in patients with ARDS, the lung compliance is reduced because much of the dorsal aspects of both lungs are collapsed and not accessible to ventilation.[18] The compliance of the aerated ventral portions of the lungs would be close to normal. If a volume-oriented approach is used, a lower tidal volume has to be used in order to prevent alveolar distention in the aerated lungs. There is evidence of acute lung parenchymal and microvascular injury histologically similar to ARDS in animals that are ventilated with high inflation pressures and tidal volumes.[19,20] To avoid ventilator induced lung injury, small tidal volumes with sufficient PEEP should be used to prevent end-expiratory collapse and tidal recruitment of alveolar units. The dependent lung regions in ARDS are subject to a superimposed pressure from the weight of the overlying lung resulting in atelectasis during expiration.[21] Providing sufficient PEEP to avoid end-expiratory atelectasis in these dependent regions probably results in higher than normal end-expiratory volume in alveoli in non-dependent regions. Thus the non-dependent ventral portions of the lungs run the risk of becoming overdistended especially if normal tidal volumes (10–12 mLs/kg) are used. Furthermore, the aerated portions of the lungs are small in ARDS even with optimum PEEP.[18] For these reasons, it is necessary to use low tidal volumes with optimum PEEP to avoid end-expiratory overdistention. Gattinioni has demonstrated that most of the recruitment of atelectatic lung units occurs with PEEP levels of 15 to 20 cm H_2O.[22] Current recommendations would include the use of PEEP levels of usually 15 to 20 cm H_2O and to set tidal volumes of between 5 to 8 mLs/kg. In addition to these measures, the adoption of permissive hypercapnia and pressure-limited rather than

volume-cycled ventilation have been advocated to prevent lung damage resulting from regional overdistention of the lung units related to high ventilatory volumes and reopening of atelectatic lung units with resultant shear injury.[23] Amato *et al.* have shown in a prospective randomized controlled trial that these measures, so-called "open-lung approach", improved oxygenation significantly there was but there was no impact on the survival rate to hospital discharge.[24]

Permissive Hypercapnia

Permissive hypercapnia is a ventilation strategy in which hypoventilation and hypercapnia are allowed so that detrimental rise in alveolar pressure is avoided. The $PaCO_2$ is allowed to rise as tidal volume and respiratory rate are adjusted to prevent increase in plateau pressure. In general, the resulting respiratory acidosis is treated with intravenous bicarbonate if arterial pH <7.2.[25] This was introduced initially to reduce the barotraunma in mechanically ventilated patients with obstructive airway disease.[26] Permissive hypercapnia may be an inevitable consequence of the "lung protective strategy" that limits airway pressure and volume adopted in ventilating patients with ARDS. Several uncontrolled studies initially showed that survival might be improved by adopting permissive hypercapnia.[26–28] However, randomized studies have shown that there was no survival advantage to permissive hypercapnia in patients with ARDS.[29,30] In adopting permissive hypercapnia, one has to be aware of the consequences of carbon dioxide (CO_2) retention. Acute elevations in CO_2 result in increase in sympathetic activity and cardiac output, impairment of musculoskeletal and central nervous system function.[31,32] Therefore, CO_2 retention may be detrimental in patients with co-existing raised intra-cranial pressure, significant cardiovascular dysfunction, beta-blockade and autonomic dysfunction. Based on the available evidence, permissive hypercania with pressure targeted ventilation may be considered in patients with ARDS developing plateau pressures in excess of 35–40 cm H_2O.[33]

Inverse Ratio Ventilation (IRV)

Inverse ratio ventilation (IRV) is another ventilation strategy employed by clinicians to improve oxygenation while attempting to maintain acceptable peak airway pressures in patients with ARDS.[34] Commonly this mode of ventilation is combined with pressure controlled ventilation,

though not always. In this mode the inspiratory time is prolonged thus reversing the Inspiratory: Expiratory (I:E) ratio. In doing so the mean airway pressure is increased without increasing the peak airway pressure.[35] Several hours may be required to achieve the maximal benefits of IRV on gas exchange as recruitment of atelectatic lung units occur with sustained inspiratory pressures.[36] IRV has been shown to improve oxygenation in patients with severe ARDS when conventional modes of ventilation have failed.[37] However in a recent randomized controlled trial using IRV, no significant improvement in oxygenation was seen.[38] Furthermore, IRV is not well tolerated by spontaneously breathing patients. As such patients are usually heavily sedated and paralyzed with neuromuscular blocking agents to allow patient-ventilator synchronization. This method of ventilation is used when arterial oxygenation cannot be maintained with conventional modes of ventilation when the use of PEEP results in excessive plateau pressure (> 35 cm H_2O).

Posture

Positional changes in ARDS patients can bring about improvement in the ventilation, perfusion mismatching that is present.[39,40] Many patients in the early phase of ARDS improve pulmonary gas exchange remarkably when turned prone.[41] In the prone position the weight of the heart rests on the ventral rib cage and sternum. This reduces the compressive atelectasis seen in the dorsal regions of the lungs. Furthermore, modifications of the chest wall compliance upon proning the patient alters the transalveolar pressure gradient.[42] These changes result in improvement in the distribution of ventilation to the dorsal atelectatic lung units. In essence the use of the prone position pursues three therapeutic goals: 1) Reduction in FiO_2 to prevent oxygen toxicity; 2) Recruitment of atelectatic dorsal lung units; and 3) Improve postural drainage of bronchial secretions.[43] Most patients show an improvement in gas exchange within the first hour of adopting the prone position while in others it may take several hours.[42] Gas exchange improves in more than 50% of patients turned prone in the early phase of ARDS allowing significant reductions in both FiO_2 and PEEP.[41] When adopting this strategy, continuous monitoring of intra-arterial blood pressure, cardiac rhythm and pulse oximetry is essential. Strict attention should be given to proper positioning of the endotracheal tube and patency of peripheral and central venous catheters during the turning

process. Hypotension, cardiac arrhythmias and hypoxemia are usually transient events. These can be minimized by providing FiO_2 of 1.0 during the manoeuvre and also by ensuring that the airway is suctioned prior to proning the patient. The optimal frequency of switching from supine to prone is yet to be determined. In general, switching from positions can be carried out twice a day. The relative duration assigned to each position is determined by the gas exchange response.[42]

Liquid Ventilation

This technique of ventilation involves the use of perfluorochemicals, which are biologically inert and non-toxic. Amongst the perfluorochemicals, perfluoro-octyl bromide (perflubron) is currently being used in the clinical and experimental settings.[44] Perflubron can dissolve 15 times the amount of oxygen per given volume as per the same volume of plasma.[45] Carbon dioxide and other gases are also highly soluble in it. There are essentially 2 methods of ventilation using perflubron: 1) Total Liquid Ventilation (TLV); and 2) Partial Liquid Ventilation (PLV).

In TLV, the entire lung (residual volume and tidal volume) is filled with oxygenated perflubron. With the aid of specialized equipment, tidal volume amounts of perflubron is removed from the lungs, pumped through an extracorporeal circuit in which gas exchange is facilitated and subsequently returned to the patient.[46] The perflubron facilitates gaseous exchange at the alveolar unit level because of the Extraordinary solubility of oxygen and carbon dioxide in it.[45]

In PLV, lungs are filled with perflubron up to a volume (20–30 mLs/kg) equivalent to the Functional Residual Capacity (FRC). Conventional mechanical ventilation is continued enabling gas exchange.[45] As the fluid evaporates from the lungs, it is intermittently replaced with small amounts (2–8 mLs/kg/hr) to maintain liquid volume equivalent to FRC. Clinical experience with this method of ventilation has been encouraging.[47] The perflubron recruits alveoli in the dependent regions by providing a PEEP-like effect.[48] Pulmonary blood flow is also preferentially redistributed to the non-dependent regions of the lungs thus bringing about an improvement in ventilation-perfusion matching.[47] The other potential benefits of liquid ventilation include facilitation of removal of cellular debris, inflammatory mediators and exudative material from the distal airways.[49,50] PLV appears to be promising in the

ventilatory management of patients with ARDS. It is currently being evaluated in a large ARDS trial in the USA.[44]

Extracorporeal Membrane Oxygenation (ECMO) and Extracorporeal Carbon Dioxide Removal ($ECCO_2R$) are two methods of supporting the respiratory system by partially substituting the gas exchange function of the lungs. Both these techniques have been evaluated in ARDS patients. A fraction of the systemic circulation in both ECMO and $ECCO_2R$ is exposed to oxygen across a semi-permeable membrane, which assists in the gas exchange function of the compromised lungs. If the oxygenation is the main reason for extracorporeal support then veno-arterial circuit is used. In this circuit, blood exiting through a large bore cannula located in central vein is allowed to traverse the extracorporeal circuit and subsequently returned to the aorta.[51] If carbon dioxide (CO_2) extraction is the main concern, then a veno-venous circuit is employed with an extracorporeal device extracting CO_2.[52] Despite its clear success in neonates with severe respiratory failure, a large trial comparing ECMO with mechanical ventilation showed no benefit.[53,54] A randomized trial comparing $ECCO_2R$ with pressure controlled inverse ratio showed no difference in outcome between the two modes of support.[55] $ECCO_2R$ and ECMO are not recommended for the routine management of ARDS.

The judicious use of fluids, diuretics and vasoactive agents are important in supporting patients with ARDS. Pulmonary edema arises in ARDS as a result of increased vascular permeability.[56] Several clinical studies have shown improvement in survival in ARDS patients in whom the pulmonary capillary occlusion pressure is kept low by the judicious use of diuretics and intravenous fluids.[57–59] Nevertheless, decreasing the intravascular pressure excessively can compromise the hemodynamic state and lead to renal impairment/failure especially in mechanically ventilated patients with PEEP. Therefore it is recommended that ARDS patients be fluid restricted and, if need be, diuresed with careful monitoring of end organ function. The use of vasoactive agents, like dopamine, should be considered in attempt to maintain adequate blood pressure especially in the setting of hypotension with euvolemia.

OPTIMIZATION OF OXYGEN TRANSPORT

It has been reported that survivors of ARDS have a significantly higher oxygen delivery (DO2) and oxygen consumption (VO2) than non-survivors.[60]

These patients with ARDS manifest a higher critical oxygen transport concentration below which they may be vulnerable to tissue hypoxia. This observation has brought forth the notion that by achieving supranormal patterns in oxygen transport in the critically ill mortality may be reduced. Such an approach was thought to reverse or prevent tissue hypoxia, which arises as a result of increased oxygen demand and maldistribution of blood flow.[61] Recent studies have shown that if such treatment measures were instituted in high-risk surgical patients during the perioperative period, a significant reduction in mortality can be achieved.[62,63] These studies support the notion that VO2 is dependent on DO2. However, this assumption has been challenged and 2 recent randomized prospective trials have failed to show any benefit in terms of morbidity and mortality in patients achieving supranormal values of oxygen transport.[64,65] At present attempts at achieving supranormal levels of oxygen transport indices is not recommended for the management of ARDS patients.

PHARMACOLOGICAL THERAPY

Steroids

Several anti-inflammatory agents have been used in the treatment of the underlying inflammation in ARDS. Corticosteroids, prostaglandin E1 and inhibitors of arachidonic acid metabolism have all been investigated for this purpose. Preliminary trials with high-dose corticosteroids in septic patients with ARDS showed possible benefit.[66,67] Multicenter placebo-controlled trials in which steroids were given to patients who were at risk of ARDS or in the early phase of ARDS did not show any reduction in the frequency of occurrence of ARDS in the at risk group or improvement in mortality in the ARDS group.[68–70]

Recent studies have suggested that corticosteroids may have a beneficial role in the fibroproliferative phase of ARDS.[71,72] The corticosteroid-treated patients had a significantly lower mortality rate than the placebo-treated group in these trials. Interpretation of this result is complicated by the small number of patients enrolled, the fact that half the placebo treated group crossed over to the corticosteroid group and that the incidence of infection in the steroid group was twice as high as the placebo group. The use of corticosteroids in late phase ARDS is currently being evaluated in the USA in an ongoing multicenter trial.

Trials have been conducted with other anti-inflammatory agents such as prostaglandin E1,[73] ketaconazole,[17] ibuprofen,[74] N-acetylcysteine[75] in patients with ARDS, but none have proven successful in reducing mortality.

Inhaled Vasodilator

Inhaled vasodilators, particularly prostacyclin and nitric oxide (NO) have been administered in patients with ARDS resulting in improvement of the ventilation-perfusion mismatching and ultimately amelioration of hypoxemia.[76,77] However, neither of these agents has been shown to improve survival in ARDS.

Surfactant

Abnormalities of surfactant function have been described in patients with ARDS.[78,79] Abnormal surfactant function, thought to result from the action of inflammatory mediators on surfactant itself, will lead to collapse of the alveolar units resulting in the creation of shunts. Administration of exogenous surfactant has been shown to improve oxygenation in ARDS patients.[80] However, in a large multicenter prospective placebo-controlled trial, there was no benefit in 30-day mortality, length of stay in the intensive care unit or the duration of mechanical ventilation.[81] Currently, there is no evidence to support the use of exogenous surfactant in patients with ARDS.

OUTCOME

Most patients who die with ARDS do so within the first 2 weeks of the illness. The cause of the death in most cases would be sepsis syndrome or multi organ dysfunction syndrome rather than respiratory failure.[82] Mortality rates of 60–70% were quoted in several studies in the first two decades.[13,14] Recent reports have indicated that with current supportive strategies the mortality rate has fallen to approximately 40%.[83,84] Predictors of mortality have been identified in a recent trial and these include[7]:

1) non-pulmonary organ dysfunction occurring between hospital admission and ICU admission;
2) sepsis syndrome; and
3) chronic liver disease.

Many survivors of ARDS can lead a normal life despite a lower health-related quality of life than the general population.[85] Lung function studies in survivors have shown that lung mechanics return to normal within the first year after hospital discharge or extubation.[86,87] However, reduction in diffusion capacity and exercise induced increment in dead space ventilation (Vd/Vt) persist.[86] The severity of the residual pulmonary defect appears to correlate with the severity of the ARDS.[88]

REFERENCES

1. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE, Acute respiratory distress in adults, *Lancet* **2**:319–323, 1967.
2. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke C, Hudson L, Lamy M, LeGall JR, Morris A, Spragg R, The American — European Consensus Conference on ARDS, *Am J Respir Crit Care Med* **149**:818–824, 1994.
3. Villar J, Slutsky AS, The incidence of adult respiratory distress syndrome, *Am Rev Respir Dis* **140**:814–816, 1989.
4. Baumann WR, Jung RC, Koss M, Boylen CT, Navarro L, Sharma OP, Incidence and mortality of adult respiratory distress syndrome: a prospective analysis from a large metropolitan hospital, *Crit Care Med* **14**(1):1–4, 1986.
5. Zaccardelli DS, Pattishall EN, Clinical diagnostic criteria of the adult respiratory distress syndrome in the intensive care unit, *Crit Care Med* **24**:247–251, 1996.
6. Moss M, Goodman PL, Heinig M, Barkin S, Ackerson L, Parsons PE, Establishing the relative accuracy of three new definitions of adult respiratory distress syndrome, *Crit Care Med* **23**(10):1629–1637, 1995.
7. Doyle RL, Szarflarski N, Modin GW, Weiner-Kronish JP, Matthay MA, Identifications of patients with acute lung injury. Predictors of mortality, *Am J Respir Crit Care Med* **152**(Pt. 1):1818–1824, 1995.
8. Lamy M, Fallat Rj, Koeniger E, Dietrich HP, Ratliff JL, Eberhart RC, Tucker HJ, Hill JD, Pathologic features and mechanisms of hypoxaemia in adult respiratory distress syndrome, *Am Rev Respir Dis* **114**(2):267–284, 1976.
9. Snow RL, Davies P, Pontoppidan H, Zapol WM, Reid L, Pulmonary vascular remodeling in adult respiratory distress syndrome, *Am Rev Respir Dis* **126**(5):887–892, 1982.
10. Crouch E, Pathobiology of pulmonary fibrosis, *Am J Physiol* **259**(4 Pt. 1):L159–184, 1990.
11. Bevilacqua MP, Endothelial-leukocyte adhesion molecules, *Ann Rev Immunol* **11**:767–804, 1993.
12. Rinaldo JE, Christman JW, Mechanisms and mediators of adult respiratory distress syndrome, *Clin Chest Med* **11**:621–626, 1990.

13. Pepe PE, Potkin R, Holtman-Reus D, Hudson LD, Carrica CJ, Clinical predictors of adult respiratory distress syndrome, *Am J Surg* **144**:124–130, 1982.

14. Fowler AA, Hamman RF, Good JT, Benson KN, Baird M, Eberle DJ, Petty TL, Hyers TM, Adult respiratory distress syndrome: risk with common predispositions, *Ann Intern Med* **98**(5 Pt. 1):593–597, 1983.

15. Hudson LD, Milberg JA, Anandi D, Maunder RJ, Clinical risks for development of the acute respiratory distress syndrome, *Am J Respir Crit Care Med* **151**:293–301, 1995.

16. Marc L, Vincent J, Pharmacologic strategies in acute respiratory distress syndrome, *Curr Opin Anaesth* **11**(2):123–128, 1998.

17. Luce J, Acute lung injury and acute respiratory distress syndrome, *Crit Care Med* **26**:369–376, 1998.

18. Gattinioni L, Pesenti A, Avalli L, Rossi F, Bombino M, Pressure-volume curve of total respiratory system in acute respiratory failure. Computed tomographic study, *Am Rev Respir Dis* **136**:730–736, 1987.

19. Hickling KG, Ventilatory management of ARDS: can it affect the outcome? *Intensive Care Med* **16**:219–226, 1990.

20. Parker JC, Hernandez LA, Peevy KJ, Mechanisms of ventilator-induced lung injury, *Crit Care Med* **21**:131–143, 1993.

21. Gattinioni L, D'Andrea L, Pelosi P, Vitale G, Pesenti A, Fumagalli R, Regional effects and mechanisms of positive end expiratory pressure in early adult respiratory distress syndrome, *JAMA* **269**(16):2122–2127, 1993.

22. Gattinioni L, Effects of positive end-expiratory pressure on regional distribution of tidal volume and recruitment in adult respiratory distress syndrome, *JAMA* **151**:1807–1814, 1995.

23. Amato MB, *et al*., Beneficial effects of the "open lung approach" with low distending pressures in acute respiratory distress syndrome: a prospective randomized study on mechanical ventilation, *Am J Respir Crit Care Med* **152**(6 Pt. 1):1835–1846, 1995.

24. Amato MB, *et al*., Effect of protective-ventilation strategy on mortality on the acute respiratory distress syndrome, *NEJM* **338**(6):347–354, 1998.

25. Menitove SM, Goldring RM, Combined ventilator and bicarbonate strategy in the management of status asthmaticus, *Am J Med* **74**(5):898–901, 1983.

26. Darioli R, Perret C, Mechanical controlled hypoventilation in status astmaticus, *Am Rev Respir Dis* **129**:385–387, 1984.

27. Hickling KG, Henderson SJ, Jackson R, Low mortality associated with low volume, pressure limited ventilation with permissive hypercapnia in severe adult respiratory distress syndrome, *Intensive Care Med* **16**:372–377, 1990.

28. Hickling KG, Walsh J, Henderson S, Jackson R, Low mortality rate in adult respiratory distress syndrome using low volume, pressure limited ventilation with permissive hypercapnia: a prospective study, *Crit Care Med* **22**(10):1568–1578, 1994.

29. Brower R, Shanholtz C, Shade D, Fessler H, White P, Weiner C, Teeter J, Almog Y, Dodd-O J, Piatadosi S, Randomized controlled trial of small tidal volume in ARDS, *Am J Respir Crit Care Med* **155**(4 Pt. 2):A93, 1997.

30. Brochard L, Roudof Thoroval F, and the Collaborative Group on VT Reduction: Tidal Volume (V/T) reduction in acute respiratory distress syndrome (ARDS): a multicenter randomized study, *Am J Respir Crit Care Med* **155**:A505, 1997.

31. Carvalho CR, *et al*., Temporal hemodynamic effects of permissive hypercapnia associated with ideal PEEP in ARDS, *Am J Respir Crit Care Med* **156**(5):1458–1466, 1997.

32. Kacmarek RM, Hickling KG, Permissive Hypercapnia, *Respir Care* **38**(4):373–387, 1993.

33. Artigas A, *et al*., The American-European Consensus Conference on ARDS Part 2, *Am J Respir Crit Care Med* **157**:1332–1347, 1998.

34. Gurevitch MJ, Van Dyke J, Young ES, Jackson K, Improved oxygenation and lower peak airway pressure in severe adult respiratory distress syndrome. Treatment with inverse ratio ventilation, *Chest* **96**:1356–1359, 1989.

35. Marcy TW, Marini JJ, Inverse ratio ventilation in ARDS: rationale and implementation, *Chest* **100**:494–504, 1991.

36. Tharratt RS, Allen RP, Albertson JE, Pressure controlled inverse ratio ventilation in severe adult respiratory failure, *Chest* **94**(4):755–762, 1988.

37. Abraham E, Yoshihara G, Cardiorespiratory effects of pressure controlled inverse ratio ventilation in severe respiratory failure, *Chest* **96**:1356–1359, 1989.

38. Mercat A, Titriga M, Anguel N, Richard C, Teboul JL, Inverse ratio ventilation (I/E = 2/1) in acute respiratory distress syndrome. A six-hour controlled study, *Am J Respir Crit Care Med* **155**:1637–1642, 1997.

39. Piehl M, Brown R, Use of extreme position changes in acute respiratory failure, *Crit Care Med* **4**(1):13–14, 1976.

40. Langer M, Mascheroni D, Macolin R, Gattinoni L, The prone position in ARDS patients. A clinical study, *Chest* **94**:103–107, 1988.

41. Pappert D, Rossaint R, Slama K, Grunning T, Falke K, Influence of positioning on ventilation-perfusion relationships in severe adult respiratory distress syndrome, *Chest* **106**:1511–1516, 1994.

42. Broccard A, Marini JJ, Position and posture changes in acute illness, *Semin Resp Crit Care Med* **18**(1):19–32, 1997.

43. Pappert D, Falke KJ, When is a patient prone for prone? *Anesth Analg* **83**(6):1139–1140, 1996.

44. Kacmarek RM, Newer ventilator strategies, *Curr Opin Anaes* **12**(2):133–141, 1999.

45. Shaffer TH, Wolfson MR, Clark LC, State of the art review. Liquid ventilation, *Ped Pulmonol* **14**:102–109, 1992.

46. Dirkes S, Liquid ventilation: new frontiers in treatment of ARDS, *Crit Care Nurse* **16**:53–58, 1996.

47. Hirschl RB, Pranikoff T, Wise C, Overbeck MC, Gauger P, Schreiner RJ, Dechert R, Bartlett RH, Initial experience with partial liquid ventilation in adult patients with acute respiratory distress syndrome, *JAMA* **275**:383–389, 1996.

48. Hirschl RB, Parent A, Tooley R, McCraken M, Johnson K, Shaffer TH, Wolfson MR, Bartlett RH, Liquid ventilation improves pulmonary function, gas exchange and lung injury in a model of respiratory failure, *Ann Surg* **221**(1):79–88, 1995.

49. Richman PS, Wolfson MR, Schaffer TH, Lung lavage with oxygenated perfluorochemical liquid in acute lung injury, *Crit Care Med* **21**:768–774, 1993.

50. Kelly KP, Partial liquid ventilation — turning back a PAGE on evolution, *Br J Anaesth* **78**(1):1–2, 1997.

51. Clowes GHAJ, Hopkins AL, Neville WE, An artificial lung dependent upon diffusion of oxygen and carbon dioxide through plastic membranes, *J Thorac Surg* **32**:630–637, 1956.

52. Gattinoni L, *et al*., Treatment of acute respiratory failure with low frequency positive pressure ventilation and extracorporeal removal of carbon dioxide, *Lancet* **2**:292–295, 1980.

53. Zapol WM, *et al*., Extracorporeal membrane oxygenation in severe acute respiratory failure. A randomized prospective study, *JAMA* **242**(20):2193–2196, 1979.

54. Bartlett RH, Andrews AF, Toomasian JM, Haiduc NJ, Gazzaniga AB, Extracorporeal membrane oxygenation for newborn respiratory failure. Forty-five cases, *Surgery* **92**(2):425–433, 1982.

55. Morris AH, *et al*., Randomized clinical trial of pressure controlled inverse ratio ventilation and extracorporeal CO2 removal for adult respiratory distress syndrome, *Am J Respir Crit Care Med* **149**(2 Pt. 1):295–305, 1994.

56. Kaplan JD, Calandrino FS, Schuster DP, A positron emission tomographic comparison of pulmonary vascular permeability during the adult respiratory distress syndrome and pneumonia, *Am J Respir Crit Care Med* **143**(1):150–154, 1991.

57. Bone RC, Treatment of adult respiratory distress syndrome with diuretics, dialysis and positive end expiratory pressure, *Crit Care Med* **6**(3):136–139, 1978.

58. Simmons RS, Berdine GG, Seidenfeld JJ, Prihoda TJ, Harris GD, Smith JD, Gilbert TJ, Mota E, Johanson WG Jr, Fluid balance and the adult respiratory distress syndrome, *Am Rev Respir Dis* **135**(4):924–929, 1987.

59. Humphrey H, Hall J, Sznajder I, Silverstein M, Wood L, Improved survival in ARDS patients associated with reduction in pulmonary capillary wedge pressure, *Chest* **97**(5):1176–1180, 1990.

60. Russell JA, Ronco JJ, Lockhart D, Belzberg A, Kiess M, Dodek PM, Oxygen delivery and consumption and ventricular preload are greater in survivors than in nonsurvivors of the adult respiratory distress syndrome, *Am Rev Respir Dis* **141**(3): 659–665, 1990.

61. Hinds C, Watson D, Manipulating haemodynamics and oxygen transport in critically ill patients, *N Engl J Med* **333**:1074–1075, 1995.

62. Shoemaker WC, Appel PL, Kram HB, Waxman K, Lee TS, Prospective trial of supranormal valuse of survivors as therapeutic goals in high-risk surgical patients, *Chest* **94**:1176–1184, 1988.

63. Boyd O, Grounds RM, Bennett ED, A randomized clinical trial of the effect of deliberate perioperative increase of oxygen delivery on mortality in high risk surgical patients, *JAMA* **270**:2699–2707, 1993.

64. Hayes MA, Timmins AC, Yau EHS, Palazzo M, Hinds CJ, Watson D, Elevation of systemic oxygen delivery in the treatment of critically ill patients, *N Engl J Med* **330**:1717–1722, 1994.

65. Gattinoni L, Brazzi L, Pelosi P, Latini R, Tognoni G, Pesenti A, Fumagalli R, A trial of goal-oriented haemodynamic therapy in critically ill patients. SvO2 Collaborative Group, *N Engl J Med* **333**(16):1025–1032, 1995.

66. Sprung C, Caralis P, Marcial E, Pierce M, Gelbard MA, Long WM, Duncan RC, Tendler MD, Karpf M, The effects of high dose corticosteroids in patients with septic shock: a prospective controlled study, *N Engl J Med* **311**(18):1137–1143, 1984.

67. Sibbald WJ, Anderson R, Reid B, Holliday RL, Driedger AA, Alveolocapillary permeability in human septic ARDS, *Chest* **79**:133–142, 1981.

68. Bernard GR, *et al*., High dose corticosteroids in patients with adult respiratory distress syndrome, *N Engl J Med* **317**:1565–1570, 1987.

69. Luce JM, Montgomery AB, Marks JD, Turner J, Metz CA, Murray JF, Ineffectiveness of high-dose methylprednisolone in preventing parenchymal lung injury and improving mortality in patients with septic shock, *Am Rev Respir Dis* **138**:62–68, 1988.

70. Bone RC, Fisher CJ, Clemmer TP, Slotman GJ, Metz CA, Early methylprednisolone treatment for septic syndrome and adult respiratory distress syndrome, *Chest* **92**(6):1032–1036, 1987.

71. Meduri GU, Chinn AJ, Leeper KV, Wunderink RG, Tolley E, Winer-Muram HT, Khare V, Eltorky M, Corticosteroid rescue treatment of progressive fibroproliferation in late ARDS: patterns and predictors of outcome, *Chest* **105**(5):1516–1527, 1994.

72. Meduri GU, Headley S, Golden E, Carson SJ, Umberger RA, Kelso T, Tolley EA, Effect of prolonged methylprednisolone therapy in unresolving acute respiratory distress syndrome, *JAMA* **280**(2):159–165, 1998.

73. Bone RC, Slotman G, Maunder R, Silverman H, Hyers TM, Kerstein MD, Ursprung JJ, Randomized double-blind multicenter study of Prostaglandin E1 in patients with adult respiratory distress syndrome, *Chest* **96**:114–119, 1989.

74. Bernard GR, *et al*., The effects of Ibuprofen on the physiology and survival of patients with sepsis, The Ibuprofen in Sepsis Study Group, *N Engl J Med* **336**(13):912–918, 1997.

75. Jepsen S, Herlevsen P, Knudsen P, Bud MI, Klausen NO, Antioxidant treatment with N-acetylcysteine during adult respiratory distress syndrome: a prospective randomized placebo controlled study, *Crit Care Med* **20**(7):918–923, 1992.

76. Rossaint R, Gerlach H, Schmidt-Ruhnke H, Pappert D, Lewandowski K, Steudel W, Falke K, Efficacy of inhaled nitric oxide in patients with severe ARDS, *Chest* **108**(4): 1107–1115, 1995.

77. Zwissler B, Kemming G, Habler O, Kleen M, Merkel M, Haller M, Briegel J, Welte M, Peter K, Inhaled Prostacyclin (PGI2) versus inhaled nitric oxide in adult respiratory distress syndrome, *Am J Respir Crit Care Med* **154**(6 Pt. 1):1671–1677, 1996.

78. Hallman M, Spragg R, Harrell JH, Moser KM, Gluck L, Evidence of lung surfactant abnormality in respiratory failure. Study of bronchoalveolar lavage phospholipids, surface activity, phospholipase activity and plasma myoinositol, *J Clin Invest* **70**(3):673–678, 1982.

79. Gregory TJ, Longmore WJ, Moxley MA, Whisett JA, Reed CR, Fowler AA 3d, Hudson LD, Maunder RJ, Crim C, Hyers TM, Surfactant chemical composition and biophysical activity in acute respiratory distress syndrome, *J Clin Invest* **88**(6):1976–1981, 1991.

80. Lewis JF, Jobe AH, Surfactant and adult respiratory distress syndrome, *Am Rev Respir Dis* **147**(1):218–233, 1993.

81. Anzueto A, Baughman RP, Guntupalli KK, Weg JG, Wedemann HP, Raventos AA, Lemaine F, Long W, Zaccardelli DS, Pattishall EN, Aerosolised surfactant in adults with sepsis-induced adult respiratory distress syndrome. Exosurf Acute Respiratory Distress Syndrome Sepsis study Group, *N Engl J Med* **334**:1417–1421, 1996.

82. Montgomery AB, Stager MA, Carricao CJ, Hudson LD, Cause of mortality in patients with adult respiratory distress syndrome, *Am Rev Respir Dis* **132**:485–489, 1985.

83. Suchyta MR, Clemmer TP, Elliot CG, Orme JF Jr, Weaver LK, The adult respiratory distress syndrome: a report of survival and modifying factors, *Chest* **101**:1074–1079, 1992.

84. Suchyta MR, Clemmer TP, Orme JF, Morris AH, Elliott CG, Increased survival of ARDS patients with severe hypoxemia (ECMO criteria), *Chest* **99**(4):951–955, 1991.

85. Weinert CR, Gross CR, Kangas JR, Bury CL, Marinelli WA, Health-related quality of life after acute lung injury, *Am J Respir Crit Care Med* **156**:1120–1128, 1997.

86. Elliot CG, Morris AH, Cengiz M, Pulmonary function and exercise gas exchange in survivors of adult respiratory distress syndrome, *Am Rev Respir Dis* **123**:492–495, 1981.

87. McHugh LG, Milberg JA, Whitcomb ME, Schoene RB, Maunder RJ, Hudson LD, Recovery of function in survivors of adult respiratory distress syndrome, *Am J Respir Crit Care Med* **150**:90–94, 1994.

88. Elliot CG, Rasmusson BY, Crapo RO, Morris AH, Jensen RL, Prediction of pulmonary function abnormalities after adult respiratory distress syndrome, *Am Rev Respir Dis* **135**:634–638, 1987.

Rehabilitative Medicine

This page intentionally left blank

81

A Clinical Approach to Rehabilitation Medicine

Peter A. C. Lim

Rehabilitation medicine is a relatively young specialty in the history of medicine and surgery that nevertheless has become increasingly important as societies everywhere develop. This is primarily related to a greater awareness of the benefits of rehabilitation, the driving factors of aging populations with extension of life expectancies, and an increase of traumatic events in a fast-moving modern world.

The US is generally accepted to be where the most organized and developed rehabilitation or rehabilitative services are available. Rehabilitation medicine there had its beginnings in wartime. But prior to the Second World War in the 1940s, the high mortality rates from infection and limitations of medical care then often meant little could be done for the victims of serious slash and stab wounds, gunshot injuries and explosions. With the availability of antiseptics and antibiotics however, there developed a large patient-base of military survivors with various impairments who needed help returning to optimal functioning and back into the society. Of great importance was public consciousness that the injuries were suffered while defending freedom and the homeland, leading to a push for development of services for these worthy veterans.

Many of the principles and techniques of modern rehabilitation medicine came from this period. Principles pioneered then such as the desirability of getting patients out of bed quickly to avoid the deleterious effects and hazards of immobility, as well as early ambulation, intensive therapy, and the therapeutic effects of activities like handicrafts are taken for granted nowadays. The comprehensive approach to rehabilitation in order to deal with wide-ranging issues from self-care ability to mobility, self-image to vocation, was also established in this era.

