

6/3.2
WEB
002

544

059322

Not to be taken outside
the Library

ASTU024062EIT

Nutrition

A health promotion approach

2nd Edition

Geoffrey P Webb BSc, MSc, PhD

Senior Lecturer in Nutrition and Physiology, University of East London,
London, UK



A member of the Hodder Headline Group
LONDON

613.2
WEB
2002

54

059322

Not to be taken outside
the Library



ASTU024062EIT

Nutrition

A health promotion approach

2nd Edition

Geoffrey P Webb BSc, MSc, PhD

Senior Lecturer in Nutrition and Physiology, University of East London,
London, UK



A member of the Hodder Headline Group
LONDON

*Dedication
For my daughter Katie*



Contents

Preface	ix
Part 1 Concepts and principles	1
1 Changing priorities for nutrition education	3
Identification of the essential nutrients	3
Adequacy – the traditional priority in nutrition	5
The new priority – diet as a means to health promotion or disease prevention	6
Is intervention to induce dietary change justified?	13
Effecting dietary change	20
Concluding remarks	26
2 Food selection	27
Introduction and aims of the chapter	27
The biological model of food	28
Dietary and cultural prejudice	29
Food classification systems	30
Non-nutritional uses of food	35
The hierarchy of human needs	36
A model of food selection	38
Physical availability	39
Economic availability	40
Cultural availability	52
'Gatekeeper' limitations on availability	58
3 Methods of nutritional surveillance and research	60
Aims and introduction	60
Nutritional assessment and surveillance	60
Methods used to establish links between diet and chronic disease	91
4 Dietary guidelines and recommendations	115
The range of 'expert reports' and their consistency	115
Variations in the presentation of guidelines and recommendations	117
'Food' recommendations	118
Energy and body weight	119

Recommendations for fats, carbohydrates, protein and salt	120
Alcohol	122
How do current UK diets compare with 'ideal' intakes?	124
Other nutrients	125
Willingness to change	126
Some barriers to dietary change	128
Aids to food selection	129
Concluding remarks	134
5 Cellular energetics	136
Aim	136
Introduction and overview	136
Metabolism of carbohydrate	138
Metabolism of fats	140
Metabolism of protein	141
The pentose phosphate pathway	142
Part 2 Energy, energy balance and obesity	143
6 Introduction to energy aspects of nutrition	145
Units of energy	145
How energy requirements are estimated	146
Variation in average energy requirements – general trends	148
The energy content of foods	148
Sources of dietary energy by nutrient	149
Energy density	151
Nutrient density	152
Sources of dietary energy by food groups	154
Starvation	155
Eating disorders	159
Cancer cachexia	162
7 Energy balance and its regulation	164
The concept of energy balance	164
Is there physiological regulation of energy balance?	166
Set point theory	166
Is energy expenditure regulated?	167
External influences that affect food intake	169
Control of energy intake – early work with experimental animals	169
Control of food intake – the discovery of leptin	173
8 Obesity	178
Defining obesity	178
Prevalence of obesity	179
The medical and social consequences of obesity	182

The causes of obesity	188
Is fat more fattening than carbohydrate?	193
Inactivity as a cause of obesity	194
Prevention and treatment of obesity in populations	196
Obesity treatment in individuals	199
More 'aggressive' treatments for obesity	202
Part 3 The nutrients	207
9 Carbohydrates	209
Introduction	209
Nature and classification of carbohydrates	210
Dietary sources of carbohydrates	211
Sugars	212
Artificial sweeteners	214
Diet and dental health	215
Starches	219
Non-starch polysaccharide	221
Resistant starch	223
Does dietary fibre/NSP protect against bowel cancer and heart disease?	225
Possible mechanisms by which diet may influence the risk of bowel cancer and heart disease	228
10 Protein and amino acids	230
Traditional scientific aspects of protein nutrition	230
The significance of protein in human nutrition	237
11 Fat	245
Nature of dietary fat	245
Types of fatty acids	246
Sources of fat in the diet	250
Roles of fat in the diet	251
Blood lipoproteins	255
Digestion, absorption and transport of dietary lipids	257
Transport of endogenously produced lipids	258
The diet–heart hypothesis	260
Current 'health images' of different dietary fats	261
The key tenets of the diet–heart hypothesis	265
Review of the evidence for the diet–heart hypothesis	267
Fish oils	272
12 Vitamins and antioxidants	277
Some general concepts and principles	277
The dietetically important vitamins	281
Free radicals and antioxidants	297

13	The minerals	307
	Introduction	307
	Iodine and iodine deficiency diseases	311
	Iron and iron-deficiency anaemia	313
	Calcium, diet and osteoporosis	317
	Salt and hypertension	327
Part 4	Variation in nutritional requirements and priorities	335
14	Nutrition and the human life-cycle	337
	Introduction	337
	Nutritional aspects of pregnancy	338
	Lactation	348
	Infancy	349
	Childhood and adolescence	358
	The elderly	361
15	Nutrition as treatment	371
	Diet as a complete therapy	371
	Diet as a specific component of therapy	374
	Malnutrition in hospital patients	379
16	Some other groups and situations	387
	Vegetarianism	387
	Racial minorities	392
	Nutrition and physical activity	395
Part 5	The safety and quality of food	405
17	The safety and quality of food	407
	Aims of the chapter	407
	Consumer protection	407
	Food poisoning and the microbiological safety of food	412
	Bovine Spongiform Encephalopathy	425
	Food processing	430
	The chemical safety of food	434
	Functional foods	441
	Glossary	443
	References	453
	Index	465

Preface

My aims for this book are largely unaltered from those of the first edition and are outlined below. The book aims to cover the whole range of nutrition so that students do not have to use a separate book for nutritional epidemiology, social aspects of nutrition, food microbiology etc. when they first start upon their nutrition studies. The level and focus of material in specialist books in these areas are often not ideal for new nutrition students. I have tried to make the book flexible enough to be used at various academic levels by students new to nutrition and perhaps with limited mathematical and biochemical background. I have concentrated on those areas that are most relevant to industrialized countries at the start of the new millennium, e.g. less emphasis on the traditional deficiency diseases and more on the role of diet in the aetiology of the chronic degenerative diseases such as cancer, heart disease and osteoporosis. This 'disease prevention' focus is the reason for the 'health promotion approach' of the title, but I have also included in this edition a short section on health promotion theory in Chapter 1. I have tried to provide an international perspective, largely by reference to the USA as a comparison with the UK. Finally, the book aims to give a critical overview of the evidence supporting and contradicting various diet-health links rather than just giving the final conclusions.

Who am I writing for? In the preface to the first edition, I picked out those taking nutrition units as part of a wider scientific or professional course as the principal audience. In this edition, I have also tried to improve the usefulness of the book for specialist nutrition students in the early part of their course and for the nutrition components of dietetics courses.

How is this edition different from the first one? All major sections have been updated and refined and many have been substantially rewritten. This is

partly due to changes in the subject of nutrition since 1994, but also reflects changes in the author since that time. Since 1994, I have written two other books, attended dozens of scientific symposia, read hundreds of scientific articles and, most significantly, had six years' experience of using the first edition as a course text. I have expanded the chapter on energy aspects of nutrition in the first edition into three chapters in this one. Obesity and weight control are one of the most important areas of modern nutrition, and the discovery of the new hormone, leptin, in 1994 has produced a flood of exciting new research into the mechanisms that control our body weight. With the benefit of hindsight, the decision not to include a short review of each of the major vitamins in the first edition was probably a mistake. In this edition, there is still a general overview of the nutrition of vitamins, but I have also given a basic outline of the nature, sources, functions and deficiency states for each of them. The sections on BSE, life-cycle nutrition, dietary guidelines, carbohydrates, fats and nutrition for sport have all been extensively revised and a short section on functional foods has been added.

Is this edition more student-friendly than the last? The most obvious change made in this respect is the inclusion of text box 'revision' summaries at the end of most major sections. I have also tried to improve the accessibility of my writing style. For example, I have made a conscious effort to reduce my average sentence length and to make my writing more direct, with much greater use of bulleted lists rather than long sentences with semi-colons. As in the first edition, there is an extensive glossary and new terms printed in bold in the text are included in this glossary.

The revisions, additions and the extensive use of text box summaries mean that this edition is consid-

erably longer than the first. There is obviously much more in this book than will be required by students taking a nutrition unit or two as part of a wider course. However, I have provided a full summary of the contents of each chapter and a full index that

should help students to select what parts of the book they need for their studies. I have taken the view that it is better for students to have to ignore sections that they do not need rather than not to have sections they do need. I urge student readers to use the index!

Part 1

Concepts and principles

Changing priorities for nutrition education

Identification of the essential nutrients	3	Is intervention to induce dietary change justified?	13
Adequacy – the traditional priority in nutrition	5	Effecting dietary change	20
The new priority – diet as a means to health promotion or disease prevention	6	Concluding remarks	26

IDENTIFICATION OF THE ESSENTIAL NUTRIENTS

Scientific research into food and nutrition was initially directed towards identifying the essential nutrients and quantifying our requirement for them. All of the essential nutrients must be present in our diets in certain minimum quantities if we are to remain healthy. A shortage of one of these essential nutrients leads to adverse symptoms, often to a characteristic **deficiency disease**. In the first half of the twentieth century, many essential nutrients were identified and their ability to cure deficiency diseases was recognized, for example:

- Niacin (vitamin B₃) was shown to cure the deficiency disease, **pellagra**, an often fatal disease that was so prevalent in some southern states of the USA in the early decades of the twentieth century that it was thought to be an infectious disease. Pellagra remained a problem in several southern states of the USA until the 1940s.
- In the late nineteenth and early twentieth centuries, up to 75% of children in some British industrial cities suffered from the disease **rickets**, which was shown to be due to lack of vitamin D. Again, it was not until the 1940s that this disease was largely eradicated in British children.

- The disease **beriberi** exacted a heavy toll of death and misery in the rice-eating countries of the Far East well into the third quarter of the twentieth century. Thiamin (vitamin B₁), a vitamin that is largely removed from white rice during the milling process, cured it.

Several of the Nobel prizes in Physiology and Medicine in this era were awarded for work on the vitamins – the prizes of 1929, 1934 and 1943. It was for the work on thiamin and beriberi, mentioned above, that Eijkman received the 1929 prize.

Such spectacular successes as these may have encouraged a 'magic bullet' image of nutrition – the expectation that simple dietary changes may be able to prevent or cure diseases other than those due to dietary inadequacy. This expectation is generally misplaced, although there is no doubt that poor nutritional status can adversely affect the course of all illnesses and some conditions do respond to restriction or extra supply of some dietary components. For example, the symptoms and the progression of both diabetes and chronic renal failure can be controlled to a degree by diet (see Chapter 15). There are even a few, relatively uncommon, diseases whose symptoms are due to an inborn or acquired intolerance to a specific component of food. In these cases, whereas diet does not strictly cure the condition, the symptoms

can be controlled by preventing or limiting the intake of the offending substance, for example:

- **phenylketonuria** (intolerance to the amino acid phenylalanine)
- **galactosaemia** (intolerance to one of the components of milk sugar – galactose)
- **coeliac disease** (intolerance to the wheat protein, gluten).

There are also a few conditions whose symptoms are alleviated by increased intake of a nutrient. For example, **pernicious anaemia** is an autoimmune condition that results in an inability to absorb vitamin B₁₂ and is relieved by injections of the vitamin.

Harper (1999) provides a list of criteria for establishing that a nutrient is essential:

- The substance is essential for growth, health and survival.
- Characteristic signs of deficiency result from inadequate intakes and these are only cured by the administration of the nutrient or a known precursor.
- The severity of the deficiency symptoms is dose dependent – they get worse as the intake of nutrient decreases.

- The substance is not synthesized in the body and so is required throughout life. (Note that a very strict application of this rule would eliminate some essential nutrients such as vitamin D and niacin.)

In addition to the essential nutrients, some additional substances may be required in premature infants, in people with certain genetic defects and in certain pathological states. Harper (1999) classifies these substances as conditionally essential. Some other substances may be desirable for health whilst not meeting the criteria for essentiality, e.g. fluoride for its beneficial effects on teeth, dietary fibre and certain substances in plants that may act as antioxidants. Some nutrients may also be used in doses that greatly exceed those that would be obtained from food to produce a pharmacological effect (i.e. they are used as drugs).

Around 40 essential nutrients have now been identified, namely:

- water
- energy sources
- protein and the nine essential amino acids
- essential fatty acids
- the vitamins A, C, D, E and K

- In the first half of the twentieth century, most of the vitamins were discovered and the ability of vitamin and mineral supplements to cure deficiency diseases was recognized.
- These discoveries may have encouraged the illusion that dietary change could cure many other diseases.
- A nutrient is classified as essential if it is needed for growth and survival, if deprivation leads to dose-dependent symptoms of deficiency, and if the substance is not synthesized in sufficient quantities to meet physiological needs.
- Some nutrients are classified as conditionally essential because they are only essential in some circumstances or for some people.
- Although dietary change can rarely cure or remove the symptoms of a disease, the progress of all diseases will be affected by poor nutritional status.
- In a few fairly uncommon conditions, diet may be the sole therapy, usually because the symptoms are due to intolerance to a component of food.
- Some nutrients may be used in pharmacological quantities, i.e. used as drugs.
- Intakes of some ‘nutrients’ may be desirable for health whilst not being strictly essential.
- Around 40 essential nutrients have been identified and estimates of average requirements have been made and published as Recommended Dietary Allowances in the USA and Dietary Reference Values in the UK.

- eight substances that make up the B group of vitamins
- 14 minerals and trace minerals.

In most cases, these nutrients have not only been identified, but good estimates of average requirements have also been made. Many governments and international agencies use these estimates of requirements to publish lists of dietary standards that can be used as yardsticks to test the adequacy of diets or food supplies. These standards, termed **Recommended Dietary Allowances (RDAs)** in the USA and **Dietary Reference Values (DRVs)** in the UK, are discussed fully in Chapter 3.

ADEQUACY – THE TRADITIONAL PRIORITY IN NUTRITION

The traditional priority in nutrition has been to ensure nutritional adequacy, to ensure that diets contain adequate amounts of energy and all of the essential nutrients. Adequacy was the traditional priority in all countries and it remains the nutritional priority for the majority of the world population. Even today, dietary inadequacy is prevalent in many countries, as illustrated by the examples below.

- Large sections of the world population still suffer from overt starvation.
- Several vitamin and mineral deficiency diseases are still prevalent in many parts of the world. Vitamin A deficiency causes hundreds of thousands of children to go blind each year and is a contributory factor in the deaths of millions of infants and children in Third World countries. Iodine deficiency is endemic in many parts of the world and retards the mental and physical development of tens of millions of people.
- Many more people in developing countries, especially children, suffer more subtle consequences of suboptimal nutrition (Table 1.1).

In the first half of the twentieth century, even in developed countries, the quality of a diet would thus be judged by its ability to supply all of the essential nutrients and to prevent nutritional inadequacy. The official priorities for improving the health of the British population during

Table 1.1 Some adverse consequences of inadequate food intake

Increased probability of suffering from a specific nutrient-deficiency disease
Reduced growth and reduced mental and physical development in children
Wasting of muscles and essential organs; wasting of the heart muscle leads to abnormal cardiac rhythms and risk of heart failure
Reduced capacity of the gut for digestion and absorption; an inability to digest lactose (milk sugar) and diarrhoea are both common features of malnutrition
Impaired functioning of the immune system leading to increased risk of infection as well as more prolonged and severe symptoms of infection
Slow healing of wounds
Reduced strength and physical capacity that may impair the ability to earn an adequate income or cultivate food
Changes in personality and other adverse psychological effects

- to reduce the consumption of bread and starchy foods
- to increase the consumption of nutrient-rich, so-called 'protective foods', like milk, cheese, eggs and green vegetables.

The following benefits were expected to result from these changes:

- a taller, more active and mentally alert population
- less of the deficiency diseases such as goitre, rickets and anaemia
- a reduced toll of death and incapacity due to infectious diseases, such as pneumonia, tuberculosis and rheumatic fever.

Britons are, indeed, much taller now than they were in the first half of the twentieth century and occurrences of overt deficiency diseases are rare and usually confined to particular high-risk sectors of the population. Children now mature faster and reach puberty earlier. The potential physical capability of the population has probably increased, even though many of us now accept the opportunity to lead very inactive lives that mechanization has given us. Infectious diseases now account for less than 1% of deaths in Britain.

In the affluent industrialized countries of North America, Western Europe and Australasia, nutritional adequacy is now almost taken for granted.

developed countries have been judged by its ability to supply all of the essential nutrients and to prevent nutritional inadequacy. The official priorities for improving the health of the British population during

- Ensuring adequate intakes of essential nutrients has been the traditional priority in nutrition and it remains the priority for most of the world population.
- Dietary inadequacy results in a number of adverse consequences that are listed in Table 1.1.
- In the 1930s, in Britain the dietary priorities for improving health were to reduce the consumption of starchy foods and to increase the consumption of so-called protective foods like milk, cheese, eggs and green vegetables.
- These changes to the 1930s' British diet were expected to lead to a taller, more active and mentally alert population, elimination of the deficiency diseases and a reduced toll from many infectious diseases. These objectives have been largely achieved.
- In the industrialized countries, deficiency diseases are largely confined to the very elderly, the infirm, alcoholics and drug addicts, or those at the extremes of social and economic deprivation.
- War and economic disruption can lead to a rapid re-emergence of the deficiency diseases in a previously affluent country.

Nutrient deficiencies only become likely if total food intake is restricted or if the range of foods that are available or acceptable is narrowed. In these industrialized countries, there is an almost limitless abundance of food and a year-round variety that would have astounded our ancestors. Even the poor in these countries should usually be able to obtain at least their minimum needs of the essential nutrients. Overt deficiency diseases are very uncommon amongst most sections of these populations. Overnutrition, manifested most obviously by obesity, is far more common in such countries than all of the diseases of undernutrition. Malnutrition and deficiency diseases in the industrialized countries are usually concentrated in groups like those listed below:

- Those who suffer chronic ill-health or some medical condition that specifically predisposes to nutrient deficiency (a high proportion of patients admitted to hospital are malnourished upon admission – see Chapter 15).
- Alcoholics and drug addicts.
- Those at the extremes of social and economic disadvantage.
- The very elderly (see Chapter 14).

War or severe economic dislocation can rapidly undermine this assumption of adequacy in developed countries, as witnessed by events in parts of the former Soviet Union and former Yugoslavia during the 1980s and 1990s. Wartime mass starvation in Holland and Germany is still recent enough to be

within living memory. In Britain, during World War II there were considerable constraints placed upon the food supply by attacks on merchant shipping. This certainly led to shortages of certain foods and a more austere and less palatable diet. However, because of very effective food policy measures and strict rationing, it also, paradoxically, led to an apparent improvement in the nutritional health of the British population.

In both the UK and USA, average energy intakes of the population have fallen substantially in recent decades, presumably as a reflection of reduced energy expenditure caused by our increasingly sedentary lifestyle. This trend towards reduced energy intake (i.e. total food intake) coupled with a high proportion of energy being obtained from nutrient-depleted sugars, fats and alcoholic drinks might well serve to increase the likelihood of suboptimal nutrient intakes. It may even precipitate overt nutrient deficiencies in some groups, such as the elderly or those on prolonged weight-reducing diets.

THE NEW PRIORITY – DIET AS A MEANS TO HEALTH PROMOTION OR DISEASE PREVENTION

The priority of health promotion and nutrition education in industrialized countries is now directed towards changing the 'Western diet and life-style' with the aim of reducing the risk of chronic disease.

is irreversible damage. Much of current health promotion, nutrition education and preventive medicine is based upon such assumptions. Dietary variables can be direct risk factors for disease, as in the examples below:

- a diet that is low in fruits and vegetables is associated with an increased risk of several types of cancer
- high sugar consumption is strongly implicated as a cause of tooth decay, whereas high fluoride intake seems to prevent it.

Dietary variables can also have a major influence upon some of the other well-established risk factors, particularly the cardiovascular risk factors, and some examples are listed below:

- A high average salt intake in a population is associated with high average blood pressure and an increased prevalence of **hypertension**.
- Changes in the types of fat eaten can have readily demonstrable effects upon plasma cholesterol concentration; a high proportion of saturated fat raises plasma cholesterol, but replacing saturated with unsaturated fat lowers it.

Much current nutrition research is focused upon these risk factors. Individual components of the diet are related to individual diseases or even to other risk factors such as high plasma cholesterol and high blood pressure. The identification of dietary and other environmental risk factors should enable health educators to promote changes in behaviour exposure to the risk factor, prevent

may be distorted and sensationalized. Health-conscious members of the public may then try to decide upon the optimal diet and lifestyle on the basis of such snippets – this is rather like trying to work out the picture on a huge jigsaw puzzle by reading inaccurate and incomplete descriptions of some of the individual pieces.

Quite a number of these individual associations are discussed and evaluated in this book. In such discussions, there are two broad aims:

1. To indicate not only where the current consensus of scientific opinion lies, but also to give a flavour of the reasoning and evidence underpinning that position and to highlight unresolved issues and/or alternative opinions.
2. To indicate diet and lifestyle patterns or modifications that are realistic, low risk, consistent and likely to produce overall health benefits.

Are recommendations to alter diet and lifestyle offered too freely?

A drug or a food additive intended for use by vast numbers of apparently healthy people would be subject to an extremely rigorous and highly structured evaluation of both its efficacy and safety. Despite this, mistakes still occur, and drugs or additives to be withdrawn, sometimes after many years. Compared to this situation, the evaluation of dietary

The author's perception is thus that there is a tendency for dietary or lifestyle intervention to be advocated before rigorous appraisal has been satisfactorily completed. This is probably because of the implicit assumption that simple changes in diet and behaviour are innocuous and that, even if they do no good, then neither will they do any harm. Intervention, rather than non-intervention, therefore, becomes almost automatically regarded as the safer option: 'If dietary change can't hurt and might help, why not make it?' (Hamilton *et al.*, 1991).

Simple changes in diet or lifestyle can't do any harm?

The assumption that simple dietary and lifestyle changes are innocuous seems logically inconsistent with the notion of diet and lifestyle as major causes of ill-health. If dietary changes are capable of producing large changes in health prospects, then there must be the potential to do harm as well as great good (see possible examples in the next section). Even if unjustified intervention does not do direct physiological harm, it may have important social, psychological or economic repercussions. Social and psychological factors have traditionally been the dominant influences on food selection and any ill-considered intervention risks causing anxiety, cultural impoverishment and social and family tensions. It has been suggested that the resources spent on unjustified health education interventions may serve only to induce a morbid preoccupation with death (McCormick and Skrabanek, 1988). Major changes in food selection practices will also have economic repercussions, perhaps disrupting the livelihoods of many people and reducing their quality of life.

Repeated changes of mind and shifts of emphasis by health educators are also liable to undermine their credibility and thereby increase public resistance to future campaigns. We saw earlier in the chapter how dietary advice changed as the nutritional priority changed away from trying to ensure adequacy and towards using diet as a means of preventing chronic disease. These changes in emphasis often reflect changing social conditions and consumption patterns, but some may be justifiably perceived as evidence of the inconsistency and thus unreliability of expert advice. The two examples below may serve to illustrate this latter point.

1. Obesity was once blamed upon excessive consumption of high-carbohydrate foods, such as bread and potatoes. These foods would have been strictly regulated in many reducing diets. Nutritionists now view bread and potatoes much more positively. They are high in starch and fibre and consumers are recommended to incorporate more of them into their diets to make them bulkier (i.e. to reduce the **energy density**). This bulkier diet may, in turn, reduce the likelihood of excessive energy intake and obesity.
2. Scientific support for nutritional guidelines aimed at lowering plasma cholesterol concentrations has now been increasing for over 30 years. Although this cholesterol-lowering objective may have remained constant, there have been several subtle shifts of opinion as to the best dietary means to achieve this goal, as illustrated below.
 - Reducing dietary cholesterol was once considered an important factor in lowering plasma cholesterol; it is now usually given low priority.
 - A wholesale switch from saturated to the n-6 (ω -6) type of polyunsaturated fatty acids prevalent in many vegetable oils was once advocated, but very high intake of these polyunsaturated fats is now considered undesirable.
 - Mono-unsaturated fats were once considered neutral in their effects upon plasma cholesterol, but are now much more positively regarded. This accounts for the current healthy image of olive oil and rapeseed (canola) oil.
 - The n-3 (ω -3) polyunsaturated fatty acids have also been viewed much more positively in recent years, especially the long-chain acids found predominantly in fish oils.
 - The current emphasis of both UK and USA recommendations is to aim for a reduction in total fat intake, with a proportionately greater reduction in the saturated fraction of dietary fat.

Attitudes to some individual foods could be strongly affected by such changes of emphasis. This is illustrated in Table 1.2. Certain fatty foods come out very differently in a 'worst' to 'best' rank order when cholesterol content, total fat content, saturated fat content or ratio of polyunsaturated to saturated fatty acids (**P:S ratio**) is used as the criterion. Sunflower oil comes out at the 'worst' end of the rankings if total fat content is the criterion, but at or near the 'best' end if cholesterol content, P:S ratio or

Table 1.2 Rank orders by different fat-related criteria of some foods

Food	Energy as fat (%)	Energy saturated fat (%)	Cholesterol content	P:S ratio
Liver	11	8	2	8
Lean steak	12	9=	6	4
Cheddar cheese	7	2	7	2=
Butter	1=	1	3	2=
Polyunsaturated margarine	1=	6	10=	13
Low-fat spread	1=	4	10=	10
Sunflower oil	1=	11	10=	14
Milk	9=	3	9	1
Human milk	9=	5	8	5
Chicken meat	13	12=	5	7
Avocado	5	12=	10=	9
Peanuts	6	9=	10=	11
Egg	8	7	1	6
Prawns	14	14	4	12

Ranks are from 1 ("worst") to 14 ("best"). P:S ratio = ratio of polyunsaturated to saturated fatty acids.
After Webb (1992a).

saturated fat content is used. Eggs and liver are worst by the cholesterol criterion, but around halfway down by all other criteria; prawns are at the bottom for all criteria except cholesterol content, they are also relatively rich in the now positively regarded long-chain n-3 (ω -3)polyunsaturated fatty acids.

Does health promotion always promote health? Examples of interventions that are now less fashionable

It would seem prudent to regard any dietary or lifestyle changes considered for health promotion as potentially harmful until this possibility has been actively considered and, as far as possible, eliminated. Below are some examples that may serve to demonstrate that this is a real possibility rather than just a theoretical one.

SLEEPING POSITION AND COT DEATH

During the 1970s and much of the 1980s, health professionals and childcare writers in the UK advised parents to put babies down to sleep on their front (prone) rather than on their back (supine) as they had traditionally done. Premature babies were reported to suffer less respiratory distress in the

prone position and babies with severe gastro-oesophageal reflux were less likely to choke upon regurgitated milk. This decreased risk of choking was thought to be important for all babies and the back sleeping position came to be regarded as unsafe:

For the first two months or so, the safest way for babies to sleep is on their fronts, head to one side, or else curled up on one side. Then if they are sick, there is no chance that they will choke.

Health Education Council (1984) *Pregnancy book*.

A young baby is best not left alone lying on his back. Because he cannot move very much he might choke if he were sick. Lie the baby in the cot on his front with his head to one side, or on his side.

Reader's Digest family medical adviser (1986).

There is now a wealth of evidence to suggest that the front sleeping position is associated with an increased risk of cot death, and a return to the traditional supine sleeping position has now been officially recommended in the UK. Compare the quote below from the 1993 edition of the Health Education Council's *Pregnancy book* with that from the 1984 version above:

Babies laid to sleep on their tummies are more at risk of cot death than babies laid down to sleep on

their backs or sides . . . Only lay your baby down to sleep on his or her front if your doctor advises it.

In early 1993, the UK Department of Health released figures that showed a dramatic decline in cot deaths in 1992 following a 'Back to sleep' campaign that was aimed at persuading parents to revert to the traditional sleeping position for infants (this reduced incidence has been maintained in subsequent years). A government minister of the time was quoted as claiming that this decline in cot deaths illustrated the benefits of health education:

The figures on cot deaths show that behaviour change can work, can save lives.

The Daily Telegraph 30/3/93.

What these figures even more obviously demonstrate, however, are the potential dangers of prematurely advocating changes in lifestyle before the assumption that they will promote health has been thoroughly evaluated. The 'success' of the 'Back to sleep' campaign was really only to reverse the damage done by the earlier advice to change the traditional sleeping position. Sleeping position is not the cause of cot death, but front sleeping does seem to increase its likelihood. No individual baby's death can be directly attributed to sleeping position. However, it does seem probable that thousands of extra cot deaths occurred in Western countries during the 1970s and 1980s as a result of the promotion of a change in sleeping position (see Webb, 1995, for a review of this example).

If one is seeking to justify intervention for a whole population, then it is not enough merely to demonstrate that a specific group will benefit from the proposed change. Intervention with a relatively small sample of 'high-risk' subjects is likely to exaggerate the benefits of intervention for the population as a whole. Such small-scale intervention will also only show up the grossest of harmful effects. It may thus fail to show up harmful effects that could be very significant for indiscriminate intervention and which might cancel out or even exceed any benefits of the intervention.

VERY HIGH INTAKES OF POLYUNSATURATED FATTY ACIDS

As noted earlier, diets containing very large amounts of polyunsaturated fatty acids were widely advocated as a means of lowering plasma cholesterol concentra-

tions. Short-term experiments had demonstrated very convincingly that switching from high-saturated to high-polyunsaturated fat diets could reduce serum cholesterol concentrations in young men (e.g. Keys *et al.*, 1959). However, there was no convincing evidence that, in the long term, such a wholesale switch from saturated to polyunsaturated fat produced any increase in life expectancy, even in studies involving 'high-risk' subjects. Concerns have been raised about the long-term safety of diets that are very high in polyunsaturated fat. Britons have been advised to limit their individual intakes of n-6 (ω -6) polyunsaturated fatty acids to no more than 10% of energy, with a population average of 6% (COMA, 1991).

Reducing an asymptomatic risk factor, such as plasma cholesterol concentration, cannot, in itself, be taken as evidence of benefit. There must be evidence that it will lead to the predicted decrease in disease risk and also that there will not be an accompanying increase in some other risk.

'THE WORLD PROTEIN CRISIS'

In the 1950s and 1960s, protein deficiency was thought to be the most prevalent and serious form of worldwide malnutrition. It was considered likely that 'in many parts of the world the majority of young children suffer some protein malnutrition' (Trowell, 1954). Very considerable efforts and resources were committed to alleviating this perceived problem. However, as estimates of human protein requirements were revised downwards, it then seemed probable that primary protein deficiency was likely to be uncommon and thus that the resources committed to alleviating widespread protein deficiency were largely wasted (Webb, 1989). See Chapter 10 for further discussion.

There is no evidence that the provision of extra protein was directly harmful. Nevertheless, scarce resources were committed to a non-beneficial measure and other, more real, problems deprived of resources. In the current edition of both the US (NRC, 1989a), and UK dietary standards (COMA, 1991), the possibility that high intakes of protein might even be directly harmful is acknowledged and an upper limit of twice the dietary standard is advised.

Resources for research, development, international aid and educational programmes are always finite. If these

resources are diverted into unnecessary measures or programmes based upon false theories, then, inevitably, worthwhile projects will be deprived of funding.

ROUTINE IRON SUPPLEMENTATION DURING PREGNANCY

Iron supplements were once routinely and almost universally prescribed for pregnant women in the UK. These supplements were considered necessary because of the perception that the risks of anaemia in pregnancy were very high, a perception that was compounded by the misinterpretation of the normal decline in blood haemoglobin concentration during

pregnancy as pathological. In the UK, routine iron supplements in pregnancy are now thought to be unnecessary. The UK panel on Dietary Reference Values (COMA, 1991) suggested that, for most women, the extra iron demands of pregnancy can be met by the utilization of maternal stores and physiological adaptation. Note that in the USA (NRC, 1989a) a different view is taken and universal iron supplementation is still recommended for pregnant women.

USE OF β -CAROTENE SUPPLEMENTS

There is overwhelming evidence that diets high in fruits and vegetables are associated with a reduced

- Average life expectancy in Britain and the USA increased by more than 50% during the twentieth century and this has led to a large increase in the proportion of elderly people.
- This increased life expectancy means that most deaths and much of the chronic illness and disability are due to the chronic diseases of industrialization that affect mainly elderly people, i.e. cardiovascular disease, cancer, diabetes, osteoporosis.
- Reducing mortality and morbidity from these chronic diseases is now seen as the priority for nutrition education and health promotion in the industrialized nations.
- The new dietary recommendations to achieve this objective are in some way almost the opposite of those of the 1930s, e.g. to moderate consumption of meat, eggs and dairy produce but to increase consumption of starchy foods, fruits and vegetables.
- A risk factor is some parameter that partially predicts one's risk of developing a disease and that is causally linked to the onset of the disease.
- Dietary factors can be direct risk factors or they can influence the level of other factors, such as plasma cholesterol, blood pressure or glucose tolerance.
- Much modern health promotion is focused upon identifying and reducing these risk factors, and the sheer number of such risk factors can cause confusion and increase the possibility of conflicting advice.
- Recommendations to alter diet and lifestyle are sometimes not fully underpinned by adequate supporting evidence. There seems to be an implicit but illogical assumption that simple changes in diet and lifestyle have the potential to do good but not harm.
- There are examples of dietary or lifestyle changes based upon past health promotion advice appear to have done net harm:

past recommendations to put babies to sleep in the prone position seem to have increased the risk of cot death

β -carotene supplements have actually increased death rates in smokers and others at high risk of lung cancer.

- Even where a non-beneficial dietary or lifestyle change does no direct physiological harm, it may have less obvious adverse consequences, e.g. the resources wasted in trying to solve the illusory crisis in world protein supplies.
- At the very least, recommendations that are changed or withdrawn will undermine confidence in nutrition education and increase resistance to future advice.

risk of several cancers. There have been active campaigns in both the UK and USA to increase fruit and vegetable consumption and to persuade consumers to take at least five portions a day. β -carotene is a plant pigment found in green and brightly coloured fruits and vegetables. It acts as a source of vitamin A and also seems to be one of a group of antioxidant substances that may protect the body from oxidative damage by **free radicals** and perhaps lessen the risks of cancer and atherosclerosis (see Chapter 12). This means that diets high in coloured fruits and vegetables are also high in β -carotene and that high carotene intakes are associated with reduced cancer risk. A number of trials of β -carotene supplementation have paradoxically found higher rates of cancer in the supplemented group than in the unsupplemented controls (e.g. Group, 1994; Omenn *et al.*, 1996). Particularly in heavy smokers and asbestos workers, there is evidence that large β -carotene supplements may increase the risk of lung cancer. β -carotene has traditionally been regarded as non-toxic and very large doses are not acutely toxic.

These five examples have been selected to try to illustrate that once very popular dietary or lifestyle interventions may in later years become regarded as unnecessary or perhaps even harmful. The eventual outcome of each debate will not alter the demonstration of how scientific opinion about the value and safety of such interventions fluctuates.

If a holistic view is taken, then it is likely that any intervention that produces no benefits will have some detrimental effects, even though this may not involve direct physiological harm. Majority support amongst scientists and health professionals for intervention based upon a particular theory is no guarantee that it will continue to be regarded as useful or even safe.

Those involved in health promotion and nutrition education need to be constructively critical of current fashions. They need to be receptive to criticism of current wisdom and objective in their evaluation of such criticism. They need to be cautious about advocating mass intervention, especially if there is no convincing, direct evidence of overall long-term benefit or where evidence of benefit is restricted to a relatively small 'high-risk' group. Any non-beneficial intervention should be regarded as almost inevitably detrimental in some way.

IS INTERVENTION TO INDUCE DIETARY CHANGE JUSTIFIED? (after Webb, 1992a)

Evidence linking diet and disease seldom materializes in complete and unequivocal form overnight. It usually accumulates gradually over a period of years, with a high probability that some of the evidence will be conflicting. The problem of deciding when evidence is sufficient to warrant issuing advice or some other intervention is thus one that will frequently need to be faced. If action is initiated too soon, there is a risk of costly and potentially harmful errors. If action is too slow, the potential benefits of change may be unduly delayed.

There is likely to be intense pressure to take action or to issue guidelines in the light of the latest, highly publicized research findings. Below is listed a set of criteria that might be used to decide whether any particular intervention is justified by available knowledge.

- Have clear and realistic dietary objectives been set and have all foreseeable aspects of their probable impact been considered?
- Is there strong evidence that a large proportion of the population or target group will gain significant benefits from the proposed intervention?
- Has there been active and adequate consideration of whether the promoted action, or any consequential change, might have adverse effects on a significant number of people?
- Have the evaluations of risk and benefit been made holistically? A reduction in a disease risk marker, or even reduced incidence of a particular disease, is not an end in itself; the ultimate criterion of success must be an increase in life expectancy or a net improvement in quality of life. Likewise, the evaluation of risks should not be confined to possible direct physiological harm, but should also include, for example, economic, psychological or social repercussions. It should also include the possibly harmful effects of any consequential changes and consider the damage ineffective intervention might have on the future credibility of health educators.
- Has the possibility of targeting the intervention to those who are likely to gain been fully explored so as to maximize the benefits-to-risks ratio?
- Has consideration been given to how the desired change can be implemented with the minimum

intrusion into the chosen lifestyle and cuisine of the target group?

Intervention that precedes satisfactory consideration of the likely risks and benefits is experimentation and not health promotion.

Such criteria are probably not, in themselves, controversial; it is in deciding when these criteria have been satisfactorily fulfilled or even whether they have been properly considered that the controversy arises.

Judging whether these intervention criteria have been fulfilled

A number of practical questions, such as those listed below, need to be addressed and satisfactorily answered before any apparent association between a dietary variable and a disease or risk factor is translated into practical health promotion advice.

- Is the reported association likely to be genuine?
- Is the association likely to represent cause and effect?
- What change in dietary composition is realistically achievable and what magnitude of benefit is this predicted to produce?
- Is this predicted benefit sufficient to warrant intervention – at the population level/at the individual level?
- Are there any foreseeable risks from the proposed compositional change for any group within the target population?
- Are there any foreseeable non-nutritional adverse consequences of these changes?
- What changes in other nutrient intakes are likely to occur as a result of the proposed advice?
- Are there any foreseeable risks from these consequential changes?
- How can the desired compositional change be brought about?