Another tragedy that gave impetus to the development of rehabilitation was the poliomyelitis epidemic of the 1940s and 50s. In a similar manner, the demand for something to be done to help the innocent children stricken may have rivaled that to help the heroes of war. Techniques in bracing, the use of thermal modalities and therapeutic exercise received much attention, research, experimentation and widespread usage. It led to the demand for physicians trained in comprehensive physical, psychosocial, and vocational rehabilitation to lead and direct the restoration efforts. The fields of physical therapy, occupational therapy, and orthotics also became areas of interest for many in healthcare.

The physicians who used physical agents and therapeutics for diagnosis and treatment were initially known as physical therapy physicians, but the term commonly used for them became physical medicine physician in 1944. They worked mainly with occupational or musculoskeletal diseases and injury, but then united with the more inpatient-centered rehabilitation medicine physicians to become one specialty in 1949, recognizing their common approaches and philosophies of care. In the US the term for the joint specialty is physical medicine and rehabilitation, and the physician in this field is referred to as the physiatrist.

In more recent times, rehabilitation medicine has played a leading role in the development of rehabilitation healthcare systems for the disabled and introducing concepts of outcome-oriented research and cost-effectiveness. It has been involved in the influencing of public policy and awareness towards the use of seatbelts, and safety helmets for motorcyclists. Accessibility for the disabled has been championed such as cuts in street curbs, installation of ramps and lifts in public buildings, as have the philosophy of independent-living, civil rights for the disabled, and the value of the individual with a disability or handicap to the society.

Singapore has acquired developed-country status in the very short period of about 30 years. It has one of the highest per-capita gross national

Table 1 Increasing Demands for Rehabilitation Services in Singapore

Singapore: Then and Now*	**1970**	**2000**		
Population (mil)	2.1	4		
Population Growth (%)	2.8	1.7		
Life Expectancy (Yrs)	65.8	78		
Infant Mortality Rate (per 1000 live births)	20.5	2.5		
Total Fertility Rate (per 1000 females)	3065	1586		
Singapore: A Comparison†	**Singapore**	**Japan**	**UK**	**US**
Population (mil)	4.1	127.1	60.0	285.7
Per-Capita GNP (US$)	24 664	37 950	23 793	35 277
Life Expectancy (Yrs)	78	80	77	77
Aged ⩾ 65 (%)	7.3	17.2		
In the year 2030 those aged 65 years and older will comprise 19% of the Singapore population.**				

*Department of Statistics, Singapore

†Asiaweek, September 2001

**Report of the Interministerial Committee on the Ageing of the Population, November 1999

product, and the infant mortality rate, a good indicator of health in general, is one of the lowest in the world. However, it has also become one of the fastest aging societies in Asia. Along with a low birth rate inadequate to even replace the population, the specter of having to care for increasing numbers of the disabled elderly looms. In conjunction with changes in cultural norms such as both the husband and wife working outside the home, smaller families to share the burden, philosophies on expectations and sacrifice, taking care of the disabled parent is often a difficult task.

Rehabilitation, which focuses on minimizing disabilities and optimizing independence, is one of the solutions. The benefits of rehabilitation includes fewer medical complications, a better functional outcome and quality of life, with corresponding lower medical costs.

IMPAIRMENT, DISABILITY, AND HANDICAP

Although these terms are often used interchangeably, awareness of the differences between them is helpful for understanding the

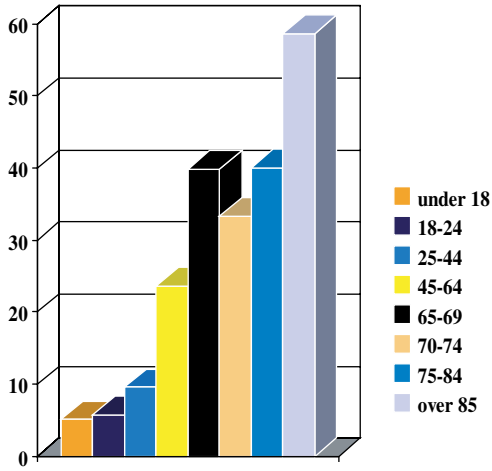


Fig. 1 Relationship between age and activity limitations of different extents. There is a strong correlation between age and limitation in functional activity or disability, and this increases significantly past the age of 65 years.
(Craus LE, Stoddard S. Chartbook on disability in the United States: an InfoUse report. U.S. National Institute on Disability and Rehabilitation Research, 1989)

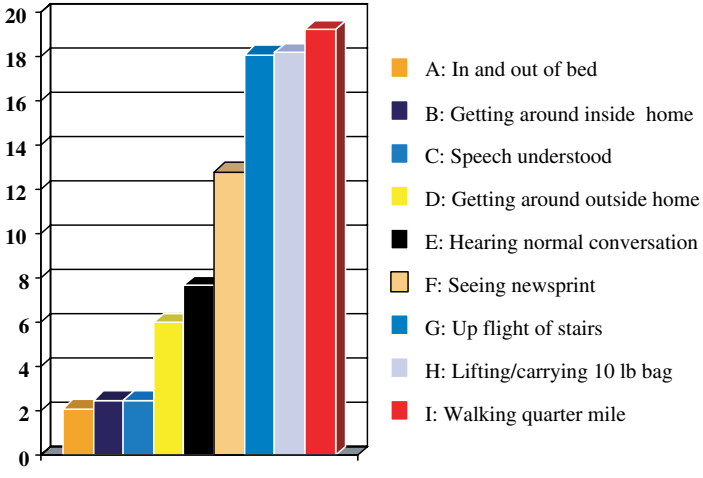


Fig. 2 Nature of functional limitations. Functional limitation can vary from the relatively mild e.g. unable to walk a quarter mile, to as severe as being unable to get in and out of bed.
(Craus LE, Stoddard S. Chartbook on disability in the United States: an InfoUse report. U.S. National Institute on Disability and Rehabilitation Research, 1989)

implications and approaches to management. As per the World Health Organization, 1980:

Impairment is a loss or abnormality of psychological, physiological, or anatomical structure or function. *Disability* is a restriction or inability due to impairment, to perform a normal activity. *Handicap* is a disadvantage resulting from impairment or disability, limiting or preventing a normal role.

Where most other medical or surgical specialties focus on the pathology or disease that leads to the impairment, Rehabilitation Medicine's job extends across the spectrum. In a stroke for instance, the rehabilitation physician's involvement begins with a referral of the patient by the neurologist or neurosurgeon, following which the patient may be taken over for rehabilitation. This then includes management of associated risk factors such as hypertension or diabetes (Pathology), the hemiplegia (Impairment) that results from the stroke, inability to walk due to the paralysis (Disability), and inability to get around in a wheelchair in public due to steps or curbs (Handicap).

Using the above definitions, it is possible to see that this same stroke survivor may have much residual impairment, yet after rehabilitation not be very disabled. This person may have significant residual hemiparesis, but be able to perform all activities of daily living (ADL) independently, using modified techniques or by compensation with equipment such as a wheelchair. He or she might still be handicapped though, if there are only stairs to get into a building, a handicap that can be removed by installing a lift. This understanding is important from the perspective that the disabled person should be very much a normal part of our society. Such a person is not necessarily a burden but is instead capable of valuable contributions to the community, yet not able to do so because of environmental or social norms and policy.

THE REHABILITATION MEDICINE TEAM

Because of the far-reaching consequences from an event such as a stroke or spinal cord injury, many deficits and issues need to be addressed and rehabilitation is often a multi-disciplinary team effort:

The *Rehabilitation Physician* who leads the rehabilitation team, is trained in neurological and musculoskeletal assessment and medical management of rehabilitation patients, including complications resulting

from disabling conditions or injuries, has training in therapeutic exercise, modalities, prosthetics, orthotics, wheelchairs/other assistive equipment, and procedures such as soft tissue injection, urodynamics and electrodiagnosis.

The *Rehabilitation Nurse* is involved in direct nursing care of physically impaired patients, assisting with the reacquisition of self-care abilities, teaching in the use of adaptive equipment, and has a major role in patient-family education.

The *Physical Therapist/Physiotherapist* works on motor restoration, posture, balance, coordination, mobility and ambulation, application of modalities, chest physiotherapy, and the evaluation of walking aids or wheelchair needs of the patient.

The *Occupational Therapist* is involved with self-care and functional activities, upper limb performance, assistive devices and technology, training in use of upper limb prostheses, home environment evaluation, and compensations for visual-spatial and cognitive disorders.

The *Speech Therapist/Pathologist* works with speech-language and communication, swallowing and dysphagia, the use of augmentative and alternative communication devices, as well as with cognitive disorders.

The *Medical Social Worker* explores patient-family relationships, resources, finances, discharge living situation, assists with vocational issues, liaison activities, provides emotional support and counseling.

The *Orthotist-Prosthetist* designs, fabricates and evaluates the orthoses (braces) and prostheses (replacement devices) used in rehabilitation.

The *Psychologist* carries out testing of intelligence, memory, cognition, counseling in body image changes, compensations for solving skill deficits, and assists the team to understand patient dynamics, helps resolve conflicts.

The *Dietitian* assesses adequacy and suitability of nutrition and hydration, gives recommendations for parenteral feeds, and counsels patient, family, or the rehabilitation team on dietary matters.

Other Members of the Rehabilitation Team

The *Kinesiotherapist/Corrective Therapist,* uses therapeutic exercise and education, adapts physical activities, fitness programs, and provides driver's education. The *Pharmacist* helps with patients on multiple medications, for drug efficacy, potential interactions and side-effects. The

Vocational Counselor assists in developing/attaining realistic vocational goals, including evaluation, counseling, training, liaison services and support. The *Podiatrist* is involved with management and care of diseases/conditions of the feet and nails, shoe inserts/devices for improving function and protection. The *Audiologist* handles hearing tests, assessment and fitting of hearing aids. The *Biomedical-engineer* designs, evaluates, customizes devices and equipment for the disabled such as wheelchairs, cushions, computers, and environmental control units. The *Therapeutic Recreation Specialist* assesses interests, uses recreational activities to assist in adjustment to disability, increase independence, and reintegration into society. The *Durable Medical Equipment Vendor* helps with knowledge of options available and pricing for devices and equipment, procurement and special orders. The *Religious Counselor/Representative* assists with spiritual aspects, including dealing with and adjusting to serious injury, disease, or dying.

The *Child Life Therapist* helps minimize interruption/disruption to the child from the hospitalization by continuation of schooling, and providing age-appropriate explanations of medical procedures needed. The *Horticultural Therapist* uses gardening as therapy, and the raising of flowers/plants for rebuilding self-confidence. The *Music Therapist* uses various forms of music including listening/performing instruments, voice, body-movements, and musical events to help with speech, paralysis, pain and palliation. The *Dance Therapist* uses the psychological/physical effects of body movements and music.

REHABILITATION MEDICINE PATIENTS

A Rehabilitation Medicine consultation can be requested for assistance with recommending, planning, managing rehabilitation options for the patient, as well as for a prognosis of functional capabilities. For example, a patient may only need a few sessions of focused therapies, or may need more comprehensive rehabilitation intervention, or sometimes may not be able to benefit from active rehabilitation. Rehabilitation can be inpatient or outpatient in nature, the former being generally more complex, with patients having associated potential for medical instability. The need for rehabilitation can result from a disease, injury or process, examples of which include:

Neurological events such as cerebrovascular accidents (stroke), spinal cord injury, traumatic or infective brain injury, Parkinson's disease, Guillain–Barre Syndrome, multiple sclerosis and motor neuron disease.

Musculoskeletal events such as fractures, joint-replacements, upper and lower limb amputation, back injury/surgery, chronic pain, arthritis and other musculoskeletal injuries.

Others Conditions include:

- deconditioning post-surgery, from prolonged immobilization, and from being bed-bound;
- cardiac rehabilitation post-MI, post-cardiac surgery such as CABG, cardiac valve surgery;
- respiratory rehabilitation such as in COPD, lung surgery; and
- cancer rehabilitation, where there are deconditioning states, spinal or cerebral metastatic lesions, pathologic fractures and peripheral nerve invasion/compression.

INDICATIONS FOR COMPREHENSIVE INPATIENT REHABILITATION

In general, a good candidate for comprehensive inpatient rehabilitation in a specialized rehabilitation unit:

- is medically stable with acute problems resolved or adequately controlled;
- has functional deficits in more than only one area, e.g. difficulties in communication, swallowing, self-care abilities (eating, grooming, dressing, toileting, bathing; bowel/bladder management), transferring from surface to surface, locomotion. Single functional deficit may not require comprehensive rehabilitation;
- has clear, achievable, rehabilitation goals; and
- is able to participate in the rehabilitation program from the perspective of cognition, medical condition, and motivation.

THE REHABILITATION ASSESSMENT

The rehabilitation medicine assessment includes elements of the following:

History

- risk factors that might need to be monitored and managed while undergoing rehabilitation;
- premorbid functioning, which is useful to assess current functional deficits and set achievable rehabilitation goals;

- social history including family relationships, the home setting, friends and significant-others, occupation, and recreational interests, to determine support systems and overall goals;
- possible impediments to community reintegration including living environment (e.g. lift landing, squatting toilets), family functioning (poor or abusive relationships), and poor social supports;
- personal and family's rehabilitation goals, as there needs to be mutual agreement and harmony between expectations and limitations; and
- discharge plans as this could affect the rehabilitation, e.g. in a cancer patient, the appropriate goal might be a very short focused period of rehabilitation rather than a longer stay. The remaining life-span might be better spent in the support and comfort of home and family. In a demented or non-participating patient whose family has committed to caregiving at home, an inpatient rehabilitation stay might be appropriate for caregiver training, equipment needs assessment/supply and home-modifications.

Physical Examination

- cognition and ability to follow directions, including the ability to "carry-over" or retain that which has been taught;
- affect/mood changes that may need counseling intervention or medications, as these could determine the patient's ability to participate and hence benefit from active rehabilitation;
- speech and language functioning;
- swallowing and dysphagia;
- sensory deficits including proprioception, that could severely limit functional activities even in the presence of good motor return, and could also lead to accidental injuries;
- neglect or lack of self-awareness on the affected side, that may result in perception deficiencies, functional difficulties, safety concerns and accidental self-injuries;
- range of motion limitations, which could impair rehabilitation due to inability to situate the limb or trunk in a functional/optimal position;
- tone, and spasticity that could restrict movements, be painful, and limit function with potential for injuries from falls during a spasmodic episode;

- motor strength, both from the perspectives of power (ability to generate a forceful movement) and endurance (ability to repetitively carry out the movement);
- skin integrity or breakdown, especially in immobile, incontinent patients with insensate skin and cognitive deficits;
- bowel and bladder functioning;
- functional issues including ability to feed oneself, dress, toilet, transfer, use a wheelchair or walk.

Additional Investigations in Rehabilitation

- *Urodynamics* is a relatively common procedure as many rehabilitation patients have neurogenic bladders, or are in the age group when obstructive pathologies such as prostatic enlargement are common. Besides a detailed pressure-volume study of the bladder in response to filling and stimulation, of particular usefulness are measurements such as post-voiding residual urinary volume, input and output monitoring, as well as tests for detrussor-sphincter dyssynergia.
- *Electrodiagnosis,* which includes the procedures of nerve conduction studies (NCS) and electromyography (EMG) are investigative procedures that may be performed by a rehabilitation physician. These studies could help in the determination of diagnosis, e.g. a neuropathy versus myopathy, could also be used for determining the extent of injury, and to monitor progress of the condition.

LEVELS AND LOCATIONS FOR REHABILITATION

Inpatient

Acute and subacaute comprehensive inpatient rehabilitation
Hospital-based Rehabilitation Medicine Unit (e.g. SGH, TTSH, CGH departments of Rehabilitation Medicine)

Generally, these patients have multiple or complex rehabilitation needs, moderate risk for medical instability, require maximal-minimal assistance with ADL (activities of daily living), physically and cognitively are able to participate and learn. They receive 2–3 hours of therapies per day, 5–6 days a week, under direct supervision of a rehabilitation physician.

Subacute or slow-stream or lower-intensity inpatient rehabilitation
Community Hospital (e.g. AMKCH, St Luke's, St Andrew's, Bright Vision, Westpoint)

Patients have minimal risk from medical instability, require maximal-minimal assistance with ADL, with cognitive or physical limitations precluding intensive rehabilitation. They receive 1–3 hours of therapy per day, 3–5 times per week. The doctor, who may not be a rehabilitation physician, sees the patient 1–3 times per week.

Outpatient

Day Rehabilitation Center (e.g. various senior citizens health care centers, Salvation Army, Apex, Muhammadiyah, Metta, HWA, SPD, or outpatient programs of acute/community hospitals)
Patients are medically stable, require minimal assistance with ADL, and have family support to manage at home. Advantages include socialization, peer support, with availability of professional personnel and equipment. Therapy sessions are 2–5 times per week. A doctor may prescribe the therapies but direct involvement is variable.

Nursing Homes

Selected Nursing Homes (various locations)

Patients need extended care, are medically stable, and often require maximal-minimal assistance with ADL. These are usually group-type therapies with limited one-on-one attention, with goals of maintenance of functioning. Therapy sessions are 1–5 days per week. Direct involvement by nursing home doctor is variable, usually monthly.

Home-based

E.g. SGH HomeCare, Tsao Foundation or Mt Alvernia Home Care Program, AMKCH

Patients cannot safely get to outpatient programs, or endurance is poor and patient is exhausted by the travel itself. There is added advantage of teaching in patient's own surroundings, especially for the cognitively impaired. Therapy may be 1–3 times per week, prescribed by a doctor, but amount of direct involvement is variable.

THE MECHANICS OF INPATIENT REHABILITATION

Rehabilitation addresses some or all aspects of the following:

Management of Immediate Care Issues:
In the acute rehabilitation patient particular attention is given to dysphagia, communication, DVT prophylaxis, nutrition, hydration, skin integrity, urinary bladder and bowel management. Patient and family involvement are intrinsic to the rehabilitation program. Emotional support, and education as to main features of stroke, treatment, and prognosis should be provided.

Training/Improving Neurological/Musculoskeletal Recovery:
As neurological recovery occurs or the musculoskeletal structures heal, rehabilitation aims for functional restoration and includes motor strengthening, sensory reeducation or compensation, balance and posture restoration, coordination drills, and strategies to improve or compensate for visuo-spatial and cognitive deficits.

Management of Complications:
The immobilized, acutely ill patient is at high risk for complications such as skin breakdown, spasticity, contractures, venous thromboembolism, musculoskeletal and neurogenic pain, depression, autonomic dysfunction, reflex sympathetic dystrophy, infections, seizures, falls, malnutrition, cardiac events, recurrent stroke, etc. which may need to be managed while on the rehabilitation medicine unit.

Compensation Strategies:
Frequently there are permanent impairments that will require teaching the patient modified techniques or sequences to carry out a routine functional task, e.g. dressing in bed instead of standing up, placing the paralyzed arm in the sleeve before the non-paralyzed. Equipment can be used to compensate for the impairment including prostheses, orthoses, mobility devices such as wheelchairs and walking aids, electronic environmental control units and computers.

Discharge Planning:
The team, patient and family, will need to work out a plan for destination after discharge, and continuity of care including arrangements for follow-up and subsequent outpatient care. A home visit and renovations/modifications may be needed. Avocational (recreational) issues are important from the perspective of quality of life and the person's interests should be assessed. Vocational issues are relevant in some patients

especially the younger ones and need to be addressed. This may involve liaison with the employers, workplace adaptations, and retraining. Finally there are many other adjustments that may be needed when a severe disability has occurred including sexuality, issues of self-worth, and identity.

THE TOOLS OF REHABILITATION

In addition to medications, injections and debridement procedures, the rehabilitation physician may also prescribe orthoses, prostheses, and:

Therapeutic Exercise

Therapeutic exercise increases strength and muscle size, may be *isotonic* (concentric or shortening exercises, also eccentric or lengthening exercises), *isometric* (static or non-shortening exercises), or *isokinetic* (constant velocity or accommodating resistance exercise). The DeLorme Axiom is an important principle in rehabilitation. This states that low-repetition, high-resistance exercises produce *power,* whereas high-repetition, low-resistance exercises produce *endurance,* and each is incapable of producing results of the other. Hence therapeutic exercises usually includes both. The concept of *specificity training* is also well-accepted in rehabilitation, e.g. although therapeutic exercises for the muscles of ambulation are important, it is still necessary to work on actual walking to regain this function well. Other aspects of therapy include restoring coordination, balance, and posture, exercises in preparation for gait retraining including crutch and cane usage if necessary, and acquiring of wheelchair skills.

Heat and Cold Modalities

Thermal modalities are often used in rehabilitation:

Heat Modalities (e.g. hot packs, paraffin wax bath, fluidotherapy, infrared radiation, shortwave diathermy, ultrasound for deep heating):

Heat has excellent properties including analgesia and sedation that decreases pain, spasms, and aids relaxation. It is a general stimulant, and dilates vessels improving blood flow resulting in increasing nutrient inflow and waste removal. It causes an increase of capillary permeability

that may either increase or decrease interstitial fluid and hence edema. An important effect is increase of non-elastic tissue extensibility, allowing elongation/stretching of tendons or scars and hence improving range of movement. The resultant decrease in joint stiffness increases speed, freedom of movement, and agility. Heat increases peripheral nerve conduction velocity and motor nerve transmission, and thus motor function as a whole.

Cold Modalities/Cryotherapy (e.g. ice massage, chemical cold packs, ice towels, ice packs, vapocoolant sprays):

Cold modalities also have properties of analgesia/sedation, decreasing pain, spasms, and aiding relaxation. The initial superficial vasoconstriction on application decreases superficial bleeding, and subsequent increase in blood viscosity with decreased flow further retards blood loss. Decrease in peripheral nerve activity from cold may reduce spasticity. Lowering of joint tissue and fluids temperature decreases activity of joint-damaging enzymes.

Hydrotherapy

Hydrotherapy in a swimming pool or whirlpool tub allows for delivery of hot or cold treatment modalities. The physical effects of buoyancy, hydrostatic pressure, surface tension, and turbulence create an ideal environment for therapeutic exercise, such as muscle strengthening, balance, range-of-motion activities, and reduced-weight ambulation. Both thermal properties and hydrostatic pressure improves circulation and reduces edema. Pressure, fluidity, and turbulence are useful for the treatment of open wounds and skin problems, including debridement, removal of dressings, and application of medications or skin lubricants. Water is also a relaxing medium in general, producing psychological benefits.

Therapeutic Electricity, etc.

Electrical Stimulation can decrease spasticity with direct stimulation of spastic or antagonistic muscles. Muscle irritation, spasm and pain may respond to high-frequency intermittent stimulation. Use of electrical stimulation to increase circulation and nutrition to denervated muscle is

controversial as it retards atrophy, but may interfere with peripheral nerve regeneration. It decreases edema by improving muscle-pumping action that increases circulation, and accelerates wound healing by retarding bacterial growth.

Iontophoresis is the use of electric current to drive ions of various substances or medications e.g. steroids, lidocaine, salicylate, zinc oxide, through the skin into underlying tissues.

Transcutaneous Electrical Nerve Stimulation (TENS) is the application of low-voltage electrical pulses to the nervous system through the skin. TENS may change pain perception as per the Gate-Control Theory whereby high-frequency stimulation of non-nociceptors or axons interferes with relay of pain sensations to higher brain centers. It may also act via natural opiates from the pituitary gland (beta-endorphins) and spinal cord (enkephalins), released by low-frequency stimulation of sensory nerves. TENS may be useful for acute and chronic, phantom, postoperative, cardiopulmonary, and neurological pain, as well as before painful stretching or debridement.

Ultraviolet Irradiation (UV) causes vasodilatation that stimulates granular tissue formation leading to tissue repair. It has an anti-rachitic effect due to its role in vitamin D formation, and by stimulation of steroid metabolism. Bacterial destruction occurs from its stimulation of reticuloendothelial cells, increasing circulatory antibodies, with associated wound healing. There are also psychological benefits such as sense of well-being.

Low-powered Helium-Neon *Lasers* (Light Amplification by Stimulated Emission Radiation) can help reduce pain reduction and accelerate tissue healing from increase in collagen synthesis, vascularization, and from decrease in microorganisms. Lasers may be used for promotion of tissue healing, and management of pain associated with muscle spasm and inflammation.

Biofeedback gives immediate information-return about physiological functions, and hence allows for active self-control over the function being monitored. It includes aural or visual feedback, skin thermal feedback, skin electrical conductance, or electromyographic (EMG) feedback for muscle activity. It is useful in general relaxation-training e.g. in hypertension, headaches, chronic pain, as well as for training improved motor control of weak muscles.

Traction/Distraction

Traction is done by application of forces to the spine or limbs to separate vertebrae, and elongate surrounding tissues such as muscles and ligaments. The increase of space between vertebrae, articulating facets, and intervertebral foramen may result in relaxation of paraspinal muscles, and diminution of bulging herniated discs or pressure on nerve roots in the intervertebral foramen. It can be helpful in nerve root impingement, subacute joint pain, subacute degenerative joint disease, discogenic pain, chronic spinal compression fracture, joint hypomobility, and paraspinal muscle spasm.

Intermittent Pneumatic Compression

Forcing air with a pump intermittently into inflatable sleeves or boots around an upper or lower extremity results in an increase of interstitial space fluid pressure that encourages fluid return to venous or lymphatic vessels, helping reduce edema. Intermittent pneumatic compression can be useful in chronic edema, venous insufficiency, lymphedema, stasis ulcers, amputated limbs, wound healing, and thrombophlebitis prevention.

Medications

Rehabilitation physicians manage certain medical complications more often than other specialties. These complications may entail the use of medications such as:

- *Spasticity*
  Baclofen (GABA analog), diazepam (benzodiazepine), gabapentin (anticonvulsant), clonidine, tizanadine (alpha2-adrenergic agonists), dantrolene sodium (skeletal muscle relaxant), carisprodol, cyclobenzaprine (brainstem neuronal depressants). Others less commonly used or controversial: cyproheptadine (nonselective 5-HT antagonist), opiates (narcotic analgesic), glycine, L-threonine (amino acid inhibitory neurotransmitters), 4-aminopyridine (potassium channel-blocker), orphenadrine citrate (antiparkinsonian agent), 9-tetrahydrocannabinol (narcotic)

- *Autonomic dysreflexia*
  Topical lidocaine (local anesthetic), nifedipine, nitrates (nitroglycerine paste), phenoxybenzamine, propanolol, terazosin, diazoxide, hydralazine, mecamylamine
- *Neuropathic pain*
  Amitriptyline, nortriptyline (tricyclic antidepressants), trazadone (nontricyclic antidepressant), gabapentin, carbamazepine (anticonvulsants), baclofen, tizanidine, clonidine (anti-spasticity agents), mexiletine, lidocaine (anti-arrthymic agents), capsaicin (substance P depletor)
- *Depression*
  Fluoxetine, fluvoxamine, sertraline, paroxetine, citalopram, escitalopram (selective serotonin reuptake inhibitors), amitriptyline, nortriptyline (tricyclic antidepressants)
- *Psycho-modulation*
  Methylphenidate (psychostimulant), bromocriptine (dopamine agonist), propanolol (β-blocker), risperidone (antipsychotic), lorazepam, chloral hydrate, zolpidem, nortriptyline, fluvoxamine (anxiolytics, sedatives, and antidepressant)
- *Bladder management*
  Oxybutinin, propantheline, tolterodine, amitiptyline, flavoxate (anticholinergics), ephedrine (sympathomimetic), bethanecol (cholinomimetic), prazosin, terazosin (adrenolytics), baclofen, dantrolene, tizanidine (antispasticity agents)
- *Bowel management*
  Psyllium, isphagula (fiber supplements), docusate, lactulose (stool softeners), senna, bisacodyl (bowel stimulants), gylcerine, paraffin (lubricants), cisapride (peristalsis stimulant)

CONCLUSION

Rehabilitation medicine is the medical field specializing in care of patients with disabling disease or injury. It utilizes a team approach, with optimal functioning and independence of the patient as the ultimate goal. Rehabilitation physicians are trained in the assessment and medical management of patients with neurological and musculoskeletal disorders, and their associated complications. They are also trained in therapeutic

exercise, modalities, prosthetics, orthotics, assistive devices for the disabled, and procedures such as electrodiagnosis. Rehabilitation is holistic in nature and addresses the restoration, or compensation where necessary, not only of the physical but also psychological, social, educational, vocational, and recreational abilities.

FURTHER READING

1. 50th Anniversary of the American Academy of Physical Medicine and Rehabilitation, Archives of Physical Medicine and Rehabilitation, 69, Special Issue, 1988.
2. DeLisa JA, Gans BM (eds), *Rehabilitation Medicine: Principles and Practice*, 3rd ed., Lippincott-Raven, Philadelphia, 1998.
3. Hecox B, Mehreteab TA, Weisberg J (eds), *Physical Agents: A Comprehensive Text for Physical Therapists,* Appleton & Lange, Norwalk, 1994.
4. Basmajian JV, Wolf SL (eds), *Therapeutic Exercise*, 5th ed., Williams & Wilkins, Baltimore, 1990.
5. Kirshblum S, Campagnolo DI, DeLisa JA (eds), *Spinal Cord Medicine,* Lippincott Williams & Wilkins, Philadelphia, 2002.

Material in "The Rehabilitation Medicine Team" extracted from DeLisa and Gans,[2] "The Tools of Rehabilitation" from Hecox *et al*.[3] and Basmajian and Wolf,[4] and "Medications" from Kirshblum *et al*.,[5] and "Levels and Locations for Rehabilitation" from Flick CL, Stroke rehabilitation. 4. Outcome and psychosocial consequences, *Arch Phys Med Rehabil* **80**(5 Suppl 1):S21–26, 1999.

Rheumatology

This page intentionally left blank

82

Diagnosis in Rheumatic Diseases

Fong Kok Yong

INTRODUCTION

Rheumatic diseases encompasses more than 100 different types of conditions associated with the musculoskeletal and/or immune system. The symptoms can be straightforward or of a very varied and protean nature. Oftentimes the manifestations can even baffle experienced clinicians. Any of the organ system in the body can be affected, though it is usually the musculoskeletal symptoms that bring the patient to the doctor. A comprehensive history taking process, with special attention to the joint symptoms, supplemented by a thorough physical examination and augmented by focused laboratory investigations, remain the foundations of rational diagnosis in rheumatic conditions. There's no good substitute for the time-tested process of history taking and physical examination. In more than 90% of the cases, a diagnosis can be reached after these steps and laboratory investigations serve the function of confirming the diagnosis or in subsetting patients for more targeted therapy. Laboratory tests can also help in evaluating the severity and disease course of the rheumatic condition.

HISTORY

The history taking process can be focused broadly into 2 categories. Firstly, those pertaining to *joint symptoms*, and secondly, those with predominantly *systemic symptoms*. Joint pain is the most common symptom among rheumatic diseases and hence plays an important part in the diagnosis of rheumatic conditions. The pattern of joint involvement, number of joints involved, characteristics of the joint pain and association with joint swellings, are useful diagnostic indicators in the history. Table 1 shows the characteristics of joint involvement of some rheumatic conditions seen locally. A symmetrical arthritis of the finger joints indicate the possibility of rheumatoid arthritis as compared to asymmetrical involvement in seronegative spondyloarthropathy patients. Ankylosing spondylitis patients usually have axial skeleton joint pain presenting with early morning low backache.

Acute onset of excruciating pain typically point toward the presence of *gouty arthritis*, with the big toe most commonly involved. *Pseudogout* or calcium pyrophosphate deposition disease (CPPD) can also present in a similar way. Differences between them are a higher frequency of knee involvement in CPPD and its usual occurrence in the elderly. Primary gouty arthritis usually occurs at a younger age group. Table 2 summarizes the broad differences between the 2 conditions.

Septic arthritis must be excluded in all patients presenting with monoarticular arthritis, even in those with a history of chronic arthritis. The consequences for delayed diagnosis in septic arthritis can be catastrophic, i.e. severe joint destruction. *Degenerative joint pains*, e.g. osteoarthritis, are usually aggravated by movement and weight-bearing stresses, and relieved by rest. *Inflammatory pain*, on the other hand, is aggravated by prolonged rest, hence the common finding of increased pain on waking. Movement of a mildly inflamed joint can help relieve stiffness. The patient's age also helps in discriminating between degeneration and inflammation as the former is predominantly a condition of the elderly while inflammatory pain, e.g. due to ankylosing spondylitis, is usually in the younger age group. Pain occurring at the enthesis (the site where the tendon joins the bone) can be suggestive of soft tissue rheumatism or seronegative spondyloarthropathies. The pain is usually localized and aggravated by certain movement of the joints. In *fibromyalgia* (generalized soft tissue rheumatism) the patient complains of characteristic painful areas, which are confirmed as localized

Table 1 Characteristics of Joint Symptoms in Some Rheumatic Diseases

Joints	Pattern of Involvement	Number Involved	Character of Pain	Associated Swelling
Osteoarthritis	Weight-bearing and DIP joints; cervical and lumbosacral spine	Variable	Aggravated by movement, relieved by rest	Due to fluid in early stages; bony hypertrophy at later stages
Gouty arthritis	Usually unilateral; big toe, foot or ankle joint	Usually monoarticular	Acute onset of excruciating pain	May be present during acute phase
Soft tissue rheumatism	Joint is normal; involved para-articular structures	Multiple tender points at the sides of joints	Localized, constant pain, stiffness on waking	None
Rheumatoid arthritis	Symmetrical involvement of synovial joints, esp MCP and PIP joints	Usually multiple	Constant, worse in the morning	Present during active phase
Ankylosing spondylitis	Axial joint involvement; usually starting from lower back	Variable peripheral joint involvement	Constant, worse in the early morning	May be present in peripheral joints
Systemic lupus erythematosus	Asymmetrical joint involvement	Variable	Constant	Usually no swelling
Psoriatic arthritis	Asymmetrical joint involvement, esp DIP joints	Variable, usually finger joints	Constant	May present with "sausage-like" fingers
Sjogren's Syndrome	Asymmetrical joint involvement	Variable	Constant	May be present

DIP: *D*istal *I*nter-*P*halangeal; MCP: *M*eta-Carpal *P*halangeal; PIP: *P*roximal *I*nter-*P*halangeal.

tender spots on palpation.[1] Referred pain due to pathology elsewhere , e.g. prolapsed intervertebral disk, can usually be picked up during history taking and confirmed by physical examination. Joint involvement in systemic lupus erythematosus is usually much milder when compared to

Table 2 Common Characteristics of Gout and Pseudogout

	Gout	CPPD (Pseudogout)
Male : Female ratio	3–7 : 1	4 : 1
Age (years)	40 to 50	60 to 70
Commonest joint involved	First MTP	Knee

CPPD: *C*alcium *P*yro*P*hosphate *D*isease.

inflammatory arthritides. Arthralgia, rather than arthritis, is the usual presentation in lupus patients. Patients with Sjogren's syndrome usually have arthralgia and only a minority has frank arthritis.

Systemic symptoms as presenting manifestations generally indicate greater severity and a more active disease process. Fever, malaise and general debility are symptoms reflective of systemic involvement. Other organ specific symptoms like coma in cerebral involvement, generalized edema in renal failure, breathlessness in alveolitis or pulmonary hypertension, Raynaud's phenomenon in vascular hyperreactivity, may also be prominent presenting symptoms.

FAMILY HISTORY

Most rheumatic diseases are sporadic in occurrence and familial cases are usually the exception rather than the rule. However, a strong family history of a rheumatic condition, e.g. ankylosing spondylitis will suggest a similar diagnosis in a young individual with persistent low back pain. Likewise, a patient with a positive anti-nuclear antibody test and had relatives with autoimmune conditions, e.g. thyroiditis, myasthenia gravis, rheumatoid arthritis, autoimmune diabetes mellitus, probably is developing an autoimmune rheumatic condition.

DRUG HISTORY

Drugs can either cause a rheumatic condition or aggravate a subclinical rheumatic condition. Drugs, e.g. procainamide can cause drug-induced lupus.[2] Recognition of this condition and stopping the drug will be beneficial to the overall management of the patient. Table 3 shows some of the

Table 3 Rheumatic Conditions Associated with Drugs Ingestion

Drug-Induced Condition	Possible Drugs
Lupus	Chlorpromazine; griseofulvin; hydralazine; isoniazid; methyldopa; penicillamine; phenytoin; procainamide; quinidine.
Gouty arthritis	Thiazide diuretics; low-dose aspirin.
Raynaud's Phenomenon	β-blockers; ergotamine; bleomycin.

drugs associated with autoimmune conditions. A patient who develops acute arthritis after administration of diuretics or low-dose aspirin, probably has drug-induced gouty arthritis. Changing the medication will alleviate the condition and prevent the patient from having unnecessary investigations.

TRAVEL HISTORY

Global travel is a common phenomenon in this modern age and travelers can go to "out-of-the-way" places easily. Hence rheumatic conditions uncommon in the traveler's home country may be acquired during overseas travel. A recent episode of bloody diarrhea may not be mentioned by the patient unless specifically asked for in the travel history. This may explain joint symptoms due to *reactive arthritis* following the bout of diarrhea.[3] Lyme arthritis[4] may not be endemic in the traveler's home country but may occur if the traveler is infected by the tick-borne spirochete, *Borrelia burgdorferi*, after a visit to endemic areas in the US or Europe.

PHYSICAL EXAMINATION

A general musculoskeletal examination followed by a more focused examination of the joints involved are very helpful in coming to a diagnosis. The general musculoskeletal examination can easily be performed expeditiously and yet yield useful diagnostic signs.[5] Examination of the hands, elbows, knees and spine is particularly useful in the diagnosis of rheumatic conditions as shown in Table 4.

The site of tenderness can also aid in differentiating the different rheumatic diseases. The tenderness can be arising from the joint as a

Table 4 Some Distinctive Signs Associated with Various Rheumatic Conditions

	Hands	Elbows	Knees	Spine
Osteoarthritis	Heberden's or Bouchard's nodes; DIP joint deformities	Usually normal	Crepitus; Genu varus	Lower back tenderness; cervical tenderness
Rheumatoid arthritis	Swan-neck, Boutonneire's, Mallet finger deformities; Ulnar deviation	Rheumatoid nodules	Genu valgus	Cervical subluxation (*Note*: Do not forcibly examine the neck if subluxation is suspected)
Psoriatic arthritis	Dactylitis; DIP joint deformities	Psoriatic plaques	Usually normal	Usually normal
Ankylosing spondylitis	Usually normal	Usually normal	Usually normal	Lower back and S-I joint tenderness and limitation of spinal movements
Gouty arthritis	Usually normal	Tophi	Usually normal	Usually normal
Progressive systemic sclerosis	Sclerodactyl; Raynaud's phenomenon; digital infarcts	Usually normal	Usually normal	Usually normal

DIP: *D*istal *I*nter-*P*halangeal; S-I: *S*acro-*I*liac.

result of fluid distention e.g. rheumatoid arthritis or localized to a specific area, e.g. lateral epicondyle in tennis elbow. In the latter, the joint space is not involved. Crepitus on joint movement indicate degenerative changes within the joint, and limited joint movement associated with pain and stiffness can be due to supraspinatus tendinitis (painful arc syndrome) or rotator cuff tendinitis (frozen shoulder).

Organ specific examination can be directed by the symptoms elicited during the history taking process, e.g. breathlessness in the patient will warrant thorough respiratory (e.g. fibrosing alveolitis) and cardiovascular (e.g. pulmonary hypertension) examinations.

LABORATORY INVESTIGATIONS

Blood investigations should not be the sole or main diagnostic determinant in rheumatic diseases. Their roles are in confirming the clinical diagnosis as well as serving as disease activity markers. In general they can be classified into 2 categories: *Diagnostic Markers* and *Disease Activity Markers*. Table 5 shows the usefulness of some common tests used in rheumatology practice. Most tests are used either for diagnostic purposes or in evaluation of disease activity. Some quantitative tests, e.g. for rheumatoid factor, Anti-dsDNA autoantibodies, are useful for both purposes.

Table 6 shows the associations of rheumatic diseases with various autoantibodies. Knowing these associations assist us in coming to a more definite diagnosis in difficult cases. However one has to bear in mind that the relationship is not exclusive and the same autoantibodies can be present in more than one rheumatic conditions.

Genetic markers are generally of limited value in the diagnosis of rheumatic condition. The known major histocompatibility complex (MHC)

Table 5 Common Tests Used in Diagnosis and Evaluation of Rheumatic Diseases

	Diagnostic Markers	Disease Activity Markers
ESR	−	+
CRP	−	+
Hb, WBC, platelet count	+/−	+
Rheumatoid factor	+	+
Anti-dsDNA	+	+
ANA	+	−
Anti- Sm	+	−
Anti-RNP	+	−
Anti-Ro	+	−
Anti-La	+	−
Anti-Scl-70	+	−
Antiphospholipid antibodies*	+	−

+ = Useful; − = Not useful; +/− = Sometimes useful.
*Anti-phospholipid antibodies include: presence of lupus anticoagulants, presence of anticardiolipin antibodies (either IgG or IgM isotype), biological false-positive VDRL test.

ESR: Erythrocyte Sedimentation Rate; CRP: C-Reactive Protein; Hb: Hemoglobin concentration; WBC: Total white cell count; ANA: Anti-nuclear antibodies test; Anti-Sm: Antibodies to the Smith antigen; Anti-RNP: Antibodies to Ribonucleoproteins; Anti-Ro: Antibodies to Ro or SSA antigen; Anti-La: Antibodies to La or SSB antigen; Anti-Scl 70: Antibodies to Topoisomerase I.

Table 6 Autoantibodies and Disease Associations

Autoantibodies	Associated Rheumatic Condition
Anti-Sm	SLE
Anti-RNP	SLE; MCTD; Raynaud's Phenomenon
Anti-Ro	Sjogren's Syndrome; SLE; subacute cutaneous lupus; congenital heart block
Anti-La	Sjogren's Syndrome; SLE
Anti-Scl 70	Progressive systemic sclerosis
Anti-phospholipid antibodies	PAPS; SLE
Anti-neutrophil cytoplasmic antibodies (c-ANCA)	Wegener's granulomatosis
Anti-neutrophil cytoplasmic antibodies (p-ANCA)	Vasculitis
Anti-Jo 1	Dermatomyositis; Polymyositis
Anti-dsDNA	SLE
Anti-histone	SLE (drug-induced)
Rheumatoid factor (IgM)	Rheumatoid arthritis; SLE

SLE: *S*ystemic *L*upus *E*rythematosus; MCTD: *M*ixed *C*onnective *T*issue *D*isease; PAPS: *P*rimary *A*nti-*P*hospholipid *S*yndrome; IgM: Immunoglobulin M.

associations are believed to confer susceptibility to development of rheumatic diseases. However, there is as yet no firm evidence to suggest associated genotypes can predict development of disease. Whether the genetic associations also predict severity is debatable. Table 7 shows some rheumatic diseases with their associated HLA antigen types.

Joint fluid examination is very useful in differentiating between sepsis and joint inflammation. The simple Gram-stain test on the fluid had enabled early and definite diagnosis of gonococcal septic arthritis in many instances. The presence of Gram-negative diplococci in joint fluid usually enables the clinician to elicit from the patient, albeit after the diagnosis is made, a history of unprotected sexual practices with multiple partners. The patient may not volunteer the information initially even if asked specifically. Detection of negatively birefringent needle-shaped crystals are diagnostic of gouty arthritis even if the joint involved may be uncommon, e.g. wrist.

Imaging studies in rheumatology help in evaluating the extent and severity of the disease. Plain radiographs are usually sufficient in determining whether erosions are present in erosive arthritis or the joint space is narrowed medially in radiographs done on weight-bearing osteoarthritic knees. "Fuzziness" or sclerosis of the sacroiliac joints can also be determined on plain radiographs. Computed tomography scans

Table 7 Genetic Markers and Rheumatic Diseases

Rheumatic Disease	HLA Type	HLA Subtype	References
Systemic lupus Erythematosus	DRB1*15 DRB1*03	N.A.	6, 7
Rheumatoid arthritis	DRB1*04	0401; 0404; 0405; 0101; 0102; 1402; 1001	8, 9
Ankylosing spondylitis	B*27	2704; 2705;	10

Table 8 ARA Revised Criteria for Classification of Rheumatoid Arthritis (1987)

	Criteria	Classified as
1	Morning stiffness	*Clinical*
2	Arthritis of 3 or more joint areas	*Clinical*
3	Arthritis of hand joints	*Clinical*
4	Symmetric arthritis	*Clinical*
5	Rheumatoid nodules	*Clinical*
6	Serum rheumatoid factor	*Laboratory*
7	Radiographic changes	*Laboratory*

RA is diagnosed if at least 4 of the criteria are present. For criteria 1 to 4, they must be present for at least 6 weeks.

and magnetic resonance imaging are usually more sensitive in "picking up" early erosions or finer changes. However the perceived advantages have to be balanced against the cost of performing these imaging studies.

Tissue biopsy may be needed to confirm or clarify a provisional diagnosis, e.g. skin — leucocytoclastic vasculitis, kidney — lupus nephritis, synovium — chronic synovitis. Renal biopsy helps in the management plan by supplying the histological classification[11] or type of the renal involvement in lupus patients. A synovial biopsy may differentiate between chronic synovitis and tuberculous infection of the synovium as the cause of recurrent monoarticular effusion.

CONCLUSION

Diagnosis of rheumatic diseases is predominantly clinical in nature. This is well-illustrated by the generally accepted guidelines for diagnoses of rheumatoid arthritis[12] and systemic lupus erythematosus.[13]

Table 9 The Revised ACR Criteria for Classification of Systemic Lupus Erythematosus (1982)

	Criteria	Classified as
1	Malar rash	*Clinical*
2	Discoid rash	*Clinical*
3	Photosensitivity	*Clinical*
4	Oral ulcers	*Clinical*
5	Arthritis	*Clinical*
6	Serositis	*Clinical/Laboratory*
7	Renal disorder	*Laboratory*
8	Neurologic disorder	*Clinical*
9	Hematologic disorder	*Laboratory*
10	Immunologic	*Laboratory*
11	Antinuclear antibody	*Laboratory*

Systemic lupus erythematosus is diagnosed if at least 4 of the criteria are present.

Tables 8 and 9 illustrate the predominantly clinical nature of the diagnostic criteria. The laboratory criteria serve as supplementary factors in helping the clinician come to a more definitive diagnosis.

Newer laboratory tests do not yet replace the clinical process of history-taking and physical examination, but they can help to subset patients into more homogenous groups for potentially more effective targeted therapy.

REFERENCES

1. Wolfe F, Smythe HA, Yunus MB, *et al.*, The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the multicentre criteria committee, *Arthritis Rheum* **33**:160–172, 1990.

2. Hess EV, Role of drugs and environmental agents in lupus syndromes, *Curr Opin Rheumatol* **4**:688–692, 1992.

3. Toivanen A, Toivanen P, Reactive arthritis, *Curr Opin Rheumatol* **9**:321–327, 1997.

4. Steere AC, Malawista SE, Snydman DR, *et al.*, Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three Connecticut communities, *Arthritis Rheum* **20**:7–17, 1977.

5. Bluestone R, The Five-minute Rheumatologic Exam, in: Bluestone R (ed.), *Rheumatology*, Houghton Mifflin, Boston, pp. 29–37, 1980.

6. Woodrow JC, Immunogenetics of systemic lupus erythematosus, *J Rheumatol* **15**: 197–199, 1988.

7. Scherak O, Smolen JS, Mayr WR, HLA-DRw3 and systemic lupus erythematosus, *Arthritis Rheum* **23**:954–957, 1980.

8. Chan SH, Lin YN, Wee GB, Koh WH, Boey ML, HLA Class 2 genes in Singaporean Chinese rheumatoid arthritis, *Br J Rheumatol* **33**:713–717, 1994.

9. Gao XJ, Gazit E, Livneh A, Stastny P, Rheumatoid arthritis in Israeli Jews: shared sequences in the third hypervariable region of DRB1 alleles are associated with susceptibility, *J Rheumatol* **18**:801–803, 1991.

10. Gonzales-Roces S, Alvarez MV, Gonzalez S, *et al.*, HLA B-27 polymorphism and worldwide susceptibility to ankylosing spondylitis, *Tissue Antigens* **49**:116–123, 1997.

11. Austin HA, Muenz LR, Joyce KM, *et al.*, Prognostic factors in lupus nephritis. Contribution of renal histologic data, *Am J Med* **75**:382–391, 1983.

12. Arnett FC, Edworthy SM, Bloch DA, *et al.*, The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis, *Arthritis Rheum* **31**: 315–324, 1988.

13. Tan EM, Cohen AS, Fries JF, *et al.*, The 1982 revised criteria for the classification of systemic lupus erythematosus, *Arthritis Rheum* **25**:1271–1277, 1982.

This page intentionally left blank

83

Connective Tissue Disease and Systemic Vasculitis

Leong Khai Pang

Introduction

The term "connective tissue diseases" is a convenient label for the group of diseases characterized by the presence of serum antibodies and other immune cells that recognize a diverse range of self-antigens. It encompasses systemic lupus erythematosus (SLE), Sjogren's syndrome (SS), dermatomyositis/polymyositis (DM/PM), systemic sclerosis (SSc), mixed connective tissue disease (MCTD) and antiphospholipid syndrome (APS). There are no pathognomonic laboratory tests for any one of these, so the diagnosis is based on recognizing the pattern of laboratory and clinical features. The American College of Rheumatology (ACR) has published classification criteria (not to be equated with "diagnostic" criteria) for almost all the CTDs. We can use them for making diagnoses as well, provided we keep their limitations in mind.

The vasculitides, or inflammation of the blood vessels, are uncommon diseases. Vessels of any size or location may be involved, which could lead to occlusion, aneurysm formation or rupture, resulting in

ischemia or necrosis of the tissues they supply. The clinical manifestations are protean and can range from mild to immediately life-threatening.

Principles of Management

There is no cure for the CTDs and primary vasculitides but we have to treat the disease manifestations. These illnesses are characterized by disease flares between periods of quiescence. Disease flares must be controlled because prolonged periods of active disease lead to organ damage and poorer outcome. The term "flare" and "activity" refer to the increased immune-mediated inflammation over the baseline. "Severity" indicates the likelihood of sustaining morbidity or mortality from the disease activity. Irreversible morbidity is also known as "damage."

Since drugs and other therapeutic modalities have side-effects, we weigh the expected benefits and risks carefully before we use them. The decision on whether, when and how to treat depends on the severity and the rate of progression of the disease.

CONNECTIVE TISSUE DISEASES

Systemic Lupus Erythematosus (SLE)

SLE is characterized by immune-mediated attack on many organs in the body in the presence of a host of different autoantibodies. The range of disease manifestations is broad but generally recognizable.

Epidemiology

Most local rheumatologists agree with the estimate that there are four thousand lupus sufferers in the island, giving us a prevalence rate of about 100 per 100 000. One of the better epidemiology studies showed that prevalence of SLE in Birmingham was 27.7 per 100 000 (49.6/100 00 for females and 3.6/100 000 for males) on 1 January 1992. The prevalence rate was highest among the Afro-Caribbeans, followed by Asians and Caucasians.

Immunogenetics and pathogenesis

Relatives of lupus patients are 20 times more likely to suffer from the same illness than the general population. The disease concordance is 24% in monozygotic twins but only 2% in dizygotic ones.

HLA DR2 and DR3 are found more often in Caucasian patients than controls. In Singapore, the frequency of DR 16(2) was statistically significantly increased in 51 Malay SLE patients. However, DR2 does not appear to be associated with SLE based on a study of 26 Chinese patients from National University Hospital.

Other genes that are linked with the pathogenesis of SLE encode proteins associated with antigen or immune complex clearance (complements and their receptors, mannose-binding protein), apoptosis (fas and fas ligand, poly ADP-ribosyl polymerase), lymphocyte signaling (HLA, TAP, IL-6, IL-10, tumor necrosis factor-α, T cell receptor ξ chain) and immunoglobulin heavy chain receptors (FcγRII, FcγRIII).

Genome-wide scans for lupus susceptibility implicate multiple genes distributed over numerous chromosomes. Indeed, SLE result from the interaction of at least 20 gene products and multiple environmental triggers (which could be drugs, viruses, ultraviolet radiation or sex hormones). The precise way in which these factors interact is unknown, but the result is a loss of self tolerance, with the production of a host of autoantibodies by B cells, helped by activated but dysregulated T cells.

Manifestations and outcome

The most common clinical features of Singapore patients presenting with SLE are hematological (73%), arthritis (57%), malar rash (43%), renal disorder (31%) and photosensitivity (30%). Most SLE patients suffer from constitutional symptoms (fever, malaise, anorexia and weight loss) when their disease is active. Their reappearance in a hitherto well-controlled lupus patient may herald a disease flare.

Mucocutaneous manifestations are common in SLE. Four of them, mouth ulcers, photosensitivity, malar or butterfly rash and discoid lupus erythematosus (DLE) are classification criteria of SLE. Mouth ulcers are typically painless. DLE is a raised, scaly, circumscribed rash that may scar; when the scales are removed, follicular plugging is found. Subacute cutaneous lupus erythematosus (SCLE) is a non-scarring rash found in sun-exposed areas, sometimes mistaken for psoriasis. SLE causes a non-scarring alopecia, but scars from discoid lesions also result in irreversible hair loss. "Lupus hair" breaks easily, giving the patient an unruly and shaggy appearance.

SLE produces a symmetrical and non-erosive arthritis. Ligamentous laxity may simulate swan-neck deformities of the fingers, producing Jaccoud's arthropathy. The complaint of weakness or muscle tenderness should prompt us to look for inflammatory muscle disease. Fortunately, myalgia (muscle pain) is much commoner problem than polymyositis or dermatomyositis (inflammation of the muscles).

Renal manifestations in SLE range from asymptomatic proteinuria to acute renal failure. In one study, maleness, higher activity index, proliferative and inflammatory changes on renal biopsy, hypertension and severe infections were associated with progression to chronic renal failure, whereas the institution of treatment and higher education achievement of patient were associated with better renal outcome.

Cytopenias, in one or more cell lines, is common in lupus patients. Anemia and thrombocytopenia are usually due to peripheral immune destruction. Before we attribute the cytopenia to SLE, we have to exclude the effects of drugs.

Abdominal manifestations include serositis, aseptic peritonitis, pancreatitis and acute abdomen. Serositis may present as abdominal pain or ascites. Pancreatitis can be the presenting feature of lupus.

Lung involvement in lupus includes pneumonitis, pulmonary fibrosis, pulmonary hemorrhage, shrinking lung syndrome and pulmonary hypertension. Chronic pneumonitis may lead to interstitial fibrosis. Pulmonary hemorrhage, due to alveolar capillaritis, usually occurs in the context of active lupus and is a life-threatening complication. Two local series including more than 50 cases demonstrated that survival ranges from 50–60%. Clues to diagnosis are sudden onset or worsening dyspnea, fleeting opacities in the chest X-ray, otherwise unexplained fall in hemoglobin and raised diffusion capacity (DLCO). Shrinking lung syndrome is an unusual but interesting complication where the lung volume actually shrinks in serial chest films due to diaphragmatic dysfunction. Pulmonary hypertension can result from micro-embolism to the pulmonary capillary bed or can be idiopathic. Severe pulmonary hypertension (systolic pressure > 70 mmHg) is a grave prognostic factor.

Pericarditis is common but frequently asymptomatic. Libman–Sacks endocarditis is the classic heart lesion of SLE. Verrucous vegetations are found on the valve leaflets, most commonly the mitral valve. They are frequently not apparent clinically, though they can cause thromboembolism or valve dysfunction.

In 1999, the ACR defined 19 syndromes in SLE that involve the central or peripheral nervous systems: aseptic meningitis, cerebro-vascular disease, demyelinating syndrome, headache, movement disorder, myelopathy, seizures, acute confusional state, anxiety disorder, cognitive dysfunction, mood disorders, psychosis, acute demyelinating polyradiculopathy, autonomic disorder, mononeuropathy, myasthenia gravis, cranial neuropathy, plexopathy and polyneuropathy. Generally, other causes such as infections or drugs must be excluded before these syndromes are attributed to SLE.

Mortality from SLE follows a bimodal distribution; patients tend to die early or late in the course of disease. Early death is due to disease activity or infection, while that occurring eight to ten years later results from complications of disease or therapy (such as ischemic heart disease). The survival rate of lupus has been improving since the 1950's when such data first became available. In the late 1990's, patient survival from a prospectively followed-up cohort in Canada was 93% at 5 years, 85% at 10 years, 79% at 15 years and 68% at 20 years.

In the past 15 years, it was shown that measuring the mortality alone, or even morbidity, is not sufficient for a chronic illness like SLE. The patients' perception of the impact of their disease and its treatment on their lives, also known as health-related quality of life, is regarded as an important outcome measure. Instruments for assessing this have been validated for Singapore.

SLE and reproduction

There is a slight risk of developing SLE from oral contraceptive (OC) use. In the Nurses' Health Study involving 121 645 women, past users of OC had a relative risk of 1.9 of developing SLE compared to those who never used them. The precise risk of using (OC) to SLE patients is unknown. It is safest for patients with nephritis to avoid OC for fear of provoking disease flare, and for those with anticardiolipin antibody (ACA) to do so because of the risk of thrombosis.

Pregnant SLE patients do flare more frequently compared to their non-pregnant selves and other non-pregnant patients. SLE patients also have a worse obstetric outcome than unaffected women. The Lupus Databank Research Program of Toronto reported that of 141 pregnancies in 73 patients, 60% resulted in live births, 23.8% in spontaneous abortions, 2.2% stillbirths and 14% therapeutic abortions. About a quarter of the live births were premature. Maternal hypertension predicted poor fetal outcome.

The best advice for SLE patients is to avoid OC and to choose barrier methods of contraception. They should plan to be pregnant when their disease is inactive. The pregnant SLE patient must be monitored closely by her physician and obstetrician for disease recrudescence and abnormal fetal development.

Investigations

Investigations are used to corroborate the clinical diagnosis of SLE, determine the extent of organ involvement and gauge disease activity.

The antinuclear antibody (ANA) is present in the serum of over 95% of lupus patients. However, 34% of 100 normal Singaporeans have ANA positivity at a titer of 1/40, 18% positivity at a titer of 1/80, 8% at 1/160, 3% at 1/320 and 1% at 1/640. This means that the positive predictive value of a positive ANA result for an unselected Singaporean, that is, the chance that the person truly has SLE, is less than 1%. The anti-double stranded DNA (anti-ds DNA) antibody is associated with renal involvement and it can be used to track disease activity in many patients. The anti-Sm antibody, though found in only 20–30% of SLE patients, is very specific (though not pathognomonic) for the disease.

Among 94 predominantly Chinese SLE patients from Singapore, 43% have anti-ds DNA antibody, 81% anti-histone antibody, 26% anti-Sm antibody, 32% anti-ribonucleoprotein (anti-RNP) antibody, 63% anti-Ro antibody, 12% anti-La antibody, 9% anti-SL/Ki antibody, 16% anti-ribosomal ribonuclear protein antibody, 5% anti-p70/p80 and 3% anti-proliferating-cell-nuclear-antigen antibody.

The history and clinical examination help us to determine the extent of organ involvement in SLE and guide further investigations. Since cytopenia is found in 70% and renal involvement in 30% of patients at presentation, full blood count, serum chemistries and urinalysis are essential for the evaluation for a newly-diagnosed lupus patient. For surveillance of the patient in the outpatient clinic, we regularly order blood count, serum creatinine and urinalysis tests.

Mild leucopenia and lymphocytopenia usually do not warrant further work-up, though significant thrombocytopenia mandates a bone marrow biopsy. Anemia is common and possible causes include autoimmune hemolysis, anemia of chronic illness and blood loss via the gastrointestinal tract.

Sterile pyuria, microscopic hematuria, cellular casts and persistent proteinuria suggest glomerulonephritis. In the presence of these abnormalities, we must determine the creatinine clearance and 24-hour urinary protein excretion. Kidney biopsy is often needed for prognostic purposes. It tells us the WHO histologic class, the amount of inflammation and of scarring (the activity and chronicity indices). Patients with class IV nephritis (diffuse proliferative) will develop renal failure without treatment. However, those with low activity and high chronicity indices are unlikely to respond to cytotoxic drugs and will eventually require renal substitution.

However, in the presence of active disease and rapidly deteriorating renal function, renal biopsy is not a prerequisite for instituting urgent treatment. We have to first rule out other possible causes of azotemia in SLE like urinary tract obstruction, nephrotoxic drugs (including NSAIDs, intravenous contrast media, cyclosporin A and aminoglycosides), infections (systemic and urinary), thrombotic thrombocytopenic purpura (TTP), disseminated intravascular coagulation (DIVC), intra-abdominal viscus perforation and renal vein thrombosis. Then, pulsed intravenous methylprednisolone followed by intravenous cyclophosphamide often controls the autoimmune assault on the kidneys. A brief period of dialysis support may be needed. After regaining control of the disease, the subsequent plan usually involves kidney biopsy.

Elevated anti-ds DNA antibody level, high erythrocyte sedimentation rate (ESR) and hypocomplementemia (usually C3 and C4) are reflective of active disease for most patients. However, serologic tests and disease activity are discordant in about 10% of lupus patients ("serologically active and clinically quiescent").

Diagnosis

The 1982 American Rheumatism Association (now the ACR) classification criteria of SLE and its 1994 revision have frequently been used to diagnose clinical cases of lupus. A patient is classified to have SLE if she satisfies four out of eleven criteria (malar rash, discoid rash, photosensitivity, oral ulcers, arthritis, serositis, renal disorder, neurologic disorder, hematologic disorder, immunologic disorder and antinuclear antibody) simultaneously or serially over any period of time.

We should be aware that this set of criteria was designed to ensure uniformity in case classification for studies, not for daily practice. Of

necessity, they tend to be relatively insensitive, so, milder SLE patients will be misclassified. Nevertheless, the criteria have found their way into clinical use as they are convenient to use and there are no available diagnostic criteria. We have to be aware of a few caveats. First, patients do not have to fulfill these criteria to be diagnosed to have SLE. For example, a 20-year old lady who has positive antinuclear antibody (ANA), malar rash and histologically-proven glomerulonephritis obviously has SLE, but does not fulfill the requisite 4 ARA criteria. Second, certain features strongly associated with SLE, like alopecia, consistent renal biopsy and hypocomplementemia are not classification criteria because they are not specific or because they are not sought for in all patients.

Treatment

The decision on when and how to treat a patient with SLE depends on our assessment of the extent, severity and rapidity of organ involvement. For example, minimal or no treatment is required for a patient with alopecia or mouth ulcers while intravenous corticosteroid and cyclophosphamide are needed to halt and reverse kidney failure due to active glomerulonephritis. See Table 1 for a checklist on the management of lupus.

Categorizing the disease manifestations as minor (mucocutaneous and musculoskeletal) or major and organ threatening (hematologic, renal, cerebral, pulmonary, gastrointestinal or cardiac) helps us to select the correct treatment.

Prednisolone at a dose of 0.5 mg/kg body weight daily is prescribed for minor manifestations while a dose of 1 mg/kg daily may be needed for major involvement. The high dose is maintained for not more than four weeks, after which it is tapered. Steroid-sparing drugs should have been started. Prednisolone in divided doses is more potent than when it is taken once a day. Some patients will flare when the prednisolone dose is reduced below a threshold level; they need a minimal intake of steroid daily.

Infection and lupus activity share many common manifestations, like fever, headache, meningism, pulmonary infiltrates, pleurisy, pleural effusion, pyuria, colitis, pericarditis and arthritis. Do not interpret these merely as disease flare until sepsis has been clearly ruled out. Even if the disease is evidently active, it does not mean that a concurrent infection is not present.

Table 1 Checklist for Managing SLE

Principles	Points to Cover	Comments
Confirm diagnosis	Consistent clinical picture Consistent laboratory investigations Consider differential diagnoses	Generally, the diagnosis is not difficult. Possible differential diagnoses include viral infection, drug-induced lupus and acute HIV infection.
Determine extent of organ involvement	History Clinical examination Investigations directed by clinical suspicion	An astute clinician will serve his or her SLE patients better than one who overly relies on investigations.
Determine disease activity	Distinguish reversible (activity) from irreversible (damage) factors	Accurately distinguishing activity from infections and irreversible organ damage is crucial, as treatment of each is different.
Choice of therapy	Corticosteroid Antimalarial drugs Cytotoxic drugs Intravenous immunoglobulin Plasmapheresis Experimental therapy	The treatment of active lupus is dictated by the extent, severity and rapidity of disease manifestation.
Collaboration with other health professionals	Medical Social Worker Occupational therapist Orthopedic surgeon Physiotherapist Rheumatology Nurse Support group volunteer	SLE is a disease that impacts every aspect of a patient's life. No one healthcare provider alone can do justice to the complexity of the total management.
Patient education	Diagnosis and relapsing/remitting nature of disease Drugs: benefits and adverse effects Need for compliance Alternative therapy Fertility and contraception Strategies to prevent flares Hospitalization Support group Coping with family and career Finances Insurance	Patients who understands their disease and actively participate in their own management has a better outcome than those who are apathetic.
Surveillance for disease recrudescence and development of new manifestations and co-morbidities	Physician's vigilance Investigations: activity markers, serum lipids, bone mass density, chest X-ray Patient education	There is no cure for lupus and every patient must form a long-term relationship with her or his physician.

Antimalarial drugs, like hydroxychloroquine (5–7 mg/kg body weight per day), are good for constitutional, mucocutaneous and musculoskeletal manifestations of SLE. NSAIDs may be needed if arthritis is a prominent feature. Photosensitive patients should be instructed to avoid sunlight (even that coming through a windowpane) and use a sunblock with a Sun Protection Factor (SPF) of at least 15.

Thrombocytopenia often respond to steroids alone. Otherwise, danazol or azathioprine may be used. Intravenous gammaglobulin (1 g/kg bodyweight daily for two days) must be used in cases of life-threatening thrombocytopenia to rapidly boost the platelet count. Splenectomy may have to be resorted to.

The value of cyclophosphamide and azathioprine in the treatment of lupus nephritis was established beyond doubt in 1986. Azathioprine can cause immune suppression and transaminitis while cyclophosphamide is associated with immune suppression, hemorrhagic cystitis, ovarian failure and malignancy. There is still no consensus on the optimal intensity and duration of therapy. In a patient with active proliferative lupus nephritis, it is reasonable to institute steroid therapy (prednisolone 1 mg/kg body weight daily) and monthly pulse cyclophosphamide (0.5 to 1.0 g/m^2 body surface area). After six months to a year, depending on the response, cyclophosphamide may be discontinued or replaced with the less toxic drug azathioprine. Good control of blood pressure is essential in retarding progressive loss of renal function. Mycophenolate mofetil and cyclosporin A are alternative treatment for lupus nephritis.

Urgent plasmapheresis is the best treatment for pulmonary hemorrhage while corticosteroid and cyclophosphamide are apparently less effective. Cyclophosphamide has also been used for neuropsychiatric, as well as aplastic anemia, thrombocytopenia, interstitial lung disease and pulmonary hemorrhage.

Very aggressive regimes have been used to treat severe lupus, including synchronized plasmapheresis and intravenous cyclophosphamide, stem cell transplant, and immunoablative doses of cyclophosphamide without stem cell rescue.

Sjogren's Syndrome (SS)

In SS, chronic lymphoctyic inflammation of exocrine glands leads to dry eyes and mouth. Fifty percent of patients have salivary gland enlargement.