Is the association genuine?

It is quite possible that an apparent association between diet and a disease may have arisen because of bias in the study or simply by chance. There is often a range of papers reporting conflicting findings; these need to be evaluated and weighted rather

than the latest or the majority view being mechanically accepted. It is quite possible that a common logical flaw, incorrect assumption or methodological error is present in all of the papers supporting one side of an argument.

Easterbrook *et al.* (1992) reported that, in clinical research, positive results are more likely to be published than negative ones. This not only means that the literature may give an unbalanced view of total research findings, but also encourages authors to highlight positive findings. In another analysis, Ravnskov (1992) reported that, when citation rates were determined for papers dealing with cholesterol-lowering trials, those with positive outcomes (i.e. suggesting a beneficial effect on coronary heart disease) were cited six times more frequently than those with negative outcomes. Thus, at the author, peer review, editorial selection and citation level, there may be bias towards positive over negative findings.

Statistical significance at, or below, the 5% level (i.e. $p < 0.05$) is, by convention, taken as the point at which a scientist can claim that, say, a correlation or a difference between two means is 'statistically significant'. It means that there is less than a 1 in 20 likelihood that the correlation or difference between the means has arisen simply by chance. This statistical significance may simply reflect flaws or bias in the design of the experiment, in the allocation of subjects or in the measurement of variables. With numerous investigators all correlating numerous dietary variables with health indicators, it is inevitable that some statistically significant associations will arise simply by chance. Significance at the 5% level is not proof of the underlying theory, merely an indication that any difference or association is unlikely to be due to pure chance. This should be borne in mind when evaluating isolated reports of improbable-sounding associations. Given the perceived onus upon authors to highlight positive results, they may be tempted to highlight the one barely significant correlation in a whole battery of otherwise insignificant tests.

Dietary guidelines often stem from reviews of the scientific literature or the conclusions of expert committees, but we have already seen that even apparent consensus amongst experts does not guarantee the long-term future of a particular position or theory. There was once an almost unanimous belief in a crisis of world protein supply, but few would support this view today. Scientists are trained to make objective

judgements based solely upon scientific evidence, but they are also fallible human beings whose judgements may, perhaps unwittingly, be affected by prejudice, political or peer pressure and self-interest. Reviewers and committee members will often be selected because of their active involvement and expertise in a particular field of study. This may sometimes hamper their objectivity, especially when they are required to evaluate material that may undermine or support their own work and reputation, or that might even jeopardize their own career or research funding. McLaren (1974) tried to identify some of the reasons for the now discredited concept of a world protein crisis or protein gap and for its persistence (see Chapter 10). He suggested that, even in 1966, when this theory was still at its height, many scientists had privately expressed sympathy with his opposition to the theory and some of the consequential measures designed to alleviate the perceived crisis. He also claims that they were unwilling to support him publicly for fear of having their research funding withdrawn. He even suggests that critical discussion of the issues was suppressed.

I have made a conscious decision to include in this book a substantial section on nutrition research methods, even though the bulk of readers are unlikely to be aiming for research careers in nutrition. Those readers who are primarily interested in the practice of promoting health and are unlikely to participate in research projects may be tempted to disregard much of this section as superfluous to their needs. I think that this would be a narrow and short-sighted decision. Some general appreciation of the research methodology, especially the strengths, limitations and weaknesses of the various methods, is essential for any critical reading and evaluation of the literature. Anyone who wishes to keep up to date with the literature and to be able to make quality judgements about conflicting evidence, or controversial new reports, will need some basic understanding of the methodology.

Is the association causal?

Epidemiological methods produce evidence of associations between diseases and suggested causes, but even a strong association does not necessarily mean that there is a cause-and-effect relationship. It is quite probable that the statistical link between a risk mar-

ker and a disease may have arisen because both are associated with a third, **confounding variable**. For example, even though alcohol may not directly contribute to causing lung cancer, a statistical association between alcohol consumption and lung cancer might arise if heavy drinkers also tend to be heavy smokers. Earlier in this chapter it was noted that high fruit and vegetable consumption is associated with reduced cancer risk. This also means that high β -carotene intake is associated with lowered cancer risk, but does not necessarily mean that high β -carotene intake prevents cancer. There are several other explanations for this finding, such as those listed below.

- People who eat lots of fruits and vegetables may have other lifestyle characteristics that help to reduce their cancer risk, e.g. different smoking, drinking and exercise habits.
- High fruit and vegetable consumption may be a marker for a diet that is healthier in other respects, e.g. lower in fat.
- Even if high fruit and vegetables intake is directly protecting against cancer, it may be some component(s) other than β -carotene that is exerting the protective effect.

As we saw earlier, some studies have shown that β -carotene supplements appear to increase lung cancer rates in smokers and asbestos workers.

Epidemiologists almost always try to correct their results for the effects of confounding variables when trying to establish a causative link between diet and disease. There is, however, no statistical magic wand that unerringly and accurately corrects for all confounding variables; it is a matter of judgement what the likely confounders are, and the process of correction itself may be imprecise, particularly if there is imprecise or limited measurement of the confounding variable (Leon, 1993). Take the examples of smoking and exercise as possible confounding variables in epidemiological studies of the relationship between diet and heart disease. Apparently reliable information on smoking habits is usually obtained in such studies and one has confidence that a reasonable correction of the results for the effects of smoking has been made. On the other hand, it is notoriously difficult to make an accurate assessment of activity level and so, in many of these studies, there is little or no attempt to correct for the effects of variation in activity levels. This problem of confounding variables is discussed more fully in Chapter 3, along

with some tests that may be applied to help decide whether any particular association is causal.

In some cases it may be that the disease itself leads to the measured changes in diet or behaviour, i.e. an effect-and-cause relationship. It is particularly important to consider this possibility when one is comparing the current behaviour of those who have or may have the disease with those who do not. Obese people are less active than lean people; is this a cause or an effect of their obesity? A low serum vitamin A concentration may be associated with an increased risk of developing symptomatic cancer within a few years, but is this a cause of the cancer or an early symptom of it?

Dietary and lifestyle changes tend not to occur in isolation; often many changes occur simultaneously. Increasing affluence and industrialization tend to bring with them numerous changes in diet and lifestyle as well as major changes in the causes of death and disease. This means that an association between one of the 'diseases of industrialization' and one of the wealth-related environmental variables needs to be treated with extreme caution until there is substantial evidence to support this association being causal.

Realistically achievable change – magnitude of benefit?

If a cause-and-effect relationship is established between a dietary factor and a disease, one should be able to predict the magnitude of benefit that is likely to result from a change in diet. For example, if there is a linear relationship between the average blood pressure of adult populations and their average daily salt consumption, one can attempt to predict the reduction in average blood pressure that would result from a reduction in average salt consumption (e.g. Law *et al.*, 1991a). Then, making certain assumptions, one can go on to predict by how much the incidence of hypertension will fall and thence the beneficial effects on mortality from the hypertension-related diseases. (See Chapter 13 for further discussion of this example.) One must then decide what degree of change is realistically achievable and what the benefits of change on this scale are likely to be. It is unhelpful to predict the huge health gains that might result from a degree of change that is impractical or unrealistic. A reduction in average UK daily

salt consumption from around 10 g to less than 1 g, as seen in some populations such as the Yanomamo Indians of Brazil, might eradicate hypertension and produce major reductions in mortality from hypertension-related diseases. There is, however, no chance that a change of this magnitude can be achieved.

Expert committees on both sides of the Atlantic have concluded that very major health benefits would accrue if the proportion of total energy derived from fat were to be reduced from a population average of around 40% to around 30% or less. Whilst such a change may be readily achievable by small numbers of highly motivated and well-educated individuals, is it a realistic goal for the population as a whole? According to Passmore and Eastwood (1986), fat provided 38% of total energy in Britain in the period 1934–1938 but, as a result of wartime shortages and rationing, this figure fell gradually in the following years to reach a minimum of 33% in 1947. Passmore and Eastwood suggested that this relatively small reduction in fat consumption caused widespread discontent. They concluded that, although the wartime diet contained much more fat than was strictly necessary, it was, nonetheless, unpalatable to British tastes and did not satisfy the social and cultural demands for fat in the diet.

This is not a good omen for hopes of bringing average fat consumption down to 30% or less of total energy. It is likely to be a difficult task to persuade an unrestricted population voluntarily to consume a diet in peacetime that is even lower in fat than the partly imposed wartime diet that caused so much discontent. Despite the apparent increase in diet consciousness of the British population over the last three decades, there has been comparatively little change in the proportion of dietary energy that is derived from fat, although there have been marked changes in the types of fat consumed. Between 1975 and 1992, the proportion of food energy derived from fat remained relatively constant at about 41%, and had shown only a modest decline to 39.7% by 1996 (MAFF, 1993, 1997). In contrast, the ratio of polyunsaturated to saturated fat in the diet more than doubled from 0.19 in 1975 to 0.47 in 1996. People seem much more willing to change the type of fat they consume than to reduce their total fat consumption (see Chapter 11).

Those people who are most refractory to change may often be the most likely to benefit from the

change. Young women from the upper social groups are more likely to try to reduce their fat consumption than young, working-class men. The latter might be expected to have much more to gain from such a change than the former. Thus, Gregory *et al.* (1990) found that British women were more likely than men to choose food items with a 'healthy image', such as low-fat milks, salad vegetables, fruit and wholemeal bread, but less likely than men to choose sausages, meat pies, chips (French fries) and fried white fish. Social class differences in food selection are discussed under economics in Chapter 2.

At what level is intervention justified?

Even when there is clear evidence of a causal association between a dietary factor and a disease, it may still be difficult to decide whether this justifies intervention at the population level. The extreme scenarios are listed below.

- If a minor dietary change could produce a major reduction in the risk of a common fatal disease, the case for intervention would appear strong.
- If a major dietary change were only predicted to marginally reduce the risk of a relatively uncommon disease, clearly population-level intervention would not be justified.

In between these extremes, the significance or 'cost' of the proposed change will need to be balanced against the predicted reduction in population risk. Even where likely benefit cannot justify intervention at the population level, if those who are particularly 'at risk' or who are particularly susceptible to the causal factor can be identified, then this may justify intervention for those high-risk groups.

There are, therefore, essentially two strategies to choose from when designing a health promotion campaign aimed at risk reduction, namely:

1. *The population approach:* advice is directed at the whole population, or large sections of it, with the aim of reducing average population exposure to the risk.
2. *The individual approach:* efforts are made to identify those most at risk from the risk factor and intervention is targeted specifically at these 'high-risk' individuals.

If one takes the example of a programme aimed at lowering plasma cholesterol in order to reduce the

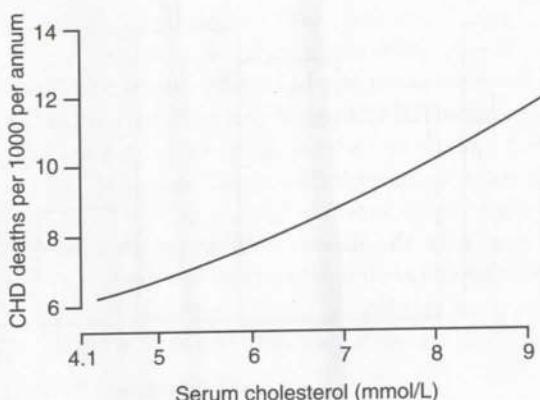


Figure 1.1 The relationship between serum cholesterol concentration and subsequent death rate from coronary heart disease (CHD) in a sample of American men. From Webb (1992a), after NACNE (1983).

risk of coronary heart disease (CHD), then one could choose either of the following options:

1. *The population approach:* direct the message at the whole adult population; if successful, this would be expected to lower average plasma cholesterol concentrations and shift the whole population distribution of plasma cholesterol concentrations downwards.
2. *The individual approach:* one could seek to identify those with the highest cholesterol concentrations and most 'at risk' from CHD and target the advice specifically at these 'high-risk' individuals.

Much health promotion/nutrition education in the UK has tended to use the population approach. If one is seeking to make a significant impact upon population mortality/morbidity from a disease, as in the 'Health of the Nation' targets (DH, 1992) or 'Health People 2000' in the USA (DHHS, 1992), then the population approach appears to offer the greatest probability of success. Consider the example of the relationship between plasma cholesterol concentration and CHD. It is generally accepted that, for young and middle-aged men, high plasma cholesterol concentrations predict an increased risk of premature death from CHD. For example, Figure 1.1 shows that, for one population of American men, mortality risk for CHD approximately doubled as the plasma cholesterol concentration doubled. It is assumed that the lowering of serum cholesterol will result in a corresponding reduction in CHD risk. Figure 1.2 shows that there are relatively few men in this population

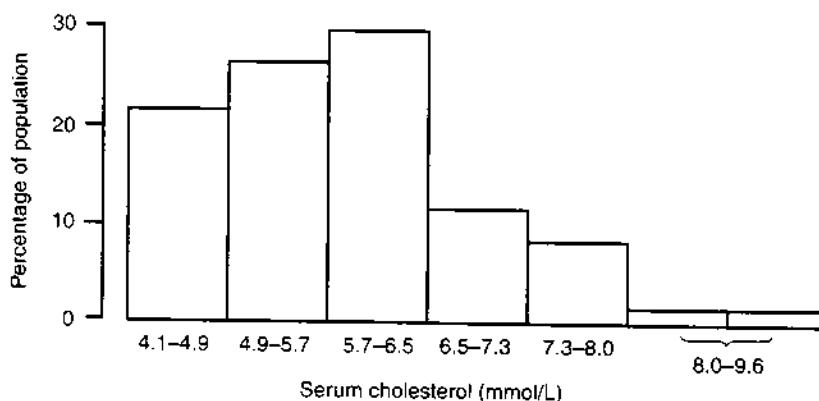


Figure 1.2 Distribution of serum cholesterol concentrations in the sample of men in Figure 1.1.
From Webb (1992a), after NACNE (1983).

with very high plasma cholesterol levels (say, over 7.8 mmol/L) and so restricting cholesterol-lowering efforts to these high-risk subjects can only have limited impact. Even a substantial reduction of risk in this small group will do little to affect population mortality from CHD. The majority of the population in Figure 1.1 (and the total population of the UK) have plasma cholesterol concentrations that are regarded as slightly or moderately elevated (say, 5.2–7.8 mmol/L). Even a small reduction in individual risks for this very large group will have a significant effect upon total population mortality.

Some of the advantages and disadvantages of these two approaches are listed below (after Webb and Copeman, 1996).

- If everyone is advised to change, then the promoted behaviour may become the norm and may create conditions that make it easier for any individual to change. For example, if everyone is urged to reduce their fat and saturated fat intake, this increases the incentive for the food industry to develop and market products that are low in fat and/or saturated fat – products such as low-fat spreads; margarine and oils that are low in saturated fat; lower-fat milks; and reduced-fat processed foods. The ready availability of such foods makes it easier to adopt a low-fat diet and low-fat choices may come to be perceived as the norm.
- The population approach removes the need to identify the 'high-risk' subjects and mark them out as 'victims' of a particular condition. The individual approach may require a mass screening programme, e.g. to identify those above some arbitrary cut-off point for plasma cholesterol. If screening is not universal, some 'high-risk' individuals
- may not be identified. Once identified, 'high-risk' individuals would probably feel threatened by their 'high-risk' classification and change might be more difficult for them to implement because the changes in product availability and social norms discussed above would be unlikely to occur.
- The population approach requires mass change and, for many, that change will offer only a small reduction in individual risk and thus the motivation to change may be low.
- If mass change is induced, there is an increased likelihood that some currently 'low-risk' individuals may be harmed by the change; it may create norms of behaviour and perhaps even changes in the range and pricing of foods that are not universally beneficial. We saw earlier that the increased use of the front sleeping position for babies in the 1970s and 1980s may have helped some premature babies and some babies with particular conditions, but for the majority it probably increased the risk of cot death.
- The population approach may create an unhealthy preoccupation with death and disease in the whole population (McCormick and Skrabanek, 1988), e.g. even people at very low risk of CHD may be made extremely fearful of this condition. This would be particularly true if health promotion focused upon the negative aspects of not accepting the message, i.e. if it used fear of death and disease and the threat of social isolation as the levers for effecting behaviour change.
- The population approach may lead to people being blamed for their condition and perhaps may be used by doctors and politicians to justify the withholding of expensive treatments: 'It is your

fault that you have heart disease/cancer because you smoke/eat badly/do not take enough exercise.'

Harmful effects of intervention?

First do no harm.

Health promotion should increase life expectancy and reduce ill-health. It is illogical to assume that dietary or lifestyle changes are an inevitable one-way bet and can only be beneficial. The possibility of harmful effects must be actively considered and tested. Several possible examples of simple dietary or lifestyle interventions having unexpected harmful effects have already been discussed in this chapter. Harmful social, psychological, economic or cultural repercussions are likely even where no direct physiological harm results from unjustified intervention. Non-beneficial intervention will, at the very least, ultimately undermine the credibility of future programmes.

Some reports published in the early 1990s suggested that very low plasma cholesterol levels might be associated with increased total mortality. CHD mortality is low in subjects with very low cholesterol levels, but there is a rise in mortality from other causes, including most cancers (Neaton *et al.*, 1992; Jacobs *et al.*, 1992). According to Neaton *et al.* (1992), the inverse association between plasma cholesterol level and most cancers remained even after a 12-year follow-up period, making it unlikely that the lowering of plasma cholesterol is simply an early effect of cancer. Jacobs *et al.* (1992) analysed the relationship between plasma cholesterol concentration and various causes of mortality in more than 600 000 people. They found a U-curve of mortality associated with a rising plasma cholesterol concentration. At very low plasma concentrations (say, < 4 mmol/L), although death rate from CHD is very low, total mortality is raised due to increased death rate from non-cardiovascular causes such as cancer. At high plasma cholesterol levels, there is increased mortality due to high

- Dietary change should only be recommended if there is clear evidence that it will yield net holistic benefit.
- The possibility that change may have adverse direct or indirect effects should be considered and tested.
- Premature action based upon insufficient evidence runs the risk of doing net harm, and undue delay means that any benefits of the intervention will also be delayed.
- Some apparent links between diet and disease may arise because of statistical freaks or may reflect bias in the study design.
- Trying to obtain a consensus view from published reports will be hampered because negative results are less likely to be published and cited than studies with positive outcomes.
- Association between diet and disease does not necessarily mean that there is a cause-and-effect relationship. It may arise because both the disease risk and dietary factor are linked to other confounding variables.
- Health promotion intervention may be directed at the whole population with the aim of inducing mass change (the population approach) or it may be targeted at those individuals regarded as at higher risk of the particular disease (the individual approach).
- The population approach may create a climate and conditions that facilitate individual change. It removes the need to identify and mark out those at high risk of the disease.
- On the other hand, mass change may only offer many people a small likelihood of benefit and it increases the risk that some will be harmed by the change. This approach may frighten even those who are at low risk of the disease, and it may foster an unhealthy preoccupation with death and disease.
- Any recommended dietary change would have other consequential changes upon diet composition that need to be considered before mass change is advocated.

rates of death from cardiovascular disease. In young and middle-aged men, there still appeared to be clear evidence of a net benefit to be gained from a lowering of average cholesterol concentration. In women, however, they found no association between total mortality and plasma cholesterol concentration: even cardiovascular mortality was not significantly associated with cholesterol level because, as cholesterol concentration fell, so the fall in CHD mortality was offset by a rise in mortality from stroke.

If very low cholesterol level does have some adverse effects, then cholesterol-lowering interventions that are aimed at the whole population will presumably tend to increase the total mortality of people with low plasma cholesterol. It would also tend to push some individuals into the 'danger-zone' at the low end of the range. Hulley *et al.* (1992), in an editorial discussion of some of these results, questioned whether universal cholesterol-lowering advice could still be ethically justified. They particularly questioned the value of cholesterol lowering in women.

There have also been several reports of increased numbers of violent deaths and suicides in several cholesterol-lowering trials (see Engelberg, 1992). Engelberg has suggested that cholesterol-lowering might increase aggression by altering brain transmission involving 5HT (serotonin).

What are the secondary consequences of change?

Dietary changes almost never occur in isolation. Changing the diet to produce an intended change will almost inevitably affect the intake of other nutrients. When dietary changes are to be advocated, the likely impact of the intended change upon the total diet should be considered. If reduced intake of a particular food or nutrient is advised, how are the intakes of other nutrients likely to be affected and what is likely to replace the missing food? If people are advised to increase their intake of particular foods, what will they displace from the existing diet?

For example, if food manufacturers are persuaded to use less salt in processing, what effect will this have on shelf-life and therefore on the cost and availability of foods? Will it compromise the microbiological safety of some foods? Or will it lead to increased reliance on other preservatives whose long-

term safety may also be questioned? In Chapter 9 there is a brief discussion of the likely effects of using artificial sweeteners: will they reduce total energy intake by reducing sugar consumption or will the lost sugar energy be replaced by other foods and so increase total fat consumption?

EFFECTING DIETARY CHANGE

It is easy to devise diets or plan menus that meet any particular set of compositional criteria, but much more difficult to make them acceptable to the clients. For example, it is easy to produce nutritionally adequate reducing diets, but the long-term dietary treatment of obesity is notoriously unsuccessful. Diets or dietary recommendations that are sympathetic to the beliefs, preferences and usual selection practices of the recipients will probably fare better than those that ignore the non-nutritional influences on food selection or try to impose the prejudices and preferences of the adviser.

Food has a host of cultural, psychological and social functions in addition to its biological function. It is these non-nutritional uses that have traditionally been the dominant influences upon our food selection. If we lose sight of these functions in the perhaps fruitless pursuit of the optimal chemical diet, any technical improvements in our nutrition may be outweighed, or at least reduced, by damage to the general emotional health of the population. Miserable, anxious people will probably suffer more illnesses and die younger than happy, contented people.

Some of these non-nutritional influences upon food selection are discussed in Chapter 2. If health promoters want to effect changes in eating behaviour, they must have some appreciation of the determinants of food selection practices. They can then offer practical advice that will bring about any desired changes in diet in ways that are compatible with the overall cultural milieu of the target group. Nutrition educators should try to effect compositional changes in the desired direction with the least interference in current cultural practices, i.e. a cultural conservationist approach. This is the most ethical approach as well as the most likely to succeed.

Some factors influencing the likelihood of dietary change

Fieldhouse (1998) listed certain parameters that will particularly influence the likelihood of dietary change occurring:

- *The advantage of change and the observability of that advantage.* If a relatively minor dietary change results in immediate and apparent benefits, this increases the likelihood of permanent adoption. However, if the benefits of change are not immediately apparent or speculative or if the change itself is seen as having major disadvantages (e.g. restricting favoured foods, increased cost, or increased preparation time), this will decrease the chances of its adoption.
- *Its compatibility with existing beliefs, cultural values and culinary style.* Advice that is compatible with existing beliefs and behaviour is more likely to be implemented than that which is contrary to these beliefs and behaviours (e.g. advising a pious person to break the dietary rules of his or her religion has a low probability of success).
- *Its complexity.* The more difficult it is to implement or understand any change, the less chance there is that it will be made. Advice aimed at lowering plasma cholesterol is often perceived as requiring major reductions in the consumption of meat and dairy produce. These high-prestige and high-flavour foods are central to the culinary practices and the food ideology of many Western consumers. Such changes may therefore be seen as requiring a complete restructuring of the diet. Promoting simple modifications of existing practices that involve relatively little 'cost' for the consumer may produce more real change than advocating much more ambitious and drastic changes that will be implemented by relatively few people – for example, simple suggestions such as: buy leaner cuts of meat; trim off visible fat; grill (broil) rather than fry; use lower-fat milk; use a spreading fat that is lower in saturated fat.
- *Its trialability.* If advice is easy to try out on a 'one-off' basis, people are more likely to try it out than if it requires more commitment, such as the learning of a new preparation method or buying new catering equipment. Getting people to try out a dietary change is the first step towards adoption of the change. Switching from butter to mar-

garine and cooking with vegetable oil rather than animal fats seem like readily triable changes. The fluoridation of water supplies is an example of a change that is not readily triable. If a change requires a period of adjustment, this would reduce its triability. People may, for example, get used to food cooked without added salt, but in order to adjust to the change, they may need an extended trial period. Increasing fibre intake by switching from white to wholemeal bread may seem eminently triable. However, sudden increases in fibre intake may produce transient diarrhoea, flatulence and abdominal discomfort and it may require some time to adapt to the higher fibre intake.

Advice to make a dietary change that is aimed at curing a deficiency disease has a good chance of being accepted. The major advantages of such a change are usually readily apparent within a short period of time and can often be observed in others who have already made it. The disadvantages of changing may be small, e.g. incorporation of a nutrient-rich or fortified food into the diet or even the consumption of a nutrient supplement. This sort of dietary modification is also easy to understand and implement. It can usually be tried out without any long-term commitment, and the rapid and obvious benefits encourage people to continue with the change in diet.

Similarly, when a therapeutic diet gives rapid symptomatic relief, this will also act as powerful, positive reinforcement and encourage clients to stick with the diet and perhaps even to stick to it more rigorously. For example, if the adoption of a low-protein diet gives symptomatic relief to patients with chronic renal failure, this may well encourage them to persevere with this diet despite its relative complexity and restrictiveness and irrespective of any possible long-term benefits on the course of the disease (see Chapter 15 for more details). Increased fibre and water intake will often relieve constipation or give symptomatic relief from other minor bowel disorders (e.g. diverticulosis or haemorrhoids). These benefits will act as a positive reinforcement for people suffering from these conditions.

It may be much more difficult to persuade healthy people to make changes that are aimed at reducing the long-term risk of chronic disease. The benefits of many of the health-promoting dietary changes

currently recommended by expert committees are often difficult to demonstrate at the individual level and in the short term. The benefits are often only expected to accrue some time after the introduction of the change; they are often perceived as speculative, with some expert opinion arguing against even the long-term benefits of change. These measures are designed to produce significant long-term reductions in the population risk of death and disease, and the individual may only be offered a relatively small reduction in his or her individual risk of suffering from some disease at some seemingly distant date. Anecdotal observations may also tend to confound the predictions of the experts; there will be examples of 'healthy eaters' who nevertheless develop diet-related diseases, and of those who ignore dietary guidelines but remain healthy well into old age. This anecdotal tendency may be compounded by the tendency of people to adopt practices they perceive as healthy when they start to become concerned about their health, i.e. the adoption of a healthy diet may apparently immediately precede the onset of overt symptoms and so may even appear causative.

The dietary changes required to meet current nutrition education guidelines may be perceived as having immediate and major disadvantages, i.e. the restriction of highly palatable sugary, salty and fatty foods and their partial replacement with bland, starchy and fibre-rich foods. Increasing fear and guilt about 'inappropriate' eating behaviours may be one way of altering the balance of advantage of implementing change. Thus, a first heart attack can persuade people to adopt strict cholesterol-lowering diets, and the onset of serious lung disease may persuade people to give up smoking. However, health promotion that tries to harness such fear of death and disease as a means of inducing dietary/lifestyle change in healthy people serves to focus people's attention upon death and disease. It may have some negative impact upon quality of life. If such 'health promotion' succeeds in generating fear and guilt without changing behaviour, or results in changes that ultimately prove not to be beneficial, then its impact will be totally negative. Similarly, increased feedback by regular monitoring of risk factors such as plasma cholesterol may improve compliance but, McCormick and Skrabanek (1988) argue, such focusing upon symptomless risk markers may serve only to induce an unjustified and morbid preoccupation with death. Health and wellness promotion is a very positive-sounding objective;

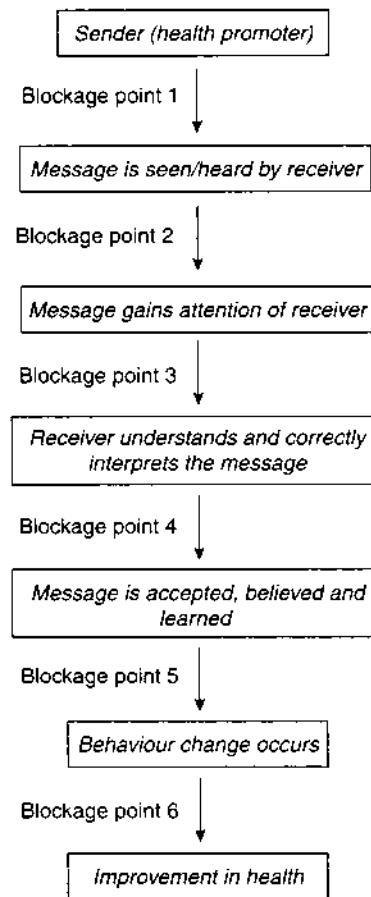


Figure 1.3 Steps in the process by which a health promotion message leads to beneficial change in behaviour and potential barriers to change. Modified from Webb and Copeman (1996).

death delaying has a much more negative ring to it. Nutrition education must be presented and implemented in ways that emphasize the former objectives rather than unnecessarily increasing fear of, and preoccupation with, death and disease.

Barriers to successful health promotion

Figure 1.3 summarizes some of the stages in the process of a nutrition education or health promotion message leading to changes in behaviour that benefit health. There are potential barriers or blockages to successful health promotion at each of these stages and these blockage points on Figure 1.3 are discussed below.

BLOCKAGE POINT 1: THE RECEIVER FAILS TO SEE OR HEAR THE MESSAGE

An inappropriate channel may have been chosen for the message. *The channel* is the vehicle for transmitting the message, e.g. radio, television, newspaper or poster advertisements, leaflets provided at a doctor's surgery or community centre, speakers addressing groups or one-to-one counselling. Those initiating the health promotion campaign should consider what channel is most likely to reach the target group and to influence their behaviour. They would also have to consider what would be the most cost effective vehicle for transmitting the message to the target group. One would choose different television or radio slots and different newspapers and periodicals if one were targeting teenagers than if elderly people were the primary target group. Some simple market research may need to be done to ensure that the message will reach the receiver.

BLOCKAGE POINT 2: THE MESSAGE MAY NOT GAIN THE ATTENTION OF THE TARGET GROUP

A poster or television advertisement may not attract or interest the target group – images of teenage pop stars are more likely to interest teenagers than elderly people. Programmes, posters, oral presentations and leaflets must be attractive and interesting; they must be appropriate for the target group.

BLOCKAGE POINT 3: THE MESSAGE MAY NOT BE UNDERSTOOD BY THE RECEIVER

This may result in a blockage at this stage or, if the message is misinterpreted, it may result in an inappropriate behaviour change. For example, a message to increase dietary fibre may result in an inappropriate behaviour change if the nature and sources of dietary fibre are misunderstood. Health promoters must ensure that their message is clear, correct, believable and realistic. They must consider the standards of education and literacy of the target group and make special provision for groups for whom English is not the first language. If illiteracy is common, a written message is of little value. Ideally, any promotional material should be piloted to identify potential problems with terminology, understanding or interpretation.

BLOCKAGE POINT 4: THE MESSAGE MAY NOT BE BELIEVED OR, IF IT IS TOO COMPLEX AND NOT REINFORCED, IT MAY BE FORGOTTEN

The source may be seen as biased or may lack credibility with the receiver. An elderly male speaker perceived as conservative might lack credibility with an audience of streetwise teenaged girls. Promotional leaflets sponsored by a commercial company and bearing the company logo may be viewed as biased. The message may be inconsistent with existing scientific beliefs, e.g. advice to increase starchy foods may be inconsistent with the belief that they are fattening. It may be inconsistent with traditional cultural beliefs about diet, such as the need of Chinese people to balance foods classified as 'hot' and 'cold' in their cultural food classification system (see Chapter 2 for details of this example).

BLOCKAGE POINT 5: BEHAVIOUR CHANGE DOES NOT OCCUR DESPITE THE MESSAGE BEING RECEIVED AND UNDERSTOOD

The required change of behaviour may be difficult or impossible for the receiver to implement. Dietary changes that increase the cost of the diet may be beyond the economic means of some groups. Some groups who are largely reliant on others to provide their food may find it impossible to implement the change, e.g. children who rely upon their parents to provide food or those living in residential homes for the elderly or other institutions.

BLOCKAGE POINT 6: BEHAVIOUR CHANGE OCCURS BUT DOES NOT RESULT IN IMPROVED HEALTH

The message may be wrong or inappropriately targeted (see earlier in the chapter), or it may be misunderstood and result in an inappropriate behaviour change (see blockage point 3).

Those designing nutrition education or other health promotion campaigns must:

- ensure that *the source* of the message will be seen by the receivers as credible and reliable
- make sure that *the message itself* is clear, correct, realistic and appropriately targeted
- ensure that *the channel(s)* chosen for disseminating the message is likely to reach the target group, attract their attention and influence their behaviour

- consider the particular characteristics of the receivers or target group before they can design an appropriate strategy. What is their level of education and literacy? What is their first language? What are their recreational interests, cultural beliefs and media habits?

The BASNEF model (Figure 1.4)

- B – beliefs
- A – attitudes
- SN – subjective norm
- EF – enabling factors

An individual's beliefs about diet and health will clearly influence his or her judgement about whether a suggested change is beneficial. If one is seeking to persuade people to modify their behaviour in order to reduce the risk of a particular disease: they must believe:

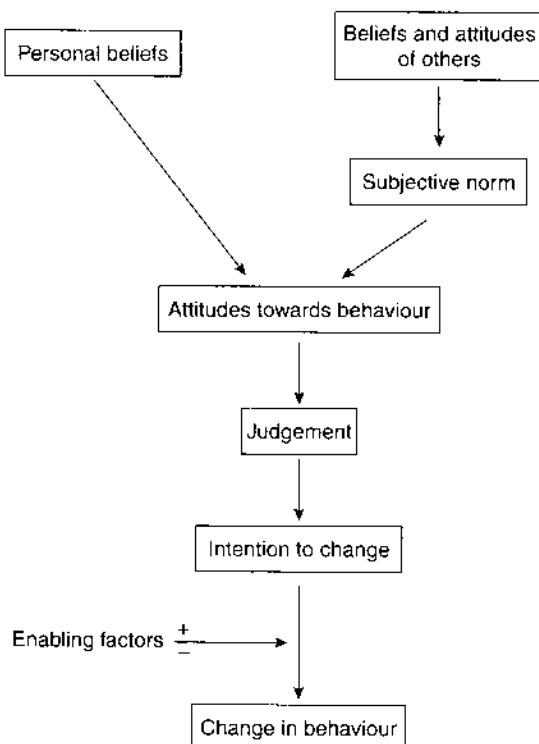


Figure 1.4 A scheme to illustrate how beliefs, attitudes, the subjective norm and enabling factors (BASNEF) interact to determine whether health promotion recommendations results in actual behaviour change. After Hubley (1993).

- that they are susceptible to the disease
- that the consequences are serious
- that it can be prevented (Becker, 1984).

If we take heart disease as a specific example, in order to make a judgement to modify their behaviour to avoid heart disease, the subjects must believe that:

- they are at risk from heart disease (i.e. susceptible)
- heart disease can kill or disable (the consequences are serious)
- if they change their diet and behaviour (e.g. stop smoking), the risk is reduced (prevention is possible).

It is not only the beliefs of the subjects themselves that will influence their judgements, but also the influences of family, friends, colleagues, teachers, religious leaders, health workers etc. Ajzen and Fishbein (1980) described these influences from the beliefs and attitudes of others as the perceived social pressure or the **subjective norm**. Health promotion has succeeded in changing several norms of behaviour (see examples below).

- Thirty years ago, butter would have been regarded by most UK consumers as the most natural, palatable spreading fat, and margarine as a cheap and inferior substitute. Those who could afford it would usually have chosen butter, and serving margarine to guests may well have been regarded as showing low esteem for them or meanness. Butter is now regarded as the least healthy of the spreading fats; it is seen as high in fat and saturated fat. The majority of UK consumers now choose either soft margarine or a low-fat spread as their everyday spreading fat and often use them for social occasions as well.
- In the 1950s, smoking was the normal behaviour in most sectors of British society and the non-smoker was often the odd one out. Now the opposite is often true: smokers are in a minority in many sectors of society and smoking is often regarded as an unacceptable activity in many public places and social situations. Smokers are often made to feel ostracized because of their habit.

Changing the subjective norm can encourage others to follow the change and may even make it easier for people to adopt the promoted behaviour, e.g. the greater range, availability and acceptance of alternative spreading fats to butter.

There are sound health reasons for trying to reduce the number of people who are overweight and obese. Nevertheless, creating a climate (subjective norm) in which fat is regarded as bad and lean good may not be a helpful health promotion approach. Most people already desperately want to become or remain lean. By focusing upon body weight *per se*, health promoters may simply increase the unhappiness and discrimination already experienced by obese people. It may also encourage some normal or even underweight young women to try to become even leaner and to adopt unhealthy dietary lifestyle practices in order to achieve this. Health promoters will instead need to try to promote dietary and lifestyle practices that encourage better weight control. (See Chapter 8 for further discussion of this example.)

Individual beliefs and the subjective norm will affect people's judgement of any particular recom-

mendation and thus affect their behaviour intentions. A host of 'enabling factors' will influence whether or not any intention or wish to change behaviour is translated into actual behaviour change. These enabling factors may come under a variety of headings, such as those listed below.

- *Physical resources:* is the change physically possible for the people concerned? Can they obtain the recommended foods? Have they the facilities necessary for their preparation? For example, the range and quality of fresh fruit and vegetables available in some rural areas may be limited. If low-income groups are temporarily housed in accommodation without cooking facilities, this may severely limit their dietary choices.
- *Financial resources:* can they afford to make the recommended changes? Fruit and vegetables

- It is relatively easy to devise diets that meet any given set of scientific criteria, but much more difficult to produce such a diet that will be complied with by the client.
- A dietary change is more likely to be implemented if there are clear and rapidly discernible benefits, if the change is consistent with existing beliefs and practices and if it is easy to understand, try out and incorporate.
- A dietary change that cures a deficiency disease produces clear and apparent benefits and often involves changes that are easily understood and implemented.
- A dietary change to reduce the risk of chronic disease:
 - usually produces no immediate benefits
 - has long-term benefits that may be small or even speculative
 - may require considerable and complex changes
 - may involve changes that markedly reduce the desirability of the diet.
- Health promotion will be ineffective if:
 - the message is wrong
 - the message is not seen and correctly assimilated by the client
 - the message does not convince the client that change is desirable
 - the client is unable or unwilling to make the recommended changes in behaviour.
- The BASNEF model suggests that beliefs, attitudes, the subjective norm and a range of enabling factors are the determinants of whether a health promotion message results in actual change:
 - people must believe that the recommended change will benefit them
 - they must be of a mind to make changes for the benefit of their long-term health
 - the subjective norm is the beliefs and attitudes of those people who influence the client and perhaps of society as a whole
 - enabling factors are a range of practical factors, such resources, skills, mobility and knowledge, that can determine the extent to which the client can change.

provide few calories per penny and so may be uneconomic choices for those seeking to satisfy appetites with a very low income.