About 60% of all patients will develop extra-glandular features, including arthritis, Raynaud's phenomenon, lymphadenopathy, lung involvement and vasculitis. SS can exist on its own (primary) or is secondary to another connective tissue disease, most commonly RA. SS patients have a risk of developing lymphoma, usually of low-grade B cell type.

There are diagnostic criteria for SS but none are universally accepted. Some are strict and insist on minor salivary gland biopsy, while others are lax, only requiring clinical findings of dry eyes and mouth and positive anti-Ro (found in 50–80% of SS patients) or anti-La antibodies (30–60%).

There are many causes of dry eyes and mouth (including drugs, aging, chronic hepatitis B or C) and SS should be diagnosed only with evidence of autoimmunity (high ESR and presence of autoantibodies) or lymphocytic infiltration of the salivary gland. There is a useful study on the long-term outcome of patients with sicca syndrome. One hundred and six patients were referred for dry eyes: primary keratoconjunctivitis sicca was diagnosed in 56, primary SS in 31 and secondary SS in 19. Three of the patients with primary SS died of lymphoma, compared to none in the other two groups.

Artificial tears and saliva supplements are very useful for symptomatic treatment of dry eyes and mouths. Oral pilocarpine 5 mg thrice daily increases tear and saliva production. Hydroxychloroquine may modulate some of the immunologic abnormalities and has beneficial effects on the sicca symptoms.

Dermatomyositis and Polymyositis

These are characterized by inflammation of striated muscles. Patient presents with muscle ache or weakness, most pronounced at the proximal muscle groups. The patient may notice that he has difficulty getting out of his car, combing hair or reaching things on the top shelf. The cutaneous features of DM are the heliotrope rash (a purplish rash around the eyelids), a photosensitive rash on the exposed skin, telangiectasia and Gottron's papules (pathognomonic papules found over the dorsal aspect of the interphalangeal joints).

The diagnosis is made on the basis of the proximal weakness, elevated levels of serum muscle enzymes (creatinine kinase, aspartatic transaminase, lactic dehydrogenase and aldolase), consistent electromyography and muscle biopsy.

After a thorough search in Singapore, 35 patients with PM and 40 with DM were found and analyzed. The combined incidence of PM/DM worked out to be 0.77 cases per 100 000 population per year. At presentation, 86.7% had proximal myopathy, 34.7% arthralgia/arthritis and 18.7% cutaneous vasculitis. DM was more likely to be associated with malignancy than PM in this series. In Singapore, we must exclude nasopharyngeal carcinoma in patients with DM or PM. Other malignancies associated with DM or PM arise from the lung, breast and gastrointestinal tract.

High-dose corticosteroid (prednisolone 1–2 mg per day) is prescribed as the initial treatment, and it is maintained for at least a month before the dose is tapered. Muscle enzymes are useful for determining the disease activity. Methotrexate, cyclophosphamide or azathioprine may be added if the response to steroid is unsatisfactory.

Scleroderma

Scleroderma, a rare disease, is characterized by fibrosis of the skin and internal organs. It is observed that there is a resemblance between the skin of scleroderma patients and that of those with chronic graft-versus-host disease. The finding of fetal cells in the skin of parous scleroderma patients suggests a target for immune attack.

Scleroderma has a diffuse form (systemic sclerosis or SSc) and a limited form (CREST). CREST is an acronym for the clinical features of calcinosis, Raynaud's phenomenon, esophageal involvement, sclerodactyly and telangiectasia. PSS is associated with the anti-Scl 70 antibody and CREST with anti-centromere antibody.

Patients often present with Raynaud's phenomenon, which is a three-color change (white, blue and red) of the digits in response to cold, as well as finger puffiness and arthritis. Convincing sclerodactyly may take some time to appear. Other skin manifestations are finger pulp resorption and telangiectasia. Patients may also have dysphagia, reflux esophagitis, malabsorption, pulmonary hypertension or systemic hypertension. At present, patients rarely die from renal crisis because of effective antihypertensive drugs (with the credit going to the ACE inhibitors) and availability of dialysis. Interstitial lung disease and pulmonary hypertension are now the main causes of mortality.

Treatment is difficult. Patients with Raynaud's phenomenon should be instructed to keep their extremities warm. Oral iloprost has

been shown to help. Hypertension, malabsorption and esophagitis are treated in the usual way. D-penicillamine, colchicine, methotrexate, cyclosporin A and relaxin (a connective tissue remodeling peptide) have been tried in SSc with minimal or no benefit.

Mixed Connective Tissue Disease

MCTD is a condition characterized by the presence of overlapping features of two or more CTDs (SLE, DM, PM and SSc) together with high titers of the anti-RNP antibody. It was reported to have a benign course and to respond well to steroids, but a longer period of follow-up proved the contrary.

The right of this syndrome to be regarded as an independent entity has been challenged ever since it was described. Detractors maintain that patients initially labeled as MCTD eventually evolve to have PSS or SLE. They also point out that the definition of MCTD has changed since its original description. Proponents argue that some patients do remain in the MCTD category since diagnosis and that there is a distinct antibody, the anti-RNP, to justify the distinct status of the condition.

There is no special treatment for this condition; it depends on the disease manifestation. Long-term follow-up is required to determine if the disease evolves to one of the recognized CTD.

Antiphospholipid Syndrome (APS)

APS is diagnosed when a patient has recurrent abortions, thrombosis or thrombocytopenia and antiphospholipid antibodies (demonstrated by the presence of lupus anticoagulant or anti-cardiolipin antibodies). The term "antiphospholipid antibodies" is a misnomer as these antibodies are actually directed against phospholipid-binding plasma proteins, one of which is beta 2-glycoprotein I.

APS can exist in isolation or together with one of the CTDs, usually SLE. Even though the partial thromboplastin time (PTT) is prolonged, the main problem is thrombosis. Thrombosis can occur in the arterial or venous system, in vessels of any calibre. Some patients may manifest clinical features not associated with coagulation like livedo reticularis, migraine and cardiac valvular vegetations.

The treatment for thrombosis is life-long anticoagulation, aiming for an international normalized ratio (INR) above 3. Patients with repeated abortions may be able to carry a pregnancy to term with aspirin and heparin (either regular or low-molecular weight).

SYSTEMIC VASCULITIDES

The vasculitides can be primary, or secondary to other diseases like malignancies and drug reactions. We can further classify the primary vasculitides according to the size of blood vessels they involve: large (aorta and its main branches), medium-sized (arteries) and small (arterioles, capillaries, venules). Primary systemic vasculitides are uncommon diseases, with a prevalence of 144.5 per million population in UK. We summarized useful information on the vasculitides in Table 2.

The ACR classification criteria for the main primary vasculitides became available in 1990. In 1994, the participants of Chapel Hill Conference published their definitions of the systemic vasculitides and proposed a uniform set of nomenclature. The most significant proposal of the Conference is that microscopic polyangiitis (MPA) (which affects medium-sized vessels and those smaller) should be distinguished from classic polyanteritis nodosa (PAN) (which involve the medium-sized vessels exclusively).

Takayasu's arteritis and temporal arteritis share the same abbreviation and involve the large branches of the aorta. Takayasu's arteritis (TA) presents in people below age 50, while temporal arteritis, also known as giant cell arteritis (GCA) is found in those above that age. GCA is found mainly in Caucasians while TA is seen in Japanese and Orientals.

Classic PAN often presents with constitutional symptoms like malaise, fever and weight loss. Testicular pain can sometimes be the presenting sign. Arthralgia and skin rash and ulceration are common. Classic PAN should spare the kidneys. Hepatitis B surface antigen is found in about 5% of patients with PAN, depending on its endemicity in that population. The clinical features of MPA are similar to those of PAN but renal disease is found in the majority of patients with MPA and 30% may present with oliguria.

One of the most recognizable forms of systemic vasculitis is Wegener's granulomatosis (WG). It is characterized by granulomatous

Table 2 The Primary Vasculitides at a Glance

	Size of Vessel Involved	Age, Sex and Racial Predilections	Classification Criteria
Large Vessels			
Takayasu's arteritis	Large branches of the aorta	< 40 years, Japanese and Orientals	3 out of 6 criteria: Age at disease onset < 40 years; claudication of extremities; decreased brachial artery pulse; BP difference > 10 mmHg; bruit over subclavian arteries or aorta; and arteriogram abnormality.
Giant cell arteritis	Large branches of the aorta like the carotids, temporal arteries	Older patients (> 50 years), usually Caucasians	3 out of 5 criteria: Age at disease onset > 50 years; new headache; temporal artery abnormality; elevated erythrocyte sedimentation rate; and abnormal artery biopsy.
Medium-Sized Vessels			
Polyarteritis nodosa	Medium-sized	Middled-age patients, male : female ratio 2 : 1	3 out of 10 criteria: Weight loss > 4 kg; livedo reticularis; testicular pain or tenderness; myalgias, weakness or leg tenderness; mono- or polyneuropathy; diastolic BP > 90 mmHg; elevated BUN or creatinine; hepatitis B virus; arteriographic abnormality; and biopsy of small or medium-sized artery containing PMN.
Small Vessels			
Wegener's granulomatosis	Small arteries and veins	Young or middle-aged patients, slight male preponderance. Mostly Caucasian patients in	2 out of 4 criteria: Nasal or oral inflammation; abnormal chest radiograph; urinary sediment; and granulomatous inflammation on biopsy.

Table 2 (*Continued*)

	Size of Vessel Involved	Age, Sex and Racial Predilections	Classification Criteria
		literature, but found in all races.	
Hypersensitivity vasculitis	Small	No special predilection	3 out of 5 criteria: Age at onset > 16 years; medication at disease onset; palpable purpura; maculopapular rash; and biopsy including arteriole and venule.
Churg–Strauss syndrome	Medium-sized to small	Middle-aged patients, slight male preponderance. Possible association with the use of leukotriene-receptor antagonists, especially zafirlukast.	4 out of 6 criteria: Asthma; eosinophilia > 10%; neuropathy, mono or poly; pulmonary infiltrates, non-fixed; paranasal sinus abnormality; and extravascular eosinophils.
All size vessels			
Behcet's disease	All sizes	Female preponderance, Middle-eastern and Japanese	Recurrence of aphthous ulceration at least 3 times during a 12-month period, plus 2 of the following: Recurrent genital aphthous ulceration; eye lesions; skin lesions; positive pathergy test result.

inflammation of the small vessel, including arterioles, capillaries and venules. Upper airway (sinusitis), lung (nodules and pulmonary hemorrhage) and kidney involvement (glomerulonephritis) is prominent. As many as 90% of the patients systemic WG will test positive for anti-neutrophilic cytoplasmic antibody (ANCA). Limited WG is anatomically restricted to airway involvement and its association with ANCA is weaker, around 70%.

It may be very difficult to distinguish WG from MPA as they have small vessel vasculitis in the lungs and kidneys and ANCA is found in both; the demonstration of granulomata in WG may allow its differentiation from MPA.

Characteristic of Churg-Strauss syndrome (CSS) is granulomatous inflammation of the small vessels with eosinophilic infiltration, together with asthma and eosinophilia. Other manifestations are fever, weight loss, hypertension, peripheral neuropathy and heart failure. Renal involvement is common, and biopsy reveals a focal segmental necrotizing glomerulonephritis, often with crescents. Asthma and eosinophilia may precede the development of vasculitis for years.

Behcet's disease manifests as recurrent mouth ulcers (in at least 97% of patients), genital ulcers and uveitis. Less frequent findings are cutaneous signs (acne-like, follicular or papular), erythema nodosum, epididymitis, CNS lesions or deep vein thrombosis (associated with higher Lp(a) lipoprotein levels). In spite of the fact that vessels of any size can be affected in Behcet's disease, there is predilection for certain organ involvment: patients may lose their sight but rarely develop renal failure. An international study group developed a unified set of diagnostic criteria in 1990 that has become widely accepted.

Eighteen new patients with primary systemic vasculitis presented to our Department between January 1994 and December 1997. There were 8 male and 10 female, their ages ranging from 17 to 81 years. Fifteen were Chinese, two Malay and one Indian. According to ACR criteria, there were two cases of CSS, two of PAN, two of Wegener's granulomatosis (WG) but eleven were unclassifiable (Lau TC, unpublished data).

Secondary Vasculitides

Vasculitis is a feature of autoimmune diseases like SLE, rheumatoid arthritis and SS. Cryoglobulinemia can manifest vasculitic lesions as well.

There are three forms of cryoglobulinemia: Type I is typified by the presence of monoclonal cryoglobulin, often due to a lymphoid malignancy; Type III is a polyclonal cryoglobulinemia; while Type II contains a mixture of the two. The vast majority of Types II and III are related to hepatitis C infection.

The use of horse-derived antithymocyte globulin, interferon-alpha, granulocyte colony-stimulating factor and streptokinase, anticonvulsants, sulphonamides and allopurinol have been associated with vasculitis. Drug-induced vasculitis is generally limited to the skin, but glomerulonephritis, interstitial nephritis, liver granuloma and necrosis and even lung, cardiac and central nervous system involvement have been reported. A common histological finding is leucocytoclastic vasculitis (also known as hypersensitivity vasculitis). This appears as a rash on the arms and legs and has a good prognosis. It can also be idiopathic or secondary to viral infections.

Pathogenesis

The pathogenesis is unknown but the classic explanation is the deposition of immune complexes below the endothelium, leading to infiltration of leucocytes (chiefly neutrophils), medial necrosis and even inflammation of the surrounding tissues. The endothelium itself is also involved. When it is activated, it can upregulate the production of cytokines and adhesion molecules and attract and activate leucocytes; this is believed to occur at vascular branches in large vessel vasculitis. ANCA and neutrophils are postulated to be involved in the pathogenesis of small-vessel vasculitis.

Differential Diagnosis

Infection, coagulopathy, intravascular lymphomas, atrial myxoma, Buerger's disease and ergot overuse can also masquerade as vasculitides. Cholesterol embolism is a particularly good mimic because the patient can present with acute renal failure and may have a typical skin rash. In addition, an elevated ESR and positive ANA and RF may be found. It is usually but not necessarily preceded by an arteriographic procedure in the recent or distant past. Cholesterol crystals may be demonstrated in the skin or kidneys. Sometimes, these may be seen in the retinal arteries on fundoscopic examination.

Investigations

Biopsy is the definitive way to confirm a diagnosis of vasculitis. The skin is a favorite site because of its accessibility, the ease of distinguishing lesional and normal areas and procedural safety. If the skin is not affected, muscle, nerve, kidneys or testes could be chosen. The biopsy is more likely to be contributory if it is taken from an affected organ than a blind site. While biopsy proves that vasculitis is present, it is the accompanying clinical manifestations that determine the type of vasculitis. In general, biopsies are preferred over angiograms, unless it is dangerous (e.g. involvement of main branches of the aorta) or impractical (e.g. suspected mesenteric vasculitis in a stable patient) to obtain them.

ANCA is present in patients with WG (90% of patients), CSS (48%) MPA (75%). Neutrophils can be activated to release leukotrienes, superoxide and elastase when incubated with ANCA, so the antibody itself is likely to be pathogenic.

The discovery of ANCA has fundamental significance for the understanding and management of the vasculitides. First, it is useful in the diagnosis of WG. Second, it may help to predict a disease flare, as 24–50% of WG patients demonstrate a rise in serum ANCA before their disease becomes active. Third, the true importance of ANCA lies in its ability to unite conceptually the three small vessel vasculitides (WG, CSS ad MPA), which led the Chapel Hill Consensus to propose that MPA be distinguished from classic PAN (which is not associated with ANCA).

ANCA that exhibits a cytoplasmic immunofluorescence pattern in alcohol-fixed neutrophil preparations is called cANCA. The main antigen target for cANCA is serine protease 3 (PR3) in 90% of cases. The other main pattern, perinuclear immunofluorescence, is produced by pANCA, which is directed mainly against myeloperoxidase (MPO) in 50% of cases but also against elastase and lactoferrin. During ethanol fixation, MPO, being cationic, translocates from the cytoplasm to the negatively-charged nuclear membrane. Thus, the perinuclear pattern is an artifact of alcohol fixation, and the same antibodies should produce cytoplasmic staining in paraformaldehyde-fixed slides.

Clinical suspicion is very important for early detection of increasing disease activity. Constitutional symptoms such as fever and weight loss suggest disease recrudescence. Signs of new organ dysfunction may develop, for example, rising serum creatinine concentration, hemoptysis,

breathlessness, digital or limb ischemia or a new vascular bruit. In active disease, the hemoglobin and albumin levels fall, the ESR and C-reactive protein level rise, and thrombocytosis and leucocytosis become prominent. If a decision about disease activity still cannot be made after carefully examining the above clinical information, a repeat a biopsy or angiogram must be resorted to even though these are invasive procedures.

Treatment

The goals of treatment of vasculitis are the induction of disease remission, maintenance of remission and close surveillance for relapse, all the while mindful of the adverse effects of drugs.

In 1973, Fauci and Wolff reported that oral steroid and cyclophosphamide dramatically altered the outcome of Wegener's granulomatosis, from an almost uniformly fatal disease to one with a good chance of survival (80% at five years). This provided a paradigm for the treatment of the systemic vasculitides: corticosteroids and cytotoxic drugs used in combination to induce remission, followed by dose tapering once control is obtained.

There are two problems with this regime: the high frequency of adverse events, and inability to keep the disease in remission indefinitely. Nowadays, it is also felt to be too inflexible as patients express a continuum of disease severity and activity.

The adverse effects of corticosteroid are well known, and may be broadly described as its effects on the immune system (susceptibility to infections), skin, endocrine system, musculoskeletal (avascular necrosis, osteoporosis), ophthalmic (cataract and glaucoma), neuropsychiatric, gastrointestinal and cardiovascular. Oral cyclophosphamide at a dose of 12 mg per kg daily results in a higher cumulative dose than monthly intravenous pulses of 0.5 g per square metre body surface area.

Methotrexate is increasingly used in the treatment of vasculitides like Takayasu's arteritis, WG and GCA. Side effects of methotrexate are bone marrow suppression, pneumonitis, stomatitis, infections, transaminitis, liver fibrosis and cirrhosis, and, rarely, lymphoma.

Disease flares of WG may be attributable to upper respiratory tract infections, particularly due to *Staphylococcus aureus*. Accordingly, cotrimoxazole has been prescribed to maintain remission and it has shown efficacy, though it is not known if this is due to its antibiotic or

immuno-modulatory (specifically antineutrophilic) effect. Other drugs and therapeutic modalities that have been used in small series or isolated reports are azathioprine, intravenous gamma globulin (IVIG), cyclosporin A, pentoxyfylline and plasmapheresis. Antivirals, principally interferon alpha, are used to treat hepatitis B and C virus-associated vasculitides.

FURTHER READING

Gaffney PM, Moser KL, Graham RR, Behrens TW, Recent advances in the genetics of systemic lupus erythematosus, *Rheum Dis Clin North Am* **28**:111–126, 2002.

Robson MG, Walport MJ, Pathogenesis of systemic lupus erythematosus (SLE), *Clin Exp Immunol* **31**:678–685, 2001.

Wallace DJ, Management of lupus erythematosus: recent insights, *Curr Opin Rheumatol* **14**:212–219, 2002.

Zimmerman R, Radhakrishnan J, Valeri A, Appel G, Advances in the treatment of lupus nephritis, *Annu Rev Med* **52**:63–78, 2001.

Hunder GG, Arend WP, Bloch DA, Calabrese LH, Fauci AS, Fries JF, Leavitt RY, Lie JT, Lightfoot RW Jr, Masi AT, *et al*., The American College of Rheumatology 1990 criteria for the classification of vasculitis. Introduction, *Arthritis Rheum* **33**:1065–1067, 1990.

This page intentionally left blank

84

Rheumatoid Arthritis

Humeira Badsha

Rheumatoid arthritis (RA) is a systemic, autoimmune disorder of unknown etiology. It is characterized by a chronic, inflammatory, symmetrical arthritis, as well as systemic features.

EPIDEMIOLOGY AND PATHOGENESIS OF RHEUMATOID ARTHRITIS (RA)

The estimates of the prevalence of RA are based on the 1987 American College of Rheumatology Revised Criteria for the Classification of Rheumatoid Arthritis (see Table 1).[1] Based on these criteria between 1–2% of the world population is affected with RA. However the prevalence in Asians is only 0.3%.[2] Women are affected two to three times more commonly than men.[3] Pregnancy appears to improve RA but the immediate postpartum period and breastfeeding worsen it.[4]

There is an association between Rheumatoid Arthritis and HLA-DR4.[5] Sub-typing of DR in Singaporean Chinese revealed that the association was DRB1*0405.[6] In Caucasians the DR association is DRB1*0401 and *0404.

Table 1 1987 American College of Rheumatology Criteria for the Classification of Rheumatoid Arthritis

1) Morning stiffness in and around the joints lasting at least 1 hour before maximal improvement.
2) At least 3 joint areas simultaneously have had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas are right or left PIP, MCP, wrist, elbow, knee, ankle, and MTP joints.
3) At least 1 area swollen (as defined above) in a wrist, MCP, or PIP joint.
4) Simultaneous involvement of the same joint areas (as defined in 2) on both sides of the body (bilateral involvement of PIPs, MCPs, or MTPs is acceptable without absolute symmetry).
5) Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxta-articular regions, observed by a physician.
6) Demonstration of abnormal amounts of serum rheumatoid factor by any method for which the result has been positive in less than 5% of normal control subjects.
7) Radiographic changes typical of rheumatoid arthritis on posterioanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized in or most marked adjacent to the involved joints (osteoarthritis changes alone do not qualify)
For the diagnosis of Rheumatoid Arthritis, 4 of the 7 criteria are required.

MCP = metacarpophalangeal; PIP = proximal interphalangeal;
MTP = metatarsophalangeal

The etiology of RA[7] appears to focus on the relationships between infectious agents, genetics and autoimmunity. Heat Shock Proteins (HSP) are major bacterial antigens which could bind to HLA-DR4, triggering the proliferation of antigen-specific T lymphocytes. One hypothesis for abnormalities in RA is that an immune response to a HSP is amplified and perpetuated because of molecular mimicry between HSP and HLA-DR4 susceptibility sequences. Several viruses have been proposed to predispose to RA (Ebstein–Barr virus, parvovirus, retroviruses), but conclusive evidence is lacking. There are reports of increased Proteus antibody titers in RA.

There is some evidence for genetic predisposition to RA (concordance in monozygotic twins 30% and in dizygotic twins is 3%) but it is felt that this predisposition accounts for only 20–30% of the risk of developing RA. There are also hypotheses that X or Y linked factors may account for the female preponderance amongst RA patients.

There is a strong role for CD4+ T cells, tumor necrosis factor alpha (TNF-α) and interleukin-1 (IL1) in the initiation and maintenance of inflammation in RA.

In the rheumatoid joint initially, there is synovial endothelial cell damage with exudation of an inflammatory cellular infiltrate. This leads to synovial tissue swelling and effusion in the joint. As the disease progresses the synovium becomes massively hypertrophic and edematous. In the hyperplastic sublining synovial stromal connective tissues there are highly activated and invasive fibroblast-like cells and new blood vessels. This highly proliferative and invasive connective tissue stroma called *pannus* invades and destroys periarticular cartilage and bone leading to bony erosions, and ultimately to deformities and cartilage loss.

CLINICAL FEATURES

There are *racial variations* in the expression of rheumatoid arthritis. K. Veerapen *et al*.[8] compared 70 consecutive patients with rheumatoid arthritis in a Malaysian hospital with a similar number of age, sex and duration matched hospital patients in the UK. The rheumatoid factor (latex) positivity was 65% in both groups of patients. Disease severity was also similar in both groups. The Malaysian patients tended to have more wrist involvement and less forefoot deformities than their caucasian counterparts. Rheumatoid nodules, vasculitis and other systemic features were also less common in the Malaysian patients. There was no difference in the clinical features of the 23 Indian, 4 Malay, and 42 Chinese patients in the Malaysian group. Studies of southern Chinese patients have also found extraarticular manifestations to be less common.[9,10]

The most common *mode of onset* of rheumatoid arthritis is the insidious development of symptoms over a period of several weeks.[11] Patients can also present with explosive, polyarticular, or a monoarticular mode of onset. Isolated less severe attacks of monoarticular synovitis lasting few days, resolving spontaneously with periods of spontaneous remission lasting few months can occur. These attacks are called palindromic rheumatism and can evolve into classical RA.

Articular manifestations of RA can be due to inflammatory synovitis (Fig. 1) with reversible signs and symptoms, or structural damage with irreversible deformities. It is important to differentiate between active

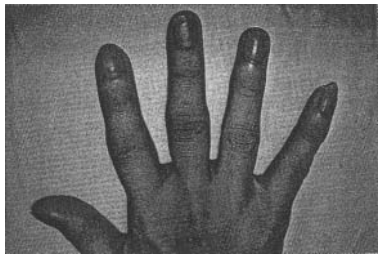


Fig. 1 Early rheumatoid arthritis [characterized by the swelling of the interphalangeal joints (PIP)].

synovitis and structural damage as the management of the two entities is different. Morning stiffness of greater than one hour or warm, swollen joints are usually indicative of active synovitis. Bony deformities and bone on bone crepitus indicate structural damage.

It is important to exclude other conditions causing synovitis, such as Systemic Lupus Erythematosus (SLE), Seronegative Spondyloarthropathy, and Gout. The characteristic clinical features, usual lack of extraarticular manifestations, and laboratory profile are helpful in distinguishing RA from these conditions. Septic arthritis should be excluded if there is a single swollen, inflamed joint in a patient with established RA. Other causes of pain such as osteoarthritis, fibromyalgia, amyloidosis, and angioimmunoblastic lymphadenopathy should be excluded. Factors typical of RA include: pain and stiffness in multiple joints; boggy joint swelling that includes soft tissue and synovial fluid; joints that are tender to touch; symmetrical joint involvement, often of the hands and feet; sparing of the DIP (distal interphalangeal) joints; decreased, painful range of motion of the joints; subcutaneous nodules, low grade fever, small enlarged lymph nodes; raised peripheral white blood cell counts, platelets and ESR (erythrocyte sedimentation rate); anemia and a positive rheumatoid factor. The diagnosis of RA is made by estabilished criteria of which the requirement for the presence of objective evidence of synovitis for 6 weeks is important. The diagnosis of RA should be confirmed within two weeks of onset of symptoms.

Involvement of the cervical spine is common,[12] manifesting as pain and loss of motion. There are several types of cervical spine instability: atlanto-axial subluxation (50–70% of patients), subaxial subluxation (20–25% of patients) and basilar invagination with or without atlantoaxial subluxation (20% of patients). Evaluation of the patient with RA

should always include a careful neurological examination including screening for symptoms such as neck pain, occipital headache, and paraesthesias in the extremities. The clinician should also seek symptoms of vertebrobasilar insufficiency (tinnitus, vertigo, visual disturbances, and dysphagia) due to basilar invagination of the cord. Care should be taken prior to extension of neck during endotracheal intubation to exclude cord compression or atlanto-axial subluxation by physical examination, as well as flexion-extension X-rays of the c-spine. A study of patients with RA[12] concluded that operative stabilization of the cervical spine should be done to minimize the risk of irreversible paralysis, regardless of whether neurological signs or symptoms are present, in patients who have atlantoaxial subluxation and a posterior atlanto-odontoid interval of 14 mm or less, patients who have atlanto-axial subluxation and at least 5 mm of basilar invagination, and patients who have subaxial subluxation and a sagittal diameter of the spinal canal of 14 mm or less.

In virtually all patients with RA, the *wrists and hands* are involved but the distal interphalangeal (DIP) joints are usually spared. Ulnar deviation of the fingers at the metacarpophalangeal (MCP) joints is associated with radial deviation at the wrists. Swan neck deformities (Fig. 2) due to contracture of the interosseous and flexor muscles and tendons, can develop with a hyperextension at the proximal interphalangeal (PIP) joints and flexion at the DIP joints. Boutonniere's deformity (Fig. 3) refers to flexion at the PIP joints and hyperextension at the DIP joints. In addition there may be symptoms of carpal tunnel syndrome due to compression of the median nerve at the wrist by inflammatory synovitis. Inflammatory tenosynovitis may lead to tendon rupture characteristically affecting the

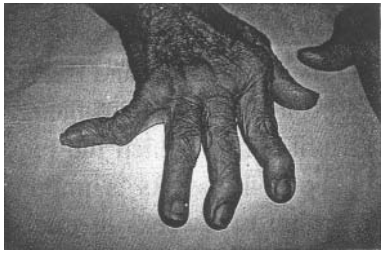


Fig. 2 Swan neck deformities [of the 2nd and 3rd fingers characterised by hyperextension at the proximal interphalangeal joints (PIP) and flexion at the distal interphalangeal joints (DIP)].

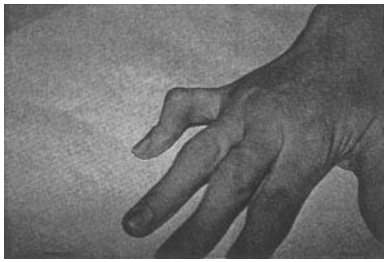


Fig. 3 Boutonierre deformity (5th finger — flexion at proximal interphalangeal joint and hyperextension at distal interphalangeal joint).

extensor tendons of the 3rd, 4th, and 5th fingers. Hip joint involvement is less common in Asian populations with RA.

Knee effusions can be easily detected and aspirated for further evaluation. Posterior herniation of the capsule into the popliteal area can create a cystic structure called the *Baker's cyst* or Popliteal cyst. This cyst can rupture into the calf mimicking a deep vein thrombosis, but can usually be differentiated by ultrasonography.

Metatarsophalangeal (MTP) arthritis can lead to characteristic cock-up deformities and subluxation of the MTP heads on the sole.

Extraarticular manifestations such as fever and fatigue are common in patients with RA during the active stage of the disease.

Pleuropulmonary manifestations are common in patients with RA.[13] Pleural effusions in RA occur at a prevalence of about 5%. Interstitial pulmonary fibrosis occurs in 20–40% of patients with RA, more commonly men and patients with rheumatoid nodules and those seropositive for rheumatoid factor. Early interstitial fibrosis may be manifested as ground-glass opacities not visible on plain chest radiographs, but diagnosed on high resolution CT scan of the lungs. Open lung biopsies may reveal one of five histological patterns in patients with RA: pulmonary rheumatoid nodules, usual interstitial pneumonitis (UIP), bronchiolitis obliterans with organising pneumonia (BOOP), lymphoid hyperplasia and cellular interstitial infiltrates.

Cardiac involvement is rare in RA although asymptomatic pericardial effusions may be present in a large proportion of patients.

Felty's syndrome[14] occurs in about 1% of patients with RA and consists of RA, leucopenia, and splenomegaly. It is also associated with rheumatoid nodules, fever, weight loss, sjogren's syndrome, leg ulcers, vasculits, peripheral neuropathy, hepatomegaly, and pulmonary fibrosis.

Pseudo-felty's syndrome[14] occurs in RA patients with neutropenia but a normal total white count due to an increased number of large granular lymphocytes (LGL). These patients have less extraarticular manifestations and may have a chronic lymphoproliferative disease. Patients with RA have an increased risk of developing lymphomas.[15] An anemia of chronic disease is almost universal in patients with RA. An iron deficiency anemia secondary to chronic gastrointestinal blood loss may have to be excluded as many patients with RA are on long-term NSAID (Non Steroidal Anti-inflammatory Drug) treatment.

Rheumatoid nodules occur in 15–40% of patients and are associated with a positive rheumatoid factor, erosive disease, vasculitis and other systemic features. These rheumatoid nodules consist of a central area of necrosis surrounded by palisading histiocytes encircled by granulomatous tissue infiltrated with lymphocytes. The nodules usually occur in extensor surfaces and also in viscera such as the lungs and heart. Methotrexate treatment can accelerate nodulosis.[16]

Cutaneous vasculitis as well as systemic vasculitis can develop. Splinter hemorrhages and necrotic areas at the finger tips may be caused by vasculitis. The presense of such lesions means that the patient should be meticulously examined to exclude systemic vasculitis.

Keratoconjunctivitis sicca manifested by dryness of the eyes and mouth, may occur as part of a secondary Sjogren's syndrome. Episcleritis can occur but usually runs a self-limited course. Scleral inflammation which may resemble a rheumatoid nodule can erode through the sclera into the choroid giving rise to *scleromalacia perforans*.

LABORATORY EVALUATION

It should be emphasized that the diagnosis of rheumatoid arthritis is based on a good history and physical examination. The presence of rheumatoid factors can aid in the diagnosis but their absence does not exclude a diagnosis of rheumatoid arthritis. Rheumatoid factors are usually IgM directed against IgG. They are detected in 65–85% of patients with RA. They are also found in about 3% of healthy people as well as in conditions such as subacute bacterial endocarditis, tuberculosis, leprosy, syphilis, cytomegalovirus, rubella, influenza, parasitic diseases, sarcoid, pulmonary interstitial lung disease, and cryoglobulinemia. A raised ESR (Erythrocyte Sedimentation Rate) and CRP (C-reactive protein) may

indicate active disease. Other laboratory abnormalities observed in RA include hypergammaglobulinemia, anemia, occasional hypocomplementemia, thrombocytosis, and eosinophilia. All of these tend to occur more often in patients with more severe disease.

In the initial stages, joint radiographs may demonstrate a peri-articular osteopenia and the characteristic marginal erosions of RA may not be seen till later in the course of the disease. Synovial fluid aspirate and biopsy usually demonstrates inflammation, and while not useful to diagnose RA, may help to exclude joint infection. Joint fluid from a rheumatoid joint may be straw colored, cloudy and contain flecks of fibrin. There are large numbers of white cells with a preponderance of neutrophils in the fluid. Synovial glucose level is usually low.

MANAGEMENT

Rationale for Early Aggressive Treatment and Prognostic Factors

RA is a potentially serious chronic disease with excessive mortality and disability. A 25-year prospective study showed that the median life expectancy was shortened by 7 years in male, and 3 years in female patients with RA.[17] In reviewing major studies after 1950, the risk of death in RA patients was found to be twice as high as that of the general population.[18] In addition, disability increased from 52% of patients with RA to 84% at the end of 12 years.[19] A large portion of the damage appears to be early, as half of RA patients experience some difficulty in the performance of activities of daily living, within two years of disease onset.[20] Hence recent consensus favors early aggressive management of RA patients, especially in those with poor prognostic factors (see Table 2).[21] Recent years have seen a proliferation of new treatments for RA.

General Approach to the Treatment of RA Patients

It is important to have a well-balanced and coordinated treatment scheme involving education, physiotherapy and occupational therapy. The usual approach to drug treatment should be to start with the Non Steroidal Anti-inflammatory Drugs (NSAID's) and add a Disease Modifying Anti-Rheumatic Drug (DMARD) early in the course of disease. The choice of

Table 2 Indicators of Poor Prognosis in RA

1)	More ACR criteria fulfilled
2)	Sustained disease activity
3)	Rheumatoid factor positive
4)	Extra-articular manifestations such as rheumatoid nodules and systemic features
5)	Heterozygosity for HLA-DRB1.0401 (DW4) and DRB1.0404 (DW14)
6)	Raised ESR and CRP
7)	Early erosions and periarticular bone loss
8)	Low socio-economic status and low formal education
9)	Female sex
10)	Insidious onset and severe initial presentation

ESR: Erythrocyte sedimentation rate; CRP: C-reactive protein; ACR: American College of Rheumatology; HLA: Human leucocyte antigen

DMARD would depend on patient and disease characteristic. NSAID's should be avoided in the elderly or in those with contra-indications to its use. Low-dose steroids are an alternative in this situation. Patients need to be monitored regularly for disease activity parameters as well as drug toxicities.

Patients should be educated as they need to understand that RA is a lifelong disease, requiring some form of treatment. Several patients seek alternative medicine looking for a cure and may fail to continue taking DMARD's or other prescribed medications. Doctors, nurse counsellors and social workers can enforce compliance to treatment as well as help disabled patients gain access to government or other sources of financial assistance. Physiotherapists can help patients understand how whole body rest can decrease the systemic inflammatory response. The level of physical activity should be increased judiciously as joint inflammation subsides. Physiotherapy and occupational therapy help patients maintain joint mobility and overcome existing disabilities.

Drug Treatment

NSAIDs (Non steroidal anti-inflammatory drugs)

The anti-inflammatory action of NSAID's is due to its actions inhibiting cyclooxygenase and decreasing prostaglandins of the E series. By shunting arachidonic acid back into the triglyceride pool it also decreases its conversion to leukotrienes.

NSAID associated toxicities

1) Gastrointestinal: Gastric and duodenal ulcers, with hemorrhage and perforation, can occur. Other problems encountered are gastritis, gastric erosions, small bowel ulcers, oesophagitis and oesophageal stricture. The risks of gastrointestinal complications are higher in older patients, and those with a history of peptic ulcer disease and other comorbid conditions. When an NSAID-associated ulcer occurs, the drug should be stopped if possible. Omeprazole was found to be more effective than ranitidine[23] and as effective but better tolerated than the prostaglandin analogue misoprostol[24] in the treatment of NSAID induced ulcers.

   Prophylaxis against NSAID-induced gastric or duodenal ulcers should be considered in patients with risk factors mentioned above. Misoprostol in doses of 100–200 μg four times a day have been effective in prevention of gastric and duodenal ulcers.[25] It is poorly tolerated, causing nausea, diarrhea, and abdominal cramps. Omeprazole is effective for the prevention of both gastric and duodenal ulcers whereas ranitidine only prevents duodenal and not gastric ulcers.[26]
2) Renal toxicitiy: NSAIDs decrease creatinine clearance and increase serum creatinine levels, hence these drugs should not be used in the elderly and in those with impaired renal function or decreased circulating blood volume. Acute interstitial nephritis and rarely nephrotic syndrome can also occur with NSAID use.
3) Other toxicities: Reversible elevation of liver enzymes and rarely fatal fulminant hepatitis can also occur. NSAIDs displace Warfarin from it's binding sites and can prolong the prothrombin time. They also interact with phenytoin and other highly protein-bound drugs. NSAIDs can interfere with platelet function and cause prolonged bleeding.

It is recommended that patients on long term NSAIDs have a full blood count and creatinine level checked every six months.

Selective cyclooxygenase 2 inhibitor (COX2)

There are at least two related isoforms of cyclooxygenase — 1 and 2 (COX1 and COX2). The toxicity of NSAID is due to the inhibition of COX1 and the therapeutic effects are due to COX2 inhibition. Selective COX2 inhibitors have been developed including celocoxib (celebrex) in the dosage of 200 mg once daily or 100 mg twice daily and rofecoxib 25 mg

daily. In two recent large trials comparing highly selective COX2 agents with traditional NSAIDs, the patients in the selective COX2 agent group had significantly fewer GI events.[27,28] There are several caveats, however. If antiplatelet therapy is indicated (e.g. as risk reduction for cardiovascular disease), an agent such as low-dose aspirin should be used because, unlike non-selective NSAIDs, the selective COX2 inhibitors have no effect on platelet adhesion or aggregation. The same precautions that apply to the use of NSAIDs in patients with renal disease should be exercised with the use of COX2 inhibitors.

DMARDs (Disease modifying anti-rheumatic drugs)

The doses and efficacies of DMARDs (Table 3) reported below are based on experiences with caucasian populations.[22,29] In a study of Singaporean Chinese, low-dose Methotrexate was found to be effective and safe.[30] Based on considerations of safety, convenience, and cost, many rheumatologists select HCQ or SSZ first, but for the patient with very active disease or with indicators of a poorer prognosis, MTX or combination therapy would be preferred. If a patient with RA has not achieved remission or a satisfactory response to the initial trial of DMARDs, and if a rheumatologist has not yet been involved in the patient's care, a rheumatology consultation should be obtained. MTX as monotherapy or as a component of combination therapy should be instituted in patients whose treatment has not yet included MTX.

Methotrexate (MTX)

Many rheumatologists select MTX as the initial DMARD, especially for patients whose RA is more active. Because of its favorable efficacy and toxicity profile, low cost, and established track record in the treatment of RA, MTX has become the standard by which all new DMARDs are evaluated.[31,32] Randomized clinical trials have established the efficacy of MTX in RA, particularly in patients with severe disease. MTX retards the progression of radiological erosions.[33]

Observational studies indicate that more than 50% of patients who take MTX continue the drug beyond 3 years, which is longer than any other DMARD.[34] RA patients taking MTX are more likely to discontinue treatment because of adverse reactions than because of lack of efficacy.

Table 3 Baseline Evaluation of Disease Activity and Damage in Patients with Rheumatoid Arthritis

Subjective
Degree of joint pain
Duration of morning stiffness
Duration of fatigue
Limitation of function
Physical examination
Actively inflamed joints (tender and swollen joint counts)
Mechanical joint problems: loss of motion, crepitus, instability, malalignment, and/or deformity
Extraarticular manifestations
Laboratory
Erythrocyte sedimentation rate/C-reactive protein level
Rheumatoid factor*
Complete blood cell count†
Electrolyte levels†
Creatinine level†
Hepatic enzyme levels (AST, ALT, and albumin)†
Urinalysis†
Synovial fluid analysis‡
Stool guaiac†
Other
Functional status or quality of life assessments using standardized questionnaires
Physician's global assessment of disease activity
Patient's global assessment of disease activity
Radiography
Radiographs of selected involved joints§

*Performed only at baseline to establish the diagnosis. If initially negative, may be repeated 6–12 months after disease onset.
†Performed at baseline, before starting medications, to assess organ dysfunction due to comorbid diseases. AST_aspartate aminotrans-ferase; ALT_alanine aminotransferase.
‡Performed at baseline, if necessary, to rule out other diseases. May be repeated during disease flares to rule out septic arthritis.
§Helps to establish a baseline for monitoring disease progression and response to treatment.

Stomatitis, nausea, diarrhea, and perhaps, alopecia caused by MTX may decrease with concomitant folic acid[35] or folinic acid[36] treatment without significant loss of efficacy. Relative contraindications for MTX therapy are

preexisting liver disease, renal impairment, significant lung disease, or alcohol abuse.

Since the most frequent adverse reaction to MTX is elevation of liver enzyme levels, liver function must be monitored, but the risk of liver toxicity is low.[37] Based on the ACR guidelines for monitoring liver toxicity in patients receiving MTX, a liver biopsy should be performed in patients who develop abnormal findings on liver function studies that persist during treatment or after discontinuation of the drug. Rare but potentially serious and even life-threatening pulmonary toxicity may occur at any time with any dosage of MTX. Lymphoproliferative disorders may rarely occur in patients taking MTX,[38] but the relationship to the medication is unclear. Since MTX is potentially teratogenic, appropriate contraceptive use is required.

Hydroxychloroquine (HCQ) and Sulfasalazine (SSZ)

In the last decade, a number of studies have documented the symptomatic benefit of HCQ and SSZ, particularly for patients with early, milder forms of the disease.[39,40] Although HCQ alone does not slow radiologic damage, early treatment with HCQ has a significant impact on long-term patient outcome. Rash, abdominal cramps, and diarrhea are infrequent adverse effects. HCQ is generally well tolerated and requires no routine laboratory monitoring, although patients need periodic ophthalmologic examinations for early detection of reversible retinal toxicity.[41] The risk of retinal toxicity is increased when the dose exceeds 6 mg/kg. The length of time to benefit may vary from 1 month to as long as 6 months. SSZ may act more quickly than HCQ, with benefit sometimes as early as 1 month after beginning therapy. More importantly, SSZ has been shown to retard radiographic progression of RA. SSZ is usually well-tolerated, with most side-effects, which include nausea and abdominal discomfort, occurring in the first few months of therapy. The incidence of these side-effects is lessened by starting at a low dosage and then gradually increasing the dosage. Leukopenia is an occasional, more serious side effect that may occur at any time, and periodic laboratory monitoring is therefore necessary. Clinical response should be apparent within 4 months, and the need for a change in therapy may be determined at that time.

Table 4 Disease Modifying Anti-Rheumatic Drugs (DMARDs)[22]

Drug	Dosage	Efficacy	Toxicity	Monitoring Guidelines
Methotrexate	7.5–20 mg once weekly oral or IM. Start at low dose.	+++	Liver function abnormalities, liver cirrhosis, interstitial pneumonitis, bone marrow suppression	Baseline FBC, Hepatitis B and C serologies and Chest X-Ray. FBC, LFT, Creatinine every 1–2 months.
Sulfasalazine	500–3000 mg/day oral in divided doses. Start at low dose.	+++	Bone marrow suppression	Baseline G6PD. FBC every 2–4 weeks for the first 3 months and then every 3 months. LFT periodically.
IM Gold	10–50 mg IM weekly	+++	Pancytopenia, nephrotic syndrome, pneumonitis	FBC, UFEME, at every injection
D-Peniciallamine	250–750 mg/day oral	+++	Bone marrow suppression, bronchiolitis, other autoimmune disease	FBC, UFEME every 2 weeks initially, then every 1–2 months
Hydroxy-chloroquine	0.6 mg/kg/day (200–400 mg/day)	++	Retinal pigmentation (rare)	Ophthalmological evaluation every 6–12 months
Azathioprine	2–2.5 mg/kg/day oral	++	Bone marrow suppression	FBC every month LFT periodically
Auranofin	3 mg twice daily	+	Rare	FBC, UFEME every month
Etanercept	25 mg subcutaneously twice weekly	++++	Infections	Monitor for injection site reactions

Table 4 (*Continued*)

Drug	Dosage	Efficacy	Toxicity	Monitoring Guidelines
Infliximab plus oral and subcutaneous methotrexate	3–10 mg IV every 8 weeks	++++	Infections, same as methotrexate	As for methotrexate
Leflunomide	20 mg/day in a single dose if tolerated; otherwise, 10 mg/day	+++	Diarrhea, alopecia, rash, headache, theoretical risk of immunosuppression infection known teratogen.	Hepatitis B and C serology in high-risk patients, CBC, creatinine, LFTs monthly for the first 6 months; every 1–2 months thereafter. For minor elevations in AST or ALT (< 2-fold ULN), repeat testing in 2–4 weeks. For moderate elevations in AST or ALT (> 2-fold but < 3-fold ULN), closely monitor, with LFTs every 2–4 weeks and dosage reduction. For persistent elevations of AST or ALT (> 2- or 3-fold ULN), discontinue leflunomide and eliminate with cholestyramine therapy; perform liver biopsy as necessary. Patients also taking MTX should have LFTs at least monthly.
Cyclosporine	2–5 mg/kg/day	+++	Nephrotoxicity, neurotoxicity, hypertension, hirsutism	Creatinine every 2 weeks then monthly. FBC, potassium, LFT 1–2 monthly

FBC: Full Blood Count; LFT: Liver Function Test; UFEME: urine analysis

Newer Drugs

Leflunomide

Leflunomide is a new pyrimidine synthesis inhibitor now FDA approved in the treatment of RA and shown to be as safe and effective as methotrexate. Several randomized, controlled clinical trials (including some conducted in Singapore and other Asian countries) have established leflunomide as an alternative to MTX as monotherapy, especially for patients who cannot tolerate MTX or are experiencing an inadequate response to MTX). The reduction in RA disease activity and in the rate of radiologic progression achieved by leflunomide appears to be equivalent to that of a modest dosage of MTX. Leflunomide is also beneficial as combination therapy with MTX, in the absence of a complete clinical response with full doses of MTX. The usual dosage of leflunomide is 20 mg per day and liver function tests should be monitored regularly.

Five percent of patients receiving leflunomide and up to 60% of patients receiving MTX plus leflunomide have elevated liver enzyme levels.[42] Since enterohepatic recirculation plays a large role in leflunomide metabolism, leflunomide has a long half-life. Without the recommended washout protocol with cholestyramine resin, elimination of the drug would take as long as 2 years. Leflunomide is a potent teratogen, and women taking leflunomide who wish to conceive must discontinue leflunomide and undergo cholestyramine washout before attempting conception. Obstructive biliary disease, liver disease, viral hepatitis, severe immunodeficiency, inadequate birth control, and rifampin therapy (which raises leflunomide serum levels) are all contraindications to the use of leflunomide.

Anti-tumor necrosis factor (anti-TNF) therapy

The development of genetically engineered biological agents that selectively block cytokines (anticytokine therapy) in the short term represents a major advance in the treatment of RA. The most clinically effective anticytokine agents studied to date are antagonists to TNF, an essential mediator of the cytokine inflammatory cascade in RA. Two anti-TNF agents are available: etanercept, a recombinant soluble TNF-Fc fusion protein; and infliximab, a chimeric (mouse–human) anti-TNF monoclonal antibody. Randomized, double-blind, placebo-controlled trials have demonstrated the efficacy of etanercept and infliximab in improving clinical symptoms and signs in

patients with RA. Patients with early RA[43] and those with active RA in whom previous DMARD therapy had failed[44] showed improvement with etanercept therapy. Both etanercept[45] and infliximab[46] have been shown to be beneficial when used in combination with MTX in patients with ongoing active RA despite adequate doses of MTX alone. Infliximab is currently recommended for use only with concomitant MTX therapy.

Concerns about the short-term and long-term safety of these agents exist. TNF-α plays an important role in host protection against infection and tumor genesis. Postmarketing experience with etanercept and infliximab shows hospitalizations and deaths from serious infections in patients treated with these agents. Many of the patients who died while being treated with anti-TNF-α had significant chronic infections or risk factors for infection. Anti-TNF agents should therefore be used with caution in patients with any susceptibility to infection or a history of tuberculosis, should be avoided in patients with significant chronic infections, and should be discontinued temporarily in all patients with acute infection.

Postmarketing surveillance has yielded reports of sepsis, tuberculosis, atypical mycobacterial infections, fungal infections, other opportunistic infections, demyelinating disorders, and aplastic anemia. Risk of latent tuberculosis should be assessed prior to initiation of a TNF antagonist. While the follow-up period with these new agents is still relatively short, thus far there have been no demonstrated increases in the incidence of malignancy in patients treated with etanercept or infliximab compared with the expected rates in the general population. At this time, there appears to be no need for routine laboratory monitoring with the anti-TNF agents, but patients should be alerted to report any signs or symptoms of infection.

In addition to the absence of long-term safety data, the disadvantages of anti-TNF agents are the need for parenteral administration and the high cost of these medications. Not all patients with RA respond to anti-TNF therapy, and disease flares occur after therapy is discontinued.

Staphylococcal protein A immunoadsorption

Extracorporeal immunoadsorption of plasma against a staphylococcal protein A column (Prosorba) was reported to be efficacious in a portion of patients with severe refractory RA.[47]

Interleukin receptor antagonists (IL1 receptor antagonists)

These have also proven beneficial but are not ready for use. Anakinra, a human recombinant form of interleukin-1 receptor antagonist (IL-1Ra) was shown in 4 trials to be efficacious as monotherapy or combination therapy, compared with a placebo, for the treatment of active RA.[48]

Combination DMARD therapy

Conventional treatment with a single DMARD often fails to adequately control clinical symptoms or to prevent disease progression and then combinations of various DMARDs have been used.

Role of steroids in RA

When RA is newly diagnosed rapid, effective suppression of inflammation may be desired with a short course of steroids. Parenteral methylprednisolone or hydrocortisone can be used for severe flares of disease or for vasculitis. There may be some evidence that long term steroids may slow the progression of joint damage.[49] The benefits of low-dose systemic glucocorticoids, however, should always be weighed against their adverse effects. The adverse effects of long-term oral glucocorticoids at low doses are protean and include osteoporosis, hypertension, weight gain, fluid retention, hyperglycemia, cataracts, and skin fragility, as well as the potential for premature atherosclerosis. These adverse effects should be considered and should be discussed in detail with the patient before glucocorticoid therapy is begun. For long-term disease control, the glucocorticoid dosage should be kept to a minimum. For the majority of patients with RA, this means < 10 mg of prednisone per day. RA is associated with an increased risk of osteoporosis independently of glucocorticoid therapy. Patients taking glucocorticoids at dosages as low as 5 mg/day have an increased risk of osteoporosis, and densitometry to assess bone loss should be performed at regular intervals for the duration of glucocorticoid therapy. Measures to prevent glucocorticoid induced osteoporosis should be instituted (1000–1500 mg elemental calcium/day $\pm$ 400–800 IU vitamin D/day $\pm$ hormone replacement therapy $\pm$ bisphosphonates).

Role of Surgery in RA

Surgery is most useful for:

- ruptured tendons at the hand and wrist (tendon repair);
- functional impairment due to end stage joint disease (joint replacement surgery);
- joint instability (fusion of thumb, wrist, ankle, C-Spine); and
- intricate hand surgery to correct deformities has fallen out of favor.

Management of Complications

Vasculitis is treated with high dose steroids and possible addition of drugs such as cyclophosphamide.

Splenectomy has only a variable effect on the leucopenia of Felty's syndrome. There have been reports of improvement in this syndrome with DMARD therapy, particularly gold, methotrexate and D-penicillamine.

Popliteal cysts are usually managed by intra-articular injection of steroids but synovectomy may be needed. Surgical removal of rheumatoid nodules may help temporarily but the nodules usually recur. Surgery is indicated for infected nodules or those that mechanically interfere with usual activities.

Therapy for rheumatoid lung disease should be instituted during the inflammatory stage and not after fibrosis has occurred. High-resolution CT scans of the thorax can demonstrate active alveolitis and these patients should be treated with high-dose steroids and possibly cyclophosphamide.

CONCLUSION

RA is an autoimmune disease affecting mainly the joints, but with significant systemic features. Early aggressive management is advocated to prevent joint deformities, disability, morbidity and mortality, especially in those patients with poor prognostic features.

REFERENCES

1. Arnett FC ES, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, *et al.*, The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis, *Arthritis Rheum* **31**(3):315–324, 1988.

2. Edmonds J, Overview of RA in the Asia Pacific region, in: EbFP (ed.), *Proceedings of the XIX ILAR Congress of rheumatology*: Communication Consultants Singapore, pp. 465–468, 1997.

3. Mitchell D, *Epidemiology, Rheumatoid Arthritis, Etiology, Diagnosis and Treatment*, JB Lippincott, Philadelphia, 1985.

4. Brun JG NS, Kvale G, Breast feeding, other reproductive factors and rheumatoid arthritis. A prospective study, *Br J Rheumatol* **34**(6): 542–546, 1995.

5. Stansky P, Association of the B-cell alloantigen DRw4 with rheumatoid arthritis, *N Engl J Med* **298**(16):869–871, 1978.

6. Chan SH LY, Wee GB, Koh WH, Boey ML, HLA class 2 genes in Singaporean Chinese rheumatoid arthritis, *Br J Rheumatol* **33**(8):713–717, 1994.

7. JJ Goronzy CW, Rheumatoid arthritis, epidemiology, pathology and pathogenesis. 11 ed: Arthritis Foundation, Atlanta Georgia; 1997.

8. Veerapen K MG, Watt I, Dieppe P, The expression of rheumatoid arthritis in Malaysian and British patients: a comparative study, *Br J Rheumatol* **32**(7):541–545, 1993.

9. Cohen MG LE, Ng PY, Chan KL, Extra-articular manifestations are uncommon in southern Chinese with rheumatoid arthritis, *Br J Rheumatol* **32**(3):209–211, 1993.

10. Moran H CS, Muirden KD, Jiang SJ, Gu YY, Hopper J, Jiang PL, Lawler G, Chen RB, A comparison of rheumatoid arthritis in Australia and China, *Ann Rheum Dis* **45**(7):572–578, 1986.

11. J Fleming A BR, Corbett M, Wood PH, Early rheumatoid Disease. II. Patterns of joint involvement, *Ann Rheum Dis* **35**(4):361–364, 1976.

12. Boden S, Rheumatoid arthritis of the cervical spine. Surgical decision making based on predictors of paralysis and recovery, *Spine* **19**(20):2275–2280, 1994.

13. McDonagh J GM, Wright AR, Heycock C, Owen JP, Kelly C, High resolution computed tomography of the lungs in patients with rheumatoid arthritis and interstitial lung disease, *Br J Rheumatol* **33**(20):118–122, 1994.

14. Rosenstein ED KN, Felty's and pseudo-Felty's syndromes, *Semin Arthritis Rheum* **21**(3):129–142, 1991.

15. Gridley G MJ, Ekbom A, Klareskog L, Adami HO, Hacker DG, Hoover R, Fraumeni JF Jr., Incidence of cancer among patients with rheumatoid arthritis, *J Natl Cancer Inst* **85**(4):307–311, 1993.

16. Kerstens PJ BA, Jeurissen ME, Fast JH, Assmann KJ, van de Putte LB, Accelerated nodulosis during low dose methotrexate therapy for rheumatoid arthritis. An analysis of ten cases, *J Rheumatol* **19**(6):867–871, 1992.

17. Vandenbroucke JP HH, Cats A, Survival and cause of death in rheumatoid arthritis: a 25-year prospective followup, *J Rheumatol* **11**(2):158–161, 1984.

18. Isomaki H, Mortality in patients with rheumatoid arthritis. *Rheumatoid Arthritis. Pathogenesis, Assessment, Outcome, and Treatment*. Marcel Dekker, New York, 1994.

19. Sherrer YS BD, Mitchell DM, Young DY, Fries JF, The development of disability in rheumatoid arthritis, *Arthritis Rheum* **29**(4):494–500, 1986.

20. Wolfe F CM, The assessment and prediction of functional disability in rheumatoid arthritis, *J Rheumatol* **18**(9):1298–1306, 1991.

21. Van Zeben D HJ, Zwinderman AH, Vanderbroucke JP, Breedveld FC, Factors predicting outcome of rheumatoid arthritis: results of a followup study, *J Rheumatol* **20**(8): 1288–1296, 1993.

22. Guidelines for management of rheumatoid arthritis. American College of Rheumatology Ad Hoc committee on clinical guidelines, *Arthritis Rheum* **46**(2):328–346, 2002.

23. Yeomans ND TZ, Juhasz L, Racz I, Howard JM, van Rensburg CJ, Swannell AJ, Hawkey CJ, A comparision of omeprazole with ranitidine for ulcers associated with nonsteroidal antiinflammatory drugs. Acid Suppression Trial: Ranitidine versus Omeprazole for NSAID-associated Ulcer Treatment (ASTRONAUT) Study Group, *N Engl J Med* **338**(11):719–726, 1998.

24. Hawkey CJ KJ, Szczepanski L, Walker DG, Swannell AJ, Yeomans ND, Omeprazole compared with misoprostol for ulcers associated with nonsteroidal antiinflammatory drugs. Omeprazole versus Misoprostol for NSAID-induced Ulcer Management (OMNIUM) Study Group, *N Engl J Med* **338**(11):727–734, 1998.

25. Raskin JB WR, Jaszewski R, Korsten MA, Schubert TT, Fort JG, Misoprostol and ranitidine in the prevention of NSAID-induced ulcers: a prospective, double-blind, multicenter study, *Am J Gastroenterol* **91**(2):223–227, 1996.

26. Committee. HCONS. Progress in prophylaxis against nonsteroidal anti-inflammatory drug-associated ulcers and erosions, *Am J Med* **104**(3A):67s–74s, 1998.

27. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, *et al.*, Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group, *N Engl J Med* **343**(21):1520–1528, 2000.

28. Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, *et al.*, Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. Celecoxib Long-term Arthritis Safety Study, *Jama* **284**(10):1247–1255, 2000.

29. Felson DT AJ, Meenan RF, The comparitive efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses, *Arthritis Rheum* **33**(10):1449–1461, 1990.

30. Thumboo J KW, Leong KH, Boey ML, Feng PH, The safety of weekly low dose methotrexate in an Oriental population with rheumatoid arthritis, *Ann Acad Med Singapore* **26**(2):205–209, 1997.

31. Andersen PA, West SG, O'Dell JR, Via CS, Claypool RG, Kotzin BL, Weekly pulse methotrexate in rheumatoid arthritis. Clinical and immunologic effects in a randomized, double-blind study, *Ann Intern Med* **103**(4):489–496, 1985.

32. Furst DE, Erikson N, Clute L, Koehnke R, Burmeister LF, Kohler JA, Adverse experience with methotrexate during 176 weeks of a longterm prospective trial in patients with rheumatoid arthritis, *J Rheumatol* **17**(12):1628–1635, 1990.

33. Williams HJ, Willkens RF, Samuelson CO Jr., Alarcon GS, Guttadauria M, Yarboro C, *et al*., Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis. A controlled clinical trial, *Arthritis Rheum* **28**(7):721–730, 1985.

34. Kremer JM, Phelps CT, Long-term prospective study of the use of methotrexate in the treatment of rheumatoid arthritis. Update after a mean of 90 months, *Arthritis Rheum* **35**(2):138–145, 1992.

35. Morgan SL, Baggott JE, Vaughn WH, Young PK, Austin JV, Krumdieck CL, *et al*., The effect of folic acid supplementation on the toxicity of low-dose methotrexate in patients with rheumatoid arthritis, *Arthritis Rheum* **33**(1):9–18, 1990.

36. Shiroky JB, Neville C, Esdaile JM, Choquette D, Zummer M, Hazeltine M, *et al*., Low-dose methotrexate with leucovorin (folinic acid) in the management of rheumatoid arthritis. Results of a multicenter randomized, double-blind, placebo-controlled trial, *Arthritis Rheum* **36**(6):795–803, 1993.

37. Kremer JM, Alarcon GS, Lightfoot RW Jr., Willkens RF, Furst DE, Williams HJ, *et al*., Methotrexate for rheumatoid arthritis. Suggested guidelines for monitoring liver toxicity. American College of Rheumatology, *Arthritis Rheum* **37**(3):316–328, 1994.

38. Moder KG, Tefferi A, Cohen MD, Menke DM, Luthra HS, Hematologic malignancies and the use of methotrexate in rheumatoid arthritis: a retrospective study, *Am J Med* **99**(3):276–281, 1995.

39. Tsakonas E, Fitzgerald AA, Fitzcharles MA, Cividino A, Thorne JC, M'Seffar A, *et al*., Consequences of delayed therapy with second-line agents in rheumatoid arthritis: a 3 year followup on the hydroxychloroquine in early rheumatoid arthritis (HERA) study, *J Rheumatol* **27**(3):623–629, 2000.

40. Williams HJ, Comparisons of sulfasalazine to gold and placebo in the treatment of rheumatoid arthritis, *J Rheumatol Suppl* **16**:9–13, 1988.

41. Easterbrook M, An ophthalmological view on the efficacy and safety of chloroquine versus hydroxychloroquine, *J Rheumatol* **26**(9): 1866–1868, 1999.

42. Mladenovic V, Domljan Z, Rozman B, Jajic I, Mihajlovic D, Dordevic J, *et al*., Safety and effectiveness of leflunomide in the treatment of patients with active rheumatoid arthritis. Results of a randomized, placebo-controlled, phase II study, *Arthritis Rheum* **38**(11):1595–1603, 1995.

43. Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC, *et al*., A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis, *N Engl J Med* **343**(22):1586–1593, 2000.

44. Moreland LW, Schiff MH, Baumgartner SW, Tindall EA, Fleischmann RM, Bulpitt KJ, *et al*., Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial, *Ann Intern Med* **130**(6):478–486, 1999.

45. Weinblatt ME, Kremer JM, Coblyn JS, Maier AL, Helfgott SM, Morrell M, *et al*., Pharmacokinetics, safety, and efficacy of combination treatment with methotrexate and leflunomide in patients with active rheumatoid arthritis, *Arthritis Rheum* **42**(7): 1322–1328, 1999.

46. Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, *et al*., Infliximab (chimeric anti-tumor necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomized phase III trial. ATTRACT Study Group, *Lancet* **354**(9194):1932–1939, 1999.

47. Felson DT, LaValley MP, Baldassare AR, Block JA, Caldwell JR, Cannon GW, *et al*., The Prosorba column for treatment of refractory rheumatoid arthritis: a randomized, double-blind, sham-controlled trial, *Arthritis Rheum* **42**(10):2153–2159, 1999.

48. Calabrese LH, Anakinra treatment of patients with rheumatoid arthritis, *Ann Pharmacother* **36**(7):1204–1209, 2002.

49. Kirwan J, The effect of glucocorticoids on joint destruction in rheumatoid arthritis, *N Engl J Med* **333**:3142–3146, 1995.

This page intentionally left blank

85

Seronegative Spondyloarthropathy

Koh Wei Howe

INTRODUCTION

The seronegative spondyloarthropathies are an interrelated group of chronic inflammatory diseases that share common clinical features, including a tendency to affect the axial skeleton, and an association with the histocompatibility antigen HLA-B27.[1] Ankylosing spondylitis is the prototype disease of the spondyloarthropathies that include reactive arthritis/Reiter's syndrome, psoriatic arthritis, chronic inflammatory bowel diseases and juvenile onset spondyloarthropathy.

The frequency of HLA-B27 occurring in the seronegative spondyloarthropathies varies greatly among the various forms and show differences between ethnic and racial groups.[2] Amongst the group of disorders, ankylosing spondylitis has the strongest association with the HLA-B27 (over 95%) and the other disease such as psoriatic arthritis and inflammatory bowel disease, bear weaker associations. Typically, patients with seronegative spondyloarthropathies are tested negative for rheumatoid factor.

Table 1 The European Spondyloarthropathy Study Group (ESSG) Criteria for Spondyloarthropathy

Inflammatory spinal pain or Synovitis
• asymmetrical; or
• predominantly in the lower limbs.
and
One or more of the following:
• alternate buttock pain;
• sacroiliitis;
• enthesopathy;
• positive family history;
• psoriasis;
• inflammatory bowel disease; or
• urethritis or cervicitis or acute diarrhea occurring within one month before the arthritis

Common to all forms of spondyloarthropathies is the greater tendency to afflict males and the development of the disease in the early twenties or late adolescent age.[1,3] Characteristic clinical manifestations may involve inflammation at sites of tendon and ligamentous insertions, sacroiliitis and inflammatory spinal disease, peripheral oligoarthritis predominantly of the lower limbs, dactylitis and extra-articular features such as iritis. However, a spectrum of clinical syndromes exists in association with the HLA-B27 antigen that may not satisfy established classification criteria. These undifferentiated subsets of spondyloarthropathy encompass syndromes such as isolated acute anterior uveitis, seronegative oligoarthritis, dactylitis and enthesitis, and even cardiac valvular lesions with conduction abnormalities. Recognition of these distinct entities have led to the development of classification criterias such as the European Spondyloarthropathy Study Group (ESSG) criteria (Table 1)[4] that attempts to unify these syndromes under the broader spectrum of spondyloarthropathies.

ETIOLOGY AND PATHOGENESIS

The role of HLA-B27 in conferring disease susceptibility to spondyloarthropathy is widely accepted but the precise mechanism by which it is involved in the pathogenesis of the disease is yet to be confirmed. It is suspected that an interplay between genetic and environmental factors, especially infections, is responsible for the development of the disease.[5,6]

For example, enteric infections by Gram-negative bacteria such as Yersina, Campylobacter, Salmonella or Shigella, or urogenital infection by Chlamydia are associated with reactive arthritis. By definition, reactive arthritis implies that the inflammatory response in the joint is sterile and that the inciting bacteria cannot be cultured from the joint fluid. However there is some evidence that bacteria-like product can be detected in synovial tissues of patients with reactive arthritis, suggesting that the persistence of microbial antigens play an important role in perpetuating the inflammatory process in the joints.[7] Common features shared by these Gram-negative organisms that cause reactive arthritis are the ability to invade mucosal surfaces and replicate intra-cellularly, and the possession of lipopolysaccharides.