- *Personal resources:* have they the skills and mobility necessary to implement the recommended changes? Many frail elderly people have mobility problems that will restrict their shopping and cooking capabilities. Some widowed elderly men or students living away from home for the first time may have had very little experience of preparing food for themselves.

In Chapter 2 many of these enabling factors are discussed using the 'hierarchy of availabilities' model of food selection (see Figure 2.2). In this model, it is assumed that a host of factors at several levels limit any individual's practical freedom of dietary choice and thus limit his or her ability to implement any recommended change.

People may be convinced that a health promotion message is applicable to them, the subjective norm may favour change and the enabling factors may permit them to make the recommended change. Despite this, change may still not occur; they may make a decision (judgement) to accept the risks associated with the current behaviour, perhaps because of the perceived adverse effect of change on their quality of life. In younger men, for example, the acceptance of risks associated with dangerous or

unhealthy practices may be seen as an expression of their virility.

CONCLUDING REMARKS

If the questions posed in this chapter are fully considered by those making nutrition or other health education recommendations, the likelihood of making changes that ultimately prove to be non-beneficial or harmful will be minimized. If a totally holistic view is taken, any intervention that is non-beneficial can probably be regarded as harmful. Observations of migrant populations show that, in general, they tend to progressively assume the mortality and morbidity patterns that are characteristic of their new homeland (Barker *et al.*, 1998).

This re-affirms the belief that many of the health differences between populations are due to environment rather than genetics and therefore also confirms that there is considerable potential for promoting health by lifestyle intervention. Paradoxically, as each population tends to have its own problem diseases, it also underlines the potential for harm from ill-considered and insufficiently tested interventions. A narrow preoccupation with changing diets to reduce the risk of one disease may simply lead to an increase in the risk of another.

Food selection

Introduction and aims of the chapter	27	A model of food selection	38
The biological model of food	28	Physical availability	39
Dietary and cultural prejudice	29	Economic availability	40
Food classification systems	30	Cultural availability	52
Non-nutritional uses of food	35	'Gatekeeper' limitations on availability	58
The hierarchy of human needs	36		

INTRODUCTION AND AIMS OF THE CHAPTER

Most of this book focuses upon food as a source of energy and nutrients; for example, which nutrients are essential and why; how much of these nutrients people need at various times of their lives; how changing the balance of nutrients eaten can affect long-term health, etc. This chapter is in some small way intended to be an antidote to the necessary scientific reductionism of much of the rest of the book. The general aim of the chapter is to remind readers that 'people eat food and not nutrients' and that nutrient content has only relatively recently become a significant factor in the making of food choices. Only in the latter part of the twentieth century has our knowledge of nutrient needs and the chemical composition of food become sufficient to allow them to be major influences upon food selection.

A host of seasonal, geographical, social and economic factors determine the availability of different foods to any individual or group, whilst cultural and preference factors affect its acceptability. Some of these influences are listed in Table 2.1.

If the promotion of health requires one to try to influence people's food choices, then some understanding of the non-nutritional uses of food, and of the way non-nutritional factors interact to influence

food choices, is essential. It is no use devising an excellent diet plan or drawing up detailed dietary guidelines unless they are actually implemented. Diets or dietary recommendations that may seem ideal from a reductionist biological viewpoint may have little

Table 2.1 *Some factors affecting food choice and eating patterns*

Availability of foods in the local environment; this is in turn influenced by several factors, such as climate, soil type, transportation links, rationing, shopping facilities
Nutrition knowledge and/or food beliefs; this is in turn influenced by such things as cultural traditions, education, and religious/ethical beliefs
Habit and previous food experience
Individual likes and dislikes
Facilities for storing, preparing and cooking
Cooking just for one or eating with and/or cooking for others
Skills in food preparation and willingness to experiment and develop new skills
Financial resources, budgeting skills and the cost of foods; these may be affected by political decisions about taxation, subsidy and welfare payments
Time available and the time needed to prepare and eat foods
State of health and appetite

After Webb and Copeman (1996).

impact upon actual food choices. Some people may have very limited freedom to make food choices, e.g. those living in institutions where all of their food is provided by a caterer. There will be some restraints upon the food choices of everyone. Dietary recommendations will almost certainly be ineffective if they are culturally unacceptable to the clients, incompatible with their personal preferences or beliefs, beyond their economic means or incompatible with what is provided by their caterer or parent.

When making dietary recommendations (or devising therapeutic diets), the advisers need to consider questions like those listed below.

- Can the clients obtain the recommended diet?
- Can they prepare it?
- Can they afford to implement the recommendations?
- Are the recommendations compatible with their cultural beliefs and practices?
- Will their caterer or parent provide a diet that complies with the recommendations?
- Will they find the recommended diet palatable?
- Can they tolerate the recommended foods?
- Are the recommendations environmentally sustainable if widely implemented? (Also discussed in Chapter 4.)

An appreciation of the non-nutritional roles of food and of the non-nutritional influences upon food selection should enable health educators to give advice that is more acceptable to clients and easier for them to act upon. Dietary recommendations that are in sympathy with existing beliefs, preferences and selection practices will fare better than those that try to impose the prejudices and preferences of the adviser. This culturally sensitive approach should also reduce the likelihood that the cost of scientifically better nutrition will be increased anxiety, social

tension and cultural impoverishment. Readers who wish to explore the social and cultural determinants of food choice and the non-nutritional roles of food are recommended to see Fieldhouse (1998).

THE BIOLOGICAL MODEL OF FOOD

The primary biological function of food is to provide the body with a sufficient supply of energy and essential nutrients to meet current physiological needs. Diet composition also influences the long-term risk of chronic disease. The reductionist scientific model of food is that it is merely a complex mixture of nutrients that should be combined in optimal proportions both to meet current needs and to maximize health and longevity in the longer term. Eating could be seen simply as a flawed and inefficient behavioural mechanism used to select and consume this mixture.

Such paradigms of food and of eating might tempt one to believe that if consumers are given extensive advice about optimal nutrient intakes coupled with full nutritional labelling of food, they will be enabled to select the ideal diet. Increasing information about healthy eating and food composition might be expected to lead to rapid and major 'improvements' in the diet of the population. (See Chapter 4 for evidence of the effect of health promotion on 'the national diet'.)

It is quite obvious that most people do not select their food solely according to biological criteria. In the past, selection on this basis would not have been possible because public awareness of nutrition and food composition is comparatively recent. Even today, few people would have the skills and knowledge to select food solely on a compositional basis. It has been suggested that biological mechanisms akin

- Nutrient content has not traditionally been a significant influence upon food selection: 'people eat food not nutrients'.
 - Social, economic and geographical factors influence the availability of food, whilst cultural and preference factors determine its acceptability.
 - Dietary guidelines or prescribed diets are more likely to be complied with if they allow for and are sympathetic to the non-nutritional factors that influence or limit food choices.

- Scientists may see food as no more than a mixture of nutrients that need to be consumed in the correct proportions and eating as merely the process by which these nutrients are selected and eaten.
- If food and eating fulfilled no other functions, then one would expect scientific guidance to lead to rapid improvements in this selection process.
- However, food does have many other functions and food choices are influenced by many factors, which may make it difficult to persuade people to make changes in their diet that seem rational from the purely scientific viewpoint.

to thirst or salt hunger operate for some important nutrients, enabling people to intuitively select a balanced diet. There is little evidence to support such mechanisms, but there is some research suggesting that satiation tends to be food specific, i.e. satiation for a particular food develops as that food is consumed, but satiation towards other foods is less affected. Such a mechanism would tend to increase the range of foods consumed, and dietary deficiencies become less likely if energy needs are fully met and a wide variety of foods consumed. In the circumstances of almost limitless abundance and variety experienced by those in the industrialized countries, such a mechanism might also tend to encourage over-eating and obesity (see **sensory-specific satiety** in Chapter 8).

Much current food evangelism seems to be encouraging a trend towards the reductionist scientific models of food and eating. This could greatly impoverish human existence without necessarily extending its duration, because food has numerous social, psychological and cultural functions in addition to its biological role of supplying nutrients.

DIETARY AND CULTURAL PREJUDICE

There is an almost inevitable tendency to regard one's own beliefs and patterns of behaviour as the norm and so preferable to those of other cultures. Foreign or unusual cultural practices tend to be regarded as wrong, irrational or misguided. The term **ethnocentrism** has been used to describe this tendency. Ethnocentrism is apparent in reactions to alien food habits. There is a widespread tendency to ridicule or even to abhor the food choices or eating habits of others. The intrepid explorer patronizingly

accepting some revolting native delicacy to avoid offending his host is a Hollywood cliché. One manifestation of this phenomenon is in the slightly derogatory names used for other races that have their origins in food habits, such as:

- *frogs*: because frog legs are a French delicacy
- *kraut*: from the traditional fermented cabbage (sauerkraut) associated with German cuisine
- *pom*: because of the perceived dominance of potatoes, or *pomme de terre*, in English cuisine
- *limey*: because of the past practice of providing lime juice to British sailors to prevent scurvy
- even the term *Eskimo* originates from a disparaging Indian term meaning 'eaters of raw flesh'.

Ethnocentric-type attitudes need not be confined to 'between-culture' judgements. It is not uncommon to hear haughty disapproval and caricaturization of the dietary practices of other regions of one's own country, of other social or religious groups or, indeed, of anyone who does not share a particular food ideology (see the examples below).

- Within the UK, it is not uncommon to hear disparaging comments made about the diets of 'northerners' which often implicitly blame northerners for the relatively poor health statistics of the North compared to the South.
- Vegetarianism is frequently denigrated and ridiculed by meat eaters and, conversely, some vegetarian propaganda makes very hostile comments about meat eaters.
- It is not uncommon to hear disparaging comments exchanged between those who do and those who do not consciously attempt to practise healthy eating.

Criticisms of other dietary and cultural practices are often based upon a prejudiced, narrow and inaccurate view of the other people's beliefs and behaviour. Nutritionists and dietitians are not immune to ethnocentrism, but it is hoped that most would be aware of this tendency and consciously try to avoid it when dealing with an alien culture. They might, however, be more unsuspecting, and thus inclined to behave ethnocentrically, when dealing with the more familiar behaviour patterns of other groups within their own culture. Although these behaviour patterns may be familiar, they may nonetheless be quite different from their own. For example, it must be very difficult for a dietitian who is a committed vegetarian to give advice to a meat eater that is totally uninfluenced by his or her own beliefs, and *vice versa* if the client is vegetarian and the dietitian a meat eater.

The opposite of ethnocentrism is **cultural relativism**. The cultural relativist tries to understand and respect other cultural practices and to accept them as normal no matter how bizarre they may at first seem or how different they are from his or her own. Only if practices are clearly and demonstrably dysfunctional does one try to change them. There would be little argument that such an approach was correct if one were dealing with an unique alien culture, but more familiar cultural practices may be handled with less sensitivity. The American's hamburger, the Briton's fish and chips and even children's sweets or candies have real cultural significance. Nutrition education need not try to totally forbid or blacken the image of such foods, but should rather attempt to use them within a diet that, in its entirety, complies with reasonable nutritional guidelines.

It is likely that many practices that are strongly and acutely dysfunctional to a group will have been select-

ed out during the cultural evolution of the group. Changing social or environmental conditions, or simply increased longevity and consequent changes in health priorities may cause traditional practices to become regarded as dysfunctional. The aim of nutrition education, under such circumstances, should be to minimize or avoid the dysfunction with the least possible cultural interference – the *cultural conservationist* approach. This is likely to be the most successful as well as the most ethical strategy.

FOOD CLASSIFICATION SYSTEMS

Nutritionists, consumers and anthropologists all categorize foods, but they use different criteria for their classification. These categorizations may be a formal and explicit classification system or a largely subconscious, practical classification.

Nutritional classification of food – food groups, plates and pyramids

Nutritionists classify foods according to the nutrients that they contain and these classifications are used to advise and guide consumers towards a healthy diet. Foods were initially classified into food groups according to their nutrient content, and consumers were advised to eat specified minimum amounts from each food group to ensure that their diet had adequate quantities of all the essential nutrients. As the priorities of nutrition education were widened to include the prevention of chronic disease, so new consumer guides were developed. These new guidance systems

- Ethnocentrism describes the tendency to regard one's own beliefs and practices as the norm and those of other cultures as wrong or abnormal.
- Ethnocentric-type attitudes can also exist within cultures and in these circumstances they may be less obvious and harder to correct.
- Cultural relativism is the acceptance of other cultural practices as normal even if they are very different from one's own.
- No matter how strange they may seem, one should only try to change other dietary practices if they are clearly dysfunctional, and then by the minimum amount to avoid the dysfunction – the cultural conservationist approach.

- Nutritionists classify foods according to their nutrient profiles. Food groups and food guide plates or pyramids are examples of nutritional classifications of food.

needed to indicate not just the minimum amounts needed for adequacy, but also the balance between the various food groups that would minimize the risk of chronic disease. A tilted plate model is used in the UK, and a food guide pyramid in the USA. Food group systems and these newer food guides are discussed more fully in Chapter 4.

CONSUMER CLASSIFICATIONS OF FOOD

Consumers also classify foods, but such classification systems have not traditionally had any theoretical basis in scientific nutrition. Despite this, such classification systems may have evolved rules that produce good practical diets, even though they may have a theoretical framework that seems incompatible with scientific nutrition theory. Nutritionists and dietitians should try to understand such classification systems and to offer advice that is consistent with them.

One of the best known and most widespread of the traditional and formal classification systems is the **hot and cold** classification that is found in various forms in Latin America, India and China. The general principle is that good health results from a state of balance and thus, to maintain or restore good health, there must be a balance between hot and cold. Foods are classified as hot or cold and should be selected and mixed to produce or maintain the balanced state. Disease results from an imbalance. Certain diseases and phases of the reproductive cycle are also regarded as hot or cold states and so certain foods will be more or less appropriate in these different circumstances. As an example, in the Chinese system, a sore throat is a hot disease and might be treated by a cold food such as watermelon in order to try to restore balance. Hot foods, such as beef, dates or chilli, are considered detrimental in such hot conditions. Past surveys of Chinese families living in London (Wheeler and Tan, 1983) and Toronto (Yeung *et al.*, 1973) both found that at that time the traditional hot and cold system was still widely adhered to and practised despite Western cultural influences. Wheeler and Tan (1983)

concluded that, despite the use of such a non-science-based food classification system, the dietary pattern of the London Chinese families in their survey was varied and nutritionally excellent. Any programme designed to increase the science-based nutrition and food composition knowledge of the 'gatekeepers' in these Chinese families runs the risk of undermining confidence in the traditional system and perhaps even worsening their diets. The 'gatekeepers' in English families living in the same area would probably have had more science-based nutrition knowledge, but these authors would probably have judged that, by their criteria, the English diet was inferior to that of the Chinese. Knowledge of food and nutrition is perhaps more loosely correlated with good dietetic practice than many nutritionists would like to think.

Most Western consumers do not use such a formal and overtly structured classification system, but they do classify foods. In any cultural group, there is clearly a classification of potentially edible material into food and non-food. Except under conditions of extreme deprivation, any cultural group will only eat some of the substances around them that would comply with any biological definition of potential food. In the UK, there are numerous plants, animals, birds, fish and insects that are edible but are rarely, if ever, eaten by Britons. They are simply not viewed as potential foods and are classified as non-food. In many cases, the idea of eating such items would be repellent to most Britons. For example, cows, chickens, cabbages and crabs are seen as potential food, but not horses, dogs, crows, nettles or frogs.

The traditional main course of a British dinner (or lunch, if that is the main meal of the day) consists of a meat or meat product, potatoes, one or more extra vegetables with sauce or gravy. Very few Britons would, however, consider eating such a meal or even some of the individual foods for breakfast. A 'cheeseburger and fries' is not (yet) on the breakfast menus of restaurant chains that have built their global empires on such fare. Clearly, some foods are seen, or classified, as more appropriate for particular meals. These classifications of edible material into food or

non-food and into foods appropriate for particular meals or occasions vary considerably, even within populations of Western Europe. In Britain, horse-meat is not classified as food and yet it has traditionally been eaten in France. Many Britons would not consider cheese and spicy cold meats as suitable breakfast foods, yet in some other European countries they would be typical breakfast fare.

Schutz *et al.* (1975) conducted a survey of 200 female, mainly white and mainly middle class consumers distributed between four US cities. They used a questionnaire in which these consumers were asked to rate 56 different foods in terms of their appropriateness for a total of 48 food-use situations; they used a seven-point scale from 1 ('never appropriate') to 7 ('always appropriate'). Their aim was to allow the respondents to generate classifications of foods based upon their appropriateness ratings. The authors identified five food categories based upon consumer usage rather than upon biological criteria, and these are listed below.

- *High-calorie treats* such as wine, cakes and pies. These were considered especially suitable for social occasions and for offering to guests. The foods in this category tended to be rated towards the inappropriate end of the scale for questions relating to healthiness, e.g. inappropriate when 'needing to lose weight' or when 'not feeling well'. Healthy, wholesome foods seemed to be considered more suitable for everyday eating and eating alone than for parties and entertaining.
- *Speciality meal items* were considered suitable only for special occasions and circumstances. The authors offer liver and chilli as examples. The foods in this category were notable for the number of food-use situations for which they were rated as never appropriate.

- *Common meal items* were considered suitable for all occasions and all ages and would be served at main meals, e.g. meats, fish and some vegetable items. They were generally rated as inappropriate 'for breakfast' and, not surprisingly, 'for dessert'.
- *Refreshing healthy foods* such as milk, orange juice and cottage cheese were considered to be nutritious but were not viewed as suitable for a main course. These scored highly on the healthiness questions 'nutritious' and 'easy to digest', but were rated low as to spiciness. Perhaps spiciness/flavour and healthiness were not seen as compatible.
- *Inexpensive filling foods* were considered cheap and filling as well as fattening, e.g. bread, peanut butter, potato chips (crisps) and candy bars. These were not considered appropriate for those trying to lose weight, but were seen to be useful to assuage hunger between meals and appropriate for hungry teenagers.

This is, of course, just one group of investigators' interpretation of the comments expressed by one small group of consumers more than 25 years ago. It does, however, highlight how even Western consumers, despite not having a formal cultural food classification system like the Chinese hot and cold system, do nonetheless have clear and, within a group, fairly consistent views on the appropriateness of different foods in different situations. They may use quite elaborate classification systems for food, even though such classification may be informal or even subconscious.

If it is to be effective, then clearly any dietary advice or any prescribed diet must recognize such views on the appropriateness of particular foods for particular occasions and situations. It will also be more likely to succeed if it uses foods that are classified as appropriate for the individual. Many cultural

- Consumers classify foods according to the ways in which they are used.
- The hot and cold system is an example of a formal classification system in which foods are classified as hot or cold and selected to maintain or restore the body's hot-cold balance.
- Even where there is no overt classification, all consumers divide potentially edible material into food and non-food. They also regard different foods as more or less appropriate for particular uses, occasions or people. These classifications are culture specific.
- Dietary advisers must be aware of these cultural classifications and make sure that their recommendations are consistent with them.

groups might classify milk as a food for babies and therefore not suitable for adults; some Western consumers might only grudgingly concede that some pulses or salads are suitable for working men.

Anthropological classification of foods

There have been several attempts to devise food categorization systems that could be used across cultures. These are useful not only for the anthropologist seeking to describe the diets and the uses made of foods in particular cultures, but they could also be of great use to the nutrition educator seeking to identify the most appropriate and effective ways of trying to bring about nutritional improvement.

One of the earliest and simplest of these systems is that of Passim and Bennett (1943), who divided foods up into the three categories:

- *Core foods* are those that are regularly and universally consumed within the society. In developing countries, these are likely to be starchy staple foods (e.g. bread, rice, millet or cassava). In industrialized countries, such as Britain and the USA, milk, potatoes, bread and meats would probably fit into this category.
- *Secondary foods* are those that have widespread but not universal use. Most fruits and vegetables would probably be classified as secondary foods in the UK.
- *Peripheral foods* are the least widely and frequently used foods. It is in this category that the most individual variation would be expected. Most shellfish and most species of fish would probably be in this category in the UK.

Such categorization would almost certainly be easier to make in societies whose range of available foods is relatively restricted, and most difficult in countries, like the USA and UK, which have a vast array of foods available to an affluent population. Affluent consumers are able to express much more individuality in their diet structure.

Any particular food may be classified differently for different cultures, it may be classified differently for different social classes within a culture, and foods may change categories over time. Rice is clearly a core food for the Japanese, but for most British groups would probably be classified as secondary. A few decades ago, rice (except pudding rice) would

probably have been classified as peripheral for most social groups in Britain. Prior to 1960, many working-class Britons would almost never have eaten a savoury rice dish. Chinese and Indian restaurants and take-away outlets and foreign holidays have transformed this situation so that there may now be some groups, even within the indigenous population, for whom rice might be approaching the status of a core food. The growth of vegetarianism in Britain might lead some readers to question whether meat should still be regarded as a core food.

A nutritionist trying to effect change in the diet of any community might expect to find most resistance to change in the core foods, more ready acceptance in the secondary foods, and most flexibility in the peripheral foods.

Some foods have acquired a cultural status beyond their purely dietary and nutritional significance; they play an integral part in the cultural life of the community and they have been termed **cultural superfoods**, e.g. rice in Japan and bread in many European cultures.

Rice has maintained a particular emotional and cultural status for the Japanese despite a marked fall in consumption since World War II and a corresponding increase in bread consumption:

- the emperor still cultivates a symbolic crop of rice
- in the Japanese language, the same word can mean either food or rice
- rice plays a part in Japanese cultural and religious rituals
- in the past, the Japanese calendar was geared to the cycle of rice production, rice was used as a medium for taxation, and some units of measurement were based upon the amount of rice necessary to feed a man for a year

Bread has declined in its cultural significance in some European countries, but television scenes of people in war-ravaged areas of Eastern Europe queuing for bread still strike a particularly poignant chord. If the bread supply is threatened, the situation is perceived as that much more desperate than if other foods are in short supply. The arrival of bread is often seen as a symbol of hope for these people. There are numerous reminders of the past importance of bread in the UK, for example:

- bread and dough are both used as slang terms for money

- in Christian communion, bread is used symbolically to represent the body of Christ
- in the Lord's Prayer, Christians pray for their daily bread.

Nutrition educators need to understand and respect the cultural significance of such foods if they want their dietary advice and guidance to be effective. Insensitive denigration of a cultural superfood may provoke incredulity or even hostility and so reduce the chances of any of the advice being taken seriously. Meat has had a particular cultural importance to many Western men, which may even have been reinforced in the past by nutrition education that presented meat as a good source of the high-quality protein that was seen as a key requirement of good nutrition. If nutrition educators now try to persuade such men that red meat and a healthy diet are incompatible, they may simply persuade them that the price for a healthy diet is too high and make them ignore all dietary advice. In a previous edition of a classic British nutrition text, Passmore and Eastwood (1986) quoted a passage that describes the African's craving for meat as:

the strongest and most fraternal bond that the continent had in common. It was a dream, a longing, an aspiration that never ceased, a physiological cry of the body, stronger and more torturing than the sexual instinct.

Gary, R. (1958) *The roots of heaven*. London:
Michael Joseph.

The speaker goes on to assume that this need for meat is a universal craving for all men. If there is even a grain of truth in this assertion, then nutrition education must not suggest that the only option for healthy eating is one that would require unreasonable sacrifice. To use a more recent example, chips (French fries) are a traditional and important element in British cuisine and culture. The following quotation from a serious national newspaper emphasizes the importance that some Britons attach to their chips:

According to the Chip Census, one person in 10 would rather give up sex, alcohol and smoking than chips!

The Independent, 2/3/94.

Once again, if healthy eating appears to require total avoidance of chips, then the chances of acceptance by a high proportion of the British population are probably doomed. Moderate amounts of chips can be included in a diet that meets current nutritional guidelines, especially if attention is paid to advice about their method of preparation and the foods that they most appropriately complement.

Cultural superfood is just one of five food categories described by Jelliffe (1967) in a food classification system that he thought had universal application, both in developing and industrialized countries. The five categories in this system are:

- Anthropologists classify foods according to their importance to a culture and the roles they play within the culture's diet.
- The simplest anthropological classification divides foods into three categories: core foods, secondary foods and peripheral foods.
- The same foods may be used differently and be in different categories in different cultural groups and a food may change categories within a group over time.
- A cultural superfood is one that plays an important part in the cultural life and identity of the group and has acquired a cultural significance that transcends its purely nutritional one, e.g. rice in Japan, maize amongst the Hopi Indians, and bread in some European societies.
- Jelliffe classified foods into five categories:
 1. cultural superfoods – usually the starchy staple of the group
 2. prestige foods – reserved for special occasions or important people
 3. body image foods – foods thought to contribute to good health
 4. sympathetic magic foods – foods that impart some special property to the consumer
 5. physiologic group foods – foods considered suitable for particular groups.

- *Cultural superfood*, as discussed above: considered by Schutz *et al.* (1975) to correspond to 'common meal item' in their consumer classification.
- *Prestige foods* are reserved for important occasions or important people. According to Jelliffe, foods within this category are usually high in protein, which is often animal protein. They are also usually difficult to obtain because of their scarcity, high cost or difficult preparation. Truffles, lobster and caviar might fall into this category in the UK. In past decades, salmon would have been in this category, until the ready availability of cheaper, farmed salmon reduced its scarcity and price.
- *Body image foods* are those that are thought to promote wellness. Jelliffe lists food that contribute to maintaining or restoring hot-cold balance as an example from developing countries. High-fibre foods would be an example in the UK.
- *Sympathetic magic foods* are thought to have some special properties that they impart to the consumer. An example given by Jelliffe is the avoidance of eggs by many women in East African countries because they believe that they cause infertility. As another example he uses underdone steak used in the training of athletes (in 1967) because it symbolized vigour, masculinity and energy.
- *Physiologic group foods* are foods reserved for or forbidden to certain physiologic groups. Taboos against certain foods for pregnant and lactating women or for young children are examples. Breast-milk, infant formula and certain cereal preparations that are normally only consumed by infants and young children are obvious examples of this category.

NON-NUTRITIONAL USES OF FOOD

The primary biological role of food is to provide an adequate supply of energy and essential nutrients, but food also has many non-nutritional functions and several examples are listed below. This list is not definitive; it is merely intended to illustrate the enormous diversity and range of potential non-nutritional uses of food.

Religion, morality and ethics

- Food and drink are used in the rituals and ceremonies of many religions, e.g. the use of bread

- and wine in Christian communion to symbolize the body and blood of Christ.
- Adherence to the dietary rules of a religion acts as a common bond for adherents and serves to differentiate them from non-believers. For example, adherence to the dietary rules and customs of the Torah helps to bond orthodox Jewish families and to set them apart from gentiles.
 - Adhering to dietary rules is an outward symbol of a person's piety and self-discipline, e.g. the dawn-to-dusk fasting undertaken by Moslems during Ramadan.
 - Policing of dietary rules provides a mechanism for religious leaders to exert control over their followers and to demonstrate their authority, e.g. the strict penalties for alcohol consumption, even for non-believers, underlines the primacy of Islam and the mullahs in some countries.
 - People may avoid certain foods to demonstrate their disapproval of animal cruelty or their ecological awareness, e.g. the avoidance of veal raised in crates or tuna caught in drift nets.
 - People may boycott food from particular countries or companies to express their moral disapproval of human rights abuses, exploitation of workers or other perceived moral transgressions, e.g. boycotts of South African produce during the apartheid era, boycotts of food companies that are accused of 'pushing' powdered infant formulae in Third World countries.

Status and wealth

- Expensive and exotic foods can be used to demonstrate one's wealth and sophistication, e.g. serving expensive, but essentially rather bland, caviar as a snack with champagne.
- Serving elaborate and expensive meals can be used to demonstrate esteem for guests, as in traditional feasts of welcome and the Western dinner party.
- People's social status may be defined by who they eat with and where they eat, e.g. some companies have different dining areas for workers of different status, and different Hindu castes have not traditionally eaten together.
- Unusual food choices may be used to express a person's individuality, e.g. serving foods associated with one's distant ethnic roots.

Interpersonal relationships

- Offerings of food and drink are commonly used to initiate and maintain personal and business relationships. Offering new neighbours a drink or food is primarily a gesture of welcome and warmth rather than an attempt to satisfy any perceived nutritional need.
- Giving of food can be a demonstration of love, and withdrawal or failure to offer food can be used to signal disapproval or even to punish, e.g. the use of bread and water punishment for prisoners. A gift of chocolate can be used as a reward, a gesture of affection or apology; any nutritional impact of the chocolate is incidental.
- Food and drink may provide a focus for social gatherings.

Political

Control of the food supply and the price of food can be a very potent method of exerting political control or of gaining political favour. Some would argue that food aid has purposes or effects beyond its obvious humanitarian one. It may save the consciences of wealthy donor nations, foster dependence and subservience in the recipients or help to bolster the political and diplomatic fortunes of the donor government.

Folk medicine

Diet is an element in many traditional treatments for disease. In the traditional hot and cold system, cold foods may be used to treat hot diseases.

THE HIERARCHY OF HUMAN NEEDS

The relative importance of any particular influence on diet may be quite different in different groups, even within the same local environment. Take, for example, the influence of religion on food choice in the UK or USA. Religion would have a major influence upon strictly orthodox Jews or Seventh Day Adventists, but minimal influence upon equally devout Anglicans or Episcopalians.

The orthodox Jew would, for example:

- avoid all pig meat and meat from any animal that does not have cloven hooves and chew the cud
- not eat the hind quarters of animals or unbled meat
- not eat flightless birds, shellfish or other fish without fins and scales
- not eat meat and dairy produce together
- not prepare food on the Sabbath day.

It would be difficult to identify any specifically religious influences on the diets of the two protestant groups except, perhaps, the occasional communion bread and wine or voluntary and self-selected abstinence during Lent.

The priority attached to the various biological and non-biological factors influencing food selection is likely to change as one moves from a situation of gross deprivation or scarcity to one of almost limitless abundance. Maslow (1943) suggested that human needs could be arranged into a hierarchy of motivation, and this hierarchy is summarized in Figure 2.1. In this very well-known theory of motivation, needs that are lower down the hierarchy must be at least partially satisfied before the higher-up needs can become significant motivating factors. In Maslow's theory, the basic biological and emotional needs come at the base of the hierarchy and only

- Food has many functions over and above its purely nutritional purpose.
- Dietary rules and traditions can be an important part of religious life and cultural identity.
- Food selection can be used to make ethical and political statements.
- Eating habits and food choices can be used to indicate a person's social status or wealth.
- Food can be used to initiate and foster interpersonal and business relationships.
- Control of the food supply can be used as a means of political control.
- Food can be used to reward or to punish.
- Foods can be used as medicines in traditional medicine.

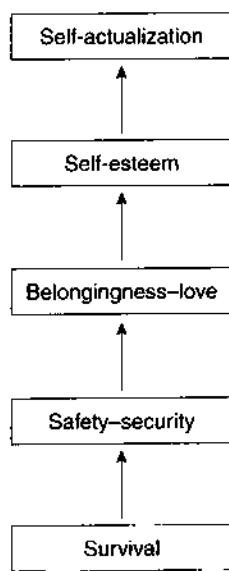


Figure 2.1 Maslow's hierarchy of human needs. After Maslow (1943).

once there has been reasonable satisfaction of these needs do the more aesthetic and esoteric needs become significantly motivating.

In conditions of extreme deprivation, survival is the priority and people may resort to eating almost anything that is remotely edible, perhaps even breaking one of the strongest and most widespread of all dietary taboos, that against the consumption of human flesh. After survival, the need to ensure future security and safety becomes motivating. Potentially edible material is classified into food and non-food and ensuring that future needs are met becomes a priority. Food hoarding, perhaps even obesity, might be

outward manifestations of this desire for security. Obesity in some cultures has been sought after and admired rather than dreaded and despised, as it usually is in the UK and USA. Once security of food supply is relatively assured, the need for love and belongingness becomes a motivating influence on food selection. This could be manifested in the extensive use of food to demonstrate group membership and affection. Then comes the need for self-esteem. This might be manifested in the selection of high-cost, prestige foods to demonstrate one's wealth and success. At the pinnacle of Maslow's hierarchy, the need for self-actualization becomes motivating. Selection of food to demonstrate one's individuality or uniqueness becomes prominent, and this may be manifested in experimentation with new foods, new recipes and non-conforming patterns of selection.

The current penchant for the simple peasant foods of the East by some educated, affluent 'Westerners' may be one manifestation of the need for self-actualization. The partial replacement of these traditional foods with a more 'Western diet' by some in the East may be symbolizing their need to demonstrate their new wealth, i.e. to improve their self-esteem. In many Third World countries, bottle feeds for infants are promoted as the modern and sophisticated Western alternative to 'primitive' breastfeeding. In many affluent countries, like the US and UK, breastfeeding predominates in the upper social classes (e.g. Foster *et al.*, 1997).

There may be some disagreement over the details of Maslow's hierarchy as it has been applied to food selection, but the underlying principle seems unquestionable, i.e. that the priority attached to different influences will change with increasing affluence and

- The importance of any influence upon food selection and eating behaviour will vary from group to group.
- Religion would be a major influence upon the choices of orthodox Jews but not upon Protestants.
- In situations of shortage or deprivation, survival and security would be dominant influences upon food choices.
- In situations of abundance, prestige and aesthetic factors would come into play.
- Maslow devised a theory of human motivation that envisaged a hierarchy of human needs (see Figure 2.1).
- In Maslow's scheme, there has to be some minimum satisfaction of a need lower down in the hierarchy before the next need becomes a significant motivating influence.

food availability. The range of factors influencing selection will tend to be greater in affluent than in poor populations, and the relative prominence of these different factors will also change with increasing affluence. The more physiological, biological drives tend to become less important in the affluent.

A MODEL OF FOOD SELECTION

Several workers have attempted to organize and integrate the various and disparate factors influencing food selection into a unified model. Without such a model, any discussion of the vast array of factors that can influence food choices can end being an amorphous jumble of ideas and facts. Such a model should be simple but comprehensive and provide a useful framework for consideration of the factors that affect food choices. Availability of a food is clearly a prerequisite for selection. Wheeler (1992) suggested that various constraints act to limit the range of foods that are, in practice, available to the individual. In effect, the image of the affluent Western consumer having almost limitless theoretical choice of food is partly an illusion. Practical limits upon the range of foods that are really available to different individuals may greatly limit the scope of the nutrition educator for effecting change. Even if an individual understands and accepts dietary advice, there may still be barriers, sometimes beyond the individual's control, that prevent or limit the implementation of change.

Figure 2.2 shows a simple model for food selection. This model uses the concept that many different types of factors will limit the practical availability of foods to an individual and thus the chances of different foods being eaten. As with Maslow's hierarchy of human needs, it is envisaged that some minimum availability at lower levels in the hierarchy must be achieved before the next level becomes a significant influence upon selection. These influences at each level can be absolute, i.e. complete unavailability to sole availability. More often in Western countries, there will be more subtle variations in the practical availability of different foods.

- **Physical availability:** a food can only be eaten if it is physically present in the local environment.
- **Economic availability:** people can only eat the foods that they can afford to buy. If a food is avail-

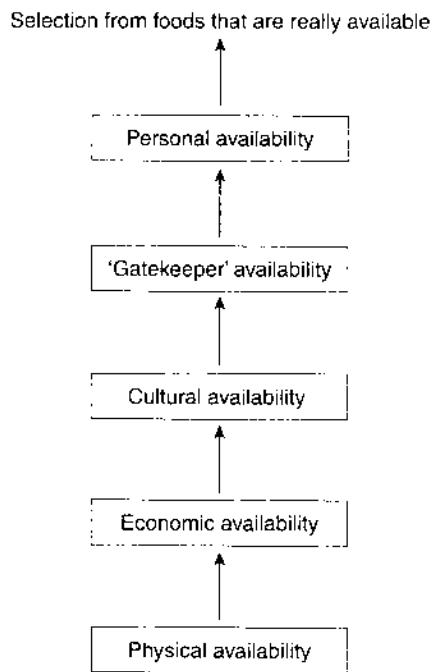


Figure 2.2 A model of food selection based upon the concept of a Maslow-type hierarchy of constraints upon food selection.

able locally but they cannot afford to buy it, then it is economically unavailable to them.

- **Cultural availability:** people will normally only eat things that they recognize as food and which are culturally acceptable to them. Foods that are present and affordable but culturally unacceptable to them are culturally unavailable.
- **'Gatekeeper' availability:** a gatekeeper is someone who controls the supply of food to others. A mother may regulate what her family eats and a catering manager may regulate what the residents of an institution eat. What the gatekeeper provides may limit the supply of particular foods, even though they may be physically, economically and culturally available to the consumer.
- **Personal availability:** even foods that have overcome all of the previous barriers to their availability may still not be consumed. Individual dislike of a food or avoidance for reasons of individual belief or because of physiological intolerance may make it personally unavailable. Someone who is (or who believes that he or she is) allergic to fish will be effectively prevented from consuming fish.

Someone revolted by the taste or texture of broccoli will not eat it, and meat will not be available to someone revolted by the idea of eating the flesh of dead animals. Someone who relies upon poorly fitting dentures may be discouraged from eating foods that require a strong bite and lots of chewing.

These availabilities are not usually absolute 'all-or-nothing' phenomena, but a continuum, varying from absolute unavailability to high desirability and availability. There can be both positive and negative influences upon the availability of foods at each level of the hierarchy, for example:

- the number and locations of retail outlets stocking a food would influence its physical availability
- subsidy would increase economic availability of a food, but a tax would reduce it
- clever advertising that raised the prestige of a food might increase its cultural availability
- increased menu choice in an institution might lessen the impact of the gatekeeper
- an 'acquired taste' might increase the personal availability of a food.