The evidence for infective trigger in ankylosing spondylitis is less clear but recent research has focussed on Klebsiella pneumoniae as a possible etiological agent especially when peripheral arthritis is present[8] and elevated IgA class anti-Klebsiella antibodies in patients with ankylosing spondylitis have been reported.[9] The association between gut bacteria and pathogenesis of HLA-B27-related diseases has been a subject of much research. The presence of silent inflammatory intestinal lesions in patients with spondyloarthropathy has been observed.[10] It is unclear whether these lesions represent a spectrum of the inflammatory bowel disease or if they provide a site for bacterial antigens to enter the circulation and stimulate immunologic inflammatory processes.

Studies with transgenic animals expressing HLA-B27 has shed more light on the importance of the gene and its interaction with gut bacteria in the pathogenesis of disease. Transgenic rats with the human HLA-B27 gene have exhibited inflammatory changes in the joints, skin, intestinal tract, genitals and heart that are similar to B27-associated disease in humans.[11] However, the exact mechanism by which the bacteria and the HLA-B27 gene interact to produce the disease is yet to be elucidated. The two hypotheses that are often proposed are that the HLA-B27 has a molecular mimicry with enterobacteria or that it presents arthritogenic peptides (derived from bacterial antigens or HLA-B27 itself) to T cells. *At least 23 different subtypes of HLA-B27 have been identified and they are named B*2701 to B*2723*.[12] The presence of various subtypes differ in racial distribution and whilst some common subtypes (B*2705, B*2702, B*2704, B*2707) are associated with spondyloarthropathy, others such as B*2706 in Southeast Asians seem to lack the disease association.[13]

ANKYLOSING SPONDYLITIS

The typical presentation of ankylosing spondylitis is an inflammatory backpain with bilateral sacroiliitis demonstrated on radiographs. The disease is more common in men but the sex ratio is much lower than was previously thought (male to female ratio of 3:1 in Western population, 7:1 in Singapore).[3] The prevalence of the disease show different ethnic distribution and is estimated to be between 0.1 to 0.2% of the caucasian population. A family history of spondyloarthropathy is present in 10–20% of patients with ankylosing spondylitis.[3] The onset of the disease is usually in the early twenties but the disease can also occur in children.[1] In juvenile spondyloarthropathy, which starts before the age of 16 years, the initial presentation is commonly a peripheral oligoarthritis and enthesitis without backpain. The New York criteria for diagnosis of ankylosing spondylitis has been used for epidemiological studies but lack sensitivity in identifying early and mild disease.[14]

Clinical Features

The sacroiliac joints are often the first sites to be affected and patients may experience unilateral or alternating buttock pain. The disease then progresses up the lumbar spine and frequently the neck. The characteristics of backache includes an insidious onset, presence of morning stiffness and a pain that is worse with prolong rest and improves with mobilization of the spine. On examination, the movement in the lumbar spine is often limited in all directions. Patients may have sleep disturbance at night because of the backache and stiffness, and wake up in the morning with great difficulty getting out of bed. Fatigue is a common problem for many patients and sleep disorder is probably one contributory cause for the symptom.[15] Osteoporosis, vertebral fractures, spondylodiscitis and pseudoarthrosis are complications that may occur in the spine. Even in the early and mild stages of the disease, the bone mineral density of patients may be low.

Peripheral joint involvement occurs in up 25–50% of patients but seems to be more frequent in Oriental patients. Hip disease tends to occur in patients with an early onset of disease, less than 20 years of age, and is usually bilateral and insidious in onset (Fig. 1). Erosion and ankylosis may result in flexion contractures of the hip joints and some patients may

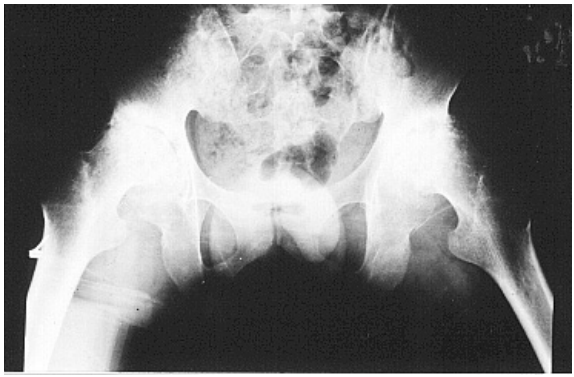


Fig. 1 Bilateral hip involvement in a patient with ankylosing spondylitis.

require hip replacement surgery later in life. The shoulders and knees are other joints that can be affected and ten percent of patients may develop temporomandibular joint disease.

Another hallmark feature of the spondyloarthropathies is the presence of enthesitis (inflammation at sites of tendon or ligamentous attachment to bones) which commonly occurs at the achilles tendons, heels, ischial tuberosities, tibial tubercles or costosternal junctions. Bony ankylosis or fusion of the joints may result from fibrosis and ossification of tissue across the affected joints and hence the name "ankylosing spondylitis".

Mild fever, weight loss and malaise are some of the constitutional symptoms that may occur in early disease. At some time during the course of their disease, 25–40% of patients develops acute anterior uveitis (iritis). Uveitis is more common in HLA-B27 positive than negative patients. It is typically acute, unilateral and results in eye pain, increased lacrimation, blurred vision and photophobia. Other extra-articular manifestations occur less frequently and these include aortitis, mitral valve incompetence, conduction abnormalities, myocardial dysfunction, apical pulmonary fibrosis, cauda equina syndrome, amyloidosis and IgA nephropathy.

Investigations and Assessment of Disease Status

Hematological investigations may show non-specific abnormalities such as raised ESR and serum IgA; and a normochromic, normocytic anemia. The rheumatoid factor and anti-nuclear antibodies are usually negative.

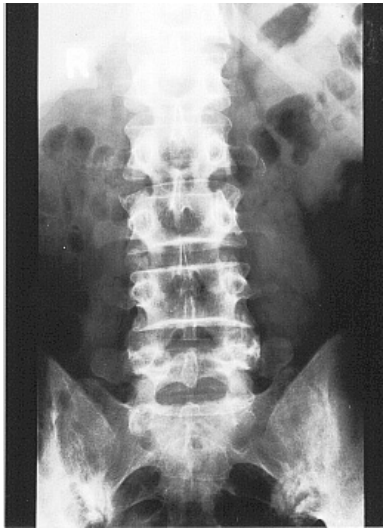


Fig. 2 Antero-posterior view of the lumbosacral spine showing bilateral sacroiliitis.

Radiographic evidence of bilateral sacroiliitis supports the diagnosis of ankylosing spondylitis and an antero-posterior view is sufficient to detect the changes (Fig. 2). Other imaging modalities such as bone scan, computerized tomography (CT) scan and magnetic resonance imaging (MRI) are more sensitive in detecting changes of sacroiliitis but are not routinely required. "Squaring" of the vertebral body seen on lateral spine radiographs is one of the earliest features present. Subsequently, ossification of the superficial layers of the annulus fibrosus results in the formation of syndesmophytes. The bridging of vertebral bodies by the syndesmophytes, fusion of apophyseal joints and ossification of interspinous ligaments can result in ankylosis of the spine and give the appearance of a "bamboo spine" on the radiograph (Fig. 3). However, the radiographic changes may take years to be apparent and the initial X-rays may appear normal. *A radiological scoring system, Bath Ankylosis Spondylitis Radiology Index (BASRI) has been devised to grade radiological changes in patients with ankylosing spondylitis. It uses a scale of 0–4 to grade the severity of change in the sacroiliac joint, lumbar and cervical spine.*[16]

The HLA-B27 test is not necessary for diagnosis in the majority of cases when the clinical and radiological features are characteristic of the disease. It should not be used as a "confirmatory" test to diagnose or exclude the

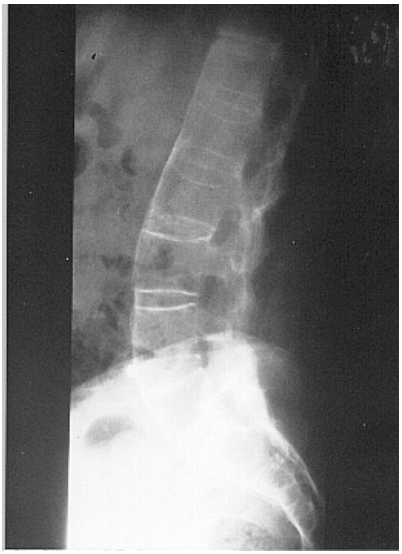


Fig. 3 Ankylosis of the vertebral spine giving the appearance of a "bamboo spine".

disease because it is present in about 7% of the normal local population, and is absent in about 5% of patients with ankylosing spondylitis.

The axial skeleton is predominantly involved in the disease and assessing the degree of inflammation in the spine clinically is more difficult compared to a largely peripheral arthritis such as rheumatoid arthritis. Various instruments in the form of self assessment questionnaires (e.g. Bath Ankylosis Spondylitis Disease Activity Index, BASDAI; Functional Index, BASFI) and metrological measurements (Bath Ankylosing Spondylitis Metrological Index, BASMI) have been devised in an attempt to assess the state of the disease.[17]

Management

At present, there is no known cure for ankylosing spondylitis but drug treatments can control the patient's symptoms whilst physiotherapy helps to reduce joint stiffness, increase mobility and strengthen the muscles. The non-steroidal anti-inflammatory drugs (NSAIDs) are often effective in treating the active inflammation. Indomethacin is probably the most common NSAID used but others such as diclofenac and naproxen are also effacious when used in the full therapeutic dose. Patients should

be advised that these drugs have anti-inflammatory properties and are not merely "pain-killers". They should be taken regularly for a period of time rather than when necessary. The use of phenylbutazone has been discouraged due to the potential risk of bone marrow suppression, although the drug is very effective in treating ankylosing spondylitis. The common side-effect of all NSAIDs is gastrointestinal toxicity and concommitant prescription of gastroprotective agents should be considered in those at higher risk of developing the side effect, such as a past history of peptic ulcer disease. *Cyclo-oxygenase-2 (COX2) inhibitors (e.g. Celecoxib and Rofecoxib) are a group of anti-inflammatory drugs that are associated with less serious gastrointestinal side-effects as compared to conventional NSAIDs, and they can also be used in the treatment of ankylosing spondylitis especially in patients who are at risk of developing peptic ulcer disease.*

Sulphasalazine is effective as a second-line agent in treating patients with active peripheral arthritis but there is little evidence for its use in purely axial skeleton disease.[18] However, like most second-line agents, there is a delay in onset of therapeutic efficacy with sulphasalazine of up to three months. Low-dose weekly methotrexate has also been used with some success in patients who do not respond well to NSAIDs and sulphasalazine. The inhibition of pro-inflammatory cytokine tumor necrosis factor (TNF) alpha has a beneficial effect on the disease activity of patients with spondyloarthropathy. Drugs such as thalidomide and pamidronate, which has some inhibitory effect on TNF-α, has been reported to be effective in the treatment of ankylosing spondylitis. Patients with ankylosing spondylitis that have been treated with biological therapies (e.g. infliximab) that specifically block the action of TNF have shown significant improvement in disease activity. However, these agents have to be given parenterally and are presently reserved for those patients who have severe disease.

A few studies have reported that pulse methylprednisolone may be beneficial but it should probably be used only in patients with severe, acute peripheral arthritis as a "bridging therapy" at the initiation of sulphasalazine treatment because of the delayed response of the latter drug. Oral systemic steroids have no role in the long-term treatment of the disease because it does not modify the disease process and may have deleterious side effects. Local corticosteroid injections are useful for treating severe refractory enthesitis or peripheral arthritis (especially after an aspiration of the joint is performed for large joint effusion) as well as for

persistent sacroiliitis. Treatment of acute iritis requires steroid eye drops and rarely systemic steroids or immunosuppressive agents when it is severe. Fatigue, sleep disturbance and joint pains may improve with the use of low dose amitriptyline as adjuvant therapy.[19]

The importance of regular physical exercises to maintain spinal mobility and limit deformity cannot be over-emphasized. Smoking is to be discouraged and daily breathing exercise serves to enhance chest expansion. Intensive group physiotherapy and hydrotherapy programs do provide substantial improvement in patient's disease, at least in the short-term.[20] Patient should be advised to adopt a proper posture, keep their spine straight and use firm mattresses when sleeping.

Patients with severe hip disease may require total hip arthroplasty that can significantly improve their functional ability. Cervical spine surgery is sometimes performed for severe kyphosis and neurological complications, and the long-term outcome is often good.[21]

REACTIVE ARTHRITIS (REITER'S SYNDROME)

Reactive arthritis (ReA) is a form of seronegative spondyloarthropathy that encompasses the more familiar term called Reiter's Syndrome. By definition, the inflammatory arthritis that occurs in ReA is sterile and develops sometime after an infection, often in urogenital or intestinal tract, or throat. The typical triad of oligoarthritis, conjunctivitis and urethritis is described in Reiter's syndrome but incomplete "forme fruste" disease have been observed. The common initiating infectious agents include Chlamydia trachomatis in genitourinary tract infection and Salmonella, Shigella, Campylobacter or Yersinia in enteric infections.

The spectrum of clinical syndromes range from mild arthralgia to severe peripheral arthritis and axial skeleton involvement. Enthesitis, frequently at the heel resulting in achilles tendinitis or plantar fascitis, is a typical feature of ReA and some patients develop diffuse sausage-like swelling or dactylitis of the fingers and toes. Mucocutaneous lesions such as circinate balanitis, painless oral mucosal ulcers and keratoderma blenorrhagica (papular, pustular and crusting skin lesion on the palms and soles of the feet) are characteristic extra-articular features that may be present.

NSAIDs are the mainstay of drug treatment in ReA and should be given at full therapeutic dose regularly for an adequate course. Patients

with large joint effusion or persistent synovitis may benefit from joint aspiration and intra-articular corticosteroid injection provided septic arthritis has been excluded. Systemic oral steroid is to be used only if the arthritis is severe and not adequately controlled by NSAIDs. The dose should be tailed down when the arthritis improves and its long-term use should be avoided. Sulphasalazine, starting at a low dosage and gradually increasing to about 2–3 g daily, is a useful second-line agent for chronic ReA. Other agents that has been tried are low dose methotrexate, gold and azathioprine. Antibiotic treatment has little role in the management of enteric ReA and it possibly has only some benefit in the treatment of chronic chlamydia-induced ReA.

REFERENCES

1. Khan MA, Ankylosing spondylitis: Clinical features, in: Klippel JH, Dieppe PA (eds.) *Rheumatology*, 2nd ed., Mosby, London, pp. 6.16.1–6.16.10, 1998.
2. Khan MA, HLA-B27 and its subtypes in world populations, *Curr Opin Rheumatol* **7**:263–269, 1995.
3. Koh WH, Boey ML, Ankylosing spondylitis in Singapore: a study of 150 patients and a local update, *Ann Acad Med Singapore* **27**:3–6, 1998.
4. Dougados M, van der Linden S, Juhlin R, *et al.,* The European Spondyloarthropathy Study Group preliminary criteria for the classification of spondyloarthropathies, *Arthritis Rheum* **34**:1218–1227, 1991.
5. Feltkamp TEW, Khan MA, de Castro JAL, The pathogenetic role of HLA-B27, *Immunol Today* **7**:5–7, 1996.
6. Lopez de Castro JA, Structure, function and disease association of HLA-B27, *Curr Opinion Rheum* **6**:371–377, 1994.
7. Granfors K, Jalkanen S, von Essen R, *et al.,* Yersinia antigens in synovial fluid cells from patients with reactive arthritis, *N Engl J Med* **320**:216–221, 1989.
8. Maki-Ikola O, Nissila M, Lehtinen K, *et al.,* Antibodies to *Klebsiella pneumoniae*, *Escherichia coli* and *Proteus mirabilis* in the sera of patients with axial and peripheral form of ankylosing spondylitis, *Br J Rheumatol* **34**:413, 1995.
9. Cooper R, Fraser SM, Sturrock RD, Gemmell CG, Raised titres of anti-klebsiella IgA in ankylosing spondylitis, rheumatoid arthritis, and inflammatory bowel disease, *Br Med J* **296**:1432–1434, 1988.
10. Leirisalo-Repo M, Turunen U, Stenman S, Helenius P, Seppala K, High frequency of silent inflammatory bowel disease in spondyloarthropathy, *Arthritis Rheum* **37**:23–31, 1994.

11. Hammer RE, Maika SD, Richardson JA, *et al.,* Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human B2-m: an animal model of HLA-B27-associated human disorders, *Cell* **63**:1099–1112, 1990.

12. Feltamp TEW, Mardjuadi A, Feng Huang, Chou CT, Spondyloarthropathies in eastern Asia, *Curr Opinion Rheum* **13**:285–290, 2001.

13. Ren EC, Koh WH, Sim D, Boey ML, Wee GB, Chan SH, Possible protective role of HLA B2706 for ankylosing spondylitis, *Tissue Antigen* **49**:67–69, 1997.

14. Bennett PH, Burch TA, Population studies of the rheumatic diseases, Amsterdam: *Exercpta Medica* 456–457, 1968.

15. Jones S, Koh WH, Steiner A, Garrett S, Calin A, Fatigue in ankylosing spondylitis: its prevalence and relationship to disease activity, sleep and other factors, *J Rheumatol* **23**:487–490, 1996.

16. Mackay K, Mack C, Brophy S, Calin A, The Bath Ankylosing Spondylitis Radiology Index (BASRI): a new, validated approach to disease assessment, *Arthritis Rheum* **41**:2263–2270, 1998.

17. Calin A, Ankylosing spondylitis: defining disease status and the relationship between radiology, metrology, disease activity, function, and outcome (Dunlop-Dottridge lecture), *J Rheumatol* **22**:740–744, 1995.

18. Clegg DO, Reda DJ, Weisman MH, Blackburn WD, *et al.,* Comparison of sulfasalazine and placebo in the treatment of ankylosing spondylitis: a Department of Veterans Affairs Cooperative study, *Arthritis Rheum* **39**:2004–2012, 1996.

19. Koh WH, Pande I, Jones S, Samuels A, Calin A, Low dose amitriptyline in ankylosing spondylitis: a short term, double-blind, placebo controlled study, *J Rheumatol* **24**: 2158–2161, 1997.

20. Seah R, Fernandez G, Phua HS, *et al.,* Ten-day outpatient spondyloarthropathy intensive management programme: The outcome of 5 sessions [abstract], in Proceedings of the 19th ILAR Congress of Rheumatology, Satellite Meeting for Allied Health Professionals Singapore, June 9–11, Communication Consultants, Singapore, 1997.

21. Koh WH, Garrett SL, Calin A, Cervical spine surgery in ankylosing spondylitis: is the outcome good? *Clin Rheumatol* **16**:466–470, 1997.

This page intentionally left blank

86

Gout

Julian Thumboo

INTRODUCTION

Gout is a common condition in which hyperuricemia leads to urate crystal deposition in joints, skin, kidneys and other extra-articular structures. Urate crystal deposition results in a characteristic clinical syndrome of episodic arthritis affecting one or more joints, associated with urate deposits in the skin (tophi), bursae and kidneys.[1]

EPIDEMIOLOGY

Gout is a common condition which usually affects middle-aged men and post-menopausal women. The yearly incidence of gout ranges from 0.20 to 0.35 per thousand, with an overall prevalence of 2.0 to 2.6 per thousand.[1] In South East Asia, the prevalence of gout (per thousand) has variously been reported as 1.6 in rural Thailand and 17 in rural Java. In Taiwan, the prevalence of gout was higher in urban than rural areas, being 6.7 versus 1.6 per thousand respectively. The incidence of gout increases with age and increasing serum urate levels. Gout is rare

before puberty and before the menopause. A diagnosis of gout in such individuals should trigger a search for secondary causes. Gout is negatively associated with rheumatoid arthritis, though the reasons for this are unclear.

Serum urate levels generally range from 180 to 240 μmol/L before puberty. At puberty, serum urate levels in males increase by 60 to 120 μmol/L, while those in females remain relatively constant. After menopause, serum urate levels in females rise steadily and approach urate levels found in males.[1] Both genetic and environmental risk factors may contribute to the development of hyperuricemia and gout. A genetic contribution is suggested by the markedly differing prevalence of gout in different populations.[2] Known risk factors for hyperuricemia and gout include obesity, renal impairment, use of drugs (diuretics, alcohol, etc.) and hypertension (detailed below).

PATHOGENESIS OF GOUT

Uric acid is a breakdown product of purine nucleotides and deoxynucleotides (which form part of DNA and RNA, are precursors of several cofactors and co-enzymes, and are involved in cellular metabolism). Enzymatic nucleotide cleavage of both dietary and endogenous purines gives rise to purine bases. These bases are metabolized to hypoxanthine or xanthine, which are in turn converted to uric acid by the enzyme xanthine oxidase (which is inhibited by allopurinol). In humans, the liver, which expresses xanthine oxidase, is the main site of purine metabolism. Two-thirds of the daily human purine load is generated from endogenous purine metabolism, with the remaining one-third coming from dietary purines.[3]

At a pH of 7.4, 98% of uric acid is ionized as monosodium urate (MSU). Uric acid therefore exists predominantly as urate in serum. In parts of the urinary tract where the pH is less than 5.7, uric acid predominates and may crystallize in this setting because of its poorer solubility, resulting in the formation of urate calculi. Saturation of MSU in human plasma occurs at 420 μmol/L.[4] Above this level, MSU crystallizes in synovium and tissues, causing the clinical syndrome of gout. The solubility of MSU is temperature dependent, and drops two-fold from 37°C to 25°C. This favors the crystallization of MSU in the cooler peripheries, and accounts for the predilection of gout for peripheral joints.

CLINICAL FEATURES

The natural history of gout passes through 4 phases: asymptomatic hyperuricemia, acute gout, inter-critical gout, and chronic tophaceous gout. Extra-articular manifestations (mainly cutaneous and renal) typically manifest in the later stages of articular disease. A family history of gout is present in 6–80% of cases. Gout is more common in males and usually presents in middle age, with a peak incidence in the 50–59 year age group. Ninety percent of females with gout present after the menopause. In a hospital-based series of patients with gout in Singapore, 19% of patients had a family history of gout, with a male: female ratio of 3.4:1 and a mean age of 43.7 years at onset (range 16–78 years).

ASYMPTOMATIC HYPERURICEMIA

Asymptomatic hyperuricemia (AH) is defined as the presence of raised serum uric acid levels without the presence of gout, tophi or renal involvement. AH is a relatively common phenomenon which often does not lead to gout. It develops in predisposed males at puberty and in females after menopause, and is typically persistent. Gout usually develops after 20–30 years of hyperuricemia, and the risk of developing gout increases with the degree of hyperuricemia (Table 1).[5] The risk of developing urate renal calculi increases with the degree of hyperuricemia and the daily urinary uric acid excretion rate. AH itself does not cause a decline in renal function, and correction of AH does not protect against development of renal insufficiency.[4] For these reasons, AH generally does not require pharmacological treatment.

Table 1 Risk of Gout in Subjects with Hyperuricemia

Serum Uric Acid Level (μmol/L)	Yearly Risk of Gout (%)
< 420	0.1–0.5
420–534	0.5–1.2
> 534	4.9–5.7

Uric acid: 1 mg = 59.48 μmol = 0.05948 mmol.

Acute Gout

Acute gout is characterized by episodic attacks of pain, swelling and erythema, usually of lower extremity joints, initially lasting less than a week. These attacks gradually increase in duration, frequency and severity and may be associated with constitutional features (fever, leucocytosis and elevation of the erythrocyte sedimentation rate). Men are more commonly affected, with attacks usually starting after the age of 30 years. Attacks before this age should suggest a secondary cause for gout (Table 2). Affected joints are exquisitely tender, with a markedly limited range of movement. Severe attacks are associated with desquamation of the overlying skin. Once the acute attack has resolved, the patient is typically free of symptoms. Gout of the big toe metatarsophalangeal joint (also referred to as podagara) occurs at presentation in 50% and eventually in 90% of subjects. Other lower extremity joints are typically affected, although any joint in the body may be involved. Attacks are triggered by factors that raise or lower serum urate levels (alcohol, drugs (especially diuretics, low dose aspirin, heparin or cyclosporin), dietary excess or fasting (e.g. before

Table 2 Causes of Hyperuricemia

Increased Urate Production	Reduced Urate Excretion
Primary (i.e. Genetic)	
Overproducers (10% of primary gout)	Underexcretors (90% of primary gout)
Defined enzyme mutations	Reduced urate renal clearance or fractional excretion
Undefined enzyme mutations	
Secondary (i.e. Acquired)	
High Purine intake	Renal disorders (renal failure, lead nephropathy)
Obesity	
Alcohol consumption	Medications (low dose aspirin, cyclosporin, thiazide diuretics, pyrazinamide, ethambutol, laxative abuse)
Fructose consumption	
Myeloproliferative disorders	
Psoriasis, haemolytic anemias, and other disorders with increased catabolism	Dehydration and other causes of decreased renal perfusion
Sustained exercise	Hypertension
	Hypothyroidism, hyperparathyroidism
	Ketones, lactate (starvation, ketoacidosis)

surgical procedures), infections, gastrointestinal hemorrhage or iodinated contrast media) or mechanical trauma. Gout may present atypically, resulting in delays in diagnosis and treatment. In postmenopausal women and the elderly, gout may be polyarticular at onset,[6] and if left untreated, may become chronic, mimicking rheumatoid arthritis. Gout also tends to occur in already pathological joints (e.g. Heberden's nodes) and may also co-exist with septic arthritis.

Inter-critical Gout

Seven percent of subjects have only 1 attack of gout: these subjects tend to have a mildly elevated serum uric acid levels ($< 420\,\mu mol/L$ (8 mg/dL)) and normal urinary uric acid excretion. The majority of subjects have recurrent attacks of gout, usually recurring within 2 years of the first attack. With increasing urate deposition in affected joints, attacks increase in duration, frequency and severity, become polyarticular and resolve more gradually. Patients remain well between attacks.

Chronic Gout

If untreated, gout eventually becomes chronic rather than episodic because of increasing urate deposition in joints. Chronic gout is characterized by a low-grade inflammatory arthritis punctuated by acute attacks. Tophi are often present.

Extra-articular Involvement

If untreated, after 20 years, 55–72% of patients with gout will develop tophi and 30% will develop nephrolithiasis. Tophi are due to deposition of urate in the skin, and appear as yellow or chalky white, firm nodular or fusiform swellings, developing after an average of 11.6 years (range 3–42 years) of gout. They are related to the degree and persistence of hyperuricemia, and are therefore a marker for more severe disease. Tophi are classically found over the helix of the ear, but are found more commonly over the toes and fingers. They contribute to the destructive arthritis of chronic gout, may ulcerate, become secondarily infected, and may occur in unusual sites (e.g. spinal cord, myocardium, eye). Tophi can be differentiated from rheumatoid nodules by their distinctive yellow or chalky white discolouration, or

by demonstrating the typical negatively birefringent needle-shaped crystals of monosodium urate in material aspirated from a tophus.

Urate deposition in the kidney may occur in the interstitial medullary tissue, resulting in urate nephropathy, or in the tubules, pelvis or ureter, resulting in uric acid calculi or acute uric acid nephropathy. Urate nephropathy is characterized by deposition of urate crystals in the medulla with a distinctive giant cell reaction on renal biopsy, and may be associated with mild albuminuria. Although a distinct entity, urate nephropathy is not thought to have a significant effect on renal function. Uric acid calculi are present in 10–20% of subjects with primary gout, related to the degree of hyperuricemia and urinary uric acid excretion. In patients with secondary gout, particularly related to rapid tumor lysis from cytotoxic drug use, uric acid calculi may develop in up to 40% of subjects. Acute uric acid nephropathy also occurs in this setting, with precipitation of uric acid in the distal tubules and collecting ducts (the site of maximal urinary acidification and concentration). Calcium containing stones, particularly calcium oxalate, are 20–30 times more common in patients with gout.

Associated Conditions

Hyperuricemia and gout are associated with several conditions. *Insulin resistance* has been associated with hyperuricemia, and may contribute to the association of gout with cardiovascular disease. Hyperinsulinemia and insulin resistance occur in 95% and 76% of patients with gout respectively.[3] *Obesity* is associated with dietary excesses, which in turn predispose to hyperuricemia and gout in both Caucasian and Chinese populations. *Hyperlipidemia*, most commonly hypertriglyceridemia, is also associated with gout, and may be related to insulin resistance[3] or reduced activity of lipases. *Hypertension* is present in up to a third of patients with gout, and may be associated with gout in several ways. Diuretic therapy for hypertension causes hyperuricemia and so predisposes to gout; hypertension *per se* reduces renal urate excretion, possibly through a reduction in renal blood flow; renal deposition of microtophi may cause renal damage leading to renal hypertension; and excessive alcohol intake predisposes to both hypertension and gout. *Ethanol* increases uric acid production by increasing adenosine triphosphate turnover and decreasing renal uric acid excretion. In a hospital-based

series of patients with gout in Singapore, hypertension was found in 36%, hyperlipidemia in 25%, ethanol consumption in 44% and ischemic heart disease in 13% of patients.

Hyperuricemia has been associated with cardiovascular disease and increased cardiovascular and all-cause mortality. For example, in a prospective study of 49 413 Japanese male railroad workers, subjects with serum uric acid exceeding 505 μmol/L (8.5 mg/dL) was associated with an increased relative risk (RR) of death from all causes (RR 1.62), coronary heart disease (RR 1.52), stroke (RR 2.33), hepatic disease (RR 3.58) or renal failure (RR 8.52), when compared with subjects with serum uric acid levels of 300 to 380 μmol/L (5.0 to 6.4 mg/dL). This observation is not surprising, given the associations of hyperuricemia with known risk factors for cardiovascular disease. It is however unclear if hyperuricemia is merely a marker for these risk factors or itself is an additional, independent risk factor, as several large epidemiological studies have yielded conflicting results. In practice, patients with hyperuricemia should be assessed for presence of cardiovascular risk factors and managed accordingly.

DIFFERENTIAL DIAGNOSIS

The diagnosis of gout should be considered in the setting of an acute monoarthritis or episodic mono- or oligo-arthritis.

Acute Monoarthritis

Gout often first presents as an acute inflammatory monoarthritis. In this setting, the differential diagnosis[7] encompasses other crystal arthropathies (e.g. pseudogout), and most importantly, septic arthritis. Diagnostic joint aspiration is the investigation of choice. Synovial fluid should be sent for cell and differential counts, polarized light microscopy, gram stain and pyogenic culture. As acute gout is uncommon in premenopausal women and patients with rheumatoid arthritis, an acute monoarthritis in these settings is more likely to be due to septic arthritis. Acute monoarthritis of the metatarsophalangeal joint of the big toe is rarely due to septic arthritis and may be treated with NSAIDs, with joint aspiration reserved for cases that do not respond to treatment. In acute gout, examination of synovial fluid from an inflamed joint yields intracellular, negatively birefringent needle-shaped crystals in synovial fluid

(typical of MSU) in virtually all patients.[4] The presence of extracellular MSU crystals is not specific for gout, as they are seen in 5% of subjects with AH and up to 20% of subjects with renal failure with no history of arthritis. Cell counts in acute gout may exceed 50 000/mm^3 and show a neutrophil predominance, a pattern commonly seen in septic arthritis. As gout and septic arthritis can co-exist, a patient with gout who fails to respond to therapy should be re-evaluated for septic arthritis.

History of Episodic Mono- or Oligoarthritis

Patients are usually well between attacks of gout. If evaluated between attacks, the diagnosis of gout is suggested by the typical history of episodic mono- or oligoarthritis affecting lower extremity joints, and confirmed by the presence of negatively birefringent crystals typical of MSU on aspiration of a joint or tophus. In patients who have not received urate lowering therapy, examination of synovial fluid from an asymptomatic but previously inflamed knee joint yields typical MSU crystals in up to 97% of cases, while synovial fluid from a previously unaffected knee joint yields MSU crystals in up to 22%. A lower yield of MSU crystals is found in patients previously treated with urate lowering medications. Recurrent self-limiting attacks of first MTP arthritis are typical of gout, and generally do not require joint aspiration.

INVESTIGATIONS

Investigations are useful to confirm the diagnosis of gout, and to determine the cause of hyperuricemia. The diagnosis of gout is established by demonstrating (using polarized light microscopy) the negatively birefringent needle-shaped crystals typical of MSU in synovial fluid. Elevated serum urate levels are suggestive but not diagnostic of gout, as uric acid levels may be normal in up to 25% of proven cases of acute gout, and are often raised in patients with arthritis from causes other than gout. Serum urate levels do, however, reflect the severity of tissue urate deposition, and are therefore useful in titrating the dose of urate lowering therapies.

Investigations to determine the cause of hyperuricemia should include a full blood count (raised cell counts suggest myeloproliferative disorders and a raised mean corpuscular volume suggests alcohol use) and a serum creatinine level (raised in renal failure). Other investigations

to determine the cause of hyperuricemia should be guided by clinical evaluation (Table 2) and might include thyroid function tests, serum calcium levels and urine lead levels etc. It is often helpful to determine if the hyperuricemia is due to uric acid over-production or under-excretion by performing a 24-hour urinary uric acid collection while the subject is on a low purine diet.[8] Overproducers of urate have a 24-hour urinary urate excretion of > 4.5 mmol while on a low purine diet, and a normal or increased urinary urate clearance (normal value 4 to 14 mL/min).[8] A specific enzyme defect can however be identified only in a minority of these patients.[9] Underexcretors of urate have a 24-hour urinary urate excretion of < 2 mmol on a low purine diet,[8] a low urate clearance or a urate/creatinine ratio of < 0.07.

Other than soft tissue swelling during acute gout, radiographic features of gout usually develop after 6 to 8 years of disease. Eccentric soft tissue prominences due to deposition of urate in soft tissue may be seen. Erosions are due to tophus or pannus formation, and are typically punched out, with sclerotic margins, an overhanging margin of bone, and may be intra-, peri- or non-articular. Other radiographic features include the relative preservation of joint space and lack of periarticular osteopenia. An intravenous urogram or sonographic evaluation is needed to identify urate calculi, which are radiolucent.