Much current nutrition education is geared towards partly reversing some of the dietary changes that accompany affluence. In many populations, lack of physical or economic availability limits the consumption of sugars, fats and animal foods, but affluence removes these constraints. These foods often have high prestige and palatability, which increase

their cultural, gatekeeper and personal availability. Nutrition education in affluent countries may be seen as trying to introduce constraints higher up the hierarchy in Figure 2.2 to compensate for the lessening of economic and physical restraints at the base of the hierarchy.

PHYSICAL AVAILABILITY

There will be absolute limitations on the range of potentially edible materials that are available in any particular local environment. Physical and geographical factors are key determinants of food availability – factors such as climate, soil type, storage facilities, water supply, and quality of transportation links. The range and amounts of food available in an isolated and arid region may be severely limited by such factors. During severe famines, physical availability may be an absolute limitation on what food most people can eat.

Even within affluent countries, some foods will only be available seasonally. In fact, many seasonal foods are now imported and available all year round, although the price will often vary considerably in and out of the local season. Some foods may be unavailable locally because of lack of sufficient demand. Shops will only stock foods that they expect to sell, e.g. kosher foods may be difficult to obtain in areas where few Jews live. Small local stores will

- A model of food selection should provide a framework for organizing and discussing the many disparate influences upon food selection and eating.
- The hierarchy of availability model is based upon Maslow's hierarchy of human needs. It envisages a hierarchy of factors that will limit the practical availability of foods to any individual.
- As with Maslow's hierarchy, there must be minimum availability at a lower level in the hierarchy before influences at the next level start to come into play.
- A continuum of availability is envisaged from absolute unavailability to limitless abundance and easy access.
- The five levels of this hierarchy of availability are:
 1. physical (can the food be obtained?)
 2. economic (can it be afforded?)
 3. cultural (is it culturally recognized as food and acceptable?)
 4. 'gatekeeper' (is it provided by the parent, caterer or other gatekeeper?)
 5. personal (does the person like the food, can they eat it and is it personally acceptable to them?).

inevitably stock a more limited range than large supermarkets and there may even be quite marked differences in the range of products stocked by shops in poor and affluent areas.

According to Henson (1992), the number of retail food outlets in Britain fell by a third between 1976 and 1987. There is a marked trend towards fewer but larger shops. The great majority of new superstores that have opened in recent years in Britain are located on the outskirts of towns and relatively few are being opened in shopping centres and high streets. These trends inevitably mean that people must, on average, travel further for their shopping. Around three-quarters of households in Britain now use a car to make their shopping trips. These new superstores have wider choice and lower prices than local shops. To take full advantage of them, consumers need good mobility (preferably access to a car), the financial resources to allow them to buy in larger quantities, and adequate storage facilities (including access to a freezer). These retailing trends may be to the disadvantage of some groups and severely affect their access to a varied and inexpensive diet, for example:

- People who do not have access to a car, especially if they live in rural areas where local shops have all but disappeared. In some rural areas, it may be difficult to obtain a reasonable variety of fresh fruits and vegetables.
- Elderly and disabled people whose mobility is limited.
- Those living in accommodation without a freezer, little or no cooking facilities and no space for storing food.

ECONOMIC AVAILABILITY

Availability of money exerts a variable limiting influence on the availability of foods. In some circumstances, lack of money may have an absolute effect in limiting the availability of particular foods or even of food in general. This is very clear in some Third World countries where having enough money may mean the difference between relative comfort and starvation. Except in times of acute famine, it is poverty rather than inadequate physical availability of food that is the dominant cause of malnutrition. People in wealthy countries with inadequate indigenous agriculture do not go hungry and wealthy people within poor countries do not usually go hungry either. People may go hungry even where food supplies are comparatively plentiful if they do not have the money to pay for food.

Even in affluent countries, financial constraints can have a marked effect upon dietary freedom of choice. In the UK and USA, there are quite marked differences between the food choices of the richest and poorest and this is partly due to differences in purchasing power. In these countries, health and longevity are strongly and positively correlated with higher socioeconomic status, and diet is probably one of the factors contributing to these inequalities in health. It is ironic that the chronic diseases of industrialization that are internationally strongly and positively correlated with increasing affluence are, within the industrialized countries, associated with relative poverty.

- People can only eat food that is physically accessible to them.
 - During famines, the physical absence of food may be an absolute limitation on food choice.
 - A range of physical and geographical factors can affect what foods are available at any time and place. However, many seasonal foods are, now available all year round, at a price.
 - The trend in Britain is for fewer, larger food shops, often located on the edge of town. People must therefore travel further to do their shopping.
 - These large superstores offer great variety at relatively low prices, but those with limited mobility or no car may not be able to take full advantage of them.
 - People who live in rural areas without a car may have limited access to some foods, especially perishable foods such as fresh fruit and vegetables.

- Poverty is the major cause of malnutrition in the world. People usually go hungry because they do not have the money to buy food.
- In affluent countries there are marked differences between the diets of the richest and the poorest and there are also major inequalities in health.

International trends

If one looks world wide at the influence of economic status upon food 'selection', some very general trends emerge. Increasing affluence leads initially to increases in total energy intake, followed by a switch away from starchy staples to other higher prestige and more palatable foods, including fats, meats, dairy produce and sugars. Poleman (1975) suggested that the percentage of dietary energy that is provided by starchy staples provided a very simple economic ranking of diets.

In 1875 starchy staples made up 55% of total energy intake in the USA. A century later, this figure had dropped to around 20%. In the early decades of the twentieth century, maize was a sufficiently dominant component of the diets of poor farmers in some southern states of the USA to trigger epidemics of pellagra (see Chapter 12).

At very low incomes, increased wealth leads to increases in energy intake, usually based upon local starchy staples. Only when income level has risen beyond the point at which energy needs are satisfied does increasing affluence lead to changes in the

nature and quality of the diet. The local starchy staple is partly replaced by higher prestige rice or wheat. This is followed by a reduced consumption of starchy staples and replacement with foods of animal origin, fattier foods and simple sugars. In some Third World countries, fat accounts for less than 10% of total energy intake and carbohydrate for more than 75%. In the UK, fat account for around 40% of food energy and carbohydrate for only 45%. In the poorer countries almost all of the carbohydrate calories are likely to be starches, whereas in richer countries close to half of the carbohydrate calories may be in the form of simple sugars. Improved palatability, higher prestige and simple increases in physical availability all compound to persuade those with the economic means to make changes in this general direction.

Paralleling these changes in food selection practices there will be changes in the general pattern of mortality and disease. Child mortality will fall, due largely to a general reduction in mortality from infectious diseases. Malnutrition and the deficiency diseases will all but disappear, life expectancy will increase, but the chronic degenerative diseases will account for an

- As populations become more affluent, their diet changes. First, there is an increase in energy consumption as starchy staple, but then the starchy food is partly replaced by animal produce, fatty foods and sugars.
- The proportion of dietary energy derived from starchy foods decreases with affluence and provides a simple economic ranking of diets.
- In some Third World countries, fats account for less than 10% of dietary energy and starches for as much as 75%.
- In some affluent countries, such as Britain, fats provide 40% of the energy and starches not much more than a quarter.
- Life expectancy increases with affluence, child mortality falls, deficiency diseases largely disappear and deaths from acute causes such as infection fall.
- The chronic diseases of industrialization account for most deaths amongst affluent populations.

increasing proportion of deaths and will be an increasing cause of illness and disability. The nutritional guidelines aimed at reducing the prevalence of these chronic diseases (discussed in Chapter 4) would partly reverse these wealth-related changes in dietary habits.

The problem of feeding the world

In 1798, Thomas Malthus published his now infamous *Essay on the principle of population*. In this treatise he predicted that, unless steps were taken to regulate the family size of the poor, population growth would outstrip growth in the food supply, leading to global famine and epidemics. In the early 1960s, when world population was growing at an annual rate of around 2%, a spate of Malthusian-type predictions of impending population disaster were also published.

Two hundred years ago, when Malthus published his essay, the world population was less than 1 billion and in 1999 it reached 6 billion (Table 2.2). This rate of growth would have astounded even the most ardent Malthusian. There has, however, been a distinct slowing in the rate of population growth since the 1960s. In 1965–70, growth rate peaked at 2% per year, but by 1998 it was down to 1.3% (still a net addition of 78 million people each year). The deceleration in the rate of population growth has been sharper than was predicted even a few years ago. Table 2.2 shows the United Nation's (UN) population estimates and projections for the period 0–2050 made in 1998. In brackets in this table are the projections made in this biennial UN publication in 1994. The projected population in 2050 has fallen from over 10 billion, predicted in 1994, to the 8.91 billion predicted in 1998. Long-range predictions are that the world population will stabilize some time during the twenty-second century at something over 11 billion.

Despite this massive increase in the world's population, calculations of the worldwide production of primary foods (mainly cereals) indicate that sufficient is produced to meet the estimated nutritional needs of the world population. Yet in some poor countries there is starvation and in other richer countries there are 'mountains' and 'lakes' of surplus food. Poleman (1975) estimated that, in the USA, *per capita* grain consumption was 1800 lb per year, but less than 100 lb of this was directly consumed, with almost all of the rest being used as animal feed. An animal consumes between 3 and 10 lb of grain to produce 1 lb of meat.

Table 2.2 The growth in world population from year 0 to 2050

Year	Population (billions)
0	0.30
1000	0.31
1250	0.40
1500	0.50
1750	0.79
1800	0.98
1850	1.26
1900	1.65
1920	1.86
1940	2.30
1950	2.52
1960	3.02
1970	3.70
1980	4.44
1990	5.27
1998	5.90
2000	6.06 (6.23)
2010	6.79
2020	7.50
2030	8.11
2040	8.58
2050	8.91 (10.02)

Data are from UN (1998); biennial estimates. The figures in brackets are predictions made by the UN in 1994.

The USA is the world's largest producer of surplus grain. In 1995, *per capita* grain consumption was 250 kg per year in the developing world, but 665 kg per year in the industrialized nations. The primary cause of worldwide malnutrition is not inadequate food production *per se*, but an imbalance in the way this food is distributed between the peoples of the world. Differences in wealth are the key determinant of this uneven distribution.

The image of rapid world population growth outstripping growth in food production and being entirely responsible for creating current food shortages is not borne out by the statistics. Wortman (1976) estimated that, in developed countries, rates of food production between 1961 and 1975 rose by almost 40% on a *per capita* basis. Even in developing countries, population growth did not entirely negate the substantial increases in absolute food production when they were expressed on a *per capita* basis. Sanderson (1975) suggested that, even though population growth had exceeded even the most pessimistic forecasts, worldwide grain production had increased

by enough to theoretically allow a 1% annual improvement in worldwide *per capita* consumption. Wortman (1976) predicted that world food production was likely to keep up with population growth for some time to come. A report in 1995 by the International Food Policy Research Institute (see IFPRI, 1996) confirmed that this trend of increasing *per capita* grain production also continued over the period 1970–1995. The world population grew by around 2 billion (55%) over this period, but was offset by a 64% increase in world cereal production. In developing countries, *per capita* grain production increased by 15%, and by 10% in industrialized countries. The *per capita* availability of food calories was also higher in the early 1990s than in the early 1970s in all of the major developing regions of the world. Over the 20-year period from 1970 to 1990, the number of chronically undernourished people in the world fell by 17%, but still amounted to 780 million people. This global improvement masks some regional disparities: the number of undernourished people rose slightly in South Asia and almost doubled in sub-Saharan Africa, where over 20% of the world's malnourished people now live. The absolute number of pre-school children who were malnourished actually

rose over this 20-year period (to 184 million), although the proportion who were malnourished fell because of the increase in the population. Any encouraging aspects of these figures should not mask the massive scale of malnutrition and food insecurity that still afflicts the developing world (see below).

- In 1995, nearly 200 million of the world's children were moderately to severely underweight and 70 million severely malnourished.
- 760 million people in developing countries are chronically undernourished.
- At least a million children die each year as a result of vitamin A deficiency and another half a million go blind.
- Billions of people across the world are threatened by iron or iodine deficiency.

How has the massive growth in cereal production over the period 1970–1995 been achieved? Most of the increase has been the result of increased yields rather than increases in the land area under cultivation. In Asia, cereal yields almost doubled and those in Latin America increased by around two-thirds. The three major factors responsible for this 'Green Revolution' are (IFPRI, 1996):

- Food production is sufficient to satisfy the nutritional needs of the world population, but this food is unevenly distributed.
- Affluent countries have almost three times the *per capita* grain consumption of developing countries. This extra grain is fed to animals to produce meat and milk.
- The rate of world population growth has fallen sharply since its peak in the 1960s and the Malthusian predictions of rapid population growth leading to worldwide famines and epidemics now look too pessimistic.
- The world population almost doubled over the period 1960–1995, but *per capita* grain production and energy availability increased, largely because of increasing grain yields.
- The number of chronically malnourished people fell during this period, but malnutrition in children and adults, including vitamin and mineral deficiencies, remains a huge problem in much of the developing world.
- The massive increases in grain yields seen during the 'Green Revolution' have been largely achieved by the increased use of modern, high-yielding varieties, greater irrigation of farmland and greater use of agricultural chemicals.
- Africa has been relatively slow to adopt these changes and the 'Green Revolution' has largely bypassed Africa, where grain production stagnated in the period 1970–1995.
- Rates of malnutrition rose sharply in sub-Saharan Africa over this period.
- The increase in food availability has led to destruction of large areas of forest, salinization of land due to overirrigation, overuse of agricultural chemicals and depletion of the world's fish stocks.

1. increased use of modern high-yielding crop varieties
2. increased irrigation of land
3. increased use of fertilizers and pesticides: the use of fertilizers by farmers in developing countries quadrupled over this period.

In contrast to the other developing regions of the world, cereal production in Africa remained stagnant over this period, largely for the reasons listed below.

- African farmers have lagged behind in the use of new crop varieties
- less arable land is irrigated in Africa: only 8% of African farmland is irrigated, compared to 37% in Asia
- African farmers make less use of agricultural chemicals: Asian farmers use seven times more fertilizer than African farmers.

Can the rate of increase in production seen in 1970–1995 be maintained in future decades? The increased production of food to feed the growing world population has had a number of adverse environmental consequences. This has heightened concern about whether the increases in food production are sustainable using current approaches. Discussion of these issues is beyond the scope of this book, but some of the important issues are listed below (for further information and discussion see IFPRI, 1996; FAO 1996; WRI, 1997).

- Destruction of forests to increase the availability of arable land, which can increase soil erosion, affect global climate and deprive forest-dwelling people of their source of food and way of life.

- Adverse ecological effects of the excessive use of agricultural chemicals.
- Overuse of irrigation, which can lower the water table, lead to salinization of land and make it agriculturally unproductive.
- Depletion of the world's fish stocks.

Effects of income upon food selection in industrialized countries

Even within industrialized countries, the food choices of the wealthiest and the poorest may vary quite considerably. Table 2.3 shows a comparison of the spending on food and the food-related spending of families with different economic means in Britain, derived from data in the UK National Food Survey (MAFF, 1997).

- Groups A and D are the highest and lowest income groups, respectively, of those in employment.
- Group E₂ is the lowest income group, where there is no earner (long-term unemployed).

Table 2.3 indicates that in 1996 most households in Britain owned a deep freeze and a microwave oven, but nearly 20% of the poor unemployed did not own a freezer and 30–40% of those in the two poorer households did not own a microwave oven. This will have implications for the economic purchasing and preparation of food, as discussed later in this section. The richer group is able to eat out more than the poorer groups: group A takes almost twice as many meals out as the two poorer groups and they spend almost three times as much on eating out.

Table 2.3 Food spending differences between rich and poor families in the UK

	A	D	E ₂
Owning a freezer (%)	96	88	83
Owning a microwave (%)	86	71	61
Number of meals eaten out	4.7	2.5	2.3
Spending on outside meals (£)	13.71	4.96	3.95
Household food expenditure (£)	17.62	12.24	12.18
Including drinks, alcohol and sweets	20.68	13.58	13.35
Total spending on food and drink	34.39	18.54	17.30
Calories per penny (household food)	6.9	10.5	10.9
Calories per penny (all food and drink)	4.6	8.2	8.7

Group A head of household earning over £395 (c. \$595) per week.

Group D head of household earning under £150 (c. \$240) per week.

Group E₂ households with no earner and an income of under £150 per week.

Data from MAFF (1997). All values are per person per week, unless otherwise stated.

The richer group spends nearly 45% more on household food than the two poorer groups, but this represents a much smaller proportion of total income in the richer group than in the poorer groups. In general, food spending becomes a smaller proportion of total expenditure as income rises. Groom (1993) suggested that the poorest UK families spent more than a third of their income on food compared, with a national average of around 12%. This large gap between the food spending of rich and poor households widens when household drinks and confectionery are included and increases to almost double when home food and drink and meals eaten out are all included. The poorer groups obtain over 50% more calories for every penny they spend on household food than group A. When all food and drink are included, this differential rises to around 80% or more. The poorer groups are, of necessity, more economically efficient in their food spending. This increased efficiency could be achieved in many ways, such as:

- less purchasing of unnecessary items such as alcohol
- buying cheaper brands
- shopping in the cheapest shops
- reducing waste
- buying lower quality (or lower prestige) items that provide more calories per penny, e.g. cheaper types of meat and fish and cheaper meat and fish products
- taking more advantage of special offers and promotions or simply taking more advantage of seasonal fluctuations in the prices of many foods
- doing more food preparation at home rather than buying relatively more expensive, prepared foods or eating out
- adopting a totally different, cheaper eating pattern, with much larger quantities of starchy foods, inexpensive vegetables and legumes but fewer animal foods – a dietary pattern that could be described as more ‘primitive’, but could also be more in line with current nutritional guidelines.

Some of these mechanisms have already been seen in Table 2.3 and several more are apparent in the data in Table 2.4, which compares some of the purchasing practices of our three selected economic groups. Some of the trends in Table 2.4 are discussed briefly below.

- Total meat purchases were slightly higher in the richer group than in the two poorer ones. However, in the higher income group, a larger

proportion of the total meat purchased was fresh carcass meat and poultry. The poorer groups got considerably more meat for each penny they spent than the richer group, implying that they buy cheaper cuts and more bulkier and lower quality meat products.

- Less of the fish purchased by the poorer groups was fresh fish and more of it was processed fish and fish products. As with meat, the poorer groups got more fish per penny than the richer group.
- The poorer groups bought three times as much sugar as the richer group and significantly more fats and oils.
- The poorer groups bought less fruit and fewer fresh vegetables (excluding potatoes) than the richer group. People in Britain and the USA are being encouraged to eat more of these foods; at least five portions each day have been suggested. However, despite all of their nutritional merits (see Chapter 4), these foods are relatively inefficient providers of energy because of their low energy density (see Chapter 6). They accounted for 14% of household food expenditure in 1996, but provided only 5% of the energy. In contrast, fats, oils, sugar and preserves accounted for only 4% of expenditure, but provided over 17% of the energy.
- The poorer groups bought more bread and more potatoes than the richer group. In general, purchase of such starchy foods tends to be inversely related to income.
- The poorer groups bought less of products that could be considered as non-essentials from a nutritional standpoint. They bought less than a quarter as much mineral water, less than half the volume of alcoholic drinks, and slightly less confectionery and soft drinks.
- In several of the examples in the table, the poorer groups could be seen as being less likely to make the healthier choice within a category, i.e.:
 1. less of the bread bought by the poorer groups is whole grain and much more of it is standard white bread
 2. less of the milk bought by the poorer groups is low fat and much more of it is whole milk
 3. fewer of the soft drinks bought by the poorer group are low-calorie versions.

⁴⁹For reasons discussed in Chapter 3, the National Food Survey is an imperfect measure of nutrient intakes. Gregory *et al.* (1990, 1994) measured nutrient

Table 2.4 A comparison of food purchases in different income groups in the UK

	A	B	C
Fresh carcass meat and poultry (g)	478	477	456
All meat and meat products (g)	887	958	951
All meat (g per penny)	1.98	2.89	2.94
Fresh fish (g)	42	24	35
All fish and fish products (g)	164	139	156
All fish (g per penny)	1.65	2.34	2.41
Sugar (g)	67	173	224
Sugar and preserves (g)	103	204	267
Fats and oils (g)	184	249	238
Green and fresh vegetables (g)	841	622	609
Fruit and fruit juices (g)	1348	795	781
Fresh potatoes (g)	625	871	920
All bread (g)	619	786	804
Bread (% whole grain)	14.2	10.1	9.8
Bread (% standard white)	28.9	53.1	53.6
Total milk and cream (mL)	1967	2123	2192
Low-fat milk (%)	56.0	45.0	38.6
Liquid whole milk (%)	27.0	43.8	52.0
Mineral water (mL)	295	61	60
Alcoholic drinks (mL)	538	228	221
Confectionery (g)	63	48	48
Soft drinks/soda (mL)	979	888	832
Soft drinks (% low calorie)	39	25	25

Group A – head of household earning over £395 (< \$950) per week.

Group B – head of household earning under £150 (< \$240) per week.

Group C – households with no earner and an income of under £150 per week.

Data are derived from MAFF (1997). All values are per person per week, unless otherwise stated.

intakes more directly in a sample of British adults aged 16–64 years. This survey can be used to compare the nutrient intakes of different socioeconomic groups, although the data were collected more than 10 years ago.

- Recorded energy intakes were lower in both sexes in households receiving welfare payments and in unemployed men. Recorded energy intakes in women declined with declining socioeconomic status, although **body mass index** (a measure of fatness) increased with declining social status. Taken at face value, this would suggest greater inactivity in the women of lower social class (see Chapter 8 for further discussion).
- In all social classes and in the unemployed, the average intakes of all vitamins and minerals (except iron) met the dietary standards of adequacy used in the UK at that time (the RDA – see Chapter 3). Even the poorest groups surveyed seem to have

enough money to obtain diets that are adequate in essential nutrients. Note that this survey did not include children or elderly people and would exclude, for example, the homeless and those living in temporary bed and breakfast accommodation.

- Satisfactory average intakes could be masking the presence of individuals with unsatisfactory intakes. The **Lower Reference Nutrient Intake (LRNI)** (see Dietary Reference Values in Glossary) is an estimate of the needs of the lowest 2.5% of the population and intakes below this are often assumed to be unsatisfactory (see Chapter 3). Gregory *et al.* (1990) found that, for men, the 2.5 percentile for all vitamins and minerals except magnesium was above the LRNI. For women, the 2.5 percentiles for riboflavin, vitamin B₆, folate, iron, calcium, magnesium, zinc and iodine were all below the LRNI.
- Unemployed people had lower intakes of many vitamins and minerals. The recorded intakes of

many vitamins and minerals were lower in social classes IV and V (the two lowest) than in social classes I and II (the highest). In women, there was a general tendency for the intakes of most vitamins and minerals to decline with social class.

- Biochemical indices of antioxidant vitamin status were also higher in the upper social classes. There were marked trends for carotene and vitamin C intakes to fall with decreasing social class.
- Intakes of fibre, total sugars (including natural sugars) and polyunsaturated fatty acids tend to be higher in the higher social classes. The P:S ratio (the ratio of polyunsaturated to saturated fatty acids) was higher in the higher social classes.
- Men and women from the higher social classes were more likely to eat fruit, fruit juices, vegetables,

especially salad vegetables, oily fish and shellfish, all dairy products, cakes, chocolate and polyunsaturated margarine.

- A dietary pattern described as 'health conscious' was more likely to be associated with the higher social classes. One pattern they described as 'a traditional meat and vegetable diet', with a tendency to high energy and alcohol consumption, was more likely to be found amongst the manual social classes.

Taking the results of the **National Food Survey** (MAFF, 1997) and the survey of Gregory *et al.* (1990, 1994) together, the diets of all socioeconomic groups within Britain seem to be nutritionally adequate, and this probably applies to most other industrialized countries. However, within any particular income

Data from the National Food Survey indicates the following.

- Poorer families are less likely to own a freezer and much less likely to own a microwave oven.
- Poorer families eat out much less often than richer ones and spend much less when they do eat out.
- Richer families spend more on household food than poorer ones, but this represents a much smaller proportion of their total income.
- Poorer families are much more economically efficient in their food purchasing and obtain more calories per penny, e.g. they tend to buy cheaper types of meat and fish and cheap meat and fish products.
- Poorer families buy more bread, potatoes, sugar and fats than richer families, but less fruit and fresh vegetables.
- Poorer households consume fewer non-essential items, such as mineral water, alcoholic drinks and confectionery.
- Poorer families are less likely to buy the 'healthier' option within a category, e.g. bread made with less-refined flour, low-fat milk and low-calorie soft drinks.
- The poorest families with several children spend less than a third as much per head on food compared to rich adult-only households.

Data from *The Dietary and nutritional survey of British adults* (Gregory *et al.*, 1990, 1995) suggest the following:

- Recorded energy intakes were lower in the unemployed and those receiving welfare payments.
- In women, recorded energy intake decreased with social class, but fatness increased.
- Average intakes of all nutrients (except iron) were adequate for all social groups in Britain. Some individuals, especially some women, had unsatisfactory intakes of vitamins and minerals.
- Intakes of most vitamins and minerals were lower in the lowest social classes, as were biochemical indicators of antioxidant nutrient status.
- The higher social classes were more health conscious in their diet selections and, for example, ate more fruit, vegetables, oily fish and polyunsaturated margarine.
- The higher social classes consumed more fibre, sugars and polyunsaturated fatty acids and had a higher ratio of polyunsaturated to saturated fatty acids.

group, the average weekly *per capita* expenditure on food falls with increasing numbers of children in the household (MAFF, 1997). This is to be expected, given the lower needs of children, the economies of scale possible in larger families and reduced wastage in larger families. The highest weekly average was for 'adults only' households in group A and this was more than treble that in households in groups D and E₂, with four children or more. The expenditure of those in groups D and E₂ with four or more children was also more than 40% less than the average for groups D and E₅. Poor families with several children are a likely high-risk group.

The minimum cost of a healthy diet

Welfare payments in the UK have traditionally been calculated on the basis that enough money should be provided to allow the purchase of the essentials of life, but not enough to allow the purchase of luxuries. The same underlying philosophy governs the welfare systems of most industrialized countries. These welfare payments should be enough to prevent claimants from starving or being deprived of basic shelter and clothing, but not make their situation so comfortable as to discourage them from seeking employment or to encourage others voluntarily to live off the state. A more cynical Marxist interpretation of this principle would be that enough money is provided to prevent social unrest, but not enough to reduce the competition for jobs and thus risk increasing the cost of labour. Pressures on public expenditure may encourage governments to economize by reducing the purchasing power of those who are dependent upon welfare payments.

Calculations of minimum subsistence levels of income will be difficult. There are a host of factors that will produce wide variations in the quality of life that different families can achieve with the same level of income. At the physiological level, people vary in their energy and nutrient requirements. Regional and individual variations in housing, transport and fuel costs may make it difficult to produce a single welfare figure for a nation. Differences in cultural and personal preferences will affect the size of the required food budget. A minimally adequate income may make it difficult to live at maximum economic efficiency. For a variety of reasons, such as those listed below, it is often true that 'the poor pay more'.

- The poorest may be forced to buy in small, expensive quantities; they can only afford to buy enough to meet immediate needs.
- They may be unable to stock up when a seasonal food is cheap or when special offers are made, e.g. because they do not own a freezer or simply lack capital.
- They may not have the transportation or personal mobility to enable them to shop at the most economic places.

Table 2.5 shows the potential magnitude of some of the differing costs of buying in small and large quantities. Webb and Copeman (1996) surveyed the differences in unit costs of the same brands of common grocery items in a large London supermarket. The general trend was for decreasing unit cost with increasing pack size, penalizing those who need to buy in minimum quantities to satisfy immediate needs. There was often a sharp decrease in unit cost between the smallest pack and the next smallest size, to the disadvantage of those living alone or buying enough for one. Special offers tended to be concentrated in the largest packs and many special offers were in the form of discounts for multiple purchases. One would generally expect all household costs (e.g. replacement of capital items, rental charges and fuel costs) to rise on a *per capita* basis as the size of the household decreases, thus increasing pressure on the food budget of poor, small households.

Those people who are forced to shop in smaller, more accessible shops may also have to pay much more. In a survey conducted in the Reading and Oxford areas of England in 1991, the cost of a basket of 21 food items was about a quarter higher in small, independent grocery stores than in the supermarkets owned by the large multiple chains (Henson, 1992). The unit cost of milk bought in half gallons from large supermarkets in the UK is only half that from the traditional doorstep delivery services. Major supermarkets are offering 'economy' white bread at less than a quarter of the cost in many small, independent grocers.

Other demands upon the household budget may greatly restrain food expenditure for some individuals or families. When financial resources are restricted, especially if there is a sudden decline in income (e.g. through loss of employment, retirement or long-term illness), expenditure on food may be sacrificed in order to maintain spending upon other essentials

Table 2.5 Differences in costs of small and large packs of a variety of grocery items obtained from a large London supermarket in early 1995

Item	Small size	Unit price (pence)	Large size	Unit price (pence)	Price difference (%)
Cucumber	Half	70	Whole	49	43
Mushrooms	1/4 lb	155	1 lb	115*	35
Peas (can)	142 g	8.3	300 g	5.4	54
Margarine	250 g	14.8	1 kg	11.9	24
Corned beef	198 g	36.4	340 g	28	30
Rice	500 g	11.8	4 kg	9.3	27
Baked beans	150 g	14	420 g	7.4	89
Chicken	1/4	99	Whole	68*	46
Eggs	6	98	12	89	10
Cooking oil	500 mL	74	1 L	59	25
Weetabix*	12	68	48	46	48
Tea bags	40	40.8	160	33.8	21
Milk	1 pt	28	6 pt	22.2	26
Instant coffee	50 g	212	300 g	179	18
Washing-up liquid	500 mL	126	1 L	99*	26
Sugar	500 g	90	1 kg	62	45
Burgers (frozen)	4	43.7	20	36	15
White bread	400 g	6	800 g	4.4	36
Flour	500 g	48	1.5 kg	27.4	75

* Denotes special offer.

All comparisons are between same brands and prices are pence per unit (variable) and so are directly comparable within items. From Webb and Copeman (1996).

such as housing, loan repayments, clothing or fuel. Food expenditure may be sacrificed to allow the purchase or retention of some prestigious item or activity that helps maintain self-esteem, e.g. Christmas presents for children, a car, a colour television, tobacco or going to the pub. Considerable savings can often be made in the food budget without obvious acute hardship, e.g. by switching to cheaper brands or varieties, by reducing wastage and by using cheaper alternatives such as margarine for butter or lower quality meat and fish products for fresh meat and fish.

Dobson *et al.* (1994) published a survey of the food-buying and dietary habits of low-income families living in the English Midlands. Almost all of these families were in receipt of welfare payments (Income Supplement) and half of them had mothers as lone parents. Money and the cost of foods were seen as the most important factors in deciding what food could be bought. Other factors, such as taste, cultural acceptability and healthy eating, could only be considered in the light of what was economically available. These parents often saw incorporating

advice about healthy eating as impractical or only partly practical in their economic circumstances. These families all reported great difficulties in managing on their incomes. Unexpected demands for money could derail their spending plans and at these times the food budget was often reduced to cope with these unexpected demands. These families needed to adopt various strategies to cope with their restricted food budgets:

- They needed to shop around for 'best buys' and this made shopping a time-consuming activity that they ceased to enjoy.
- Parents often reported going without food themselves to minimize the impact on their children.
- The need to avoid waste of food and money meant that children's preferences were given high priority. They did not have the means to provide foods that reflected the individual tastes of family members; parents often ate what their children would eat irrespective of their own likes and dislikes. They all had to eat the same food at the same time.

The burden of coping with the money available was largely borne by the mothers, who did most of the shopping and food preparation and tried to maintain a culturally usual 'mainstream' diet. This meant that they had to eat cheaper versions of mainstream meals. Parents went to great lengths to make sure that their children were seen to be eating the same things as their peers, even if it meant making sacrifices themselves. They resisted radical changes to their diets.

To subsist on a minimum income is easier for those with some degree of budget management skill. Unfortunately, those with the least of these skills may often be those most likely to be required to exercise them: poor educational attainment and poor socio-economic status are strongly correlated. Strictly theoretical calculations of minimum subsistence levels cannot allow for food choices that appear irrational or even irresponsible using strictly biological criteria. People whose self-esteem has been dented by unemployment, disability or social deprivation may buy foods to boost their spirits and self-esteem rather than out of strict scientific necessity. In *The road to Wigan Pier*, George Orwell gives an example of an adequate but dull and low-prestige diet that in 1937 could be obtained for 20 pence (30 cents). He also gives an example of the real expenditure of an unemployed miner's family at the time. This real family spent only 4 pence each week on green vegetables and sweetened condensed milk, nothing at all on fruit, but 9 pence on sugar (representing 4 kg of sugar), 5 pence on tea and 3 pence on jam. Clearly, white bread spread with margarine and jam and washed down with sweet tea was an important part of this family's daily diet. Orwell describes this mixture as being almost devoid of nutritional value and it certainly was low in several vitamins and minerals.

When you are unemployed, which is to say when you are underfed, harassed, bored and miserable, you don't want to eat dull wholesome food. You want something a little bit tasty... Unemployment is an endless misery that has got to be constantly palliated, and especially with tea, the Englishman's opium. A cup of tea or even an aspirin is much better as a temporary stimulant than a crust of brown bread.

George Orwell (1937) *The road to Wigan Pier*.
London: Victor Gollancz Limited.

Readers who would like a more scientific and detailed review of trends in the British diet between

1841 and 1994 should see Charlton and Quaife (1997). Orwell goes on to describe the poor people of a British industrial city as being small of stature and having universally bad or missing teeth – he suggests that it was rare to see a working-class person with good, natural teeth. He suggests that the death rate and infant mortality rate of the poorest area of any city were always at least double those in the wealthy areas, in some cases much more than double. He was writing about mid-twentieth-century peacetime conditions in one of the wealthiest industrialized countries of the world. These few pages of Orwell's book underline how far the social conditions and nutrition of most poorer people in Britain have improved over the last 50 years, despite the disparity between the health and nutrition of rich and poor that still remains.

One practical consequence of the sort of effect described in the above quotation will be to make it even more difficult to persuade the less affluent to make dietary changes that they perceive as reducing either the palatability or prestige of their diet. It is important to their self-esteem that they, and especially their children, are seen to be eating like their peers (Dobson *et al.*, 1994). Diet-related disease is more prevalent in the poorer groups, but they also seem to be more resistant to nutritional and other health education advice. The decline in the prevalence of breastfeeding with social class (Foster *et al.*, 1997) is a very clear example of this. This increases the onus upon those seeking to effect dietary change amongst all social groups to suggest changes that are simple, economically feasible, easy to incorporate permanently in usual culinary practices, and not overly restrictive of highly palatable and high-prestige foods. Change must not require the consumption of a diet that, although wholesome by current nutritional criteria, is perceived as dull and unpalatable, uses low-prestige ingredients or is simply different from the mainstream.

Changes must also be within any group's economic means if they are really expected to be implemented. Groom (1993) and Leather (1992) review a number of attempts to devise 'healthy' diets that comply with current UK dietary guidelines and are also within the economic means of poorer families. In 1991, the UK Ministry of Agriculture, Fisheries and Food (MAFF) produced a 'low-cost, healthy diet' that could be purchased for £10 (\$16) per person per week at 1991 prices. However, it was suggested by

Leather (1992) that this diet would be so different from the mainstream British diet as to be totally unrealistic. It would require, for example:

- almost total exclusion of meat and most dairy products from the diet
- a very large increase in bread consumption and most of this bread would have to be eaten dry, without butter or margarine
- large increases in consumption of tinned fruit and frozen vegetables.

It is easy to produce lists of ingredients that meet any set compositional criteria within a very small budget, but it is much more expensive to purchase a collection of ingredients that can also be combined into meals that make up a culturally acceptable diet. The survey of Dobson *et al.* (1994), discussed earlier,

found that maintaining a mainstream diet was very important to low-income families. The Family Budget Unit at York University estimated the cost of a 'modest but adequate' diet which would not only be adequate and comply with current nutrition education guidelines, but also be broadly in line with the usual diet of the UK population. Using these criteria, their minimum cost for feeding an adult couple was about 70% more than the minimum-cost diet produced by MAFF in 1991 (see Leather, 1992). This modest but adequate diet cost slightly less than average *per capita* food expenditure, but significantly more than that spent by low-income families. To stay as close as possible to current eating patterns but to choose 'healthier' alternatives such as leaner meat, wholemeal bread and fruit leads to a substantial increase in food costs which may be beyond the practical means of many low-income families.

- Welfare payments aim to provide enough money for the essentials of life, but not enough to discourage people from working.
- Many factors will affect how much money any given family needs to provide a modest lifestyle, which makes setting the level of welfare payments a difficult and flawed process.
- Poverty may reduce the practical purchasing power of those on low incomes, e.g. if they:
 - cannot reach the cheapest shops
 - lack the storage facilities or capital to buy more of special offers
 - have to buy in small quantities to meet immediate needs.
- For low-income British families, financial considerations were the most important factors governing food choice. For these families, incorporating advice on healthy eating was seen as impractical.
- Poor families maximized economic efficiency by shopping around and avoiding waste. To avoid waste, children's preferences were given high priority.
- Poor families tried to maintain a mainstream diet and parents made personal sacrifices to prevent their children being stigmatized by eating differently from other children.
- Poor families resisted radical changes to their diets.
- During times of financial pressure, the food budget may be pared to allow spending on other essential items or bills.
- People struggling with difficult economic circumstances do not always make selections that are scientifically logical. They may buy foods that are palatable or help to maintain their self-esteem rather than those that are cheap and nutritious.
- It is possible to produce very low-cost lists of ingredients that meet particular nutritional criteria, but these may involve eating a diet that is very different from the mainstream.
- The minimum cost of a healthy diet increases markedly if the ingredients must combine to make a culturally acceptable mainstream diet.
- Lack of cooking facilities will reduce the food choices of poorer people and raise the cost of a healthy diet.

Families living in 'temporary' bed and breakfast accommodation may face yet another extra financial burden. They may have few or no cooking facilities and thus be forced to rely on eating out, which is expensive, or to rely largely upon cold foods that require no further cooking. Relying largely upon cold foods also has its disadvantages: the diet is less varied and palatable and quite different from the diet of the culture. The minimum cost of eating a healthy diet is likely to be higher (perhaps around 15% higher, according to Leather, 1992).

CULTURAL AVAILABILITY

Beef is physically and economically available to devout Hindus in the UK, but religious conviction makes it culturally unavailable. Under normal circumstances, people only eat things that are culturally recognized as foods. People choose things that, in their culture, are considered appropriate for the particular meal or occasion and that are suitable for the life-cycle group to be fed. (This is discussed above in the section dealing with the consumer classification of food.) The cultural acceptability of a food may change as it becomes physically more available and more familiar. The indigenous population may initially view unfamiliar immigrant foods with suspicion and even distaste but, with time, these foods may become more familiar and acceptable: Chinese foods and curry are now almost as much a part of British culture as fish and chips. Similarly, immigrants tend to gradually adopt some of the foods and dietary practices of the indigenous population. Exposure to alien foods through foreign travel may have a similar effect.

Dietary taboos

The word taboo is derived from the Polynesian name for a system of religious prohibitions against the use of sacred things. It has come to mean any sacred prohibition and, with common usage, the religious connotation is no longer obligatory. One could define a dietary taboo as any avoidance that is maintained solely because failure to do so would generate disapproval, ostracism or punishment within one's own cultural group or because it would compromise one's own ethical standards. A food taboo should be

distinguished from a food avoidance that is based upon sound empirical evidence of harm.

Many taboos are aimed at the avoidance of flesh and these are often rigorously adhered to. It is said that the initial spark that ignited the Indian Mutiny against British rule in 1857 was the introduction of cartridges supposedly greased with animal fat. Hindu troops were unwilling to bite the ends off of these cartridges or, indeed, even to handle them. On May 6th, 1857, 85 out of 90 members of the 3rd native cavalry in Meerut refused direct orders to handle cartridges, which, according to a contemporary account, they mistakenly believed to be contaminated with animal fat. They were subsequently sentenced to terms of imprisonment of between 6 and 10 years. These men would undoubtedly have been aware of the likely consequences of their disobeying a direct military order, which underlines the potential strength of such taboos.

The taboos of other cultures tend to be viewed as illogical restraints, often imposed upon individuals by authoritarian religious leaders. This model leads to the belief that taboos should be discouraged. This is almost certainly a misreading of the situation in most cases. Most taboos need little external compulsion to ensure adherence; people feel secure when maintaining cultural standards that have been ingrained since childhood. There are, nonetheless, a few examples of essentially religious taboos being incorporated into the secular law and of severe penalties being imposed upon transgressors: in many Indian states, the cow is legally protected, and there are severe penalties for alcohol consumption in some Moslem countries.

Most people of Western European origin would not consider dietary taboos to have any significant influence upon their food selection, but no society is completely free of dietary prohibitions. It is not too long ago that Roman Catholics were required to abstain from meat on Fridays, and some Christians still avoid meat on Good Friday. In the UK, cannibalism is clearly prohibited but, equally, the consumption of animals such as cats, dogs and horses would result in widespread hostility and probably ostracism by other members of society. Most British people would not recognize these as taboos because they do not classify these creatures as potential food, despite their being highly regarded as foods in some other countries. In France, for example, horsemeat is eaten, sometimes even meat from British horses!

In the UK, vegetarianism would seem to be an obvious candidate for an example of a dietary taboo that may have negative nutritional implications. Certainly, many omnivores would like to think so, as the vegetarian's rejection of flesh is often perceived as a threat to highly desired and high-prestige foods. There are some problems associated with a strictly vegetarian (**vegan**: See Vegetarian in Glossary) diet (see Chapter 16). However, a well-constructed vegetarian diet can be healthy and in the current climate of nutrition opinion may even be regarded as healthier than the typical omnivorous diet. Vegetarianism must be socially inconvenient where most of the population is omnivorous and where caterers may make only nominal provision for vegetarians. These inconveniences may, however, be more than compensated for by a comradeship, akin to religious fellowship, seemingly shared by many vegetarians and by enhancement of self-actualization.

Perhaps taboo also played some part in the dramatic decline in breastfeeding that occurred in Britain in the decades following World War II, or at least in thwarting current efforts to encourage breastfeeding. The inconvenience of breastfeeding and its 'prestige' can only have been worsened by a very strong taboo against breastfeeding amongst strangers, sometimes even amongst friends and relatives. The lack of provision of a suitable environment for feeding babies in most public places increases this problem. The arrival of a modern, scientifically developed and socially acceptable alternative provided women with an ideal opportunity to free themselves of the need to continue with this 'inconvenient and embarrassing process'. A survey conducted by the Royal College of Midwives in 1993 found that more than a third of the 700 restaurants surveyed either would not permit breastfeeding at the table or would ask women to stop if another customer complained. In a second survey, half of the men interviewed disagreed with women breastfeeding 'in any place they choose'. They gave reasons for their objections such as 'it's unnecessary', 'it's a form of exhibitionism', 'it's disgusting behaviour' (both of these surveys are briefly summarized in Laurent, 1993). White *et al.* (1992) reported that 28% of women said that one reason they chose not to breastfeed was because they were embarrassed or found it distasteful. Clearly, breastfeeding is seen by many men (and many women too) as a distasteful or embarrassing task that needs to be performed

furtively. In Britain, images of bare female breasts are ubiquitous in popular newspapers, television, films and even on some beaches, and yet the sight of a woman breastfeeding her baby apparently still has the potential to shock and embarrass a high proportion of British adults.

Many of the best-known taboos are permanent, but some are only temporary. Some of the temporary ones may apply only during certain phases of the life cycle or during illness. Some taboos may be restricted to certain groups of people. The familiar permanent religious taboos generally cause few nutritional problems. The temporary ones will also not usually give rise to adverse nutritional consequences unless they are imposed at times of rapid growth or other physiological stress. An extreme example quoted by Fieldhouse (1998) is the varied food avoidances amongst the women of the southern Indian state of Tamilnadu; during lactation, for example women should avoid: meat, eggs, rice, dhal, chilies, cow's milk, sardines, fruits, potato, yam, cashew nuts and onions. These avoidances would compound with previous restrictions during pregnancy and this could have serious nutritional consequences. There is a widespread belief amongst various peoples that sick children should not be fed or should only be offered very restricted diets, thus potentially precipitating **protein energy malnutrition**. There is also a widespread belief that many adult foods, especially meat and fish, are unsuitable for young children. This leaves only the starchy staple foods with their problems of low energy and nutrient density as the major permitted food at this time of high physiological demand for energy and nutrients.

If most dietary taboos are not usually associated with any major adverse nutritional consequences, then why should nutritionists and nutrition educators study them? They provide an insight into a culture, and some understanding is necessary if the general philosophy of giving advice that is culturally compatible is to be maintained. If advice is given that involves breaking an important dietary taboo, all of the advice may be ignored and, perhaps, the future credibility of the adviser destroyed. Maybe, even worse, if certain individuals are persuaded to conform to this advice, this may provoke friction and divisions within a community or family.

If a taboo is nutritionally only marginally detrimental, neutral or perhaps even beneficial in its impact, then, no matter how bizarre or irrational it

may seem, the cultural choice of the client should be respected. Some of the practices of the adviser may seem equally bizarre to the client. If a taboo is manifestly harmful, the aim of the adviser should be to eliminate the harmful impact with minimum disruption to the cultural life of the client.

The conclusion that a taboo is harmful should be based upon a wide analysis of its impact. Fieldhouse (1998) used the example of the Hindu sacred cow to make this point. There are nearly 200 million cows in India and yet their slaughter for food is generally forbidden, even in times of famine – illogical? harmful? He points out that cows provide milk and their dead carcasses provide leather and are used as food by the lower castes. Cows provide the oxen traditionally vital for Indian agriculture; cow dung is also a major source of fuel and fertilizer in rural India. Yet they scavenge much of their food, which is basically inedible to humans. The conclusion that this taboo is harmful is now less secure and the value of the taboo in preventing the destruction of these valuable and well-adapted animals in adverse conditions is, at least, arguable.

Taboos are often derived from some religious commandment, be it in the religious book of the group, the oral mythology or the edict of some historical

religious leader or teacher. Many theories as to their origins exist, such as in the list below; each can give plausible explanations of some taboos and often these different theories can be used to give equally plausible explanations of the same taboo.

- *Aesthetic:* particular animals are rejected because of their perceived unpleasant lifestyle, e.g. some people consider the flesh of the mackerel to be low prestige or inferior because they regard the mackerel as a scavenger. All meat may be rejected because of the perceived cruelty involved in rearing and slaughter.
- *Health and sanitation:* it is argued that there is, or was, some inherent logic underlying the exclusion of particular foods on health grounds. In effect, it is suggested that taboos originate from avoiding foods that are harmful. For example, the Jewish prohibitions against shellfish consumption may have been prompted by the risk of poisoning by the toxin produced by the plankton species *Gonyaulax tamarensis* that the shellfish may have consumed (see Chapter 17).
- *Ecology:* there is said to be some underlying environmental logic behind a particular prohibition. The exclusion of meat and fish at particular times

- People will normally only eat things that they recognize as food and things that they see as culturally acceptable and appropriate.
- Many religions have taboos against the consumption of particular foods, but a religious context is not always present.
- People will often go to great lengths to avoid breaking a taboo and there is little need for external compulsion to ensure compliance.
- Some examples of taboos for Western Europeans might be: vegetarianism, the avoidance of horsemeat in Britain, and perhaps even the avoidance of breastfeeding in public.
- In general, the permanent religious taboos seem to have few adverse nutritional consequences, but multiple taboos imposed at times of physiological stress, e.g. childhood, pregnancy or illness, may have greater potential to do harm.
- Dietary advisers need to be aware of taboos and avoid giving advice that would require breaking a taboo.
- There are several theories about the origins of taboos:

an animal or fish becomes taboo because of its unpleasant or 'dirty' lifestyle
 they originate because the avoidance confers some biological or ecological benefit, e.g. the avoidance of a food-borne disease or ecological damage
 they are a mechanism for binding a religious group together and distancing them from non-believers.

to conserve stocks is an obvious example. The avoidance of meat by some present-day vegetarians to prevent the inefficient use of grain and thus to increase world food availability would be another.

- *Religious distinction:* taboos serve to demonstrate the separateness of a religious group from non-believers, and the self-restraint required to obey them may serve as a symbol of piety and obedience to the religious leaders.

The avoidance of pork by Jews is a very well-known taboo and each of the above theories can be used to explain its origins (see below).

- *Aesthetic:* the pig is widely viewed as a dirty and scavenging animal, even by people who value its flesh. Jewish dietary law states that only animals that have cloven hooves and that chew the cud are clean and therefore edible. The pig is considered unclean and therefore inedible.
- *Health and sanitation:* pork is a source of a parasitic worm that causes the disease trichinosis. Perhaps this is the origin of the Jewish taboo against pork, which is shared by other Middle Eastern peoples. Opponents of this theory argue that this is merely a later attempt at scientific rationalization. The risk of trichinosis has been used as an argument against horsemeat consumption in France.
- *Ecology:* it is suggested that supernatural prohibitions against the consumption of pork arose because the desert conditions of the Middle East were unsuitable for efficient pig rearing. Proponents of this theory argue that taboos rarely involve abundant species that can be eaten with no threat to total food supplies.
- *Religious distinction:* it is suggested that the pork taboo was originally part of a relatively lax and general prohibition against 'imperfect' animals, including flightless birds and shellfish. Its symbolic importance was heightened when Syrian invaders forced Jews to eat pork as a visible sign of submission and thus, once Jews regained their independence, strict pork avoidance came to symbolize their Judaism and their opposition to pagan rule.

tal and genetic influences upon the disease patterns of populations (see Chapter 3). Migrants may also suffer from nutrition-related problems that may be much rarer both in their native homeland and amongst the indigenous population of their new homeland. For example, migrants to the UK from the Indian subcontinent have suffered from mini-epidemics of rickets, a condition that had been largely eliminated in the white population. It is, therefore, useful to try to establish and understand trends and patterns to the changes in dietary habits that inevitably occur after migration in order to answer questions such as:

- What factors influence the speed of change?
- Why is the health record of migrant groups often worse than that of the indigenous population?
- Are there any measures that might facilitate the healthful assimilation of migrant groups into their new environment?

Why should change occur at all? Culture, including food habits, is a learned phenomenon rather than something that is innate or biologically determined. Culture is transmitted between generations by the process of socialization. There is an inherent tendency of cultures to change over time. Although conflicts will inevitably arise between traditional influences and influences from the new culture, this process of change is almost inevitably greatly hastened by migration and exposure to the different cultural practices of the indigenous population. **Acculturation** is the term used to describe this acceleration of cultural change that occurs when different cultures interact. Both indigenous and migrant cultures are likely to be changed by their interaction, but migrants may feel the need to adopt aspects of the indigenous culture to facilitate their acceptance and assimilation into their new country. They will usually be in a minority and be dependent upon the goodwill of the indigenous population; they may therefore feel great pressure to conform in order to be accepted and to succeed in their new homeland.

Paradoxically, every culture also has a built-in resistance to change. Culture is mostly internalized: routine activities done unthinkingly in a particular way because that is the way they have always been done. After migration, however, there is repeated exposure to different culture practices, and familiarity may eventually lead to the acceptance of initially strange and alien practices as normal. The dietary practices of the indigenous culture that are inconsistent with the

Effects of migration upon eating habits

Migrant groups are frequently used by epidemiologists when they try to distinguish between the environmen-

values and beliefs of the migrant's culture may be the slowest to be adopted and also those most likely to cause social divisions within migrant communities or families. For example, migrants may be most reluctant to absorb dietary practices that would involve breaking the food rules of their religion. The older, more conservative, migrants will probably be most resistant to such changes and may also be hostile to changes in the behaviour of younger, less ethnocentric members of their community or family. Dietary practices of migrants that are inconsistent with the values of the indigenous culture may also be a source of considerable friction between migrants and their fellow citizens. For example, Moslem and Jewish rules concerning the slaughter of animals have, in the past, provoked hostility in Britain from people concerned about animal welfare.

Bavly (1966) analysed changes over three generations in the diets of immigrants to Israel. She considered that several factors had a major accelerating effect upon the speed at which change occurred:

- marriage between different ethnic groups
- the homemaker working outside the home
- children receiving nutrition education at school
- children having school meals
- nutrition education via the media for immigrants of European origin.

Although this work was done some time ago, these specific influences given by Bavly may be generalized to factors that increase interaction with the indigenous population and increase familiarity with, and understanding of, indigenous dietary practices. Conversely, any factors that tend to isolate the migrant from the indigenous population and culture may be expected to slow down acculturation, such as those listed below:

- Inability to read or speak the new language restricts interaction at the personal level and restricts access to the media.
- Cultural beliefs that discourage the homemaker from independent socializing outside the family.
- Religious beliefs and dietary rules that are at odds with the dietary practices of the indigenous majority.
- Living within a fairly self-contained immigrant area with shops that are run by fellow migrants and where familiar foods and even native language media are available.

Many of these isolating influences would apply to many new immigrants from the Indian subcontinent to Britain.

Migration is often prompted by the attraction of improved economic opportunities in industrialized countries. Thus, large-scale migration is often accompanied by a complete change of social structure from a rural agrarian to a Western industrial type of society. In the rural agrarian community, society is likely to be based upon extended and close-knit family groupings, with a constant flow of food between families and individuals. In Western industrialized societies, this informal family and community support may no longer be available because they are organized into relatively isolated family groups, food is normally shared only with the immediate family, and this sharing may help to define the family. These changes in social organization may mean that, in times of hardship, the missing informal neighbourhood and family support will have to be replaced by formal charitable or state welfare support. The impact of such changes may be ameliorated where movement is into an established ethnic community.

Migrants from a rural, food-orientated economy may suddenly be confronted with a cash-dominated economy. The family incomes of new immigrants are likely to be relatively low (inappropriate skills, language problems, discrimination etc.). This combination of low income and lack of experience in cash budgeting may make them unable to cope adequately, even though the income may be technically adequate. An educational induction programme might reduce the impact of this potential problem. Food selection for migrants may be complicated by the unavailability of recognizable traditional foods; even where they are available, they may be relatively expensive because of their specialist nature. Food selection by immigrants may be particularly influenced by advertising pressures that encourage the excessive consumption of foods that are high in palatability and prestige but of relatively low nutritional value. Social and cultural pressures may discourage migrant women from breastfeeding their infants.

Migration from a rural agrarian society in a developing country to a Western industrial society will probably be associated with overall trends that are similar to the worldwide changes that accompany increasing affluence. The traditional starchy and predominantly vegetarian diet is likely to diversify and become more omnivorous, and sugars and fats will

progressively replace some of the starch. The diseases associated with nutritional inadequacy will decline, but the chronic diseases of industrialization will almost certainly become increasingly prevalent.

Wenkam and Woolff (1970) surveyed the changes in the dietary habits of Japanese immigrants to Hawaii and their descendants. Japanese migrants began arriving in Hawaii at the end of the nineteenth century to become plantation workers. The typical Japanese diet at this time was a high-carbohydrate, predominantly rice and plant food diet. The major animal foods would have been fish and other seafood. Initially, these migrants maintained their traditional diet; many regarded their stay in Hawaii as temporary and ate

frugally in order to save for their return to Japan. After the annexation of Hawaii by the USA in 1898, these Japanese migrants started working outside of the plantations, and businesses specializing in the importation of traditional Japanese foods sprang up. Initially, this resulted in the increased consumption of imported and high-status Japanese foods. As these Japanese foods became cheaper, their status diminished and they came to be regarded as old-fashioned by many younger Japanese, who preferred to eat American foods.

The traditional Japanese social structure, with its strong family tradition, worship of ancestors and subordination of personal desires to the welfare of

- The disease patterns of migrants provide information about the relative importance of genetic and environmental factors in the aetiology of diseases.
- Migrants often have poorer health than other members of the new community.
- Migrants tend to gradually adopt elements of the diet and lifestyle of their new country – they tend to acculturate.
- The indigenous population may also adopt some elements of the migrant culture and diet.
- Changes will be slowest where the dietary practices of the indigenous population are in conflict with the values and beliefs of the migrants, e.g. where they would involve breaking a religious taboo.
- The adoption of change by some migrants may be a source of conflict with their parents and elders.
- Anything which increases interaction between the migrants and the indigenous population and familiarizes the migrants with the indigenous diet will tend to speed up the rate of change, e.g. women working outside the home and children eating school meals.
- Anything that prevents interaction will slow down the rate of acculturation, e.g. language barriers, restrictions on women working and socializing outside the home, or living within a fairly self-contained immigrant community.
- Migrants to an industrialized country from a rural area with an agricultural food-based economy may find it difficult to cope with the new cash-based economy, particularly as their incomes are likely to be relatively low.
- Several stages in the acculturation process have been recognized:

People who have recently migrated may initially try to stick closely to their familiar diet and people who migrate later in life may remain in this category for the rest of their lives.

Younger adult migrants who enter employment will become familiar with the new diet and tend to eat both the new and traditional foods.

Children of migrants who attend school in the new country may become so acculturated that they prefer the new diet and resist eating their traditional foods, perhaps to increase their sense of belonging and acceptance by their peers.

Descendants of migrants born or brought up in the new country may seek out their ethnic roots and eat foods associated with these roots to reaffirm their ethnic and cultural identity.

the family, was replaced by a more fragmented and personal-freedom-orientated society. The traditional Buddhist restrictions on the consumption of meat weakened as meat was plentiful in Hawaii.

Public education seems to have been a catalyst for change amongst the second generation as the schools encouraged them to change their food and other habits. There also seems little doubt that World War II accelerated the 'Americanization' of Japanese in Hawaii. Wenkam and Woolff describe a Japanese wedding in Hawaii before the war as being typically Japanese in character, with no wedding cake, but by 1945 the food at such a wedding was likely to be much more cosmopolitan, but with a Western wedding cake as the highlight.

Wenkam and Woolff concluded that, in 1970, the Japanese in Hawaii could be classified into three groups:

1. a traditional group of mainly older people who maintained a very Japanese cuisine
2. a group that, despite being relatively acculturated, still consumed Japanese foods on some occasions
3. an 'Americanized' group that rejected its Japanese heritage and the foods associated with it.

There may now exist yet a further category of Japanese-Americans who are moving back to traditional foods for health reasons and to reaffirm their ethnic identity (i.e. for reasons of self-actualization).

Williams and Qureshi (1988) suggested a similar 'four-generation' concept that could be a useful way of explaining some of the variations in food choices within any particular ethnic minority group in the UK:

- *First generation A:* made up of migrants who arrived as dependent relatives of migrant workers; they are usually retired and prefer to eat their native foods.
- *First generation B:* migrants who arrived as workers; they are usually aged 21–65 years and they accept both native and British foods.
- *Second generation:* young people aged 7–21 years who went to school in Britain and in many cases were born there; they prefer the new diet and may strongly resist attempts to force them to consume their native diet, which is alien to them.
- *Third generation:* children who were born in Britain, who feel British, but who have become interested in tracing their ethnic roots; they accept both British and native foods.

'GATEKEEPER' LIMITATIONS ON AVAILABILITY

The housewife has traditionally been regarded as the family gatekeeper, regulating the availability of food to her family. If most food is eaten within formal family meals and if food purchasing and meal construction are largely the province of the housewife, then she has the potential to impose limitations on the availability of foods to the other family members. Personal beliefs and preferences may greatly affect the choices of the gatekeeper, and thus the diet of the whole family. A vegetarian gatekeeper may effectively impose vegetarianism on other family members or, conversely, a non-vegetarian gatekeeper might refuse to provide a vegetarian alternative to a family member who would wish to be vegetarian. Convenience of preparation may affect the choices of the gatekeeper. Appreciation or criticisms by other family members will feed back and also be an important influence on her future choices. A number of social changes in industrialized countries might be expected to have undermined the traditional gatekeeper role of wives and mothers:

- More food tends to be eaten outside the home or is purchased by the individual consuming it.
- Ready-made versions of dishes that have traditionally been home made shift control over composition from the housewife to the food manufacturer.
- Other family members are more likely to participate in the shopping and food preparation.
- Within the home, there is more snacking or 'grazing' upon self-selected foods available in the larder or refrigerator and fewer formal family meals.

Despite a diminution in this gatekeeper role of the mother, even within a traditional two-parent family, it may still be quite considerable. Certainly, much food advertising is still directed towards the gatekeeper housewife. Advertisers stress factors such as the convenience of foods to a busy housewife and provide images of wives and mothers who have served the promoted food receiving acclaim from an appreciative and therefore loving family. Nicolaas (1995) conducted a survey of a representative sample of 2000 British adults on their attitudes to cooking and their cooking behaviour. This survey showed a marked divergence between attitudes and actual behaviour

and suggested that the traditional gatekeeper role of the oldest adult female may be still largely intact in British families. Almost all the participants thought that it was important that both sexes had basic cooking skills and to teach their children how to cook. Almost all agreed with the proposition that boys should be taught how to cook. When questioned about their actual behaviour, however:

- 80% of women but only 22% of men claimed to prepare every meal in their household
- 3% of women and 18% of men admitted to never preparing a meal.

Most of the women who never prepared a meal were in the 16–24-year age group, and the frequency in this age group was similar to that of young men. The complete data set strongly suggests that the majority of young people in this age group who live with their parents rely upon their mothers to cook for them. In older age groups there is an even greater difference between male and female responses than those summarized above for the total sample. When adult men and women live together as partners, it is usually the female who does most of the cooking.

Catering managers may exert a considerable gatekeeper influence over those living in institutions

such as hospitals, prisons, retirement homes, boarding schools and, to a lesser extent, those eating lunches at school or in their workplace. Older people living at home may have inexpensive weekday main meals delivered to them by a 'meals on wheels' service. The Caroline Walker Trust (1995) has suggested nutritional standards and sample menu plans for 'meals on wheels' and meals provided in residential homes for the elderly.

If someone is subject to the gatekeeper influences of someone from a different cultural background, they may effectively be subject to a double dose of limitation due to cultural availability. The gatekeeper may only offer foods that are culturally available to him or her and only some of these may be culturally available to the consumer. Unless caterers make special provision for cultural/ethnic minorities, this may severely restrict the real choices of such people, e.g. hospital patients from minority groups.

Even those people seemingly immune to the influence of gatekeepers, such as the affluent person living alone, may still use staff dining facilities in the workplace. They may have limited time available for food preparation or low incentive to cook just for themselves. They may thus rely heavily upon commercially pre-prepared meals or restaurants.

- Wives and mothers have traditionally been the gatekeepers in Western European families, but anyone who caters for someone else is acting as a gatekeeper.
- Recent surveys suggest that the gatekeeper role of women is largely intact in British families and women are still largely responsible for selecting and preparing the family meals.
- Gatekeepers can have almost total control over what some people eat and their scope for making dietary changes may be dependent upon the goodwill and co-operation of their gatekeeper.
- The range of food that is truly available may be considerably narrowed if a person is reliant upon a gatekeeper from a different cultural background, e.g. someone from an ethnic minority relying upon a white gatekeeper.
- Almost everyone is subject to some gatekeeper influences.

Methods of nutritional surveillance and research

Aims and introduction	60	Methods used to establish links between diet and chronic disease	91
Nutritional assessment and surveillance	60		

AIMS AND INTRODUCTION

The methods used to study nutrition range from the high-precision measurements of the analytical chemist through to the inevitably much less precise methods of the social scientist and behavioural psychologist. All approaches can be equally valid and rigorous and the complete range of approaches is necessary to study all of the various aspects of nutrition.

The aims of this chapter are summarized below.

- To make readers aware of the range of methods that are available for nutritional surveillance and research, how they are performed, their strengths and limitations, their general theoretical basis and the circumstances in which different methods are more or less appropriate. This is a first step towards the successful use of these methods.
- To give readers a sufficient insight into the methods of nutritional surveillance and research to enable them to read the literature critically. An understanding of the relative strengths and the limitations of the methods used is essential for any critical evaluation of the conclusions drawn from reported data. Some perception of the errors inherent in methods of measurement is necessary for realistic appraisal of the results that they produce.

NUTRITIONAL ASSESSMENT AND SURVEILLANCE

Strategies for nutritional assessment

There are two ways of approaching the assessment of dietary and nutritional status:

1. One can try to determine whether the nutrient intake (or nutrient supply) is adequate to meet the expected needs of the individual or group under investigation. In order to make this assessment, one needs the three key tools listed below:
 - (a) a method of measuring the amounts of foods being eaten (or perhaps the available food supply)
 - (b) a way of translating the amounts of foods being consumed into quantities of energy and essential nutrients, i.e. tables of food composition
 - (c) some yardstick of adequacy that can be used to determine whether the measured intakes of energy and essential nutrients are sufficient for that individual or group. These standards of dietary adequacy are called **Dietary Reference Values (DRVs)** in the UK and **Recommended Dietary Allowances (RDAs)** in the USA and elsewhere.
2. One can make observations and measurements on the individuals concerned to try to determine

whether they have been receiving adequate amounts of energy and essential nutrients. The three types of indicator that can be used to make this assessment are:

- (a) **clinical signs**: one can look for clinical signs or the symptoms of a nutrient deficiency
- (b) **anthropometry**: one can make anthropometric measurements, such as of body weight and height, and then either compare these measurements to appropriate standards, or monitor changes within the individual or group over time
- (c) **biochemical assessment**: one can make biochemical measurements, usually upon blood or urine samples, that indicate either the general nutritional status of the donor or the status with regard to any specific nutrient.

Measurement of food intake

Measures of food intake have a variety of purposes, such as those listed below.

- They are the first step towards assessing the adequacy of the diets of populations, groups or individuals and to identifying the problem nutrients.
- They can be used to relate dietary factors in populations or individuals to disease incidence or **risk factors** such as plasma cholesterol concentration.
- One can compare the diets of different countries, regions, socioeconomic groups and different age and ethnic groups.
- They can be used to monitor changes in the diets of populations, groups or individuals over time and thus, for example, to monitor the effectiveness of nutrition education programmes.

POPULATION OR GROUP METHODS

Food balance sheets

These are used to estimate the average *per capita* food and nutrient consumption of nations (see Figure 3.1)

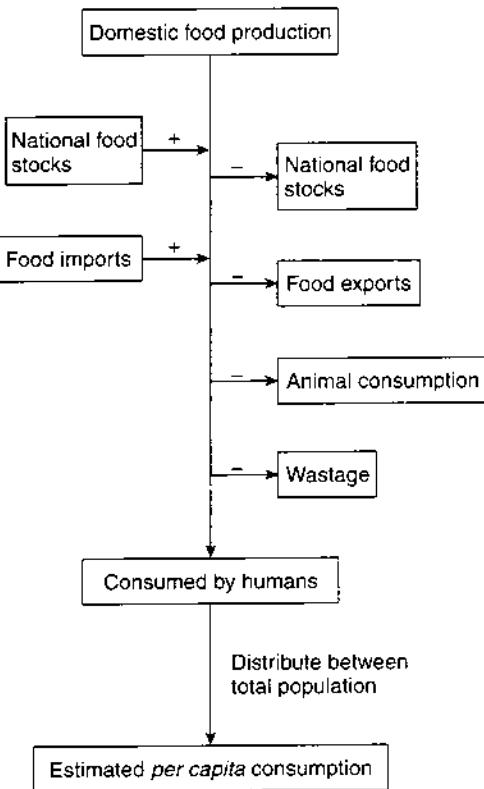


Figure 3.1 A diagram to illustrate how food balance sheets can be used to estimate *per capita* food consumption in a country.

for summary). Food balance sheets usually yield no information about distribution within the population. They will not show up, for example, regional differences in nutrient intakes, socioeconomic influences upon nutrient intakes or age and sex differences in nutrient intakes. Domestic food production is first estimated. Governments and international agencies (e.g. the Food and Agriculture Organization of the United Nations, FAO) routinely make estimates of the total crop yields and total agricultural production of countries. Any food imports are added to this total and any exports are subtracted from it.

- Dietary and nutritional status can be determined either by assessing the adequacy of the nutrient intakes of the subjects or by monitoring clinical, anthropometric or biochemical indicators of nutritional status.

Allowance is made for any change in food stocks within the country during the period. This estimate of food available for human consumption must then be corrected for wastage and any food not consumed by humans, e.g. animal fodder and pet foods (see Figure 3.1). Such methods allow crude estimates of population adequacy and crude comparisons of actual intakes with nutrition education guidelines to be made. They also enable international comparisons to be made, e.g. one could relate average *per capita* sugar consumption to rates of dental caries in different populations.

Home-produced food and small-scale production sites will be one potentially large source of error in these estimates. Wastage will be difficult to estimate accurately and is likely to depend upon the affluence of the population – the rich are likely to waste more than the poor. Within the European Union (EU), the construction of national food balance sheets has been complicated by the introduction of the single market, with the consequent reduction in checks upon imports and exports between member states. Similarly, the average intakes of residents of an institution can be estimated from the amount of food entering the institution over a period of time.

Household surveys

The **National Food Survey (NFS)** of the UK is an example of a national household survey that has now been undertaken annually for almost 60 years (e.g. MAFF, 1997). One member of a nationally representative survey of 7000–8000 households is asked to record in a diary all food entering the household for human consumption during 1 week. Records from different families are collected at different times throughout the year. A trained interviewer checks the diary at the end of the survey period. Home-grown food is included in the log. Until recently, detailed information about food eaten outside the home was not collected, merely the number of external meals taken by each family member. Details of meals eaten outside the home are now recorded and reported in a separate section within the survey report. Up until 1992, no information was collected about alcoholic drinks, soft drinks (soda) or confectionery (candies) as these are often purchased by individual family members. Those purchased for home consumption are now recorded but, in order to make historical comparisons easier, the nutritional analyses are presented both including (total household food and

drink) and excluding these products (household food excluding soft drinks, alcohol and confectionery).

Information is also recorded concerning the composition of the family, the ages and sexes of the family members, its income and socioeconomic status and the region of the country in which it lives. There are a number of errors and uncertainties when the NFS is used as an indication of what people in Britain eat, and some of these are listed below.

- The NFS does not try to measure what people eat in a week – it measures what they buy. Differences in wastage rates and preparation practices will not show up in the NFS. For example, if a family buys 500 g of meat but trims the fat off before they eat it, this will give the same record as another family that buys this meat and eats all of the fat.
- There is evidence that families buy more food than usual during the survey period and that this tendency may be proportionately greater in lower income groups.
- The survey may not be truly representative of the whole population. Some groups are excluded from the survey, such as the single homeless and those low-income families housed ‘temporarily’ in bed and breakfast accommodation. Some poorer groups may be deterred from participating in the survey because of embarrassment about the diet they can afford (see Leather, 1992).
- No attempt is made to record distribution of the purchased food between family members, although assumptions about wastage, distribution and losses of nutrients during cooking and preparation are made in order to estimate individual intakes of food and nutrients. The recorded average *per capita* expenditure upon food decreases with increasing family size, partly because children tend to eat less than adults (it may also be more economic to buy in bulk).
- Individual family members may buy snacks, drinks and sweets, which they consume outside the home and which go unrecorded in the household log.

Despite these limitations, the NFS provides an invaluable record of the changing dietary habits of the British population over more than half a century (see MAFF, 1991 for a review of the first 50 years of the NFS). Many other countries conduct similar expenditure surveys, but not with the same regularity and/or duration as the British survey. The NFS can be used for the following purposes.

- It is an invaluable source of information about changes in food-purchasing practices over time. It can thus, for example, be used to monitor the impact of health promotion, 'food scares' and price movements on the national diet. In Chapter 4, the NFS is used to quantify the switch from animal-derived cooking and spreading fats to those made from vegetable sources since the 1970s. The NFS gave clear evidence of the scale of the decline in beef sales (especially burgers) in the wake of the bovine spongiform encephalopathy (BSE) crisis in Britain.
- It gives information about regional differences in dietary habits within Britain. These data may be of use in trying to explain some of the regional differences in the health and disease pattern in Britain.
- It provides information on seasonal patterns of food purchasing.
- It gives information about differences in the diets of different socioeconomic groups. One can compare the diets of rich and poor, the diets of young and old, the diets of the employed, unemployed and retired.
- It gives us information about how food-purchasing and eating habits differ according to family structure. One can compare the diets of single-adult or two-adult families with those of families with varying numbers of children.
- It is possible to make crude estimations of the supply of energy and essential nutrients in British diets and thus of the probable levels of dietary adequacy and to highlight nutrients whose intakes may be less than satisfactory. One can also make crude estimations of the degree to which current diets comply with current dietary guidelines. One can monitor how intakes of nutrients are changing with respect to both standards of adequacy and dietary guidelines for the prevention of chronic disease.

INDIVIDUAL METHODS

Individual methods of assessing food intake may either involve prospective recording of food as it is eaten or may assess intake retrospectively by the use of interviews or questionnaires. These methods, almost inevitably, rely to a large extent upon the honesty of the subjects.

Retrospective methods

Twenty-four-hour recall is very frequently used to retrospectively assess the food and nutrient intakes of

subjects. An interviewer asks subjects to recount the types and amounts of all food and drink consumed during the previous day. This recalled list is then translated into estimates of energy and nutrient intakes by the use of food tables. The method has often been used in large-scale epidemiological surveys where large numbers of subjects need to be dealt with quickly and cheaply and where precision in assessing the intake of any individual is deemed not to be essential. The method requires relatively little commitment from the subject and thus co-operation rates tend to be good. As the assessment is retrospective, subjects do not alter their intakes in response to monitoring.

Some of the limitations and sources of error in this method are listed below.

- *Memory errors:* subjects are likely to forget some of the items that have been consumed. This method tends significantly to underestimate total calorie intake and seems particularly to underestimate, for example, alcohol intake. Even when used with probing to try to improve memory, it probably under-records energy and most nutrients by at least 20%. This makes any assessment of the prevalence of dietary inadequacy using this method highly suspect.
- *Quantification errors:* it is difficult to quantify portions in retrospect. It may be particularly difficult for the very young and the very old to conceptualize amounts. Food models or photographs are sometimes used to aid this quantification process. Nelson *et al.* (1997) have published a photographic atlas of food portion sizes.
- *Intra-individual variation:* intake may not be representative of the subject's usual intake. The day chosen for the record may not be typical of the subject's usual daily intake. Weekend and weekday intakes may be very different in some subjects. Illness or stress may affect the appetite of a subject. Some social occasion or event may exaggerate the subject's usual intake. Even where subjects are of regular habits and no particular factors have distorted that day's intake, the intake of many nutrients tends to fluctuate quite markedly from day to day and thus one day's intake cannot be taken to represent the subject's habitual intake.
- *Interviewer bias:* any interview-based survey is liable to bias. The interviewer may encourage subjects to remember particular dietary items more than others and thus may obtain a distorted

picture of the real diet. If the interviewer indicates approval or disapproval of certain foods or drinks, this may also encourage distorted recall. It requires skill and training to take a proper diet history.

It might seem logical to extend the recall period if one wishes to get a more representative picture of a subject's usual dietary pattern but one would, of course, expect memory errors to increase exponentially as the recall period is extended. It is possible to use a succession of 24-hour recalls to get a more representative picture of an individual's habitual intake. In order to try to get a more representative picture of dietary pattern, detailed **diet histories** may be taken by interview or, more frequently, **food frequency questionnaires** may be used (see 24-hour recall in Glossary). The investigators may be interested in the intakes of particular nutrients, and thus the frequency of consumption of the types of foods that contain these nutrients may give a useful indication of the usual intake of that nutrient. For example, assessing the frequency, types and amounts of fruits and vegetables consumed would be a useful guide to vitamin C or β -carotene intakes. Food frequency questionnaires are not suitable for assessing actual intakes of nutrients, but are useful for categorizing individuals into low, medium and high intakes for any selected nutrient.

Prospective methods

The **weighed inventory** requires that the subjects weigh and record all items of food and drink consumed over a predetermined period (a week is often used). The operator must then translate this food inventory into nutrient intakes with food tables. Household measures (e.g. cup, spoon, slice, etc) can be recorded in a food diary rather than subjects being required to weigh everything. This would clearly involve some loss of accuracy in determining the size of portions of many food items. These prospective methods have the advantage of being direct, accurate, current and of variable length, enabling more representative assessments of average intakes to be made.

Some disadvantages of these methods are listed below.

- Subjects may still forget to record some items that they consume, especially snacks.