TREATMENT

The aims of treating gout are to abort acute attacks rapidly, prevent recurrent attacks of gout and prevent or reverse complications of urate deposition in the joints and kidneys. Effective treatments are available for gout. Modification of dietary habits and reversible causes of hyperuricemia apply in the treatment of all patients. Pharmacological therapies are tailored to treat the acute attack, to prevent attacks and to lower serum urate concentrations.

Modification of Diet and Reversible Causes of Hyperuricemia

Dietary modification alone may lower serum urate levels by 60–120 μmol/L[3,5] and may reduce the frequency of gouty arthritis.[10] General measures to reduce hyperuricemia include reduction in dietary purines, treatment of obesity, cessation of alcohol and drugs associated

with hyperuricemia and control of co-existing hypertension. Traditionally, dietary advice in patients with gout has emphasized a low purine/protein diet. Table 3 lists foods with high and low purine content. Patients are generally advised to avoid organ meats, excessive amounts of meat and seafood, and should be informed that consuming large amounts of food with low purine content may result in a higher purine load than consuming small amounts of food with high purine content. The value of this traditional low purine diet in gout has been called into question recently.[3] Low purine diets tend to be rich in carbohydrates and saturated fat,[3] which worsen insulin resistance in subjects with concomitant gout and insulin resistance. Increasing insulin resistance leads to higher serum insulin levels, which in turn result in increased serum urate levels by increasing renal resorption of urate. Interestingly, in a pilot study in 13 male patients with gout, a diet designed to lower insulin resistance and reduce weight (1600 kcal/day, 40% from carbohydrates, 30% each from protein and fat) but with no restriction in purine-rich foods reduced both serum uric acid levels and the frequency of gouty arthritis.[10]

Asymptomatic Hyperuricemia

AH itself does not cause a decline in renal function, and treatment of AH with urate lowering drugs does not protect against development of renal insufficiency.[4] The aim of treating AH is to reduce the risk of developing gout by correcting reversible causes of hyperuricemia (Table 2) using non-pharmacological interventions. Urate lowering drugs (usually

Table 3 Purine Content of Common Foods

Moderate/High Purine Content
Meats in excess, especially organ meats and seafood
Yeast and meat extracts
Alcoholic beverages
Legumes: beans, peas, lentils, oats
Spinach, mushrooms, asparagus, cauliflower
Low Purine Content
Fruits
Lettuce, green vegetables (other than those listed above)
Cereals and cereal products (e.g. bread)
Dairy products
Eggs

allopurinol) should only be considered if there is a high risk of urate nephrolithiasis (i.e. when 24-hour urinary uric acid excretion exceeds 6.5 mmol/day and serum uric acid persistently exceeds 600–720 μmol/L), especially in the context of decreased renal function.[4] In other situations, the adverse effects of urate lowering drugs outweigh the benefits.

Treating the Acute Attack

Non-pharmacological treatments include resting and cooling the affected joint/s. In an unblinded, randomized trial of 19 patients with acute gout, application of ice packs for 30 minutes, 4 times a day to affected joints lead to greater reduction in pain and joint swelling than was seen in a control group who received identical medications but did not use ice packs.

Pharmacological treatments (Table 4) include the options of non-steroidal anti-inflammatory drugs (NSAIDs), oral colchicine or corticosteroids (oral or intra-articular). The choice of which drug to use depends on the risk to benefit ratio in a particular patient. NSAIDs improve symptoms within 1 day and are the treatment of choice if not contraindicated (Table 4), or if the patient is not at high risk of developing a peptic ulcer (i.e. age $<$ 75 years; no past history of peptic ulcer disease, gastrointestinal bleeding or cardiac disease; no concomitant steroid or anti-coagulant use; and no functional disability). Oral colchicine or corticosteroids may be used if NSAIDs are contraindicated. Colchicine is effective in 90% of patients if used within 24 hours of the onset of an attack, and in 75% of patients if used after more than 24 hours.[11] It relieves pain in most patients within 18 hours of ingestion, but the majority of these patients develop diarrhea within 24 hours of taking colchicine. Because of this narrow therapeutic index compared with NSAIDs, colchicine should only be used when NSAIDs are contraindicated. Intravenous colchicine is highly effective in the treatment of gout but has potentially fatal side-effects, and should only be given by those experienced in its use. It is important to remember that both colchicine and NSAIDs are contraindicated in patients with renal or hepatic insufficiency and in the very elderly. In such patients, corticosteroids may be used. When 1 or 2 joints are affected, intra-articular corticosteroids provide rapid relief of symptoms and avoid the need for oral corticosteroid administration. When more than 2 joints are affected, a short course of oral corticosteroids (0.25 to 0.5 mg/kg/day depending on the severity of the attack and the health of

Table 4 Medications Used in the Treatment of Gout

Medication/Dosage	Adverse Effects	Contraindications	Pharmacokinetics	Comments
Non-steroidal anti-inflammatory drugs (NSAIDs)	Peptic ulcer disease, renal insufficiency	Peptic ulcer disease, renal, liver, cardiac failure	Varies among NSAIDs	Treatment of choice for acute gout
Colchicine *Acute attack*: 1 mg initially, then 0.5 mg every 1 to 2 hours till pain is relieved, a total of 8 mg has been administered or GI symptoms develop. Maximum of 10 mg/week of colchicine. *Prophylaxis*: 0.5 to 1.0 mg/day	Vomiting, diarrhea, hemorrhagic gastroenteritis, bone marrow suppression, confusional state, seizures, myopathy, alopecia	Renal or hepatic dysfunction (decreased clearance of colchicine leads to accumulation and toxicity)	Inhibits phagocytosis of urate crystals by neutrophils Hepatic and renal excretion; Long half life (because of enterohepatic circulation)	Increased toxicity if drugs inhibiting cytochrome p-450 are taken concomitantly Myopathy may be more common with concomitant use of statins or cyclosporin.
Allopurinol 100 to 800 mg/day as a single dose (as half-life of oxypurinol is 30 hrs)	Rash in 3%, at times heralding Allopurinol Hypersensitivity syndrome (see text) Marrow suppression may result if used with Azathioprine and 6-mercaptopurine Renal and hepatic toxicity (rare)		Inhibits xanthine oxidase Metabolized to oxypurinol, which also inhibits xanthine oxidase	Allopurinol allergy is increased 10-fold in penicillin allergic patients Maximum effect on urate levels seen after 4–14 days Oxypurinol excretion increased by uricosuric agents

Probenicid (half-life 6–12 hrs) Start with 250 mg bd, increasing to 0.5 to 1.0 gm tds.	Urate renal calculi, gastrointestinal symptoms, rarely peptic ulceration, hemolytic anemia (if G6PD deficient)	Overexcretors (urate calculi formation) History of renal calculi	Uricosuric	Drug interactions (see text) Salicylates decrease effectiveness Reduce dose if 24-hr urate excretion $>$ 4.5 mmol/day to prevent urate calculi Ineffective if creatinine clearance $<$ 50 mL/min
Sulfinpyrazone (half-life 3–6 hrs) Start at 50 mg bd, increasing to 100 mg tds- qds. Maximum dose 800 mg/day.	Nausea, vomitting, leucopenia, interstitial nephritis	Overexcretors (urate calculi formation) History of renal calculi	Potent uricosuric	Also inhibits platelet function, preferred uriscouric in patients with coronary artery disease Ineffective if creatinine clearance $<$ 50 mL/min
Benzbromarone 40–200 mg/day (half-life 3 hrs)	Diarrhea		Potent uricosuric	Effective if creatinine clearance $>$ 20 mL/min, used in transplant gout Limited availability

the patient, given in divided doses) with a rapid taper over 1 to 2 weeks once the acute attack is subsiding is generally effective.

Practice Points in Treating an Acute Attack. Medications are most effective in reducing inflammation when given early in the course of an attack, and should be given as soon as acute gout is suspected. If a patient is on urate-lowering drugs when an acute attack occurs, these drugs *should not* be ceased, as the resultant rise in urate levels may worsen or prolong the acute attack. If a patient is not on urate-lowering drugs when the acute attack occurs, these drugs should only be started two weeks after an acute attack has subsided, as the fall in urate levels which occurs when urate-lowering therapy is initiated may worsen or prolong an acute attack. Attacks of gout, which are refractory to one of these medications, often improve with the addition of a second medication.[12] The combination of NSAIDs and oral corticosteroids should be used with caution, especially in the elderly, because this combination carries a 15-fold increased risk of peptic ulceration in the elderly. If the acute attack does not show signs of subsiding after 48 hours of treatment, the patient should be re-evaluated for other possible conditions mimicking an acute gouty flare, especially septic arthritis and other crystal arthropathies.

Acute attacks of gout may be prevented by reducing urate deposition in joints through the use of urate-lowering drugs. Urate-lowering drugs are indicated for the prevention of acute attacks if these occur frequently (some authorities suggest >5 times per year) or severely affect a patient's lifestyle.

Preventing Attacks (Prophylaxis)

Both colchicine and low dose NSAIDs are effective in preventing recurrent attacks of gout. Colchicine is thought by some authors to have less adverse effects than NSAIDs with long-term use, and is less costly, but no comparative studies have been performed to date. Attacks of gout are prevented by colchicine 0.5 mg once a day in 27% of patients, by colchicine 0.5 mg twice a day in an additional 63% of patients and by colchicine 0.5 mg three times a day in the remaining 10% of patients.[13] A combination of colchicine and NSAIDs may be useful if attacks persist while on either drug. The initiation of urate-lowering therapy increases the frequency of gouty attacks in 10–24% of subjects. Prophylactic colchicine or NSAID therapy is often given to prevent these attacks. The optimal

duration of such therapy is unclear, but most authors recommend a period of 6–12 months, as acute attacks generally cease 6–12 months after the serum urate level has normalized.

Lowering Serum Urate Concentrations

Urate-lowering drugs act by reducing uric acid production (allopurinol) or increasing urate excretion (uricosuric agents, e.g. probenecid, sulfinpyrazone, benzbromarone). The resultant fall in serum urate favors dissolution of monosodium urate crystals in joints and other tissues, reducing the tissue load of urate and thus the frequency and severity of attacks of gout. Urate-lowering drugs are indicated in the presence of tophi, recurrent urate renal calculi, urate nephropathy or frequent attacks of gout (>5 per year). Clinical experience suggests that attacks of gout cease after 6–12 months when the serum uric acid level is kept below a target level of 360 μmol/L. A target serum uric acid level of less than 300 μmol/L is generally necessary for resorption of tophi. In general, uricosuric agents should be used in subjects who under-excrete uric acid, and allopurinol in subjects who over-produce uric acid, who have renal impairment or uric acid renal calculi or in whom urate levels are elevated despite treatment with uricosurics.

Uricosurics are generally started in low doses to minimize the risk of urate calculi associated with increased urate excretion. This risk can be further reduced by alkalinizing the urine (e.g. with mist sodium bicarbonate or Shohl's solution) and maintaining adequate urine flow by ingesting ≥ 3 L of fluids a day. By reducing the transport of other organic acids across cell membranes, probenicid increases serum levels of penicillins, indomethacin, sulphonylureas, heparin, dapsone, rifampicin and reduces serum levels of allopurinol.[14] Aspirin was previously thought to negate the uriscouric effect of probenecid, but more recent data has shown that the addition of aspirin 325 mg per day in subjects taking stable doses of probenecid did not significantly influence serum or urinary urate levels. *Allopurinol* is associated with a rare but often fatal allergic reaction termed the allopurinol hypersensitivity syndrome (AHS). This syndrome is characterized by fever, a rash (toxic epidermal necrolysis, erythema multiformae or generalized exfoliative dermatitis), eosinophilia, leucocytosis, hepatitis, worsening renal function, other visceral involvement and vasculitis, and typically develops after 6 weeks

(range 1–728 days) of allopurinol use. It is fatal in up to 27% of patients, and is more common in patients with renal insufficiency, possibly due to the accumulation of oxypurinol. It has therefore been suggested (but not universally accepted) that allopurinol doses should be adjusted for renal function (100 mg/day if creatinine clearance is 30 mL/min, 200 mg/day if creatinine clearance is 60 mL/min and 300 mg/day if creatinine clearance is 100 mL/min). Allopurinol should be ceased in patients who develop a drug rash. Desensitization of patients with allopurinol allergy has been successfully performed, but should only be attempted if other options to lower urate levels have failed. In Singapore, AHS-related mortality occurred in 13% of patients in one series, and the indication for allopurinol use was unclear in over 50% of subjects with AHS.

In practice, hyperuricemia in many patients is due to the combination of over-production and under-excretion of urate. In this situation, allopurinol is effective in lowering serum urate levels. Urate-lowering therapy should be continued in patients whose symptoms have been controlled by such therapy. Cessation of urate-lowering therapy in 21 patients with tophaceous gout whose tophi had resolved with such therapy, resulted in the recurrence of gout and tophi in 17 and 9 patients respectively after a mean of 19 months (range 4–52 months) and 38 months (range 4–107 months) respectively.

Extra-articular Manifestations

Tophi, depending on their size, generally resolve over 6 or more months with maintenance of serum urate levels below 300 μmol/L. Surgical removal of tophi is generally inadvisable, and in a series of 45 patients, was associated with delayed wound healing (>1 week) in 50% of patients, especially those with infected tophi at the time of surgery. Urate nephropathy at times improves with urate-lowering therapy. Acute uric acid nephropathy can be prevented by prophylactic administration of allopurinol before chemotherapy is given; in the event that acute uric acid nephropathy still develops, hemodialysis is preferred to peritoneal dialysis as the former is 10–20 times more effective in removing uric acid from the circulation. Urate calculi are treated with allopurinol (which decreases urinary uric acid excretion), hydration and alkalinization of urine to a pH of 6.0 to 6.5.

Indications for Specialist Evaluation

Gout is common, and the vast majority of patients can be managed by primary care physicians. Referral for specialist evaluation may be required if attacks of gout persist despite standard therapy, if renal impairment or renal calculi are present, or if a secondary cause for hyperuricemia (e.g. a myeloproliferative disorder) is identified.

SPECIFIC CLINICAL SITUATIONS

Gout in the Elderly

Gout in the elderly may present atypically,[15] with a higher prevalence of polyarticular gout, a propensity to affect the small joints of the hands, and a higher rate of occurrence in females. These features, if associated with the ongoing low-grade inflammation seen with long-standing gout, may mimic late-onset rheumatoid arthritis. Other atypical features are the early development of tophi in unusual locations (especially the finger pads), a strong association with diuretic use and the development of gout in preexisting osteoarthritis of the hands. In elderly patients with gout, colchicine is favored over NSAIDs because the latter have increased side effects in this age group, especially GI bleeding.[9] Allopurinol should be used with caution in the elderly if renal impairment is present, because of the increased risk of developing the allopurinol hypersensitivity syndrome.

Gout in the Young (Onset $<$ 30 Years of Age) and in Premenopausal Women

The development of proven gout in these groups of patients suggests the presence of a genetic defect leading to increased purine production or decreased renal excretion of uric acid.[9] A family history of gout is found in 50% with onset under 25 years of age and in 80% with onset between 12 and 19 years of age. Early identification of these subjects is useful, as treatment may prevent joint and renal damage.

Transplant Gout

Post-transplant hyperuricemia and gout are related to diuretic use and cyclosporin-mediated inhibition of uric acid excretion. They are common

after renal and cardiac transplants but are uncommon after liver transplants.[12] Transplant gout may be severe and rapidly progressive, with polyarticular involvement and early development of tophi, which classically appear like drops of chalky material under the skin ("liquid tophi"). NSAIDs and colchicine (if not contraindicated because of renal impairment) may be used to treat acute attacks. Urate-lowering therapy is often required, and is complicated by resistance to uricosuric agents (due to cyclosporin or renal disease) and marrow suppression related to allopurinol inhibition of azathioprine metabolism. Benzbromarone may be the drug of choice in such situations, as leucopenia is common with allopurinol therapy even if the azathioprine dose in such patients is reduced by two-thirds (as is generally recommended).

ACKNOWLEDGMENTS

The author would like to thank SRN Thio Szu-tien and Ms Wee Hwee Lin for thoughtful comments on the manuscript.

REFERENCES

1. Kelly WN, Schumacher HR, Gout, in: *Textbook of Rheumatology*, Kelly WN, Harris ED, Ruddy S, Sledge CB (eds.), Philadelphia, pp. 1291–1336, 1993.
2. Wortmann RL, Gout and hyperuricemia, *Curr Opin Rheumatol* **14**:281–286, 2002.
3. Fam AG, Gout, diet and the insulin resistance syndrome, *J Rheumatol* **29**:1350–1355, 2002.
4. Kelly WN, Wortmann RL, Hyperuricemia, in: *Textbook of Rheumatology*, Kelly WN, Harris ED, Ruddy S, Sledge CB (eds.), Philadelphia, pp. 498–506, 1993.
5. Emmerson BT, The management of gout, *New Eng J Med* **334**:445–451, 1996.
6. Annonymous, *Lancet*: 703–704, 1993.
7. Hudson NP, Acute monoarticular arthritis, in: *Diagnostic Strategies for Common Medical Problems*, Robert J. Panzer, Edgar R. Black, Paul F. Griner (eds.), American College of Physicians, Philadelphia, PA, PP. 318–324, 1991.
8. Emmerson B, Identification of the causes of persistent hyperuricemia, *Lancet* **337**:1461–1464, 1991.
9. Dieppe PA, Investigation and management of gout in the young and elderly, *Ann Rheum Dis* **50**:263–266, 1991.
10. Dessein PH, Shipton EA, Stanwix AE, Joffe BI, Ramokgadi J, Beneficial effects of weight loss associated with moderate calorie/carbohydrate restriction, and increased

proportional intake of protein and unsaturated fat on serum urate and lipoprotein levels in gout: a pilot study, *Ann Rheum Dis* **59**:539–543, 2000.

11. Aherns MJ, Reid C, Gordon TP, McCredie M, Brooks PM, Jones M, Does colchicine work? The results of the first controlled study in acute gout, *Aust N Z J Med* **17**:301–304, 1987.

12. Rosenthal AK, Ryan LM, Treatment of refractory crystal-associated arthritis, *Rheum Dis Clin North Am* **21**:151–161, 1995.

13. Wallace SL, Singer JZ, Therapy in gout, *Rheum Dis Clin North Am* **14**:441–457, 1988.

14. Fox IH, Antihyperuricemic drugs, in: *Textbook of Rheumatology*, Kelly WN, Harris ED, Ruddy S, Sledge CB (eds.), Philadelphia, pp. 822–831, 1993.

15. Agudelo CA, Wise CM, Crystal-associated arthritis, *Clin Geriatr Med* **14**:495–513, 1998.

This page intentionally left blank

87

Osteoarthritis

Colleen Kim Thomas

DEFINITION AND INTRODUCTION

Osteoarthritis (OA) refers to the joint condition of degenerative articular cartilage damage. Some favor the term osteoarthrosis to reflect the generally non-inflammatory nature of the condition. Others prefer the phrase degenerative arthritis, since the prefix osteo- has little meaning with regard to descriptive value of the pathological process.

In terms of disease burden, it is by far the most common joint disorder and is a major cause of pain and poor mobility in the geriatric population. At least one-third of those above 65 are affected radiologically.

The weight bearing joints, mainly knees and hips, as well as finger joints, have a propensity to be affected. Lumbar or cervical spondylosis refer to OA affecting the facet or apophyseal joints of the low back and neck respectively. Spinal degenerative disease, however, is a large topic by itself that is outside the scope of this chapter and will not be discussed.

Table 1 Secondary Causes of Osteoarthritis

Post injury
Chronic overuse
Joint dysplasias
Hypermobility
Hemarthrosis, e.g. hemophilia
Post-inflammatory damage: 1) Inflammatory arthropathies, e.g. rheumatoid arthritis
2) Septic arthritis
3) Crystal arthropathies, e.g. gout, pseudogout
Osteonecrosis (avascular necrosis)
Osteochondritis
Endocrine disorders affecting bone growth, e.g. acromegaly
Metabolic disorders leading to abnormal joint deposition, e.g. hemachromatosis, ochronosis, mucopolysacharidoses

CAUSES

OA may be "primary", that is, degenerative change that is consistent with age, and predisposed to by genetic differences in collagen quality, obesity and usage, or "secondary", that is accelerated by a superimposed pathological process or traumatic event that leads to premature cartilage loss (Table 1).

RISK FACTORS FOR PRIMARY OSTEOARTHRITIS

The most simplistic consideration of primary osteoarthritis would be age-related cartilage degeneration worsened by abnormal loading such as increased use and body mass. However, striking epidemiological/gender specific differences in incidence, age of onset and joints affected indicate genetic/hormonal factors affecting cartilage resistance to degradation.

Age

OA prevalence increases markedly with age; from 1% in those under 30 to over 30% of those above 65 years.[1]

Obesity

Correlation of obesity with osteoarthritis is well demonstrated especially for knee OA.

Occupation

Jobs entailing overuse of a particular joint will predispose to OA. Sportsmen fall in this group and may have secondary OA from injury as well. Another well-quoted example is that of jackhammer operators suffering OA in the elbow, a joint uncommonly affected to a significant degree.

Gender/estrogen status

The pre-menopausal privileged status of estrogen adequacy, besides preserving the heart and bone density, is also protective to joints. OA occurs more frequently in women in the post menopausal age group compared to men of similar age, and more so in those not on estrogen replacement therapy.[2] Women tend to have more joints affected, and generalized OA is also more common in women. Overall prevalence is equal for males and females because of an excess of males in the younger age group contributed by post-traumatic causes, whereas females in the post-menopausal age group outnumber males of a similar age. OA of the knees is more common in women, while hip OA is more common in men.

Joint dysplasia

Genu varum or valgum, slipped capital femoral epiphysis, Perthes disease, acetabular dysplasia are all causes of secondary OA. However undetected minor subclinical degrees of these conditions may be responsible for some cases designated idiopathic or primary hip OA.

Race

Hip osteoarthritis is rare in Southern Chinese[3] and Africans, thought to be related to the rarity of congenital hip disorders, slipped upper femoral epiphysis and Perthes'disease in these populations.

Hypermobility

Patients with various degrees of hypermobile joints are predisposed to OA later in life, due to cartilage damage caused by minor degrees of frequent subluxation.

Family history

Generalized osteoarthritis is noted to run in families. OA of the hands has a heritability of up to 65%[4] and is considered a familial disease with a polygenic inheritence.

Bone density

Patients with osteoarthritis have higher bone density than those without. One reason could be that obesity predisposes to OA but has a positive effect of bone mass.

PATHOGENESIS OF OSTEOARTHRITIS

From a simplistic understanding of OA being caused by chronic mechanical stress leading to cartilage disintegration, there has been a surge of research in recent years towards the elucidation of the complex degenerative cascades at molecular level.

Normal cartilage tissue consists of chondrocytes which are embedded within the extracellular matrix they secrete. Matrix components comprise collagen and proteoglycans chief of which is aggrecan. Proteoglycans have the property of retaining water due to their content of chondroitin sulphate and keratan sulphate and are responsible for resistance to loading and shock-absorbing properties.

Proinflammatory cytokines such as interleukin-1 (IL-1), tumor necrosis factor-(TNF-α), prostaglandins, free radicals and damaged matrix components are released from chondrocytes and synovial cells as a result of cartilage damage.[5] These trigger the synthesis of the cartilage degrading proteinases upon binding to specific chondrocyte receptors. These enzymes belong mainly to the zinc containing matrix metalloproteinase (MMP) family. A major player in this family is the MMP aggrecanase. The other group of important proteinases are the cathepsins.

Natural inhibitors of these proteinases (tissue inhibitor of metalloproteinases TIMP) keep this process in check.

In OA, there is a shift towards increased cartilage degradation, due to factors such as chronic joint pressure increase, hormonal changes after menopause, as well as age-related changes in chondrocyte metabolic activities.

There is some evidence that limited cartilage repair activities do take place in OA affected joints, mediated by growth factors produced by chondrocytes, synovial cells and subchondral bone.

These are all being intensely studied in the hope of identifying the pathways to switching off cartilage degradation and promoting cartilage synthesis, thus moving towards a true cure for OA at the molecular level.

PATHOLOGY OF THE OA JOINT

With micro-injury caused by inordinate stresses and age related chondrocyte changes, cartilage quality declines. The damaged cartilage retains water and swells. Microfissures appear, which deepen and are manifested as surface fibrillation, the first obvious arthroscopic feature in early OA. Loose bodies may result from fragments that become totally detached. The defect left by the loss of these fragements may give rise to periarticular microcysts. Closer apposition of bony surfaces causes subchondral sclerosis. The final outcome of this process is ulceration to the underlying bone.

Besides cartilage changes, there is also ligamentous strain and bursitis from deranged joint biomechanics. Mild synovial inflammation with lymphoplasmocytic and histiocytic infiltration is also found. Degenerative ligamentous and meniscal tears can occur.

Osteophytosis occurs in areas of increased biomechanical strain in the weakened joint.

THE CASE FOR AN INFLAMMATORY COMPONENT IN OA

Joint fluid analysis, while generally non-inflammatory, with cell counts less than 200/mm^3, may be mildly inflammatory with counts up to 2000/mm.^{3} This, together with the fact that some cases respond to intra-articular steroid injection and/or NSAID use, is indicative of a minor

inflammatory component in at least some cases. Synovial hypertrophy is found in nearly 3/4 of patients, and is also presumed to be an inflammatory response.

In cases where there seems to be an increased inflammatory component, hydroxyapatite crystal-induced inflammation may play a role. Calcium hydroxyapatite, being visible only by electron microscopy, does not figure in the routine detection of crystal arthropathy using conventional polarizing light microscopic techniques; however this type of crystal arthropathy may account for mildly inflammatory cases diagnosed as osteoarthritis.

The role of inflammatory cytokines in perpetuating osteoarthritic joint destruction is recent thinking that is different from traditional concepts of a purely passive mechanical joint degradation. However, still the basic pathological starting point of cartilage degeneration caused by mechanical injury beyond the tensile strength of cartilage still holds. The local nature of the disease is evidenced by the lack of constitutional symptoms and markers of systemic inflammation such as C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR).

PATTERN OF JOINT INVOLVEMENT IN OA

Primary OA affects the following joints in a bilateral fashion:

1) *Knees*
2) *Hips*
3) *Hands*: distal interphalangeal, proximal interphalangeal joints, carpometacarpal joints of the thumb
4) *Feet*: metatarsaophalangeal joint of the big toe
5) *Spine*: facet joints of the cervical and lumbar spine

Involvement of other joints is to be regarded as atypical and secondary sources should be sought; e.g. post traumatic ankle OA, rheumatoid shoulders, calcium phosphate dihydrate deposition in wrists.

HAND OA

Middle-aged women are usually affected, who present with pain and bony hard swelling of the distal interphalangeal joints called Heberden's

nodes, and/or proximal interphalangeal joints called Bouchard's nodes. Involvement of the carpometacarpal joint of the thumb produces a characteristic squaring of the hands. Cystic outpouchings along the joint line may be seen which when aspirated yield a clear viscous joint fluid. This type of OA has a familial tendency.

Occasionally, there may be a fairly rapid progression of these joint changes within months, beyond what can be explained by chronic wear and tear alone. This has been termed "inflammatory OA" by some, although there is no evidence of acute phase reactant elevation. Deformities that develop not uncommonly get misconstrued as rheumatoid arthritis. This little understood process may well constitute a separate subset of OA.

SYMPTOMS

Pain

The main and eventually most disabling symptom in OA is the pain that develops gradually often over years. The joint pain is worsened by using the involved joints and relieved by rest. For hip and knee osteoarthritis, pain might be exacerbated on prolonged standing and walking, and especially using stairs. Pain at rest and night symptoms are not typical, but suggest an inflammatory, septic or neoplastic lesion.

There has been much research and postulation as to the genesis of pain in OA. Subchondral microfractures have been demonstrated, as well as venous congestion and increased intraosseus pressure. Osteophyte irritation of nerve endings and direct periosteal loading due to cartilage denudation are other possibilities.

The pain from hip arthritis is generally located in the groin and may radiate to the anterior thigh or knee. Pain over the lateral aspect of the hips, however, comes from local causes like trochanteric bursitis and not the hip joint itself.

Radicular distribution of pain, weakness and numbness caused by foraminal stenosis from osteophytes at facet joints in lumbar spondylosis can mimic hip OA. Furthermore radiographs in the elderly may reveal degeneration in both regions. It is important to distinguish the two especially where surgical intervention is planned, since imaging abnormalities

alone may not correlate with symptoms. Hip pain generally does not radiate below the knee and is reproduced during range of movement testing. Pain on spinal movements and neurogenic claudication point to a spinal pathology.

A sudden increase in pain in an osteoarthritic joint may be due to an acute event such as osteonecrosis. An area of localized pain could be from bursitis caused by abnormal joint mechanics, such as anserine bursitis of the knee and trochanteric bursitis of the hip. These should be identified and separately managed, with much relief obtained from intrabursal steroid injections as discussed in the treatment section.

Stiffness

Stiffness in the early morning, or after a period of rest and relieved by activity, is a hallmark of an inflammatory arthropathy. OA patients can also have a similar experience of a transient nature, lasting a few minutes or seconds. This is called the gel phenomenon. More than 30 minutes of stiffness, however, suggests an inflammatory pathology. Joint stiffness and pain of any pathology is often worse in cold and damp weather.

Instability

Buckling of the knee, especially with steps, may occur due to ligament degeneration, pain or muscular deconditioning.

Locking

This happens occasionally due to loose cartilage fragments or torn mensicus fragment jamming motion.

Limitation of Function

Loss in the range of motion is apparent in hip and knee OA when squatting and sitting cross legged, postures frequently assumed in the daily lives of Asian patients, becomes difficult. Dexterity in fine movements becomes affected in hand OA, especially when the first carpometacarpal or scaphotrapezoid joint are affected, making thumb opposition less efficient.

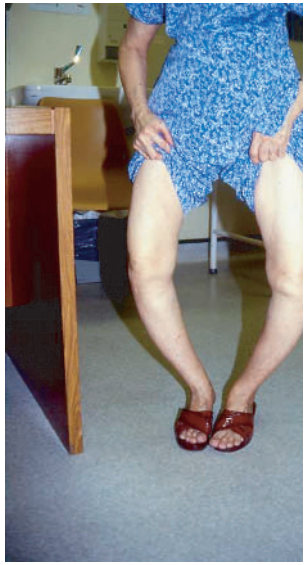


Fig. 1 An elderly lady with severe OA knees, joint remodeling and varus deformity. Photograph courtesy of A/Prof Lo Ngai Nung, Dept of Orthopedic Surgery, Singapore General Hospital.

Motion restriction is thought to be due to osteophyte formation and joint surface remodeling.

Signs

Deformity, typically, a varus deformity, occurs in the knee joint due to cartilage loss in the medial compartment (see Fig. 1).

Bony hard swelling, e.g. nodal enlargement of finger joints and bony enlargement of the knee joint, is characteristic, and is caused by bone remodelling.

Tenderness occurs at joint margins and tendon attachments.

Limitation of movement is confirmed on passive range of motion testing.

Crepitus is palpated on passive motion.

Periarticular muscle wasting, typically the quadriceps for knee OA, is due to disuse.

Effusion and synovial thickening reflect a mild inflammatory reaction to the degenerative damage. However, a markedly swollen and tender joint

may be due to crystal arthropathy or infection, both of which can occur on top of the degenerative process.

IMAGING

Plain Radiography

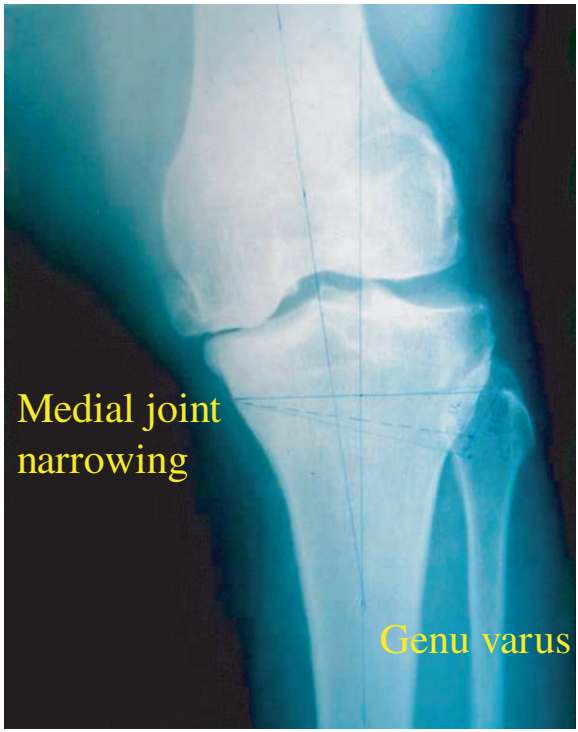


Fig. 2 Radiograph of OA knees showing osteophyte formation, subchondral sclerosis and medial joint space narrowing. Radiograph courtesy of Mr John Chen, Dept of Orthopedic Surgery, Singapore General Hospital.

Features seen in OA are:

- *Osteophyte formation* or bony proliferation at joint margins. This osteophytic lipping is a diagnostic feature of osteoarthritis.
- *Assymmetric joint space narrowing* affecting the compartment exposed to greatest stress e.g. lateral aspect of the hip joint, medial compartment

of the knee joint. This feature is an indirect reflection or cartilage volume loss. In contrast, the inflammatory arthropathies tend to affect joint spaces symmetrically.

- *Subchondral sclerosis* caused by abnormal bone loading at areas of cartilage thinning.
- *Subchondral cyst formation* with bony walls in the stressed area.
- *Alteration in the shape of bony ends* caused by bony remodeling.

Views to request:

- *Weight bearing films* especially of the knee joint reveal more information and should be requested rather than supine anteroposterior views. Joint space narrowing and varus/valgus deformities may not be otherwise demonstrated.
- *Tunnel views* through the knee joint are the alternative to best demonstrate abnormalities where the patient has difficulty standing.
- *Skyline views* are used to demonstrate patellofemoral osteoarthritis.