- They are labour intensive for both subject and operator; considerable motivation and skill on the part of the subject are required for accurate and complete recording. Participation rates may therefore be low, and this may make the population sampled unrepresentative of that being surveyed. This type of recording usually requires that subjects are numerate and literate. In order to obtain records from subjects who do not have these skills, a variety of methods have been tried. For example: subjects can photograph their food and drink; subjects have been given balances that have a tape recorder incorporated so that their oral description of the weighed food can be recorded. Even though it is very labour intensive, the 7-day weighed inventory has been used on a mass scale in Britain to assess the diets of representative samples of the British population. For example, Gregory *et al.* (1990) used this method for a dietary and nutritional survey of a representative sample of 2200 British adults aged 16–64 years. More recently, Finch *et al.* (1998) have reported the results of a 4-day weighed inventory carried out by a sample of 1275 free-living elderly British people.
- Prospective methods may be invalidated if subjects modify their behaviour in response to monitoring, e.g. in an attempt to impress the operator, in order to simplify the recording process, or simply because their awareness of their food and drink intake has been heightened by the recording process. There is evidence, for example, that subjects involved in clinical trials may conform to the recommended diet more closely on recording days.

All of the methods of recording food intake so far described require that the estimated amounts of food eaten be translated into amounts of energy and nutrients by the use of some food composition database, i.e. paper food tables or a computerized version of the database. With **duplicate sample analysis**, a duplicate sample of all food eaten is prepared and may be subject to direct chemical analysis. This method does not, therefore, have to depend upon the use of a food composition database. The use of food tables may, for example, overestimate the intake of labile vitamins like vitamin C in people relying upon hospital or other institutional food that has been kept hot for extended periods. For example, Jones *et al.*

(1988) found very low vitamin C intakes amongst a group of long-stay hospital patients when assessed by the analysis of representative samples of the food eaten. These intakes were much lower than would have been obtained using conventional food tables. They indicated that prolonged warm holding of food was an important factor in this low intake and suggested it might also apply to other heat-labile nutrients. This method is obviously labour intensive, expensive and requires good analytical facilities. It is most appropriately used in a metabolic research setting.

Table 3.1 gives a summary of the relative advantages and disadvantages of retrospective and prospective methods of dietary assessment.

Tables of food composition

Food tables are an essential tool of both nutritional surveillance and research. Printed food tables contain lists of thousands of foods, and the amounts of each nutrient in a standard amount (100 g) of each food are given. They are based upon standard databases such as that maintained by the US Department of Agriculture (USDA) or, in Britain, by the Royal Society of Chemistry on behalf of the Ministry of Agriculture, Fisheries and Food (MAFF). To translate food intakes into nutrient intakes with such printed tables is very tedious and time consuming. The content of each nutrient in the consumed portion of each food has to be calculated from the tables and

Table 3.1 *The relative advantages and disadvantages of retrospective and prospective methods of measuring food intake*

Problems with all methods

Honesty: most methods rely upon the honesty of subjects. Records that are clearly invalid may need to be identified and excluded, e.g. recorded energy intakes that are below $1.2 \times$ the basal metabolic rate in non-dieting subjects

Under-recording: for one reason or another, most methods tend to under-record energy intakes and there is considerable evidence that this under-recording is more pronounced in the obese

Intra-individual variation: one assumes that the intake of nutrients during the monitored period is representative of habitual intakes; it will require several days of monitoring to obtain a reasonable assessment of habitual intakes for most nutrients; for some nutrients (e.g. vitamin A), for which intra-individual variation is particularly great, it may require longer

Population sampling bias: when sampling populations, one assumes that the sampled group is representative of the total test population; the more commitment and skill a method requires, the less likely this is to be true; some groups may be difficult to access, e.g. homeless people, disabled people; when sampling some groups, e.g. young children or the very elderly, one may need a proxy to carry out the recording

Food table errors: all methods except duplicate sample analysis rely upon the use of food tables; this may be a particular problem in determining intakes of residents in hospitals and care institutions where prolonged warm holding may destroy much of the heat-labile vitamin content

Coding errors: when subject records are analysed, some entries may be ambiguous and misclassified

Retrospective methods

Advantages

Tend to be quick and cheap

Require little commitment from subjects so tend to get high participation rates

Honest subjects cannot change their habits in response to monitoring because the food has already been eaten

Disadvantages

Very prone to memory errors

Retrospective quantification can be difficult

Interviewer may influence the subject

Prospective methods

Advantages

Greater accuracy because less prone to memory and quantification errors

Disadvantages

Require commitment and competent record keeping by subjects and this can reduce participation rates

Subjects may deliberately or unwittingly modify their eating in response to monitoring e.g. to simplify recording, to deceive/impress the recorder or simplify because of heightened awareness.

- Measurements of food intake can be used to assess dietary adequacy, to monitor the degree of compliance with nutritional guidelines, to make comparisons between groups or populations and to relate dietary variables to disease risk.
- The food intakes of populations can be determined by the use of food balance sheets, by representative surveys of household food purchases (like the UK National Food Survey) or even by assessing the individual diets of representative population samples.
- Individual methods of assessing food intake can be retrospective or prospective. One can either ask subjects to provide information about what they have already eaten (e.g. 24-hour recall or a food frequency questionnaire) or ask subjects to record what they eat over a period, e.g. 7-day weighed inventory.
- The choice of method of measuring food intake will depend upon the time and resources available and also upon the purpose of the measurement.
- The advantages and disadvantages of the various methods are listed in Table 3.1.

eventually the total consumption of each nutrient determined.

Nowadays, the information contained within these printed tables is available on computer software; the investigator merely has to key in the foods and portion sizes and the computer will then automatically translate these into nutrient intakes. The programme will not only accept weights of foods, but will translate household measures or descriptive portions into estimated weights of food. These programmes will also calculate the nutrient intakes as proportions of the standard reference values appropriate for a subject whose size, age, sex, activity level and reproductive status have also been entered. In Britain, a program known as Comp-Eat is widely used and is obtainable from Carlson Bengston Consultants Ltd, 21 Craven Hill, London W2 3EN. In America, a large number of good and inexpensive programs are available (the USDA database is 'public domain' and can be used without cost by software designers).

Both forms of these tables are the result of thousands of hours of meticulous analytical work. To give the nutrient contents of 1000 foods requires more than 30 000 individual determinations. I have excluded discussion of these analytical methods from this section because it would be inappropriate for a book of this type. A few of these methods are discussed in other sections of the book, e.g. the determination of the energy content of foods is outlined in Chapter 6 and the protein content of foods in Chapter 10.

The analytical methods used to determine the nutrient content of foods are often extremely precise, with a very low level of measurement error. However, several sources of error are inherent in the use of food tables and these may produce errors and uncertainties that are orders of magnitude greater than the errors in the analytical methods. Food tables can only give an indication of the typical or average composition of a food, and that may in some cases be very different from the actual composition of the sample of food that has been eaten. Some of these errors are listed below.

FOOD TABLE PROBLEMS AND ERRORS

Limited range of foods covered by tables

The major databases mentioned above each contain several thousand different food items, including many menu items from the major restaurant chains. Despite this, there will inevitably be foods eaten that are not specifically listed in the printed tables or computer software that uses them. In such circumstances, the operator either has to find alternative sources of information (such as nutritional information on the food label) or has to judge which is the nearest equivalent in the food tables. Continued innovation on the part of food manufacturers and increased availability of foods from different parts of the world mean that this will be an ongoing problem. For example, in version 5 of the Comp-Eat program mentioned above, there is

no entry for goat meat or star fruit. In fairness, it must be said that it is now quite difficult to find gaps in the British database and the range of foods listed has increased very substantially over the last couple of decades.

Recipe variations

Many dishes are extensive mixtures of foods and it often not practical in diet records or histories to obtain the precise mixture of ingredients used in the preparation of a particular dish. Even though the food tables may list this dish, the recipe used for the analysis in the tables may be quite different from that consumed. Thus, for example, bolognese sauce is something that will be listed in the tables but will vary markedly in its composition from cook to cook. The range and proportions of the main ingredients used, leanness of the meat, the amount of thickening agent (e.g. flour) and salt, and the proportion of water may all vary very considerably. When estimating the nutrients in something like a slice of pie, the proportions of crust and filling used by the analyst may be very different from those that were actually eaten.

Brand variation

Different brands of the same commercially prepared foods will also vary, and the composition of any single brand may vary significantly from batch to batch. There are several brands of 'baked beans canned in tomato sauce'; food tables can only give an average or typical composition.

Nutrient supplementation

Some foods may be supplemented with nutrients. This could be a source of considerable error if the supplementation in the analysed and eaten food is different, e.g. if only some brands of a food are supplemented or if American food tables are used to analyse British bread.

Biological variation

Just as different brands and recipes of prepared foods will vary in their composition, so different varieties of natural foods will vary in their composition. There may even be marked variation, particularly seasonal variation, within the same variety. Rump (buttock) steak from different cattle may have markedly different proportions of fat and lean. The

vitamin C content of apples varies greatly among different varieties. The vitamin D content of milk varies between summer and winter. Once again, food tables can only indicate an average or a typical composition.

Method and duration of storage

The nutrient content of foods changes during storage. Milk left in strong sunlight loses much of its riboflavin (vitamin B₂). The vitamin C content of main-crop potatoes after 8–9 months of storage will be only about a quarter of that when they were freshly dug, and stale vegetables generally may have lost much of their original vitamin C. Foods upon which mould has grown may acquire dietetically significant amounts of vitamin B₁₂. Warm holding of food prior to its being eaten may lead to the destruction of most of the heat-labile vitamins in it.

Vitamin precursors and variants

Some vitamins may have several different chemical forms or may be present in the form of a precursor. When the vitamin activity of foods is assessed, the activities contributed by the different chemical forms must be estimated and added together. Vitamin A activity in food is due to vitamin A itself (retinol), the plant-derived pigment β-carotene and other plant pigments that have vitamin A activity, such as cryptoxanthine and α-carotene (see Chapter 12). The niacin (vitamin B₃) derived from food may either come directly from vitamin in the food or indirectly by conversion of the amino acid tryptophan to niacin.

Bioavailability of vitamins and minerals

Food tables can give a precise estimate of the amount of a chemical in a food, but they give no indication of its bioavailability, e.g. how much of it is absorbed in the gut. Iron from meat is much better absorbed than that from vegetable sources, and iron absorption is increased by vitamin C and alcohol but decreased by fibre and tannin. The calcium in milk is much better absorbed than that in most other foods. There may also be doubt about the biological activity of some forms of a vitamin in foods, e.g. some of the conjugated forms of folic acid.

Use of table and cooking salt

Salt is frequently added to food during cooking and at the table. The use of a weighed inventory and of

- Food tables contain lists of the nutrient contents of thousands of foods, either in paper form or on computer software.
- They can be used to convert amounts of food into amounts of energy and nutrients.
- They can only give a typical or average composition and are prone to many limitations and sources of error, such as:

some foods may be missing from the database

there are biological variations in the composition of plants and animals and brand and recipe variations in the composition of prepared foods

variation in the method and duration of food storage will affect its composition

they take no account of variations in the bioavailability of nutrients

the mineral content of foods may be affected by the addition of salt, the presence of adhered bone, the mineral content of the cooking water, or leaching of minerals into food from utensils and storage vessels.

standard food tables is therefore regarded as an unreliable means of estimating an individual's total salt intake.

Calcium from bone

Pieces of bone in meat and fish may greatly increase the calcium content. With canned fish, the bones are softened and made more readily edible, and cutting meat with a saw produces powdered bone that sticks to the meat and is eaten.

Contamination from utensils and cooking vessels

Contact of foods with processing machines or with cooking utensils may significantly affect the content of metallic elements. Cooking or brewing in iron pots may enhance the iron content of foods very considerably, e.g. beer brewed in iron vessels has been a cause of frequent toxic iron overload amongst the Bantu people of South Africa.

Minerals from cooking water

Cooking food in hard tap water may significantly increase the content of calcium and other minerals. Variation in the content of drinking water may greatly affect the intake of some minerals, e.g. fluoride.

Minerals from adhered soil

Contamination of food with small amounts of soil may affect its mineral content.

Dietary standards and nutrient requirements (after Webb, 1994)

ORIGINS OF DIETARY STANDARDS

Almost 60 years ago, the National Research Council in the USA established a committee whose brief was to produce a comprehensive set of dietary standards. These standards were intended to be a yardstick against which diets or food supplies could be assessed to determine their likely adequacy. The first printed version of these standards appeared in 1943 and they were called Recommended Dietary Allowances (RDAs). These comprised the first official and comprehensive set of dietary standards. Many other governments and international agencies now regularly publish their own dietary standards, and the American RDAs have been revised and republished regularly since 1943. The first edition of these American RDAs (NRC, 1943) covered just six pages and dealt with only ten nutrients; the current British and American versions of these standards (COMA, 1991, and NRC, 1989a, respectively) each covers more than 200 pages and deals with more than 30 nutrients.

The first British RDAs were published by the British Medical Association in 1950. In 1969, the first truly official set of UK standards was published by the then Department of Health and Social Security (see Webb, 1994).

	Nutrient 1	Nutrient 2	Nutrient 3	Nutrient 4.....etc.
Age group (examples)				
0–3 months				
7–10 years				
11–14 years(female)				
19–50 years(male)				
50+ years(female)				
Pregnant				
Lactating				

Figure 3.2 A plan of the layout of tables of dietary standards (e.g. RDAs or RNIs).

DEFINITIONS AND EXPLANATIONS

For the purposes of setting dietary standards, the population is divided up into subgroups. Children are divided up into bands according to their age and sex. Adults are subdivided according to their age and sex, with separate standards for pregnant and lactating women. Then standards are set for energy and each nutrient for each of these population subgroups (Figure 3.2). These standards are intended for use with healthy people and they make no allowance for the effects of illness and injury upon nutrient needs.

The RDA in the USA is the suggested average daily intake of that nutrient for healthy people. It represents the best estimate of the requirement of those people in the population with a particularly high need for that nutrient. The RDA does not represent the minimum requirement when it is used to assess the diets of individuals. Rather, it should be thought of as lying within a 'zone of safety'; the further intake is below the RDA, then the greater is the risk of deficiency, and the further above the RDA, then the greater is the risk of toxic effects.

Until 1991, the British standards were also termed RDAs and this term is still to be found on British food labels, where it refers to the RDA set by the European Union. However, in the latest version of the British standards (COMA, 1991), the general term Dietary Reference Values (DRVs) is used to cover a range of differently defined values. The word 'recommended' has been specifically avoided as it was felt to wrongly imply that the RDA represented the minimum desirable intake for health and thus that intakes below the RDA represented inadequacy. Instead of a single RDA, three reference values are offered for protein, vitamins and minerals in these new British standards. The highest of these three values is called the **Reference Nutrient Intake (RNI)**. It is essentially equivalent to the old RDA, as it also represents the estimated

requirement of those people with the highest need for the nutrient. In practice, it is still the value that is used in most circumstances. The other two DRVs offered for these nutrients are the **Estimated Average Requirement (EAR)**, which is self-explanatory, and the **Lower Reference Nutrient Intake (LRNI)** (see Dietary Reference Values in Glossary). The LRNI is the best estimate of the requirement of those individuals with a low need for the nutrient. The requirement of almost everyone should lie within the range covered by the LRNI and the RNI.

COMA (1991) assumed that the variation in nutrient requirements of individuals follows a **normal distribution** (Figure 3.3). Approximately half of the population should require more and half less than the EAR. The **standard deviation** is a precisely defined statistical measure of the variation of individual values around the mean or average in a normal distribution.

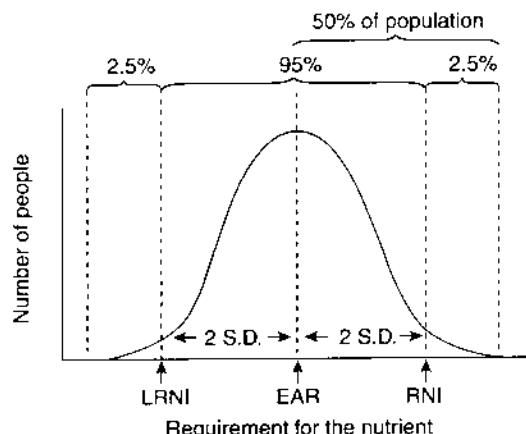


Figure 3.3 A normal distribution of individual nutrient needs within a population with the theoretical positions of the UK Dietary Reference Values. LRNI, Lower Reference Nutrient Intake; EAR, Estimated Average Requirement; RNI, Reference Nutrient Intake; S.D., standard deviation. From Webb (1994).

The RNI is set at a notional two of these standard deviations above the mean and the LRNI a notional two standard deviations below the mean (Figure 3.3). The characteristics of a normal distribution mean that the requirements of all but 5% of the population should lie within the range covered by two standard deviations on either side of the mean. Thus, the RNI and LRNI should theoretically satisfy the needs of 97.5% and 2.5% of the population, respectively. The RNI should represent an amount sufficient, or more than sufficient, to satisfy the needs of practically all healthy people (essentially the same definition as the American RDA). The panel considered that the RNI would be sufficient for everyone despite the theoretical risk that 2.5% of the population are not provided for. There is considerable lack of precision in most of the estimations used to set dietary standards and so they tend to be set generously.

COMA (1991) suggested that this more complex system of standards would allow more meaningful interpretation of measured or predicted intakes that fall below the RNI (i.e. the old RDA). At the extremes, if an individual's habitual intake is below the LRNI, then the diet is almost certainly not able to maintain adequacy as it has been defined for that nutrient. If intake is above the RNI, then it is safe to assume that the individual is receiving an adequate supply. In between these two extremes, the chances of adequacy fall, to a statistically predictable extent, as the intake approaches the LRNI, e.g. an individual consuming the EAR has a 50% chance of adequacy. Note that the iron requirements of women are an example of requirements for a nutrient being known not to be normally distributed; high iron losses in menstrual blood in some women skew or distort the distribution.

The RDAs for energy have traditionally been set at the best estimate of *average* requirement, rather than at the upper extremity of estimated requirement. This is true for the current energy RDA in the USA. Whereas for most nutrients a modest surplus over requirement is not considered likely to be detrimental, this is not so with excessive energy intake, which may lead to obesity. Consistent with this traditional practice, the latest UK standards give only an EAR for energy. Any individual's energy requirement will depend upon many factors, particularly on the person's size and activity level, and appetite should ensure that adequate intakes of energy are consumed by healthy non-dieting subjects.

The COMA panel considered that, for eight nutrients, they did not have sufficient information to estimate the rather precisely defined set of values discussed above. In these cases, therefore, they merely suggested a **safe intake** – 'a level or range of intakes at which there is little risk of either deficiency or toxic effects' (see Dietary Reference Values in Glossary). The **National Research Council (NRC)** panel in the USA also used this term.

The latest UK standards also, for the first time, set reference values for the various fat and carbohydrate fractions, including dietary fibre. These values are clearly directed towards reducing the risk of chronic disease. This breaks new ground for these standards because their traditional function has been solely to set standards of nutritional adequacy. COMA (1991) attempted to integrate the functions of standards of adequacy and of nutritional guidelines aimed at reducing the risk of chronic disease. The reference values for fat and carbohydrate fractions are discussed in Chapter 4.

The formal and precise statistical definitions used by COMA (1991) should not be allowed to obscure the fact that judgement plays a major part in the setting of these values. The criterion that any dietary standards committee uses to define adequacy is almost inevitably a matter of judgement and opinion. There may also be considerable errors and uncertainties in estimating the average amount required to satisfy this criterion and in estimating the standard deviation of requirement.

Any particular reference value is the consensus view of one panel of experts based upon the information available to them and in the prevailing social, political and economic climate. There may be genuine differences of opinion among different panels of experts and probably even between the experts within any given panel. Any particular panel's reference values are likely to be the result of compromises between the differing views of its individual members. Some non-scientific considerations may also influence different committees to differing extents. Thus, the dairy industry might lobby for the calcium standard to be high, fruit growers lobby for a high vitamin C standard, and meat and fish suppliers may be keen to keep the protein standard high. Where fruit is cheap and plentiful, a committee may be generous with vitamin C allowances because the cost of erring on the high side is perceived as minimal. However, where fruit is, for much of the year, an expensive

Table 3.2 A comparison of some current UK RNIs for an adult man, with their equivalent US RDAs (US value as a percentage of UK value)

	UK RNI	US RDA
Energy (kcal) (MJ)	2900 (114)	2550*
Protein (g)	63 (114)	55.5
Vitamin A (µg)	1000 (143)	700
Vitamin C (mg)	60 (150)	40
Riboflavin (mg)	1.7 (131)	1.3
Iron (mg)	10 (115)	8.7
Calcium (mg)	800 (114)	700

* EAR.

Data sources: COMA (1991) and NRC (1989a).

imported luxury, a panel may be more conservative in their vitamin C standard.

There have been some very significant variations in these dietary standards over the years and there are still quite marked variations in the dietary standards used in different countries. Historical differences can be partly explained by the differences in scientific information available to different panels, but current international differences are largely due to variations in the way the same information has been interpreted by different panels of experts. Comparison of the current US RDAs and UK RNIs, which are essentially equivalent, reveals numerous and sometimes substantial differences. In general, US RDAs are higher than UK RNIs; some examples are shown in Table 3.2. For example, the US RDAs for vitamins A and C are around half as much again as the UK RNIs. In Chapter 10, the huge historical variations in protein standards for children are reviewed and discussed. In Chapter 14, the large differences in the current UK and US standards for pregnant women are highlighted and discussed.

THE USES OF DIETARY STANDARDS

A set of dietary standards enables nutritionists to predict the nutritional requirements of groups of people or of nations. Governments and food aid agencies can use them to identify the needs of populations, to decide whether available supplies are adequate for the population to be fed, and to identify nutrients whose supply is deficient or only

marginally adequate. They thus provide the means to make informed food policy decisions, for example decisions about:

- the amount and type of food aid required by a population or group
- the priorities for agricultural production and food imports
- which foods might be beneficially subsidized or, if foods need to be rationed, the size of rations that are needed.

Similarly, they are used by institutional caterers (e.g. in prisons or schools) and those catering for the armed forces to assess the food requirements of their client population and also to check proposed menus for their nutritional adequacy. Those devising therapeutic or reducing diets can check any proposed diets for adequacy using these standards.

These standards provide the means to assess nutritional adequacy after an intake survey. They are ideally used as a yardstick for assessing the adequacy of groups rather than individuals. They can nonetheless also be a useful tool for evaluating individual intakes, provided it is recognized that the RDA or RNI is not a minimum requirement but a generous estimate of the requirement of those with a particularly high need. When assessing the nutritional adequacy of individuals, it should be remembered that there are very substantial fluctuations in the intakes of some nutrients from day to day. The habitual intake of the individual over a period of some days should ideally be the one that is compared to the standard.

Most consumers will be exposed to these standards when they are used on food labels to give a meaningful indication of nutrient content. Absolute numerical amounts of nutrients will be meaningless to most consumers, but when expressed as a percentage of the RDA, they become more meaningful. COMA (1991) suggested that the EAR was the most appropriate reference value to use for food labelling in Britain. As Britain is a member of the EU, the RDAs seen on British food labels are actually set by the European Directive on food labelling and these differ slightly from the recommendations in the UK DRVs. In America, a Reference Daily Intake (RDI) is used to indicate essential nutrient content on food labels. It is set at the higher of the two adult RDA values and referred to on the label simply as the 'daily value'.

INACCURATE STANDARDS

Setting dietary standards too low will negate their purpose. A yardstick of adequacy that does not meet the criterion of adequacy is of no use and is probably worse than no standard at all, because it will tend to induce a false sense of security and complacency. A serious underestimate of the standards for a nutrient could result in a nutrient deficiency being falsely ruled out as the cause of some pathology. Perhaps more probably, it could result in a suboptimally nourished group being reassured as to their dietary adequacy.

Such arguments about the obvious hazards of setting standards too low means that there will be a strong temptation to err on the side of generosity when setting them. This may be especially true in affluent countries, where the need to avoid waste of resources is less acute. There are, however, potentially adverse consequences that may result if these standards are set too high, and some of these are listed below.

- Some nutrients (e.g. iron and some fat-soluble vitamins) are toxic in excess. An unrealistically high standard might encourage sensitive individuals to consume hazardous amounts.
- High standards may encourage the production or supply of unnecessary excesses, which will result in the wasteful use of resources.
- High standard values may create the illusion of widespread deficiency and result in unnecessary and wasteful measures to combat an illusory problem.
- Unreasonably high 'target' values for particular nutrients may lead to distortion of the diet, with deleterious effects on the intake of other nutrients.
- If an unreasonably high standard results in the classification of large numbers of people as deficient, but there are no other manifestations of deficiency, this may discredit the standards and may, in the longer term, undermine the credibility of nutrition education in general.

Some of the consequences of earlier exaggeration of human protein requirements are discussed in Chapter 10 and this may serve as a case study to illustrate most of these points.

DEFINING REQUIREMENT

The first problem that has to be confronted when devising a set of dietary standards is to decide upon

the criteria that will be used to define adequacy. Ideally, one would like to determine the intake that maximizes growth, health and longevity, but these are not a readily measurable set of parameters. Overt deficiency of a nutrient will often produce a well-defined deficiency syndrome, so that the minimum requirement will be an intake that prevents clinical signs of deficiency. Optimal intakes are assumed to be some way above this minimum requirement. Subclinical indications of impairment or depletion of nutrient stores may occur long before overt clinical signs of deficiency. COMA (1991) decided that the EAR should allow for 'a degree of storage of the nutrient to allow for periods of low intake or high demand without detriment to health'. This is actually a rather loose (vague) definition and the committee had to translate this into a quantitative and measurable criterion case by case.

This problem of defining an adequate intake is well illustrated by the example of vitamin C. It is generally agreed that, in adults, around 10 mg/day of this vitamin is sufficient to prevent the deficiency disease **scurvy**. Below 30 mg/day, negligible levels of the vitamin are detectable in plasma; at intakes of between 30 and 70 mg/day, plasma vitamin levels rise steeply, and they start to plateau at intakes of between 70 and 100 mg/day. COMA (1991) chose an adult RNI of 40 mg/day because, at this intake, most individuals would have measurable amounts of vitamin C in their plasma that is available for transfer to sites of depletion. Using essentially the same available data, NRC (1989a) chose an adult RDA of 60 mg/day. With an intake of 60 mg/day, concentrations in the plasma are relatively high and starting to approach the point of saturation. Some people advocate daily intakes of gram quantities of this vitamin in order to maximize resistance to infection. COMA (1991) listed numerous suggested benefits of very high vitamin C intakes and yet none of these influenced their DRVs. The underlying criterion in setting the DRVs (and RDA in the US) is dietary adequacy, even though the additional requirement for adequate stores represents a large safety margin over that required for minimal adequacy, i.e. the prevention of scurvy.

The following sections cover the various methods that have been used to assess nutrient requirements and to set dietary standards for these nutrients. Examples of the use of each approach are also explained. Many readers may find it sufficient to select a small number of nutrients to use as illustrat-

ive case studies of how standards are set. Note that the method used to set the EAR (RDA) for energy is explained in Chapter 6.

DEPRIVATION STUDIES

The most obvious and direct way of assessing the minimum requirement for a nutrient is to use an experimental diet that lacks the nutrient and see how much of it needs to be added to this diet to prevent or cure the signs of deficiency. Experiments in Sheffield, England, during World War II demonstrated that volunteers started to develop signs of the deficiency disease scurvy after a couple of months on a vitamin C-free diet. Intakes of around 10 mg/day were shown to prevent the development of scurvy and to cure the clinical signs (see COMA, 1991). Thus, 10 mg/day is the minimum requirement for vitamin C and COMA chose this as the LRNI for vitamin C.

Deprivation studies of this type may require the consumption of very restricted diets for long periods before clinical signs of deficiency develop in previously well-nourished adults. For example, it takes up to 2 years of depletion before such adults develop even limited signs of vitamin A deficiency. This is because their livers contain large amounts of stored vitamin A. The volunteers need to consume a very restricted and unpalatable diet for a long time and this may also have some adverse consequences upon their long-term health. It would be ethically unacceptable to deliberately subject vulnerable groups to this type of deprivation, e.g. children and pregnant or lactating women. This means that the extra requirements of these groups usually have to be inferred from less direct methods.

As an alternative to controlled experiments, it is possible to use epidemiological data relating average intakes of a nutrient to the presence or absence of clinical signs of deficiency in populations. For example, it has been observed that the deficiency disease beriberi occurs when the average population intake of thiamin falls below 0.2 mg/1000 kcal (0.2 mg/4.2 MJ), but when average intake is above this level, beriberi does not occur (Passmore and Eastwood, 1986).

RADIOACTIVE TRACER STUDIES

If a known amount of radioactively labelled vitamin, or other nutrient, is administered to a volunteer, then, assuming that this labelled vitamin disperses

evenly in the body pool of that vitamin, the dilution of the radioactivity can be used to estimate the total size of that body pool. A sample of plasma is taken and the amount of vitamin and the amount of radioactivity in the sample are measured. The **specific activity** of vitamin in the plasma sample (i.e. the radioactivity per unit weight of vitamin) can be used to calculate the total pool size provided that the amount of radioactivity administered is known (see below).

For example:

- administer 1 million units of radioactivity
- specific activity in sample is measured as 1000 units of radioactivity per milligram of vitamin
- so radioactivity has been diluted in 1000 mg of vitamin, i.e. in a 1000-mg body pool of the vitamin.

If, after the administration of a radioactive vitamin load, the body losses of radioactivity are monitored, this will enable the rate of vitamin loss or depletion to be determined.

Using this approach, Baker *et al.* (1971) found that the average body pool of vitamin C in a group of healthy, well-nourished American men was around 1500 mg. On a vitamin C-free diet, this pool depleted at a rate of around 3% per day. This 3% depletion rate is termed the **fractional catabolic rate** and was found to be independent of the pool size, i.e. 3% of whatever is in the body is lost no matter how much or how little is present in the body. When the body pool fell below 300 mg, symptoms of scurvy started to appear. Baker *et al.* estimated that, in order to maintain the body pool above 300 mg and thus to prevent scurvy, their subjects needed to take in 9 mg/day (i.e. 3% of 300 mg). This agrees very well with the approximately 10 mg/day to prevent scurvy found in the much earlier depletion study in Sheffield.

BALANCE STUDIES

These methods rely upon the assumption that, in healthy, well-nourished adults of stable body weight, the body pool size of some nutrients remains constant. Healthy adults are, for example, in approximate daily balance for nitrogen (i.e. protein), calcium and sodium. Over a wide range of intakes and a suitable measurement period, the intake is approximately equal to the output. Any variations in intake are compensated for by changes in the rate of

absorption from the gut, changes in the rate of excretion, or changes in the rate of metabolism. If, say, calcium intake is progressively reduced, initially losses of calcium in urine and faeces will also decline and balance will be maintained. Eventually, however, a point will be reached when balance can no longer be maintained and output starts to exceed input. It would seem reasonable to propose that the minimum intake at which balance can be maintained represents the subject's minimum requirement for calcium. Such short-term experiments do not exclude the very real possibility that long-term adaptation to chronically low calcium intakes will occur.

COMA (1991) assumed that the average daily loss of calcium via urine and skin in British adults was 160 mg/day. In order to replace this daily loss, they estimated that an intake of 525 mg/day would be required, assuming that around 30% of dietary calcium is absorbed. They chose this as the adult EAR for calcium. They added or subtracted 30% to allow for individual variation and therefore came up with 700 mg/day and 400 mg/day as the RNI and LRNI, respectively.

FACTORIAL METHODS

Factorial calculations are essentially predictions of the requirements of particular groups or individuals, taking into account a number of measured variables (or factors) and making a number of apparently logical assumptions. For example, during growth or pregnancy, certain nutrients will be retained and accumulate in the growing body or pregnant woman. Knowing the rate at which these nutrients accumulate during pregnancy or growth, one can then make predictions of the amount required (see the example below):

Estimated requirement for pregnancy = amount to achieve balance (from value for non-pregnant women) + [daily accumulation rate for nutrient during pregnancy × factor to allow for assumed efficiency of absorption and assimilation]

COMA (1991) predicted the increase in EAR for energy of women during lactation using the following factorial calculation:

Increase in EAR for energy during lactation =
average energy content of daily milk production
 $\times 100/80$ (assuming 80% conversion of dietary energy to milk energy) — an allowance for the contribution from the extra maternal fat stores laid down during pregnancy

It should always be born in mind that, no matter how logical they may seem, such values are theoretical predictions and they may not represent actual physiological need. Physiological adaptations may occur that will reduce the predicted requirement, e.g. the efficiency of calcium absorption from the gut increases during pregnancy (see Chapter 14).

MEASUREMENT OF BLOOD OR TISSUE LEVELS

COMA (1991) defined some reference values according to the intake required to maintain a particular circulating level or tissue level of the nutrient. As we have already seen, the LRNI for vitamin C is set at the intake that prevents scurvy (10 mg/day in adults); the RNI is set at a level that maintains a measurable amount of vitamin C in plasma in most adults (40 mg/day); the EAR (25 mg/day) is set halfway between the LRNI and the RNI.

The reference values for vitamin A in the UK are based upon the intake that is estimated as necessary to maintain a liver concentration of 20 µg vitamin A per gram of liver. In order to estimate the intake of vitamin required to maintain this target liver concentration, the panel had to perform quite an elaborate factorial calculation (summarized in Figure 3.4).

First, they had to predict the size of the body pool required to achieve this liver concentration. To do this they had to make assumptions about what proportion of the body is liver and also about how the total body pool of vitamin A partitions between the liver and other tissues. The fractional catabolic rate of vitamin A has been measured at 0.5% of the body pool lost per day, and so an amount equivalent to 0.5% of this estimated pool would have to be replaced each day. Finally, assumptions had to be made about the efficiency with which ingested vitamin A is stored in the liver in order to convert this replacement requirement into an intake requirement.

BIOCHEMICAL MARKERS

COMA (1991) used the intake required to 'maintain a given degree of enzyme saturation' as another

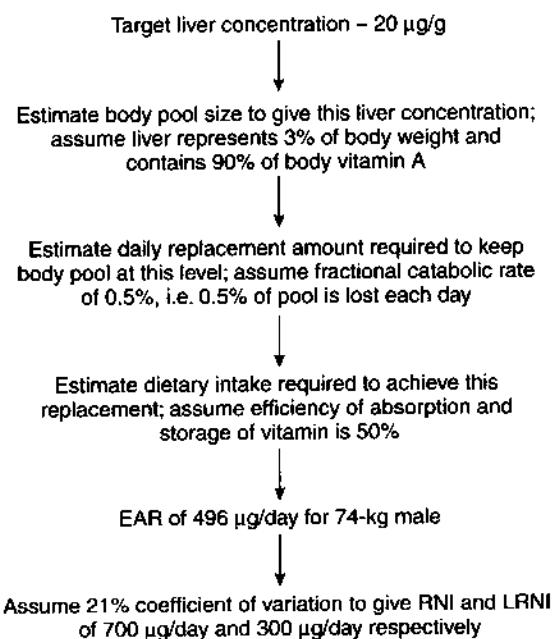


Figure 3.4 A scheme to illustrate the calculations and assumptions required to estimate the vitamin A intake necessary to maintain a designated liver concentration and thus to set the Dietary Reference Values for vitamin A. After COMA (1991).

criterion for determining reference values. An example of this is the use of the **erythrocyte glutathione reductase activation coefficient** (EGRAC) to define nutritional status for riboflavin (vitamin B₂).

Glutathione reductase is an enzyme present in red blood cells whose activity is dependent upon the presence of a cofactor **flavin adenine dinucleotide** (FAD) that is derived from riboflavin. The enzyme cannot function in the absence of the cofactor. In riboflavin deficiency, the activity of this enzyme is low because of reduced availability of the cofactor. In red blood cells taken from well-nourished subjects, the activity of this enzyme will be higher because it is not limited by the availability of the cofactor. To perform the activation test, the activity of glutathione reductase is measured in two samples of red cells from the subject – one has had excess FAD added, the other has not had FAD added. The ratio of these two activities is called the **erythrocyte glutathione reductase activation coefficient**. It is a measure of the extent to which enzyme activity has been limited by riboflavin availability in the unsupplemented sample and thus is a measure of the subject's riboflavin status. The RNI is set at the intake that maintains the EGRAC at 1.3 or less in almost all people.

Similar enzyme activation tests are used to assess status for thiamin (vitamin B₁) and vitamin B₆:

- activation of the enzyme transketolase in red cells is used to determine thiamin status – a thiamin-derived cofactor is necessary for transketolase to function
- activation of the enzyme glutamic oxaloacetic transaminase in erythrocytes can be used to assess vitamin B₆ status.

BIOLOGICAL MARKERS

Blood haemoglobin concentration has been widely used in the past as a measure of nutritional status for iron. It is now regarded as an insensitive and unreliable indicator of iron status for reasons such as those listed below.

- Haemoglobin levels change in response to a number of physiological factors such as training, altitude and pregnancy.
- Iron stores may be depleted without any change in blood haemoglobin concentration. (See Chapter 13 and 'Biochemical assessment' in this chapter for further discussion of iron status assessment.)

Vitamin K status is frequently assessed by functional tests of prothrombin levels in blood. Prothrombin is one of several clotting factors whose synthesis in the liver depends upon vitamin K as an essential cofactor. Thus, in vitamin K deficiency, prothrombin levels fall and blood clotting is impaired. In order to measure the **prothrombin time**, excess calcium and tissue thromboplastin are added to fresh plasma that has been previously depleted of calcium to prevent clotting. The time taken for the plasma to clot under these conditions is dependent upon the amount of prothrombin present and thus upon the vitamin K status of the donor. Anticoagulant drugs, such as warfarin, work by blocking the effect of vitamin K. Prothrombin time is thus a useful way of monitoring vitamin K status and of regulating drug dosage during anticoagulant therapy.

ANIMAL EXPERIMENTS

Animal experiments are of limited value in quantifying the nutrient needs of human beings. They may even encourage widely erroneous estimates to be made. It is extremely difficult to allow for species differences in nutrient requirements and to scale between species as different in size as rats and people.

The examples below illustrate some of the difficulties of predicting human nutrient needs from those of laboratory animals.

Most rapidly growing young animals need a relatively high proportion of their dietary energy as protein, but human babies grow more slowly and thus are likely to need proportionally less than most other young mammals. Rat milk has around 25% of its energy as protein, compared to only about 6% in human milk. Predicting the protein needs of human children from those of young rats is likely to exaggerate the needs of children.

Pauling (1972) used the measured rate of vitamin C synthesis in the rat (which does not require dietary vitamin C) to support his highly controversial view that gram quantities of the vitamin are required for optimal human health. He scaled up the rat's rate of production on a simple weight-to-weight basis and estimated that rats of human size would make 2–4 g/day of the vitamin. He suggested that this gave an indication of human requirements. This procedure seems extremely dubious on several grounds, for example the decision to scale up the rat's rate of vitamin C synthesis on a simple weight-to-weight basis. Vitamin needs may be more related to metabolic needs than to simple body weight. If one scales according to relative metabolic rate, one might predict that the human size rat would only make around a quarter of the amount predicted by body-weight scaling.

The expected nutritional burdens of pregnancy and lactation are also relatively much greater in small laboratory animals than in women. Laboratory animals have relatively larger litters, short gestations and rapidly growing infants when compared to human beings. Extrapolating from laboratory animals is thus likely to exaggerate any extra nutritional requirements of pregnant and lactating women (see Chapter 14).

Despite these reservations about the use of animal experiments to quantify human nutrient needs, they have played a vital role in the identification of the essential nutrients and their physiological and biochemical functions. Several of those awarded Nobel prizes for work on the vitamins used animals in their work. The need for essential fatty acids was, for example, demonstrated in the rat 40 years before unequivocal confirmation in an adult human being.

Animal experiments may also be very useful in providing in depth information on the pathological

changes that accompany prolonged deficiency and in determining whether prolonged marginal adequacy is likely to have any long term detrimental effects.

Clinical signs for the assessment of nutritional status

Nutrient deficiencies ultimately lead to clinically recognizable deficiency diseases. Identification of the clinical signs of these deficiency diseases usually requires little or no specialized equipment, is cheap, simple and quick, thus enabling assessment surveys to be conducted rapidly and cheaply, even in the most inaccessible places. Even non medical personnel can be trained to conduct clinical surveys of nutritional status by recording the presence or absence of various clinical signs from checklists of clinical signs that are likely to be associated with nutrient deficiencies. Some of these clinical signs that may indicate a nutritional deficiency are listed in Table 3.3.

The potential problem of population sampling bias (as mentioned in Table 3.1) is one that can affect any type of dietary or nutritional survey. Those conducting the survey must take steps to actively ensure that those assessed are representative of the whole population under investigation. Those most badly affected by deficiency diseases may well be those least accessible to the survey team – the weak, the frail, the elderly, pregnant and lactating women and babies are the least likely to be out and about. Sampling people in the street or those people who are mobile enough to attend a centre may well overestimate the nutritional status of the population. This may be particularly important when clinical signs are used, because these are often present only in those most severely affected by dietary deprivation.

The clinical signs of deficiency usually become recognizable only after severe and prolonged deficiency and thus they are relatively insensitive indicators of nutritional status. It is generally assumed that suboptimal intakes of nutrients produce subclinical impairment of physiological functioning long before any deficiency disease becomes clinically apparent. In surveys in affluent countries, very few cases of clinical deficiency would be found in most sectors of the population. Surveys that use clinical signs would therefore tend to provide only limited information about the nutrient status of the population, e.g. that there are no

- Dietary standards are yardsticks of nutritional adequacy, called Recommended Dietary Allowances (RDAs) in the USA and Dietary Reference Values (DRVs) in the UK.
- Reference values are set for each nutrient for each of the various age and sex groups within the population.
- The RDA in the USA and the Reference Nutrient Intake (RNI) in the UK represent an estimate of the requirements of those healthy individuals in the age group with the highest need.
- The RNI is set at a notional two standard deviations above the Estimated Average Requirement (EAR).
- The Lower Reference Nutrient Intake (LRNI) is set at two standard deviations below the EAR.
- The LRNI is the estimated requirement of those with the lowest need for the nutrient and thus is assumed to be insufficient to meet most people's needs.
- The standards for energy are set at the EAR (called the RDA in the USA).
- The setting of dietary standards depends upon the judgement of panels of experts and so they vary from country to country and over time.
- Reference values are very dependent upon the definition of adequacy used.
- COMA (1991) defined the EAR as an amount that would prevent deficiency and allow some degree of storage.
- These standards can be used to:

assess whether food supplies or nutrient intakes are adequate
estimate the needs of groups or populations
check whether menus or prescribed diets are adequate.

- The use of RDAs on food labels makes nutrient contents more meaningful to consumers.
- On labels in the USA, the higher of the two adult RDAs is present as the 'daily value', whereas in the UK, the RDA set by the EU is used.
- Standards that are set too low are obviously of no use, but setting values too high may also produce other, more subtle, problems.
- Many different methods can be used to define and determine what is an adequate intake of a particular nutrient, such as:

direct measurement of the amount needed to cure or prevent deficiency
the use of a radioactively labelled nutrient to measure body pool size and fractional catabolic rate or rate of excretion
estimation of the amount necessary to maintain balance between input and losses of the nutrient
factorial estimates of the amount of nutrient required, e.g. to produce milk in lactation to support growth or growth of the products of conception in pregnancy
estimates of the amount necessary to achieve a particular blood or tissue level, a specified level of enzyme activity or some biological marker.

- Animal studies have been invaluable in identifying essential nutrients and the effects of deficiency, but are of little value in determining human nutrient requirements.

deficiencies severe enough or long standing enough to induce overt clinical deficiency. Clinical signs are therefore not useful as early indicators of nutritional problems that can warn of the need to implement preventative measures. They may be much more useful in

providing a quick guide to the extent of nutritional deficiency in less affluent countries, during famine and even in high-risk groups in affluent countries, e.g. the homeless, alcoholics, underprivileged children and the frail elderly.

Table 3.3 A list of some clinical signs that may indicate a nutritional deficiency

Loose, hanging clothes or a wasted appearance (energy/protein deficit and weight loss)
Loss of hair pigment and easy 'pluckability' (energy/protein deficit in children? – kwashiorkor)
White foamy spots on the cornea (Bitot's spots – vitamin A)
Dry and infected cornea (vitamin A)
Oedema (thiamin, B ₁ – beriberi)
Several types of dermatitis (in skin exposed to sunlight, niacin – pellagra)
Enlargement of the liver (energy/protein deficit – kwashiorkor)
Loss of peripheral sensation (thiamin)
Spongy, bleeding gums (vitamin C – scurvy)
Angular stomatitis – spongy lesions at the corners of the mouth (riboflavin)
Pale conjunctiva (iron-deficiency anaemia)
Red, inflamed tongue (riboflavin)
Spontaneous bruising (vitamin C)
Tiny subdermal haemorrhages or petechiae (vitamin C)
Swelling of the thyroid gland in the neck or goitre (iodine)
Bowed legs (vitamin D – rickets)
Mental confusion (water – dehydration)
Confusion or dementia (niacin)

One possible dietary cause is given in brackets.

Note that several of these symptoms can be indicative of more than one nutritional deficiency or they may indicate other non-nutritional conditions. They can be profitably used to construct screening checklists, which can be used to identify individuals in need of further investigation or to assess the level of nutrient deficiency in a population.

Clinical signs tend to be qualitative and subjective. Any attempt to grade or quantify clinical signs is likely to depend upon subjective judgements on the part of the operator. For example, grading the severity of **goitre** (swelling of the thyroid gland due to iodine deficiency) depends upon the judgement of the assessor about the degree of thyroid enlargement; different clinicians may produce considerably different grades for the same population.

Clinical signs are not very specific indicators of nutrient deficiencies. Some symptoms are common to several deficiency diseases and also to non-nutritional causes. Some form of dermatitis is, for example, common to several deficiency diseases and may also be induced by a variety of non-nutritional causes. Oedema may be a symptom of beriberi, protein energy malnutrition, heart failure, kidney disease, etc.

Anthropometric assessment in adults

Table 3.4 gives a list of some of the anthropometric measures that may be used to assess nutritional status. Forbes (1999) has reviewed methods of measuring human body composition, including some methods involving the use of modern high-technology apparatus, which are not discussed here. This referenced review also provides a good entry into the specialist literature for readers who may wish to find out more about particular methods.

USES OF ANTHROPOMETRIC ASSESSMENT

Anthropometry is the scientific study of the measurement of the human body. Anthropometric assessment means making nutritional assessment by means of physical measurements of body weight and dimensions. Body composition may be estimated from anthropometric measurements and this can have a variety of uses:

- It allows one to make an assessment of nutritional status. Anthropometric measurements can be used to detect undernutrition or obesity in adults and to indicate whether the growth of children has been satisfactory.
- It enables one to make more useful comparisons of metabolic rate between individuals or groups. Adipose tissue has a low metabolic rate and so expressing metabolic rate per unit of lean body mass is more meaningful than simply expressing it per unit total weight.
- Certain drug dosages may be calculated per unit of lean body weight.
- Changes in body weight may be due to gain or loss of water, lean tissue and/or fat. Longitudinal measurements of body composition may help to decide the composition of any weight change. Most of the measures currently available for assessing body composition are relatively insensitive and thus could only be reliably used for this purpose if the weight change were substantial.

HEIGHT AND WEIGHT

Body weight alone can only really be a useful indicator of nutritional status in adults if it is measured repeatedly over a period of time. It can then, for example, be used to monitor changes in the nutritional status of patients in hospital or those living in

Table 3.4 Some anthropometric measures used for nutritional assessment**3 months' unintentional weight loss**

Up to 5%	Mild depletion
5–10%	Moderate depletion
Over 10%	Severe depletion

Useful for screening new admissions to hospital or residential care homes

Body mass index (BMI)

$BMI = \frac{\text{Weight (kg)}}{\text{Height (m)}^2}$	
Under 20 kg m^{-2}	Underweight
$20\text{--}25 \text{ kg m}^{-2}$	Ideal
$25\text{--}30 \text{ kg m}^{-2}$	Overweight
$30\text{+} \text{ kg m}^{-2}$	Obese
$40\text{+} \text{ kg m}^{-2}$	Severely obese

The standard way of classifying people on the basis of their height and weight

Demiquot index

$$\frac{\text{Weight (kg)}}{\text{Demi-span (m)}^2}$$

Alternative to BMI for people who cannot stand erect, e.g. the elderly; alternatively, demi-span or knee height can be used in regression equations to predict height and then BMI can be estimated

Measures of mid-upper arm circumference

Mid-arm circumference (MAC)

Useful as a simple indicator of nutritional status when compared to standards or as an indicator of weight change when measured sequentially; can be used for those confined to bed

Mid-arm muscle circumference (MAMC) = MAC – ($\pi \times$ triceps skinfold)

Indicator of lean body mass or to monitor changes in it

$$\text{Arm muscle area} = \frac{(\text{MAMC})^2}{4\pi}$$

Skinfold calipers

Measure skinfold thickness at several sites to indicate the amount of fat stored subcutaneously in these sites and then translate into an estimate of percentage body fat with a calibration chart or table

A more direct assessment of body fat than weight; particularly useful for athletes who have unusually high amounts of muscle for their weight

Bioelectrical impedance

Pass a weak alternating current between two points on the body and the impedance (resistance) depends upon the relative amounts of fat and lean in the body; not strictly an anthropometric method but included here for convenience

Gives a rapid and direct estimate of body fat using an internal calibration, with little technical skill required

Body density

Fat is less dense than lean tissue: if one assumes that fat has a density of 0.9 kg/L and lean tissue 1.1 kg/L, then it is possible to estimate the percentage of body fat from the body density

Not suitable for routine use; traditionally, the method used for calibration of other methods of measuring fatness, e.g. skinfold measurements and bioelectrical impedance; volume is usually measured by underwater weighing, which is technically difficult and requires specialist facilities

residential homes for the elderly. Unintentional weight loss can be an indicator of disease or deteriorating nutritional status as a result of adverse socio-economic status. Many hospital admission forms have a question about weight change in the previous 3 months (see Table 3.4) designed to help identify patients who are nutritionally 'at risk'.

An individual's 'ideal' weight will obviously be very dependent upon how tall he or she is; one would expect tall people to weigh more than short ones. In the past, weight-for-height tables were the most common way of assessing people's weight status. These tables list the desirable weight ranges for men and women of any particular height and age. Three desirable ranges would usually be given for each height for each sex depending upon frame size, i.e. whether the person is of light, medium or heavy build. Obesity was then defined as more than 20% over the person's ideal weight range and underweight defined as more than 10% below their ideal weight range. The tables produced by the Metropolitan Life Insurance Company have been widely used for this purpose. Note that the original purpose of these tables was for commercial use in assessing actuarial risk in people taking out life insurance policies. They were produced by recording the heights and weights of large numbers of life insurance applicants and then relating initial weight to risk of dying (and thus claiming on the policy) in the succeeding years. These ideal ranges were those associated with the lowest death rates. People above or below these ranges (i.e. underweight or overweight) had higher death rates.

These tables are relatively cumbersome and inconvenient to use, especially as they require some assessment of frame size to be made. There are objective measures of frame size, such as wrist circumference and elbow breadth, which are not difficult to measure in a clinic or laboratory, but they add to the inconvenience of weight-for-height tables, particularly for studies with large groups or populations.

THE BODY MASS INDEX

In recent years, the **body mass index (BMI)** has become the standard way of classifying adults using their heights and weights. This measure has been shown empirically to be the best simple and quantitative anthropometric indicator of body composition and thus of nutritional status.

$$\text{BMI} = \frac{\text{body weight (kg)}}{\text{height (m)}^2}$$

Note that BMI can be calculated from heights and weights in Imperial units using the following approximation:

$$\text{BMI} = \frac{\text{weight (pounds)}}{\text{height (inches)}^2} \times 705$$

The normal range for the BMI is set at 20–25 kg/m²; any value significantly below this range is taken to indicate underweight and values above this range would be considered to indicate varying degrees of overweight or obesity (see Table 3.4).

When the BMI is being used to classify people as underweight or too fat, then it is being implicitly assumed that variations in weight for a given height are primarily due to variations in adiposity, i.e. that heavy means fat and that light means lean. This assumption generally holds reasonably well, but there are some circumstances in which it clearly does not hold, such as those listed below.

- *Body builders*: people with very well-developed musculature, such as body builders, will be heavy but have very low fat content. In this case, their BMI may falsely indicate obesity. In the past, some very muscular American footballers were rejected for military service because they were heavy and so mistakenly classified as obese.
- *The elderly*: as people become elderly, so the mass of muscle and bone in the body tends to fall and is replaced by fat. A group of elderly people will have a higher average body fat content than a group of young adults of the same height and age. In elderly women, in particular, loss of lean tissue and bone mass may mean that BMI remains within the normal range even though more direct measures of fatness may indicate excess adiposity.
- *Oedema and dehydration*: changes in body fluid content can produce major changes in body weight without any change in body fat content. Excess accumulation of tissue fluid (oedema) can lead to a substantial increase in body weight and may obscure the person's malnourished state. Oedema is a common symptom of malnutrition and of many other conditions, such as heart failure, kidney disease, liver disease and pregnancy. Oedema

may be a major hindrance to the anthropometric assessment of nutritional status in hospital patients and malnourished Third World children if body weight plays any part in that assessment.

ALTERNATIVES TO HEIGHT

Accurate measurement of height requires that the subject be able to stand erect. In subjects who cannot stand erect, some alternative measure of skeletal size is needed. **Demi-span** is one such measure; it is the distance from the web of the fingers (between the middle and ring fingers) and the sternal notch when the subject's arm is held horizontally to the side (with support if necessary). Some disabled people and many very elderly people are unable to stand upright for height to be accurately measured. COMA (1992), in their report on *The nutrition of elderly people*, recommended that demi-span measurements should be included in all nutritional surveys of older people.

White *et al.* (1993) derived the following equations for estimating height (cm) from demi-span (cm) measurements in people aged over 55 years:

$$\text{Height for men} = (1.2 \times \text{demi-span}) + 71$$

$$\text{Height for women} = (1.2 \times \text{demi-span}) + 67$$

This enables BMI to be estimated from measurements of weight and demi span. The Royal College of Nursing (RCN) in Britain has published a set of nutritional standards for elderly people (RCN, 1993), which includes a ready reckoner for estimating BMI from body weight and demi-span. Webb and Copeman (1996) have also compiled a list of anthropometric norms for elderly people. Demi span measurements can also be used to derive indices with body weight, which are analogous to the BMI, the **mindex** and **demiquet index**.

$$\text{Mindex} = \frac{\text{weight (kg)}}{\text{demi-span (m)}}$$

$$\text{Demiquet index} = \frac{\text{weight (kg)}}{\text{demi-span (m)}^2}$$

White *et al.* (1993) found that both mindex and demiquet correlated very highly with BMI, but concluded that mindex was probably a better index of obesity than demiquet.

Knee height is another way of estimating skeletal size that is used less frequently than demi-span. It is measured in a seated position and is the height from the floor to the knee joint space. The following equations can be used to estimate height (cm) from knee height (cm):

$$\text{Male height} = 64.19 + (\text{knee height} \times 2.03) - (0.04 \times \text{age})$$

$$\text{Female height} = 84.88 + (\text{knee height} \times 1.83) - (0.24 \times \text{age})$$

SKINFOLD CALIPERS

The traditional method for more direct estimation of fatness in people is the measurement of skinfold thickness using skinfold calipers. Spring-loaded calipers are used which exert a constant pressure on a fold of skin and the thickness of the skinfold is indicated on a meter. The thickness of the skinfold will be largely dependent upon the amount of fat stored subcutaneously in the region of the skinfold. Skinfold thicknesses are measured at several sites and the assumption is made that the amount of fat stored subcutaneously at these sites (as measured by the skinfold thickness) will be representative of the total amount of body fat. Using the method of Durnin and Womersley (1974), skinfold thickness in millimetres is determined at four sites: over the triceps muscle; over the biceps; in the subscapular region; and in the supra-iliac region. The total of these four skinfolds is then translated into an estimate of percentage body fat using a calibration table or formula. Figure 3.5 shows the relationship between the sum of these four skinfolds and percentage body fat (estimated by body density) in 17–29-year-old men and women, and Table 3.5 shows a table that can be used to estimate percentage fatness from these four skinfolds. The single triceps skinfold thickness is sometimes used in nutritional surveys. It has the obvious advantage, in such circumstances, that it can be measured quickly and without the need for subjects to undress.

Some of the problems associated with use of skinfold calipers are summarized below.

- It is a relatively time-consuming method and subjects need to undress to have their skinfolds measured.
- It requires a great deal of skill and care to obtain accurate and reliable skinfold measurements.

- The process of translating skinfold measurements into estimates of fatness relies upon calibration tables derived from body density measurements and so any errors in the use of density are also inherent in this method.
- There is considerable interindividual variation in the distribution of body fat, as well as pronounced inter-racial differences. Standards need to be appropriate for the racial group being assessed.

BIOELECTRICAL IMPEDANCE

This method of measuring body composition is now being used in many health centres and fitness clubs. It is included in the anthropometry section for convenience. The devices that use this principle are becoming much cheaper and their operation is fairly simple. These devices give an almost instant estimate of the amount of fat and lean in the body and so their use is likely to grow, at least in the short term. **Bioelectrical impedance (BIA)** relies upon the fact that fatty tissue is a much poorer conductor of electricity than lean tissue. Electrodes are placed on one of the subject's hands and feet, a current is generated by the machine's battery in

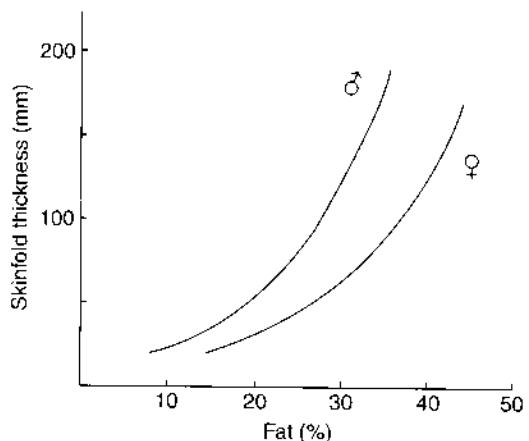


Figure 3.5 Relationship between caliper measurements of skinfold thickness at four sites and body fat content estimated from body density in young adults. Data from Durnin and Womersley (1974); figure reproduced from Webb and Jakobson (1980).

one limb, passes through the body and is picked up by the electrode on the other limb. The machine uses the resistance to the current, or the 'impedance' (because it is alternating current), to estimate lean body mass and

Table 3.5 A table for converting the sum of four skinfolds into estimates of percentage body fat

Skinfolds (mm)	Males (age in years)				Females (age in years)			
	17-29	30-39	40-49	50+	16-29	30-39	40-49	50+
20	8.1	12.2	12.2	12.6	14.1	17.0	19.8	21.4
30	12.9	16.2	17.7	18.6	19.5	21.8	24.5	26.6
40	16.4	19.2	21.4	22.9	23.4	25.5	28.2	30.3
50	19.0	21.5	24.6	26.5	26.5	28.2	31.0	33.4
60	21.2	23.5	27.1	29.2	29.1	30.6	33.2	35.7
70	23.1	25.1	29.3	31.6	31.2	32.5	35.0	37.7
80	24.8	26.6	31.2	33.8	33.1	34.3	36.7	39.6
90	26.2	27.8	33.0	35.8	34.8	35.8	38.3	41.2
100	27.6	29.0	34.4	37.4	36.4	37.2	39.7	42.6
110	28.8	30.1	35.8	39.0	37.8	38.6	41.0	43.9
120	30.0	31.1	37.0	40.4	39.0	39.6	42.0	45.1
130	31.0	31.9	38.2	41.8	40.2	40.6	43.0	46.2
140	32.0	32.7	39.2	43.0	41.3	41.6	44.0	47.2
150	32.9	33.5	40.2	44.1	42.3	42.6	45.0	48.2
160	33.7	34.3	41.2	45.1	43.3	43.6	45.8	49.2
170	34.5		42.0	46.1	44.1	44.4	46.6	50.0
180	35.3					45.2	47.4	50.8
190	35.9					45.9	48.2	51.6
200						46.5	48.8	52.4
210						49.4	53.0	

Data of Durnin and Womersley (1974) who used density measured by underwater weighing to estimate fatness in a sample of Scottish people

body fat. If the subject's age, sex, height and weight are keyed in, this machine gives a rapid prediction of the amount and percentage of fat in the body.

According to Forbes (1999), there is considerable debate and disagreement about the accuracy and validity of this method and even some doubts about the theoretical basis for the technique. An extensive series of reviews of this technique may be found in Yanovski *et al.* (1996).

ESTIMATION OF FATNESS FROM BODY DENSITY

The calibration of the skinfold method of determining fatness (see Table 3.5) has been obtained by using body density as a reference method. The density of fat (0.9 kg/L) is less than that of lean tissue (1.1 kg/L) and so the measurement of whole-body density enables one to estimate the proportion of fat in the body:

$$\text{Density} = \frac{\text{mass}}{\text{volume}}$$

Mass may be determined by simple weighing. Volume is measured by comparing the weight of the body in air and its weight when fully immersed in water. The difference between these two values is the weight of water displaced by the body (Archimedes principle). Knowing the density of water (1 kg/L), one can thus calculate the volume of water displaced, i.e. the volume of the body. The principle is very simple, but the technical procedures required to obtain accurate and valid values are quite elaborate. For example:

- Subjects wear a weighted belt to ensure that they are fully immersed and correction has to be made for this.
- Corrections will need to be made for the residual air in the lungs during the immersion, which will greatly distort density measurements.

This method requires sentient and co-operative subjects. It is not suitable for routine use or for use with young children. It has traditionally been used as the reference method against which other methods are calibrated and validated, e.g. skinfold calipers and bioelectrical impedance.

BODY WATER CONTENT AS A PREDICTOR OF BODY FAT CONTENT

The measurement of body water content should enable good estimates of body fatness to be made. The

body can be seen as consisting of two compartments: lean tissue with a constant proportion of water (around 73%) and water-free fat. A measure of body water allows one to estimate proportions of lean and fat in the body. In order to estimate body water content in a living person or animal, one needs some chemical substance that is non-toxic, will readily and evenly disperse in all of the body water and is readily measurable. Water that is labelled with the deuterium isotope of hydrogen (heavy water) or the radioactive isotope of hydrogen (tritiated water) seems to fulfil these criteria.

After allowing time for dispersal, the concentration of the chemical in a sample of body fluid is measured. Knowing the amount of the substance introduced into the body, one can estimate the amount of diluting fluid (i.e. the body water). The principle is illustrated by the hypothetical example below.

- Administer 100 units of chemical.
- Allow time for dispersal in body water.
- Measure concentration in body water sample, let us say, 2 units per litre.
- Volume of body water estimated at 50 L, i.e. 50 kg of water.
- Estimate lean body mass assuming this 50 kg of water represents, let us say, 73% of lean tissue, i.e. $50 \times 100/73$.
- Total weight – lean weight = weight of body fat.

This method should yield reliable and valid measures of body water and thus good estimates of body fat.

MID-ARM CIRCUMFERENCE MEASURES

In immobile adults who cannot be weighed (e.g. unconscious hospital patients), **mid-arm circumference (MAC)** of the upper arm can be a useful anthropometric indicator of nutritional status. A single measurement needs standards for interpretation, but longitudinal changes in this measure are a reasonably sensitive indicator of changes in body weight. MAC has been widely used as a simple and rapid means of assessing nutritional status in children. A piece of previously calibrated and colour-coded tape can be used. If the circumference lies within the green region, this can be taken to indicate normality; in the yellow region, mild malnutrition; and a circumference within the red region indicates frank malnutrition (Passmore and Eastwood, 1986).

Mid-arm muscle circumference (MAMC) has been widely used as a simple measure of lean body

mass. The circumference of the mid-arm is measured with a tape and the triceps skinfold is measured with calipers. Then:

$$\text{MAMC} = \text{MAC} - (\pi \times \text{triceps skinfold})$$

Changes in MAMC are taken to indicate changes in lean body mass and, once again, may be a useful longitudinal measure of nutritional status in hospital patients.

Anthropometric assessment in children

Anthropometric measurements are cumulative indicators of the nutritional status and general health of children. Low values for height or length provide evidence of the chronic effects of malnutrition; weight for height is a more acute measure and indicates current nutritional status. These simple measures require little equipment or professional training and so even non-specialists can easily and cheaply produce reliable results. Interpretation of these results requires that they be compared with standards. The need for reliable standards is a recurrent problem when interpreting anthropometric measurements. If standards of height and weight for children are derived from children in affluent countries, poor children in developing countries may not correspond well with such standards. These differences between rich and poor are largely a reflection of chronic adaptation to undernutrition by the poor, which may make the interpretation of results and deciding upon appropriate responses difficult. An anthropometric survey of children in rural Bangladesh is likely to show that their average heights and weights are some way below those of typical American children of the same age, but they may be typical of children living in rural Bangladesh and thus these children are not specifically 'at risk'. This point is well illustrated by considering some of the changes that have occurred in the anthropometric characteristics of affluent populations during the twentieth century. Britons are, on average, several inches taller than they were at the start of the twentieth century and they mature quicker. Britons in the higher socioeconomic groups have always been taller than those in the lower groups, but this gap narrowed considerably during the twentieth century.

Sexual maturation occurs earlier now than at the start of the twentieth century, e.g. girls start menstruating 2–3 years earlier and boys reach their adult stature at 19 years rather than in their mid-twenties. The normal or average values of height and weight of British children thus changed very significantly during the course of the twentieth century. Note that large cross-sectional measurements of height in British adults have found a decrease in average height with increasing age. Part of this is a real tendency for people to decrease in height as they grow old, but most of it is due to the tendency for younger generations to be taller than their parents and grandparents (e.g. White *et al.*, 1993).

Body size is not only dependent upon environmental factors, such as diet, but is also genetically determined. Some races of people may be genetically smaller than others and this is another factor that makes the use of non-local standards problematical. Within races, there is a natural biological variation in the genetic component of height. This will obviously create difficulties in assessing the nutritional status of an individual child from a single set of anthropometric measurements. Is a child short because it is genetically small or because it is chronically malnourished? Growth curves are more useful ways of monitoring the nutritional status of individual children. Provided the child remains on its predicted growth curve, then whether the absolute values are low or high is in most cases unimportant. If a child on a high growth curve dips significantly below that curve, this may indicate a nutrition or health problem, even though it may be close to average for its age. A child who remains on a low growth curve will be assessed as growing satisfactorily even though it may be below average size for its age.

Centiles are frequently used to interpret anthropometric measures in children. If the growth curves of a population are plotted, one can produce growth curves that are typical of each per cent (or centile) of the population, i.e. the largest 1% through to the smallest 1%. If, for example, point measurements on a group indicate an average height for age that is substantially below the 50th centile, this would suggest impaired growth in the sample group. If a point measurement on an individual shows a value at the extremes of the range, e.g. below the 5th centile, then this indicates a probability that growth has been unsatisfactory. Figure 3.6 shows the World

Health Organization (WHO) standards of height for age in girls up to 5 years of age; the 3rd, 10th, 50th, 90th and 97th centiles are given. If a girl has a height for age below the 3rd centile, there is a less than 3% probability that she has been growing satisfactorily.

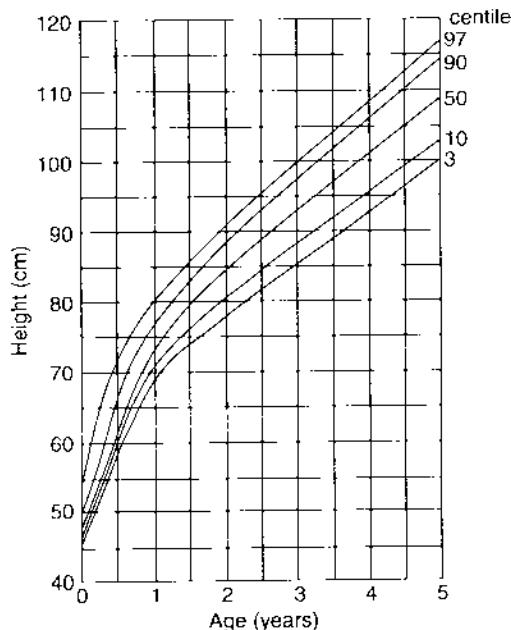


Figure 3.6 WHO standards of height (length) for age, in girls aged up to 5 years. The 97th, 90th, 50th, 10th, and 3rd centiles are shown. After Passmore and Eastwood (1986).

Estimating fatness in animals

The fat content of dead animals can be measured directly by extracting the fat with a suitable solvent (e.g. chloroform). Animal experimenters thus have the option of using this precise analytical method at the end of their experiment. This analytical method also provides an absolute standard against which other methods of estimating fatness can be calibrated and validated. In humans, of course, any method of estimating fatness can only be validated and calibrated against another estimate.

Experimental animals can be used to show that there is an almost perfect negative correlation between the percentage of body water and the percentage of body fat (Figure 3.7). This validates the use of body-water measurements to predict body fatness. In the mice used to derive the data in Figure 3.7, when body water was expressed as a percentage of the fat-free weight, all values were within the range 74–76%. This validates the earlier assumption that the body is made up of lean tissue with a constant proportion of water plus water free fat. Other methods, similar to those described for people, are also available to estimate the fatness of animals: weight/length indices; density; and, in animals that store fat subcutaneously, such as pigs, skinfold thickness. Rodents do not store much of their fat subcutaneously, but dissection and weighing of one fat storage organ may be used to predict total fat content in dead animals (Webb and Jakobson, 1980).

- Weight for height is a measure of acute nutritional status in children, but oedema may distort weight measurements.
- Height for age (or weight for age) is a cumulative or chronic measure of past status.
- Failure to grow and gain weight may be directly due to nutritional inadequacy or a secondary consequence of ill-health.
- Genetic variation in size hinders the interpretation of height and/or weight measurements in children.
- Centiles (hundredths) are one way of allowing for genetic differences in size.
- If an individual value is below the appropriate 5th centile, this means that there is only a 5% chance that that child has been growing satisfactorily.
- The average for a well-nourished population of children should be close to the 50th centile.
- If serial measurements are made, a child should approximately follow its centile line.

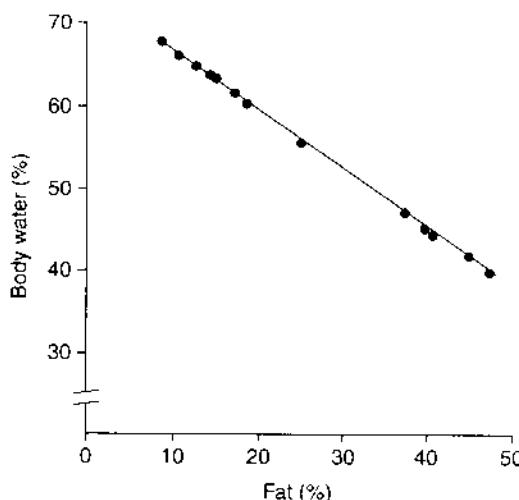


Figure 3.7 The relationship between percentage body fat and percentage body water in mice to illustrate the value of body water measurements in predicting body fat content. Data from Webb and Jakobson (1980).

Biochemical assessment of nutritional status

Biochemical measures of nutrient status yield objective, quantitative and sensitive indicators of nutritional status. They can usually be measured with great precision, as is common with such analytical procedures. A number of these biochemical measures of nutrient status are discussed in the section dealing with the methods used to assess nutrient requirements, for example:

- measures of blood and tissue levels of nutrients or metabolites
- measures of nutrient excretion rates
- enzyme activation tests

- functional biochemical tests such as prothrombin time.

Table 3.6 gives a short list of some commonly used biochemical tests of nutritional status and the range of normal values.

Biochemical measurements need laboratory facilities and they are relatively expensive and time consuming to perform. The interpretation of the values obtained will depend upon the limits of normality that have been set, and this will often be a matter of judgement. For example, blood haemoglobin concentration can be measured with considerable precision, but deciding upon the level that should be taken to indicate anaemia will be a matter of judgement. Thus, translating surveys of haemoglobin measurements into assessments of the prevalence of anaemia will depend very much on the limits of normality that are used. A figure of 12 g of haemoglobin per 100 mL of blood at sea level was traditionally taken as the level below which there is a progressively increasing risk of anaemia. If this value is taken as the lower limit of normality, the prevalence of iron-deficiency anaemia in menstruating women is high, even in industrialized countries (10–20%). However, clinical symptoms of anaemia may only become apparent with haemoglobin levels below 10 g per 100 mL of blood. Conversely, substantial depletion of iron stores (as measured by serum ferritin concentration) and some symptoms of iron deficiency may occur without any decrease in blood haemoglobin concentration. The assessment of iron status and prevalence of deficiency is discussed further in Chapter 13.

Measurements made upon urine samples have the obvious advantage of requiring no invasive procedures and of giving relatively large samples and thus avoiding the need for special micro-analytical techniques. They can be useful if the excretion rate of an intact nutrient or readily identifiable derivative can be

- The body composition of animals can be measured accurately and directly by chemical analysis of the carcass.
- Carcass analysis can be used to calibrate and validate other less direct methods of estimating body composition. In humans, any method of estimating body composition can only be calibrated and validated by comparison with other estimates.
- Chemical analyses of animal carcasses show that percentage body water is highly negatively correlated with percentage fat and that the water content of the fat-free mass is constant.

Table 3.6 Examples of biochemical tests of nutritional status

Parameter	Test	Critical value	Normal range
Protein/energy	Serum albumin		35–45 g/L
Protein/energy	Serum transferrin		2.5–3.0 g/L
Vitamin C	Serum (or plasma) vitamin C	<2 mg/L	>4 mg/L
Thiamin (B_1)	Erythrocyte transketolase activation coefficient	>1.25	<1.15
Riboflavin (B_2)	Erythrocyte glutathione reductase activation coefficient	>1.3	<1.2
Vitamin B_6	Serum pyridoxal phosphate	<5 μ g/L	
Niacin	Urinary N' methyl nicotinamide excretion (mg nicotinamide/g creatinine)	<0.5	>1.6
Folic acid	Serum folic acid	<3 μ g/L	>6 μ g/L
Vitamin B_{12}	Serum B_{12}		>100 ng/L
Vitamin A	Plasma vitamin A (clinical tests are often used)	<10 μ g/L	>20 or 30 μ g/L
Vitamin D	Plasma 25-OH cholecalciferol (D_3)	<10 μ g/L	
Vitamin E	Total tocopherol in serum	<5 mg/L	
Iron	Blood haemoglobin		>115 g/L
	Serum ferritin	<12.5 μ g/L	>25 μ g/L
Zinc	Serum zinc	<10 μ mol/L	>18 μ mol/L

Critical values for micronutrients in adults are largely taken from Groff *et al.* (1995).

related to nutritional status. Rates of urine flow, and thus the concentration of solutes in urine samples, are very variable and greatly influenced by, for example:

- recent fluid consumption
- physical activity
- consumption of substances that have diuretic or antidiuretic activity (e.g. caffeine or nicotine).

Twenty-four-hour samples may give more meaningful results, but their collection will be impractical in many survey situations. Even where they are feasible, checks will be required to ensure that the samples are complete in non-captive subjects. As an example, it is generally agreed that the best method of assessing salt intake on any particular day is to measure the sodium excreted in the urine in a 24-hour period (see Chapter 13). In order to ensure that the 24-hour urine sample is complete, subjects may be asked to take doses of a marker substance (para amino benzoic acid) with their meals. Only where there is practically complete recovery of the marker in the urine sample will the sample be accepted as complete. Note that the sodium excreted in a single 24-hour sample may not be a reliable indication of habitual salt intake (see Chapter 13).

Urinary levels of the B vitamins have traditionally been used as measures of nutritional status for these vitamins. The values are often expressed as amount of nutrient (or nutrient derivative) per gram of creatin-

ine in the urine sample. It is assumed that the excretion of creatinine is approximately constant and thus this gives a more meaningful value than the amount of nutrient per unit volume of urine, because the latter is so dependent upon the rate of urine flow.

Blood samples give much greater scope for assessing nutrient status. Plasma, serum or blood-cell concentrations of nutrients are extensively used as specific measures of nutritional status. Enzyme activation tests using red blood cell enzymes (as discussed earlier in the chapter) are also used. Plasma albumin concentration drops in severe protein energy malnutrition and it has been widely used as a general biochemical indicator of nutritional status. Many medical conditions also result in a reduced plasma albumin concentration, e.g. liver and kidney disease, some cancers and infection. Albumin has a relatively long half-life in plasma and so it is now regarded as too insensitive and non-specific to be a good biochemical indicator of nutritional status. Other plasma proteins with a shorter half-life in plasma may give a more sensitive indicator of protein energy status e.g. **transferrin** – the protein that transports iron in plasma – and **retinol-binding protein** – the protein that transports vitamin A (retinol).