There may not be correlation between the degree of bony abnormalities and symptoms. Marked degenerative changes on radiographs may be associated with little pain and minimal changes and the reverse is also true. There are many theories as to the cause of osteoarthritic pain, as discussed in the clinical features section. Indications for joint replacement surgery must be based on symptoms and function rather than severity of bony changes.

Magnetic Resonance Imaging (MRI)

Plain radiographs delineate pathology after significant bony changes have occurred in moderately advanced disease. However, for soft tissue abnormalities, MRI is an excellent imaging modality that has revolutionized musculoskeletal imaging. It is superior to CT in its image quality and capacity to select any plane.[6]

In selected cases of OA, MRI is indicated and can replace diagnostic arthroscopy in delineating pathology, especially in cases of increased pain or locking in knee OA. This may be due to degenerative meniscal tears, which tend to occur in the posterior horn of the medial meniscus.

INVESTIGATIONS

Blood tests are unneccessary for the diagnosis of OA, which is can generally be made on clinical grounds supported by radiography. Nevertheless, panel tests for joint pain that comprise rheumatoid factor, ESR and CRP are often ordered. These should not be elevated in OA. If ESR and CRP are raised and the clinical diagnosis of OA is clear, other causes should be sought. A common mistake is to diagnose RA when the acute phase elevation is found in a patient with OA of the hands. Similarly, low titre RA factor may be seen in the elderly and RA should not be diagnosed when there are no clinical features for this.

Synovial fluid analysis in OA should usually return a cell count of below 2000 cells/mm^3, and is usually in the region of 200–300 cells/mm^3. If higher cell counts are obtained, septic, crystal or inflammatory arthritis need to be excluded.

TREATMENT

Weight Reduction

This is often difficult as the severe OA patients have a limitation to the amount of exercise tolerated. Major dietary changes may need to be made to address the low caloric demands of the sedentary lifestyle.

Physiotherapy and Occupational Therapy

With increased pain on movement and excessive loading contributing to OA, it may seem surprising that exercise is recommended therapy for joint maintenance.[7]

A healthy neuromucular system provides 30% greater shock attenuation. OA joints result in deconditioning and atrophy of associated muscles leading to greater loading on the joint. An exercise programme can improve muscle bulk, mobility and stability. Passive excercises such as quadriceps strengthening are combined with active low impact excercises, of which swimming is an excellent from, as the joint is supported in a weightless environment. Stair climbing and jumping are considered to be activities of maximal joint stress and should be minimized.

Walking aids and knee guards for joint instability, wedged insoles to reduce medial compartment knee pressure, patella taping, splints for OA of the thumb can be provided by a trained occupational therapist.

The aim of all these efforts is to maximize function while the patient learns to tailor living activities to the capacity of the OA joint.

PHARMACOTHERAPY

Treating OA pain can be a most difficult task for the clinician, who is faced with an indefinite chronic pain issue in a non-terminal patient who needs to be functional. Often these individuals are compelled to take daily analgesics on a long term basis, not without complications, especially with NSAIDs.

Topical Agents

Preparations such as topical NSAID gels and capsicum creams are useful in mild cases and can reduce the dosage of oral drugs needed.

Paracetamol

This safe drug should be maximized as a first-line oral agent.

NSAIDs and Coxibs

Non-steroidal anti-inflammatory agents (NSAIDs) prove superior in analgesic properties compared with paracetamol in some; as such, the clinician is often compelled to prescribe these on a long-term basis. Peptic ulcers, gastrointestinal hemorrhage and nephropathy are known complications. Before commencing these drugs, the patient should be clearly counselled as to the potential toxicities and to be alert for symptoms of melaena, hematemesis or epigastric pain. A baseline creatinine and hemoglobin level is advisable. NSAID's should be avoided in significant renal impairment.

Appropriate measures to reduce the adverse effects of NSAIDs include: emphasizing to the patient to use them only when absolutely necessary; minimizing the amount prescribed to the patient each visit, cycling the drugs with parecetamol, narcotics and drug-free days, and prescribing them with a H2 blocker, proton pump inhibitor or misoprostol. The use of antacids has no protective value against peptic ulcers. It is important to obtain a detailed drug history, as often patients have visited several practitioners and are unknowingly consuming multiple NSAIDs.

The new generation of selective COX-2 inhibitors (COXibs) are less toxic to the gastrointestinal tract but are much more costly.

Narcotic Analgesics

Drugs such as codeine phosphate provide a stronger analgesic effect than paracetamol. Constipation, sedation and tolerance are effects that must be monitored, especially in the elderly. Tramadol is useful agent in pain management with the advantage of less likelihood of tolerance.

Glucosamine and Chondroitin Sulphate

Manufacturers claim that these molecules[8,9] build up cartilage by providing the building blocks of matrix conponents. It has not been proven by radiolabelled studies that they are actually absorbed as an intact molecule or localize to the diseased joint. However, glucosamine has been shown to decrease the joint space narrowing.

These are classified as nutritional supplements and thus can be bought over the counter. As these are not controlled drugs, the quality and drug content of different brands vary widely. They are well-tolerated and both topical and oral preparations have been shown to reduce pain. The bovine source of chondroitin should be communicated to those who avoid beef for religious reasons.

These preparations also incur additional cost; and if no analgesic effect occurs after 6 months, it would be prudent to discontinue them.

MISCELLANEOUS SUPPLEMENTS

In recent years there has been interest in the search for drugs that could possibly retard the process of cartilage degeneration, as opposed to pain relief only. Candidate agents have been classified as disease modifying osteoarthritic drugs DMOAD's or chondroprotective agents. Drugs reported to have these effects include tetracyclines, chloroquine, tamoxifen, tranexemic acid, vitamin C and D. While these have not been proven to have sufficient therapeutic benefit for routine clinical use, research into potential DMOAD's may bring hope to future generations of OA sufferers.

ROLE OF INTRA-ARTICULAR INJECTIONS FOR OA

On occasion, intra-articular steroid injections give significant and prolonged benefit. This is especially true if inflammation is present and if there is little cartilage loss. If one or two injections do not give benefit, there is no reason to continue injections.

Hyaluronan is a glycosaminoglycan that is found in normal synovial fluid but reduced in OA joints. "Viscosupplementation"[10] refers to the injection of similar molecules in an attempt to improve lubrication and shock absorbing properties of synovial fluid. There are 2 products available commercially: Synvisc® (Hylan G-F 20) given as 3 injections weekly, or Hyalgan® (sodium hyaluronate) given as 5 injections weekly. Studies give conflicting results in terms of efficacy. It is not certain that these compounds confer lubricative properties as their intra-articular half-life is short; therefore their mechanism of action is not clear. These injections are costly, retailing at approximately SGD$110 for Hylan and SGD$180 for Synvisc. If response is equivocal, there is no valid reason to continue. The main side-effect is that of injection site reactions, especially if there is extravasation of the injected fluid.

TREATING BURSITIS AS A CAUSE OF PAIN IN OA

Due to abonormal mechanical forces and gait disturbances in OA, secondary bursitis may develop, causing pain over and above that attributable to the osteoarthritic joint. These localized causes of chronic pain and sometimes acute tenderness often go unrecognized. Pain can be dramatically alleviated with a localized steroid injection into the inflamed bursa by a trained practitioner, such as a rheumatolgist or orthopedic surgeon.

Anserine bursitis

The anserine bursa is located below the medial joint line of the knee and is related to the conjoint tendons of the semimenbranosus, gracilis and sartorius muscle (anseri or goose feet tendon). Inflammation of this structure can cause severe pain, sometimes acute in onset and difficulty walking. Localizing and palpating the structure reproduces the pain and confirms the diagnosis.

Trochanteric bursitis

True hip joint pain will be localized in the groin or radiating down the thigh to the knee. Pain located on the lateral aspect of the hip, not arising from the joint itself but often from an inflamed trochanteric bursa. Sometimes patients will complain of pain at night upon lying on the affected side.

SURGERY

There are several operative strategies used in the surgical management of OA:

1) *Arthroscopic measures*. Lavage, debridement, partial meniscetomies are procedures that can improve symptoms.
2) *High tibial osteotomy and femoral osteotomy* are techniques to realign joints thereby reducing asymmetric pressures on articular surfaces. They are employed in younger people with salvageable joints in order to delay or avoid the need for joint replacement.
3) *Joint replacement*, or the resurfacing of joint surfaces with metal, plastic or ceramic prostheses, has revolutionized the management of OA in the last 30 years, giving new hope to those who would otherwise be crippled by a severely degenerate joint.

Venous thrombosis is a post-operative complication that can be reduced by prophylactic heparin. Prosthesis infection is the most feared complication and occurs in less than 1%.

Long-term joint cement loosening is a major technical issue. Prostheses nowadays can last for 15 years or more. Nonetheless, therefore the operation should be delayed for as long as possible in the younger active age group to avoid the need for revisions in later life.

Post joint replacement, intensive physiotherapy is important in the restoration of mobility and muscle strength to avoid rapid muscle deconditioning that can arise from the immobility due to post-operative pain.

Operative techniques and hardware have improved over the years such that post-operative recovery is faster and protheses last longer. Patients nowadays demand better range of motion. The latest versions of

knee replacement inplants even allow for squatting and sitting cross-legged, postures valued in Asian lifestyles.

In the case of knee replacements, a new advance has been the less invasive technique of unicompartment replacement, or resurfacing of one joint compartment only. Either the medial or lateral compartment is replaced, preserving the greater part of the joint that is relatively less damaged. This allows a more natural propioception and full knee flexion.

CONCLUSION

The view that OA is a benign condition compared to the potentially crippling inflammatory arthropathies is not a fair reflection of the morbidity caused by the severe pain, loss of mobility and decreased ability to perform aerobic exercise posed by weight-bearing joint involvement. The disease burden in communities of the ubiquitous OA will be even greater as life expectancy increases, pushing up health care costs with increased demand for joint replacements. It is ironic that excellent disease control can now be achieved in most cases of inflammatory arthritis but no disease modifying agent can reliably reverse osteoarthritis. In recent years, a surge of research interest in the pathogenesis of osteoarthritis will hopefully lead to new therapeutic modalities. Meanwhile, there is little else to offer besides physiotherapy and analgesia, before joint replacement becomes necessary for severe cases.

ACKNOWLEDGMENTS

The author is grateful to: A/Prof Lo Ngai Nung, senior consultant orthopedic surgeon, Singapore General Hospital, for contributing the photograph on OA knees and his helpful comments on the section dealing with surgical aspects of OA; and Mr John Chen, consultant orthopaedic surgeon, Singapore General Hospital, for contributing the radiograph of OA knees.

REFERENCES

1. Felson DT, Epidemiology of hip and knee osteoarthritis, *Epidemiol Rev* **10**:1, 1988.
2. Spector TD, Perry LA, Jubb RW, Endogenous sex steroid levels in women with generalised osteoarthritis, *Clin Rheumatol* **10**:316, 1991.

3. Hoagland FT, Yau ACMC, Wong WL, Osteoarthritis of the hip and other joints in southern Chinese in Hong Kong, *J Bone Joint Surg Am* **55**:545, 1973.

4. TD, Cicuttini F, Baker J, Loughlin J, Hart D, Genetic influences on osteoarthritis in women: a twin study, *BMJ* **312**:940–943, 1996.

5. Martel–Pelletier J, Alaaeddine N, Pelletier JP, Cytokines and their role in the pathophysiology of osteoarthritis, *Front Biosci* D694–D703, 1999.

6. Waldschmidt JG, Braunstein EM, Buckwalter KA, Magnetic resonance imaging of osteoarthritis, *Rheum Dis Clin North Am* **25**:451–465, 1999.

7. Minor M, Exercise in the management of osteoarthritis of the knee and hip, *Arthritis Care Res* **7**:198, 1994

8. McAlindon TE, LaValley MP, Gulin JP, Felson DT, Glucosamine and chondroitin for treatment of osteoarthritis: a systematic quality assessment and meta-analysis, *JAMA* **283**:1469–1475, 2000.

9. Brief AA, Maurer SG, Di Cesare PE, Use of glucosamine and chondroitin sulfate in the management of osteoarthritis, *J Am Acad Orthop Surg* **9**(2):71–78, 2001.

10. Simon LS, Viscosupplementation therapy with intra-articular hyaluronic acid: fact or fancy? *Rheum Dis Clin North Am* **25**:345, 1999.

88

Soft Tissue Rheumatism

Feng Pao Hsii

INTRODUCTION

Musculo-skeletal problems account for approximately 18% of primary care consultations. Half of such consultations were related to soft tissue rheumatism. Hence soft tissue rheumatism is an important cause of morbidity in the community. Primary care and general medical practitioners should be familiar with the more common form of this.

Soft tissue rheumatism may be due to localized or regional pathology, or present with generalized symptoms (Table 1).

The Upper Limb

Shoulder pain

Soft tissue disorders of the shoulder as a group comprise the most common causes of shoulder pain in the general population. Of these, conditions that affect the rotator cuff are the commonest especially rotation cuff tendinitis followed by bicipited tendinitis, adhesive capsulitis or frozen shoulder occurring in decreasing order of frequency. The clinician should

Table 1 Soft Tissue Rheumatism

The Upper Limb

Shoulder Pain
- Rotator Cuff Tendinitis
- Bicipital Tendinitis
- Adhesive Capsulitis (Frozen Shoulder)

Elbow Pain
- Lateral Epicondylitis (Tennis Elbow)
- Medial Epicondylitis (Golfer's Elbow)
- Olecranon Bursitis

Hand Pain
- Carpal Tunnel Syndrome
- De Quervain's Tenosynovitis
- Digital Stenosing Tenosynovitis (Trigger Finger)
- Ganglia
- Dupuytren's Contracture

The Lower Limb

Hip Pain
- Trochanteric Bursitis
- Ischiogluteal Bursitis
- Iliopsoas Bursitis
- Piriformis Syndrome

Knee Pain
- Osgood–Schatter Disease
- Prepatellar Bursitis
- Infrapatellar Bursitis
- Anserine Bursitis
- Popliteal Cyst (Baker's cyst)
- Internal Derangement
- Patellofemoral Pain Syndrome

Ankle and Foot
- Heel Pain
- Achilles Tendinitis
- Retrocalcaneal Bursitis
- Plantar Fasciitis
- Metatarsalgia

Generalized Aches and Pains
- Polymyalgia Rheumatica
- Fibromyalgia

Low Back Pain
- Biomechanical Causes
- Medical Causes
- Management

also be mindful of referred pain to the shoulder from pathology in the neck, the chest or abdomen.

1) *Rotator Cuff Tendinitis*

This is the most common cause of shoulder pain. It can either be acute, subacute or chronic, and the pain is localized to the anterolateral aspect of the shoulder often with radiation to the lateral upper arm. The pain is increased with abduction — the painful arc — but in severe cases pain occur with abduction past 80 or 90 degrees.

Possible causes of rotator cuff tendinitis include trauma, repetitive use, age-related decreased vascularity, degeneration in tendons, osteophytes on the inferior part of the acromio-clavicular joint or inflammatory processes such as rheumatoid arthritis. Acute tears of the rotator cuff tendon either following trauma or spontaneously especially in an older person also present with similar symptoms. Very often, there is marked loss of shoulder abduction.

Treatment should include non-steroidal anti-inflammatory drugs (NSAIDs), local heat and gentle range of movement exercise. Local corticosteroid injection into the sabacromial space is the single most effective therapy in most cases. Repeat injections may be necessary if improvement is transient or suboptimal.

2) *Bicipital Tendinitis*

Bicipital tendinitis results from inflammation of the sheath of the long head of the biceps along the course proximal near the insertion in the bicipital groove. Most often, the condition results from impingement of the long head by the acromion and is frequently associated with rotator cuff tendinitis, glenohumeral instability or overuse of the affected limb.

The clinical picture consists of anterior shoulder pain and is increased with overhead activities. Passive range of motion is normal. Bicipital pain can be reproduced by supination of the forearm against resistance with the elbow flexed to 90 degrees or by flexion of the elbow against resistance.

Treatment of bicipital tendinitis is similar to that of rotator cuff tendinitis.

3) *Adhesive Capsulitis (Frozen Shoulder)*

Adhesive capsulitis or frozen shoulder is often idiopathic but be associated with other shoulder problems. It also occurs more commonly in

diabetics, after cerebrovascular accidents, immobility or hypothyroidism. In this condition the capsule becomes thickened and contracted. The pain is felt in the shoulder and over the lateral upper arm.

The onset of frozen shoulder is often insidious and is associated with initial loss of external rotation. Eventually motion of the shoulder is restricted in all three planes. Both active and passive motion are impaired in contrast to other shoulder disorders such as rotation cuff tendinitis. Treatment consists of symptomatic measures with analgesics and local heat, an active exercise program stopping short of aggravating pain and local steroid injection into the subacromial bursa and glenobumeral joint.

Elbow pain

Soft tissue disorders in the elbow may be the result of lateral epicondylitis, medial epicondylitis or olecranon bursitis.

1) *Lateral Epicondylitis*
   This is the most common soft tissue disorder of the elbow and is also known as tennis elbow. It presents as pain over the lateral aspect of the elbow and is exacerbated by wrist and finger extension. It occurs primarily in the middle-aged and is common among those who do not play tennis.

   Risk factors include repetitive movements of the forearm especially repetitive wrist turning, hand gripping, tool use or hand-shaking. Tennis elbow therefore is an occupational hazard among carpenters, gardeners, dentists and even politicians.

   It is of course a common injury among tennis players especially beginners. Another common cause nowadays is typing on the computer keyboard — especially if the keyboard is elevated above the level of the elbows.

   Treatment of lateral epicondylitis consists of rest from the precipitating activity, application of heat or ice and nonsteroidal anti-inflammatory drugs. A local steroid injection adjacent to the tendon attachment to the epicondyle may help.
2) *Medial Epicondylitis*
   Medial epicondylitis also known as golfer's elbow is less common and less disabling than lateral epicondylitis. It commonly occurs in

those who engage in cumulative repetitive strain of the common flexors of the fore-arm such as occurs in golfing, baseball pitching or work-related activities. Pain onset is typically insidious and localized to the region of the medial epicondyle. Tenderness is present around the region and increases with resistive flexion of the wrist.

Treatment consists of rest, local ice or heat, NSAIDs or local corticosteroid injection. Surgical release of the common forearm flexor tendon is rarely necessary.

3) *Olecranon Bursitis*

A variety of conditions may affect the olecranon burea. These include traumatic olecranon bursitis, septic olecranon bursitis and nonseptic inflammatory olecranon bursitis. The latter may be associated with conditions such as rheumatoid arthritis, gout or psoriatic arthritis. Aspiration may be needed if there is doubt about the presence of infection or crystal. In nonseptic olecranon bursitis treatment may not be needed or may occasionally be with a local steroid injection. Septic olecranon bursitis is treated with the appropriate antibiotics and repeated closed aspiration.

Systemic inflammatory disorders such as gout and rheumatoid arthritis may involve the olecranon bursa. Treatment is directed at the underlying cause. Occasionally local corticosteroid infections may be helpful.

Hand pain

The most common soft tissue disorders of the hand and the wrist include carpal tunnel syndrome, de Quervain's tenosynovitis, trigger finger or digital stenosing tenosynovitis, ganglia and Dupuytren's contracture.

1) *Carpal Tunnel Syndrome*

Carpal tunnel syndrome is most frequent in the middle-aged or the elderly. The condition can result from a variety of possible anatomic (e.g. narrowing of the carpel tunnel), physiologic (diabetes, pregnancy or hypothyroidism) as well as patterns of use including repetitive wrist flexion and extension. This condition is increasingly seen in users of computer keyboards.

Compression of the median nerve produces numbness, tingling or pain in the thumb, forefinger and middle finger although frequently dysaesthesias are felt in all the fingers. Patients frequently complain

of nocturnal exacerbation of symptoms (frequently relieved by shaking the hand) and of dropping objects. Reproduction of the patient's symptoms occur with percussion over the volar wrist distal to the skin crease, (Tinel's sign) or by asking the patient to flex both wrists against each other for 1 minute (Phalen's sign). Reduced pin-prick or more commonly two-point discrimination in the median nerve distribution is found. A late sign is atrophy of the thenar eminence or weakness of thumb abduction. Treatment of the condition includes nocturnal wrist splinting, anti-inflammatory drugs, corticosteroid injection and correction of any underlying cause, e.g. hypothyroidism. Persistent symptoms despite these measures may require surgical decompression.

2) *De Quervain's Tenosynovitis*

This condition is frequently confused with acute arthritis of the wrist. Synovitis of the tendon sheath of the abductor pollicis longus and extensor pollicis brevis presents with pain and swelling over the radial side of the wrist distal to the radial styloid process. This condition most frequently results from repetitive wrist motion in the radial or ulnar direction. The pain is reproduced with the Finkelstein's test — flexing the thumb and passively ulnar deviating the wrist.

Treatment includes splinting of the wrist and thumb, application of local heat and anti-inflammatory drugs. Local steroid injection into the tendon sheath is effective in most persistent cases.

3) *Digital Stenosing Tenosynovitis (Trigger Finger)*

The most common repetitive soft tissue injury of the hand is the so-called trigger finger. The thumb and middle finger are most commonly affected although the condition can affect any finger. It results from fibrosis of the flexor tendon sheath and is associated with repetitive fine motor tasks. Increased friction of the tendon passing through the stenotic sheath leads to pain and/or fibrotic thickening of the sheath. This leads to formation of a nodule, thus locking the finger with flexion. Frequently, the finger must be forcibly extended by the patient. Diabetes, hypothyroidism and other systemic conditions should be considered when multiple digits are involved. Treatment with rest, splint or NSAIDs may be effective but corticosteroid infiltration of the tendon sheath is the most effective treatment.

4) *Ganglia*
Ganglia are benign cystic lesions of the hands and wrists that appear to result most commonly from synovial out-pouching. Most (70%) occur on the dorsum of the wrist. They may present as painless, slowly growing masses or they may be painful and rapidly growing lesions.

Treatment with aspiration and steroid injection is indicated for symptomatic ganglia. Surgical resection is occasionally necessary.

5) *Dupuytren's Contracture*
This condition is the result of nodular thickening of the palmar fascia producing flexion deformities of the involved digits most commonly the fourth digit. Identified risk factors include alcoholism and diabetes. The contractures are asymptomatic and progression is quite variable. Progressive deformity and impaired hand function should prompt surgical referral for fasciectomy.

Lower Limb

Hip pain

1) *Trochanteric Bursitis*
Patients with trochanteric bursitis present with a deep, aching pain over the lateral upper thigh. The pain is often intensified by waking and may be worse at night especially when the patient is lying on the affected side. The diagnosis is confirmed with palpation of the posterior aspect of the greater trochanter, which elicits local tenderness in this area. The actiology of trochanteric bursitis is unknown. However, potential risk factors include local trauma, overuse activities such as jogging and leg length discrepancies. Treatment consist of avoiding aggravating activity, anti-inflammatory drugs, physiotherapy and local injection of steroids.

2) *Ischiogluteal Bursitis*
The ischiogluteal bursa lies over the ischial tuberosity and facilities gliding of the gluteus maximus muscle cover the tuberosity.

Prolonged sitting on hard surface or repeated leg flexion and extension in the sitting position may lead to ischiogluteal bursitis. The diagnosis is made by the characteristic history and by eliciting tenderness on palpation of the ischial tuberosity with the patient lying supine and with the hip and knee flexed.

Treatment consists of using a cushioning "doughnut" seat and performing knee-to-chest stretching exercises while lying on the cushion. A local injection of corticosteroids is useful in refractory cases, although care should be taken to avoid the sciatic nerve.

3) *Iliopsoas Bursitis*
The iliopsoas bursa lies over the anterior surface of the hip joint under the iliopsoas muscle. Typically patients present with an inguinal mass that may be painful. The condition has been described in association with a number of hip disorders including osteoarthritis, rheumatoid arthritis, osteonecrosis and septic arthritis. CT and MRI are considered the optimal imaging methods to diagnose iliopsoas bursitis. Treatment of the underlying hip disease generally is sufficient although surgical excision sometimes is required.

4) *Piriformis Syndrome*
The piriformis muscle is an abductor and external rotator of the hip that occupies the greater sciatic foramen. Inflammation of the piriformis muscle insertion produces pain in the region of the sacro-iliac joint often extending into the buttock causing pain with ambulation. Examination disclose pain and often weakness on resisted abduction and external rotation of the thigh. Straight leg raising frequently is positive but other neurological signs are negative. Treatment include stretching exercises, anti-inflammatory drugs and in resistant cases local injection of steroids.

Knee pain

Many soft tissue disorders affect the knee joint including various forms of bursitis (involving the prepatellar and infrapatellar bursa, anserine bursa, the popliteal bursa) and a variety of internal derangements within the joint itself such as tears of the meniscal cartilage or the collateral or cruciate ligaments. Other common causes of knee pain include the patellofemoral syndrome (formerly called chondromalacia patella), tendinitis and medial and lateral ligament strain.

A variety of knee pain syndromes occur in the adolescent and are unique to this age group.

1) *Osgood–Schlatter Disease*
Osgood–Schlatter disease is one such condition that generally presents with activity-related pain around the tibial tubercle. The

condition is believed to result from submaximal repeated stress on the junction of the patellar ligament, tibial tubercle and tibia. The tibial tubercle is swollen and tender.

2) *Prepatellar Bursitis*

The condition presents as a painful, red swelling anterior to the knee cap and is seen most often in people who kneel a lot — e.g. roofers and carpet fitters. Active knee extension is often quite painful. Infection of the joint should be excluded by aspirating any bursal fluid.

3) *Infrapatellar Bursitis*

The deep infrapatellar bursa lies between the upper position of the tibial tuberosity and patellar ligament.

Infrapatellar bursitis presents in a similar fashion as prepatellar bursitis although the location of swelling and tenderness is on either side of the patellar ligament.

4) *Anserine Bursitis*

The anserine bursa lies under the insertion of the thigh adductor muscle on the medial side of the knee. Anserine bursitis is a term loosely applied to pain and associated tenderness in this region. Nocturnal pain often leading to use of a pillow between the knees is characteristic. Obesity and osteoarthritis of the knees appear to be predisposing factors. Optimal treatment is injection of local steroid with local anesthesia.

5) *Popliteal Cyst (Baker's Cyst)*

Any cause of synovitis in the knee can lead to a popliteal cyst. Symptoms consist of fullness and tightness in the popliteal space that increases with walking. It can be detected clinically by palpation of fullness in the popliteal fossa. Rupture of a Baker's cyst leads to swelling of the calf and should be confirmed by Doppler ultrasound.

Treatment of the condition require treating any underlying condition. Direct aspiration of the cyst should rarely be attempted because of the proximity of neurovascular structures.

6) *Internal Derangement*

Internal derangement of the knee refer to disruption of the unusual functioning of the ligaments and menisci. In the young adult, it is often the result of athletic or sports trauma and in the elderly degenerative tears are more frequent. Symptoms include knee pain, knee locking and giving way.

MRI is widely used to diagnose suspected meniscal tear and to help plan the therapeutic approach. These include meniscal repair, partial menisectomy or total menisectomy. Non-operative treatment can be considered in the older patient with degenerative meniscal tear.

7) *Patellofemoral Pain Syndrome*
Chondromalacia patella has been used synonmously with patellofemoral pain but this term should be abandoned because it refers to a pathological rather than a clinical entity.

Patellofemoral pain syndrome is used most commonly to describe poorly localized anterior knee pain. The condition is believed to result either from anatomic abnormalities or from repetitive microtrauma to the patella surface. Non-operative treatment modalities include avoidance of overuse, exercise, orthoses and anti-inflammatory medication. Surgery may occasionally be necessary to correct any anatomical defects.

Ankle and foot pain

Soft tissue disorders of the ankle and foot are extremely common and often result from ill-fitting foot wear (e.g. high heels, extremely narrow toe box), overuse injuries (e.g. running) or abnormal diomechanical factors (e.g. excessive pronation, flat feet).

1) *Heel Pain*
The most common causes of heel pain are Achiller tendinitis, retrocalcaneal bursitis and plantar fasciitis. Calcaneal spur can also present with heel pain.

2) *Achilles Tendinitis*
This is generally caused by repetitive trauma and microtears of the tendon at its insertion on the calcaneus. Occasionally the condition can be seen in patients with seronegative spondyloarthropathies. The usual presentation is gradual onset of pain with foot pushoff and examination reveals tenderness and occasional thickening of the tendon. Rest, anti-inflammatory drugs, a heel lift, gentle stretching exercises and local heat application are effective treatment measures.

The Achilles tendon is vulnerable to rupture in the elderly. In resistent cases, carefully applied local injections along the sides of the tendon can be attempted.

3) *Retrocalcaneal Bursitis*
The retrocalcaneal bursa resides between the Achilles tendon and a pad of fat posterior to the talus. Retrocalcaneal bursitis is associated with posterior heel pain made worse with passive dorsiflexion of the ankles. The condition may be associated with tender swelling on both sides of the insertion of the tendon.

Cause include repetitive trauma due to athletic activity, rheumatoid arthritis and all of the seronegative spondyloarthropathies. Treatment is the same as for Achilles tendinitis.

4) *Plantar Fasciitis*
Plantar fasciitis is attributed to repetitive microtrauma to the attachment of the plantar fascia at the calcaneus. Local pain with weight bearing on the undersurface of the heel is the typical presentation. Examination shows local tenderness over the anteromedial position of the plantar surface of the calcaneus made worse on dorsiflexion of the toes. Associated conditions include enthesopathy due to any of the seronegative spondyloarthropathies. Subcalcaneal bursitis may be difficult to distinguish from plantar fasciitis although passive dorsiflexion of the toes does not increase symptom. Treatment consists of a heel pad or cushion, rest, application of local heat, stretching exercises and anti-inflammatory drugs. Local injection of corticosteroids may be useful. Rarely, surgical approaches such as fasciotomy may be required in resistant cases.

5) *Metatarsalgia*
Metatarsalgia is the symptom of pain across the plantar surface of one or more MTP joints. It may be due to diverse causes including muscle imbalance, fat pad atrophy, Morton's neuroma, hallux vulgus, callosities, flat or cavus foot, arthritis of the MTP joints, intermetatarsophalangeal bursitis, tarsal tunnel syndrome and arterial insufficiency. Treatment consists of treating the underlying cause, use of metatarsal pads and flexion exercises. Arch support are recommended for patients with flat or pronated feet.

Generalized Aches and Pains

Polymyalgia Rheumatica

This condition is rare in patients under 60 years of age. The onset is often acute with severe proximal girdle aching and stiffness. Patients have

great difficulty turning over in bed at night and dressing in the morning. There may be pain on shoulder movement, but often there is little to find on examination. The ESR is usually higher than 50 mm/hr. The differential diagnoses include degenerative joint disease, early rheumatoid arthritis, polymyositis and hypothyroidism. Myositis due to drugs like statins is another common cause.

Sometime the condition may affect the pelvic girdle and patients complain of stiffness and pain in getting into or out of a car. The condition often responds dramatically to a course of steroid therapy.

Fibromyalgia

This is the most common rheumatic cause of chronic diffuse pain characterized by the clinical features of chronic, widespread musculoskeletal pain and diffuse soft tissue tenderness. Formerly known as fibrositis, the term fibromyalgia has become popular more recently in the absence of inflammatory signs. Although the pathophysiology of the condition remains unknown the weight of evidence suggests that psychological factors play a prominent and probably crucial role in particular depression.

The diagnosis of fibromyalgia has been fascilitated by classification criteria developed by the American College of Rheumatology (ACR). They are:

1) History of chronic widespread pain-chronic — longer than 3 months and widespread — present above and below the waist and on both sides of the body.
2) Pain in 11 of 18 tender points on digital palpation — occiput, low cervical, trapezius, supraspinatus, second rib, lateral epicondyle, gluteal, greater trochanter, knee. For classification purposes patients are said to have fibromyalgia if both criteria are satisfied.

The condition is typically insidious in onset with the patient complaining of "pains all over". There often is a prior history of depression, anxiety disorder, and work problems, and females are more commonly affected than males.

Physical examination is essentially normal. Moderate pressure over muscles and tendon insertions (tender points) typically elicits marked tenderness with grimacing and/or withdrawal. The condition may be

associated with irritable bowel syndrome, tension headaches, migraine. Occasionally fibromyalgia can co-exist with systemic rheumatic diseases like SLE.

The treatment of fibromyalgia includes reassurance that the condition is not a progressive, crippling or life-threatening entity. Depending on the severity of the condition, a combination of low-aerobic exercise program and antidepressants (like amitriptyline, fluoxetine, paroxetine, sentraline) is effective in alleviating symptoms in most patients. Patients should also be encouraged to take an active role in the management of their condition.

Low Back Pain

Low back pain is among the most common symptoms to affect human beings. Up to two-thirds of all individual will experienced low back pain during their life time.

1) *Biomechanical Causes*
   Those account for over 90% of low back pain. These include structural spine abnormalities like kyphosis and scoliosis, muscle spasm and degenerative disorders like intervertebral disc disease, osteoarthritis, spinal stenosis ad spondylolithesis.
2) *Medical Causes*
   These include inflammatory conditions like infection and spondyloarthropathy, and metabolic conditions like osteoporosis and osteomalacia. Occasionally, low back pain may also be caused by psychogenic conditions like depression and fibromyalgia and one must remember referred pain due to abdominal aneurysm, retropertoneal lesion, pyelonephritis, pancreatic disease and occasionally prostatitis.

   The management of low back pain include a careful history and physical examination, relevant laboratory work-up and imaging techniques like plain radiography, bone scanning or MRI.

   Optimal constructive therapy include a brief period of rest, symptomatic medication therapy tailored to the individual and physical modalities such as stretching and flexiblility exercises. Occasionally, referral to an anesthesiologist, orthopedic surgeon or neurosurgeon may be necessary.

This page intentionally left blank

Index