Blood tests require facilities and personnel capable of taking and safely handling, transporting and storing blood samples. The analysis of blood samples often requires special micro-analytical techniques because of the limited volumes that can be taken for analysis.

Measurement of energy expenditure and metabolic rate

Metabolic rate is the rate at which energy is expended by the body. Energy expenditure may be rather crudely and indirectly assessed from measurements

of energy intake. As energy can neither be created nor destroyed, energy expenditure over any given period must be equal to the intake plus or minus any change in body energy stores:

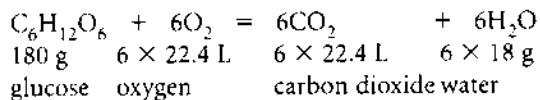
$$\text{energy expenditure} = \text{energy intake} \pm \text{change in body energy}$$

- Clinical signs, anthropometric measurements and biochemical analyses of body fluid samples can all be used to assess nutritional status. Examples of these methods are given in Tables 3.3, 3.4 and 3.6.
- With any nutritional assessment of groups or populations, steps must be taken to ensure that the sampled group is representative of the whole population under investigation.
- Clinical signs can be incorporated into checklists that can be used for screening and assessment by non-specialists and without the need for specialist equipment. They have the following merits and disadvantages:
 - they are quick, simple, and cheap to record
 - they are insensitive indicators of malnutrition and do not give early warning of an impending nutritional problem
 - some signs are common to several nutrient deficiencies and to non-nutritional problems and so they are not specific
 - they are qualitative and subjective because they depend upon the opinion of the assessor and often yield just a yes/no response or a descriptive grading of severity.
- Anthropometric methods of assessment have the following merits and disadvantages:
 - they are quantitative and objective
 - many are quick and simple to perform, require only simple equipment and do not require specialist personnel
 - accurate assessment of fatness with skinfold calipers does require care and technical expertise
 - they often need reliable standards to interpret their results – some are more useful if measured sequentially to monitor changes in nutritional status
 - previous obesity can reduce their sensitivity in detecting deteriorating nutritional status
 - oedema can substantially distort body weight and undermine any of these methods that relies upon accurate body weight measurement.
- Biochemical methods have the following merits and disadvantages:
 - they are quantitative and objective
 - they are the most sensitive methods of nutritional assessment
 - in many cases, they are highly specific indicators of the status for a particular nutrient
 - they are time consuming, expensive and they need specialist laboratory facilities and personnel
 - standards of normality may be difficult to set in some cases
 - if concentration of a nutrient or metabolite in urine is measured, either some mechanism has to be found to correct for the wide variation in rates of urine flow or a complete 24-hour sample should be collected
 - if blood analyses are used, they require an invasive collection procedure and often a micro-analytical technique, because of the small volumes that are available for the analysis.

If expenditure is measured in this way, the errors in measuring intake and changes in body energy will compound. The methods of measuring changes in body energy are insensitive, making this method unreliable for the measurement of expenditure over short periods. Tens of thousands of calories may be added to or lost from body energy without them being quantifiable by the available methods of measuring body composition. This error is minimized if the time period is long, but this also means that subjects will be required to monitor intakes for extended periods.

All of the energy expended by the body is ultimately lost as heat. It is theoretically possible to measure this directly in a whole-body calorimeter. This is a chamber surrounded by a water jacket in which the heat released by the subject in the chamber is measured by a rise in temperature in the surrounding water. It is more usual to predict energy expenditure from measures of oxygen consumption and/or carbon dioxide evolution. Foodstuffs are metabolically oxidized to carbon dioxide, water and, in the case of protein, nitrogenous waste products such as urea. This oxidation process consumes atmospheric oxygen and produces carbon dioxide. The chemical equations for the oxidation of the various foodstuffs can be used to predict the **energy equivalent of oxygen**, which is the amount of energy released when 1 L of oxygen is used to metabolically oxidize food.

- For example, the equation for the oxidation of glucose is:



- Oxidation of 180 g of glucose yields around 665 kcal (2800 kJ).

Thus, 1 L of oxygen yields approximately 4.95 kcal (20.8 kJ) when it is being used to metabolize glucose, i.e. 665 divided by (6 × 22.4). Similar calculations can be performed using the other substrates and they yield energy equivalents for oxygen that are not very different from that for glucose. If a mixture of substrates is being oxidized, a figure of 4.86 kcal/L (20.4 kJ/L) of oxygen can be used as the approximate energy equivalent of oxygen. If more precise estimation is required, the ratio of carbon dioxide evolution to oxygen consumption, the **respiratory**

quotient (RQ), gives an indication of the balance of substrates being metabolized. The RQ is 1.0 if carbohydrate is being oxidized, but only 0.71 if fat is being oxidized. Tables that list the energy equivalent of oxygen at various RQs are available.

The simplest way of measuring oxygen consumption (and carbon dioxide evolution) is to use a Douglas bag. This is a large plastic bag fitted with a two-way valve so that the subject, who breathes through a mouthpiece attached to the valve, sucks in air from the atmosphere and then blows it out into the bag. All of the expired air can be collected over a period of time and then the volume and composition of this expired air can be measured. Knowing the composition of the inspired atmospheric air means that the subject's oxygen consumption and carbon dioxide evolution can be calculated. Note that the collection period is limited to a few minutes even if the capacity of the bag is large (100 L).

The static spirometer is essentially a metal bell that is filled with oxygen and is suspended with its open end under water. The subject breathes through a tube connected to the oxygen-containing chamber of the bell. The bell rises and falls as the volume of gas inside it changes, and thus it will fall and rise with each inspiration and expiration. If the system contains a carbon dioxide absorber, the bell will gradually sink as the subject uses up the oxygen in it. This change in volume of oxygen inside the bell can be recorded on a calibrated chart and thus the rate of oxygen consumption can be determined. This is a standard, undergraduate practical exercise. It is limited to a few minutes of measurement of resting oxygen consumption or very short periods of consumption with static exercise, e.g. on an exercise bicycle.

Portable respirometers have been developed that can be strapped onto the back and thus used and worn whilst performing everyday domestic, leisure or employment tasks. These respirometers, such as the Max Planck respirometer, monitor the volume of gas expired by the wearer and divert a small proportion of the expired gas to a collection bag for later analysis. They are, once again, only suitable for relatively short-term recording, but they can be used to quantify the energy costs of a variety of tasks.

If subjects keep a detailed record of their activities during each 5-minute period over the day (**activity diary**), total energy expenditure can be roughly estimated using the measured energy costs of the various activities.

A number of research units have respiration chambers that allow relatively long-term measurement of energy expenditure of subjects performing any routine tasks that may be done within the chamber. The respiration chamber is essentially a small room, which has a controlled and measured flow of air through it. Oxygen consumption and carbon dioxide production by the subject can be measured by the difference in the composition of air as it enters and leaves the chamber. These pieces of equipment are, in practice, complex and expensive and thus restricted to a relatively few well-funded research units. They can be used to monitor the long-term energy expenditure of subjects who can live in the chamber for several days and perform everyday tasks that simulate normal daily activity.

Over the last few years, the **doubly labelled water method** has been widely used to determine the long-term energy expenditure of free-living subjects going about their normal daily activities. Subjects are given a dose of doubly labelled-water ($^2\text{H}_2^{18}\text{O}$), i.e. water containing the heavy isotopes of hydrogen (^2H) and oxygen (^{18}O). The subjects lose the labelled oxygen more rapidly than the labelled hydrogen because the hydrogen is lost only as water, whereas the oxygen is lost as both water and carbon dioxide. This is due to the action of the enzyme carbonic anhydrase, which promotes the exchange of

oxygen between water and carbon dioxide. The difference between the rate of loss of labelled hydrogen and labelled oxygen is a measure of the rate of carbon dioxide evolution and thus it can be used to estimate the long-term rate of carbon dioxide evolution and thus long-term energy expenditure. Comparisons of total energy expenditure measurements by respiration chamber and by the doubly labelled-water method suggest that the latter can give an accurate measure of energy expenditure. Using this method, investigators at last have an acceptably accurate means of estimating the long-term energy expenditure of free-living people.

COMPARISONS OF METABOLIC RATES BETWEEN INDIVIDUALS

Basal metabolic rate is the minimum rate of energy expenditure in a conscious, resting animal or person. It is the metabolic rate measured in a rested, fasted individual who has been lying in a room at thermoneutral temperature for half an hour. Absolute BMR obviously tends to increase with increasing body size; large animals have higher absolute BMRs than small ones; fully grown adults have larger BMRs than small children. However, relative BMR (i.e. per unit body mass) declines with increasing body size. Small animals have larger relative metabolic rates

- Metabolic rate is the rate of energy expenditure of the body, and basal metabolic rate (BMR) is the minimum metabolic rate measured in a conscious person who is rested, warm and fasted.
- Metabolic rate can be measured directly by measuring heat output, but it is usually calculated from the rates of oxygen consumption and carbon dioxide production.
- Conventional methods of measuring metabolic rate by oxygen consumption only allow it to be monitored for a few minutes at a time. Even in a respiration chamber, subjects are restricted to the activities that can be carried out in a small room.
- The doubly labelled water method allows long-term energy expenditure to be measured in free-living subjects.
- Absolute BMR increases with body size, but relative BMR decreases with size. Small animals or people have smaller absolute but larger relative BMRs than larger ones.
- BMR (in kcal/day) can be predicted by multiplying the body weight (in kg) raised to the power 0.75 by a constant (70).
- Metabolic rates of people have traditionally been expressed per unit of body surface area and, if this method is used, men have higher relative BMRs than women.
- It is better to express relative metabolic rate per unit of lean tissue mass and then the sex differences in relative BMR disappear.

than large ones; small children have higher relative metabolic rates than adults. The equation given below is found to approximately predict the BMR of mammals of different sizes both within and between species:

$$\text{Basal metabolic rate} = \text{constant (k)} \times \text{weight}^{3/4}$$

$$(\text{kcal/day}) \quad (70) \quad (\text{kg})$$

More information on BMR and its prediction from body weight can be found in Chapter 6. Human metabolic rates have traditionally been expressed per unit of body surface area, e.g. kcal (kJ)/hour per m². Even in human adults, if BMR is expressed per unit of body weight, large variations between individuals are found, particularly between fat and thin individuals. Adipose tissue has a much lower metabolic rate than lean tissue and so increasing adipose tissue mass has relatively much less effect in increasing whole body metabolic rate than increasing lean tissue mass. Increasing adipose tissue mass also has relatively little effect in increasing body surface area and so expressing BMR per unit surface area tends to reduce the individual variation and has been a useful way of comparing BMR in different individuals. Nomograms are available that enable surface area to be predicted from measurements of height and weight.

Nowadays, it is considered that the best way of comparing and expressing the BMR of people is to express it per unit weight of lean tissue, e.g. kcal (kJ)/hour per kg lean tissue. When expressed in this way, the BMRs of men and women are not different, whereas when expressed per unit surface area, those of women are about 10% less than those of men. This is an artefact due to the much higher body fat content of women.

METHODS USED TO ESTABLISH LINKS BETWEEN DIET AND CHRONIC DISEASE

Strategic approaches – observational versus experimental

There are two major strategies that can be used for trying to investigate putative causal links between diseases/risk factors and dietary/lifestyle variables.

- Observational, epidemiological methods are used to relate differences in the diets or lifestyle charac-

teristics of populations or individuals to their risk of developing a disease or their measured level of a disease risk factor.

- In the experimental approach, experimenters try to produce a controlled change in the diets or other behaviour of individuals or populations to determine the effect that this has upon the disease risk or the measured level of a risk factor such as blood pressure or plasma cholesterol.

Table 3.7 gives a list of the methods available for trying to investigate possible links between diet and disease. In the following sections, the principles and problems of these methods are discussed, using illustrative examples. The examples have been chosen to illustrate the methodology and its problems. Often, well-known classical studies and older studies have been chosen for illustrative purposes. Many of these classical papers were very influential upon the later direction that research took and greatly influenced the scientific opinion of their time. Some studies have been chosen because they illustrate some of the pitfalls of their methodology.

Some general characteristics and problems of epidemiological methods

ASSOCIATION DOES PROVE CAUSE AND EFFECT

Epidemiological methods produce evidence of an association between two variables – in this context, evidence of an association between some measured dietary variable and a disease or risk factor such as high blood pressure. Such epidemiological association does not necessarily mean that there is a ‘cause and-effect’ relationship between the dietary factor and the disease. The association may be coincidental and dependent upon a relationship between both tested variables and a third unconsidered or **confounding variable** (or many confounding variables). It is even possible that in some cases that there may be an ‘effect and-cause’ relationship, i.e. the disease results in a change in the dietary variable. This very important distinction has already been discussed in Chapter 1 and several examples are described there. The problem of confounding variables pervades all types of epidemiological investigation, and epidemiologists use various means and statistical devices to correct for variables they suspect to be confounding, but this is an imperfect

Table 3.7 Some of the research methods that can be used to investigate possible links between diet and disease

Observational methods

Descriptive epidemiology

Cross-cultural comparisons: correlation is made between the level of a dietary factor and the frequency of a disease in different populations; these may be comparisons between nations or between special groups within a nation (e.g. between Seventh Day Adventists and the rest of the population in the USA)

Time trends: changes in the dietary habits of populations are correlated with changes in disease frequency

Migration studies: the changes in dietary habits that follow migration are correlated with changes in disease frequency in the migrant population

Experiments of nature: these involve the use of people with some genetic mutation or other accident of nature that increases (or decreases) their exposure to some suggested disease risk factor – does the increased exposure to the risk factor result in increased rate of disease?

Analytical epidemiology

Cohort studies: the dietary characteristics of a large sample of individuals are measured; the samples are selected and the measurements made in order to test an hypothesis; the deaths and the causes of mortality and/or morbidity of individuals within the sample are then monitored over the succeeding years; disease risk is then correlated with the level of exposure to possible dietary risk factors

Case-control studies: established cases of the disease are identified and matched to a sample of non-sufferers, the controls; the past diets (or other exposure) of the cases and controls are then compared to see if this provides evidence of greater exposure to the suggested risk factor in the case group (or greater exposure to a possible protective factor in the control group)

Experimental methods

Animal experiments: dietary changes are tested in short-term or long-term controlled experiments with laboratory animals to see if they result in the postulated change in risk factor or disease frequency

Short-term human experiments: dietary changes are tested in short-term controlled experiments to see if they result in the postulated change in risk factor measurement

Clinical trials: a dietary change that is expected to lead to improved prognosis in patients is tested under controlled conditions to see if it yields the expected benefits; ideally, the treatment is compared to the effects of a dummy treatment (placebo) and neither the investigators nor the patient know who has received the real and dummy treatments until after the study is completed (double-blind trial)

Intervention trial: a population-level trial; an attempt is made to change the exposure of a test population to the postulated dietary factor, e.g. by encouragement or possibly by a nutrient fortification programme; the change in disease frequency is compared to that in a control population that has not received the intervention

- Observational (epidemiological) and experimental approaches are used to investigate links between diet and disease.
- Descriptive epidemiological methods rely upon looking at the occurrence and distribution of diseases in groups or populations.
- With analytical epidemiological methods, data are collected from individuals with the aim of testing an hypothesis. Analytical studies may involve the use of a control group.
- In the experimental approach, there are experimental and control groups and the investigators impose some 'intervention(s)' on the experimental group, which is designed to test an hypothesis.
- A list of the various methods available and a summary of their principles are shown in Table 3.7.

process and the quality of this correction is likely to vary considerably from study to study.

If, as claimed by McCormick and Skrabanek (1988), there really are 250 risk markers for coronary heart disease, then the task of identifying any one of them as independent of all of the others is a formidable one. An apparent protective effect of moderate doses of alcohol against coronary heart disease has been reported many times (see Chapter 4), but how many authors will have considered or corrected for level of physical activity as a potential confounding variable? One large English survey reported a strong positive association between alcohol consumption and level of physical activity (White *et al.*, 1993).

Barker *et al.* (1998) list six characteristics which would make it more likely that a demonstrated association was causal, and these are outlined below.

- *Strength of the association:* the stronger the association, the lower the likelihood that it is due to some unforeseen confounding factor.
- *Is it graded or dose dependent?*: there is usually a graded effect of true causes rather than a threshold effect.
- *Independence of the association:* the relationship remains significant even after correction for potentially confounding variables.
- *Consistency of the finding:* if the association is found in a variety of studies using different investigative approaches, this makes it more likely that it is due to cause and effect.
- *Reversibility:* if reduced exposure to the suspected cause leads to reduced incidence of the disease, this increases the likelihood of it being a true cause.
- *Plausibility:* the association is more likely to be 'cause and effect' if there is a believable mechanism to explain the causality of the association, which can be supported by laboratory studies. A plausible mechanism unsupported by corroborating laboratory studies probably adds little weight to the argument. It is quite possible to devise plausible mechanisms for most observations, often equally plausible mechanisms for opposing findings. The most plausible and intellectually satisfying mechanism may not be the one that ultimately proves to be correct.

SOURCES AND QUALITY OF DATA

Information about the diets of individuals and populations is derived from the methods discussed earlier

in this chapter. Food balance sheets or household expenditure surveys (e.g. the UK National Food Survey) are often the basis for international dietary comparisons. Methods that correlate dietary factors with disease risk in individuals clearly require individual methods of dietary assessment. When dietary surveys of large representative samples are made, these can be properly used for international comparisons, e.g. *The dietary and nutritional survey of British adults* (Gregory *et al.*, 1990). If small, unrepresentative survey results are extrapolated and taken to represent whole populations, these can be major source of error. The many sources of error and uncertainty in estimating dietary intakes have been discussed earlier in this chapter. Acceptably accurate and representative measures of habitual dietary intake are clearly a prerequisite for detecting associations between diet and disease risk.

Information about rates of mortality from specific diseases is usually obtained through some analysis of information recorded on death certificates. There have now been efforts made to standardize death certificates internationally and to standardize the nomenclature and diagnostic criteria for the various causes of death. The WHO publishes an International Classification of Diseases, Injuries and Causes of Death (ICD) which is regularly revised. Death certificates in the UK list both the underlying (main) cause and the contributory causes of death. There will, however, still be differences between individual physicians, and this variation is likely to be even greater between those in different regions or nations. There may be straightforward errors of diagnosis, different diagnostic criteria or terminology may be used at different times or in different centres, even the accuracy of the age recording on death certificates may vary. When autopsies have been used to confirm the accuracy of previously recorded causes of death, they show that errors in diagnosing the causes of death in individual patients may be frequent. Both false positives (i.e. people falsely diagnosed as dying of the disease) and false negatives (i.e. people dying of the disease but not recognized as such) tend to occur. If individuals are the unit of comparison, these two errors compound, but they do tend to cancel out if population mortality is calculated. Even though the number of individual errors of diagnosis may be quite large, provided false negatives and false positives occur at similar rates, then the accuracy of the estimate of population mortality may be little affected. If there is

systematic over-recording or under-recording of deaths from any cause, this will seriously distort population comparisons – for example, any attempt to play down the prevalence of a disease to protect the tourist industry or for other commercial or political reasons.

Accurate and representative measures of morbidity in populations are likely to be more difficult to obtain than mortality data. Investigators will almost certainly have to rely upon less rigorous and uniform methods of collecting data than death certificates. Some sources of information about disease frequency in populations are listed below.

- Notification or registration of particular diseases: any variations in the efficiency of these collection procedures in different areas will distort the estimates of disease frequency.
- Analysis of reasons for hospital admission or general practitioner consultation: multiple recordings of the same individuals is one source of error here.
- Social security data on reasons for absences from work.
- Extrapolation from relatively small-scale and possibly unrepresentative survey samples.

Incidence and **prevalence** are terms used to indicate disease frequency. Incidence is the number of new cases of the disease occurring within a specified time period. Prevalence is the number of cases of the disease existing at any particular point in time. Prevalence is dependent upon both the incidence and the duration of the disease. A chronic disease will have a much greater prevalence than an acute one of similar incidence. As an example, the incidence of **cystic fibrosis** in the UK remained fairly constant for many years at about 300–400 cases per year, but the prevalence has increased because the survival time of sufferers has increased. Therefore, the chronic diseases of industrialization, almost by definition, have high prevalence. Not only do they cause reduced quality of life for large numbers of people, but they also exert a considerable drain on health and social care budgets.

If an asymptomatic risk factor, such as plasma cholesterol level or blood pressure, is being related to diet, investigators will have to rely upon data obtained from their own or published surveys. If these surveys are used to make international comparisons, they may have been conducted in the various countries by different investigators using different protocols. Throughout the 1990s, there were annual

health surveys of representative samples of English adults that included anthropometric measurements and measurements of major risk factors (e.g. Prescott-Clarke and Primatesta, 1998) as a result of the *Health of the Nation* white paper (also similar Welsh and Scottish surveys).

AGE STANDARDIZATION

Whenever comparisons of the death (or disease) rates between different populations or groups are made, some allowance has to be made for any differences in their age structures.

The simplest measure of mortality is the **crude death rate** (see Death rate in Glossary) for the specified cause:

$$\frac{\text{Number of deaths (by cause) in one year}}{\text{Number of people in the population}} \times 1000$$

This is often of limited value because mortality from many causes is so highly age dependent and because of the widely varying age structures of some populations. Higher birth rates and shorter life expectancies mean, for example, that populations in developing countries tend to have a lower average age than in the Western industrialized countries. A more useful measure for international or regional comparisons would be the **age-specific death rate** (see Death rate in Glossary), for example:

$$\frac{\text{Annual number of deaths (by cause) in males aged } 45\text{--}54 \text{ years}}{\text{Number of men in this age group in the population}} \times 1000$$

To make more complete comparisons, the **standard mortality ratio (SMR)** is a useful and widely used measure. In order to calculate the SMR, it is necessary to use as a standard or reference population a population in which all of the various age-specific mortality rates for the cause under investigation have been established. Then, for each age band of the test population, one must calculate how many deaths would be expected or predicted:

- given the number of people of that age in the test population;
- assuming the same age-specific death rate as in the reference population.

For example:

number of males aged 45–49 years in the test population × death rate from the cause for males aged 45–49 years in the reference population = number of predicted deaths in 45–49 year old males in the test population

This calculation is made for each age group in the population and the total number of predicted deaths in the test population is calculated. The total number of actual deaths in the test population is then

expressed as a percentage or ratio of the predicted number of deaths from that cause in that test population:

$$\text{SMR} = \frac{\text{actual number of deaths in test population}}{\text{predicted number of deaths}} (\times 100)$$

The SMR can be used for international comparisons, to make regional comparisons within any given country, or to make comparisons between other groups. When used for regional comparisons

- Epidemiological methods provide evidence of association between dietary variables and disease risk.
- Epidemiological association between two variables does not necessarily mean that they are causally related.
- Association may arise because both of the tested variables are linked to other confounding variables.
- An association is more likely to be causal if it is strong, consistent between studies, dose dependent, independent of likely confounding variables, reversible, and can be explained by a plausible mechanism that can be supported by experiment.
- Information about causes of death is usually obtained directly or indirectly from death certificates.
- There is an international classification system for causes of death (ICD) and a standardization of the format of death certificates.
- Errors do occur in the diagnosis of cause of death – both false negatives and false positives. These tend to compound when individuals are the unit of study, but tend to cancel out when population mortality rates are calculated.
- Accurate information on the frequency of non-fatal diseases or disease risk factors is likely to be more problematical than calculation of mortality rates.
- Information on non-fatal disease frequency is obtained from the following sources:

disease notifications or registers
analysis of reasons for hospital admission, physician consultations or absences from work surveys.

- Incidence is the number of new cases of a disease that occur within a specified time period.
- Prevalence is the number of cases of a disease that exist at any point in time. It is a product of the incidence and duration of the disease.
- Comparisons of crude death rates between groups or populations may be of little value if they are not age (and sex) standardized, because the age and sex structures of different populations may be very different.
- Age-specific death rate is the rate of death from a disease within a specified age and sex group.
- Standard mortality ratio (SMR) is the ratio of predicted deaths to actual deaths in test populations.
- The predicted number of deaths is calculated as those that would occur in a population of this age structure if the age-specific death rates were the same as those in a reference population.

within a country, the whole national population can be used as a reference population. For example, comparisons of the SMRs for gastric cancer in different regions of Britain in the early 1980s showed that the SMR was over 115 in Wales but less than 85 in parts of East Anglia. People in Wales have more than a 15% greater risk of dying of this disease than the national population, whereas those in East Anglia have less than 85% of the risk of the whole population. Regional differences in diet, lifestyle and environment within Britain can be compared to see if any are associated with this variation in gastric cancer mortality. This might provide important clues about the aetiology of gastric cancer.

Cross-cultural comparisons

An early indication that diet may be implicated in the aetiology of a disease often comes from cross-cultural associations between a population measure of the dietary variable and population mortality or morbidity rates from the disease. Some examples are given below.

- A positive association between the average saturated fat intakes of populations and their mortality rates from coronary heart disease points towards high saturated fat intakes as a causal factor in coronary heart disease.
- A positive association between *per capita* sugar consumption in different countries and the average number of decayed, missing and filled teeth in young people implicates sugar in the aetiology of dental caries.
- A positive association between *per capita* salt consumption and blood pressure in various adult populations implicates high salt intake in the aetiology of hypertension.
- A negative association between *per capita* fibre consumption and bowel cancer mortality in various populations has been used to implicate the fibre depletion of Western diets in the aetiology of bowel cancer.

There are large variations between nations in mortality and morbidity rates from the diet-related diseases of industrialization. It is, therefore, common to see cross-population associations between a dietary variable and one of these diseases used to implicate diet as a major cause of the disease, as in the four

examples above. It is, of course, very premature then to assume a cause-and-effect relationship between that dietary factor and the disease. As populations become more affluent, numerous changes in diet, lifestyle and environmental conditions tend to occur simultaneously, opening up the likelihood that there will be many confounding variables. For example, all of the following changes may accompany increasing affluence and industrialization:

- increased fat and saturated fat consumption
- lower consumption of starches and fibre
- increased use of simple sugars, salt and perhaps alcohol and tobacco
- increased consumption of animal protein
- reduced incidence of nutritional deficiencies
- reduced physical activity
- increased proportions of overweight and obese people
- improved medical care
- many other changes in diet, lifestyle and environment.

Accompanying these trends for change in diet and lifestyle, there will be general trends in patterns of mortality and morbidity. Life expectancy increases, mortality from infectious diseases and other acute causes declines, but there will be increased prevalence of and mortality from the various chronic diseases of industrialization. Cross-cultural studies tend to focus upon individual associations between one of these lifestyle or dietary changes and one of these chronic diseases. Thus, the association between saturated-fat intake and coronary heart disease mortality is used to support the proposition that high saturated-fat intake causes heart disease, and the negative association between fibre intake and bowel cancer is used to support the proposition that low fibre intakes cause bowel cancer. As high-fat diets tend to be low in fibre and *vice versa*, it would certainly be possible to produce equally convincing correlations between fibre intake and coronary heart disease (negatively correlated) and between saturated-fat intake and bowel cancer.

The classic paper of Gleibermann (1973) on the relationship between the average salt intake and blood pressure of adult populations illustrates several of these problems of cross-cultural studies. She gleaned from the literature estimates of the salt intakes and blood pressures of middle-aged people in

27 populations from various parts of the world. The data for the male populations have been used to construct a simple scatter diagram (Figure 3.8). The correlation coefficient and line of best fit have also been calculated. There is a highly significant, positive and apparently linear relationship between the average salt intake and the mean blood pressure of these populations, consistent with the hypothesis that high salt intake is causally linked to the development of hypertension.

The two variables in Figure 3.8 were measured by numerous different investigators. The methods, conditions and quality of measurement are therefore not consistent across the groups. As has already been noted earlier in this chapter, salt intake is a particularly problematical measurement. The methods used to determine salt intakes in these 27 populations are variable and, in some cases, unspecified; they range from the high-precision 24-hour sodium excretion method to 'educated guesses' made by Gleibermann herself. Recorded blood pressure may be very much influenced by the method used and the care taken by the measurer to standardize the conditions, e.g. blood pressure may be affected by how long the subjects are allowed to relax before measurement and it may be raised if measured in anxious subjects.

Although the correlation shown on Figure 3.8 is highly statistically significant, there is a considerable scatter of points around the line of best fit. The correlation coefficient (r) is 0.61, suggesting that only about 37% (r^2) of the variation in population blood pressures is accounted for by variation in their salt intake. It is possible that this association is due to some confounding variable or variables, e.g. the populations with the lowest blood pressures and lowest salt intakes would also tend to be active, have low levels of obesity, relatively low alcohol intakes, high potassium intakes etc. All of these factors are thought to favour lower blood pressure. More recent investigators have suggested that the inclusion of different races may have increased the scatter with Gleibermann's data because black populations may be genetically more sensitive to the hypertensive effects of salt (e.g. Law *et al.*, 1991a).

The range of salt intakes shown in Figure 3.8 is huge, from around 1 g/day to well over 25 g/day. The average salt intakes in the UK and USA are some way below the middle of this range. If one takes the linear relationship at face value, one could predict that specified falls in any population's average salt intake would be associated with specified falls in average blood pressure. One could then go on to make predictions about the expected benefits for that population (reductions in the prevalence of hypertension and hypertension-related diseases). It would be hard to demonstrate conclusively that this association was linear over the whole range. This is what one is assuming when predicting that small or moderate reductions in salt intakes of the populations around the centre of the range will have predictable effects on average blood pressure.

There have been numerous studies conducted in Western industrialized countries that have measured blood pressure and salt intakes in large samples of individuals living in particular towns or regions. In many of these studies, no association is found between individual salt intake and individual blood pressure. One explanation for this disparity is that the gold standard method for measuring salt intake (24-hour sodium excretion) is probably a poor measure of an individual's habitual salt intake. This fact alone would make it improbable that one would find a correlation between individual salt intake measured in this way and blood pressure, even if habitual salt intake is a major determinant of individual blood pressure. (See Chapter 13 for further discussion.)

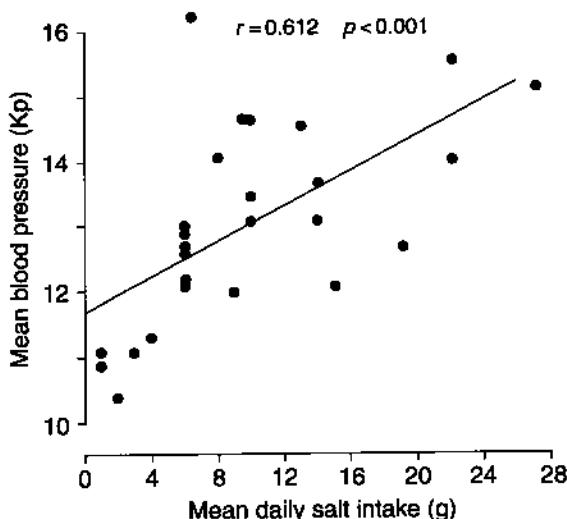


Figure 3.8 A simple scatter diagram showing the relationship between estimated salt intake and mean blood pressure in 27 male populations. Data from Gleibermann (1973); figure reproduced from Webb (1992a).

Such disparity between cross-cultural and within-population studies is common in epidemiological studies and also tends to be found, for example, when dietary saturated fat intake and plasma cholesterol are correlated. Taking these results at face value, one might conclude that, whilst average salt intake is an important determinant of average population blood pressure, individual differences in salt intake are not a major factor in producing the differences in blood pressure within a single population. This may initially seem like a contradiction because, of course, the influence of salt on population blood pressure would have to be a summation of its effects on the blood pressure of individuals. However, numerous other factors may also affect an individual's blood pressure, such as:

- individual genetic variation
- levels of activity and obesity
- alcohol consumption
- genetic variation in susceptibility to the hypertensive effects of salt.

These other influences on blood pressure may mask any effect of salt on individual blood pressure. Some people may have high blood pressure because they are overweight, inactive and consume large amounts of alcohol, yet their salt consumption may be low. Other people may have low blood pressure despite high salt intake because they are fit, abstemious, and relatively salt insensitive. Significant numbers of such people would, of course, make it difficult to find a statistically significant association between salt intake and blood pressure, especially as most people within any given population will be concentrated around the mean value and the total range of salt intakes may be relatively narrow. This problem is discussed further in the section dealing with short-term human experiments later in this chapter.

ANOMALOUS POPULATIONS

Cross-cultural studies may highlight anomalous populations that deviate very significantly from the general trend. These anomalous populations may provide useful information that can lead to modification of existing theories or to additional areas of investigation. Observations that Greenland Eskimos had low rates of heart disease despite a diet high in fats of animal origin were important in focusing attention upon the potential dietary and therapeutic

benefits of fish oils (Dyerberg and Bang, 1979; see Chapter 11 for further discussion).

In the UK, the Asian population has higher rates of coronary heart disease than the white population. This is despite their apparently lower exposure to many of the traditional risk markers for coronary heart disease (McKeigue *et al.*, 1985), e.g. they have lower serum cholesterol levels, smoke less, drink less alcohol, consume less saturated fat and cholesterol, and many are lactovegetarian. (See Chapter 16 for further discussion.)

SPECIAL GROUPS

There are 'special groups' within populations who, because of their religion or occupation, behave differently from the majority of the people around them. Such groups are often intensively studied because there should be fewer confounding variables between these groups and the rest of their own population than are seen in other cross-cultural studies. For example, Seventh Day Adventists in the USA are a religious group who abstain from alcohol and tobacco, and around half of them are ovolactovegetarian. They have lower rates of coronary heart disease and many types of cancers (including bowel cancer) than the American population as a whole (e.g. Fraser, 1988).

Despite the use of such groups, because of better matching than in most cross-cultural studies, there may still be numerous differences in the diet and lifestyle of these groups compared to, say, the average American. Picking out one dietary or other lifestyle difference between such groups and the rest of population may still produce false assumptions about cause and effect. A religion may 'impose' numerous lifestyle and dietary constraints upon its adherents, religious groups may be socially, intellectually or even genetically unrepresentative of the population as a whole and, of course, their religious faith may provide emotional comfort that lessens the impact of stresses upon mental and physical health. One report even suggests that being religious may in itself protect against the development of bowel cancer (Kune *et al.*, 1993).

Time trends

Populations' diets and lifestyles tend to change with the passage of time. These changes may be particularly rapid and pronounced if there has been rapid

- The use of cross-cultural comparisons to identify associations between diet and disease risk is particularly prone to confounding effects.
- Increasing affluence and industrialization bring with them multiple changes in diet composition and lifestyle, as well as major changes in life expectancy and disease frequencies.
- Several associations between diet and disease risk factors that are consistently found in between-population studies are not apparent in samples of individuals from within the same population, e.g. the association between daily salt intake and blood pressure. Some of the possible reasons for this dichotomy are:

measures of the average nutrient intakes of populations may be much more valid than measures of the habitual intakes of individuals
 within populations there is likely to a much narrower range of nutrient intakes and probably of risk factor levels
 many factors other than the dietary variable may influence an individual's risk factor level and so obscure the relationship.

- In cross-cultural studies, anomalous populations that deviate markedly from the general trend may open up new areas of investigation or generate new hypotheses.
- Special groups within populations whose diet and behaviour differ markedly from the bulk of the population are of special interest. There are likely to be fewer confounding variables than with other cross-cultural comparisons.

economic development and industrialization. Epidemiologists may look for associations between the changes in a population diet and its mortality and morbidity rates for particular diseases. Seasonal changes may also be important if an acute event is being studied.

In the USA and Western Europe, salt consumption declined during the twentieth century as salting was largely replaced as a means of preserving food by refrigeration and other preservation methods. This decline in salt consumption in these countries was associated with reduced mortality rates for stroke and gastric cancer (note that high blood pressure is a risk factor for stroke). These observations have been used to support the proposition that salt is an aetiological influence for both of these diseases (Joossens and Geboers, 1981). Figure 3.9 shows a very strong correlation between average death rates from stroke and gastric cancer in 12 industrialized countries between 1955 and 1972. This is consistent with the hypothesis that a change in exposure to a common causal influence was responsible for the decline in mortality from both diseases, i.e. reduced intake of dietary salt. The apparent, relative acceleration in the rate of stroke mortality after 1972 is attributed to the

population impact of antihypertensive drugs, which would be expected to reduce mortality from stroke but not gastric cancer. MacGregor and de Wardener (1999) give several further examples of associations between falling salt consumption and falling blood pressure, but the decline in salt intake in these studies was partly induced by active publicity campaigns rather than being wholly spontaneous, and so they cannot be classified as descriptive epidemiological studies.

The Japanese are frequently used by those investigating links between diet and disease as an example of a nation that, though affluent and industrialized, has a diet very different from the typical American and Western European diet. The Japanese diet has much less meat and dairy produce, but much more vegetable matter and seafood. Since World War II, however, the Japanese diet has undergone considerable 'Westernization'. For example, since 1950 total fat intake has trebled; the ratio of animal to vegetable fat has doubled; and American-style fast food outlets have mushroomed. The mortality patterns of the Japanese have also changed over this time scale. Two-thirds of deaths in Japan are now due to cardiovascular diseases, cancer and diabetes, whereas in 1950

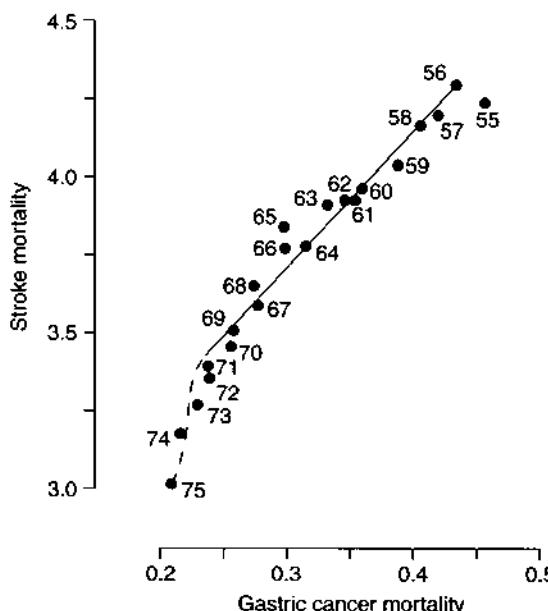


Figure 3.9 Time trends in mortality rates from gastric cancer and stroke over 20 years. Each point represents average death rates from these diseases in 12 industrialized countries for that year, e.g. 55 represents rates in 1955. From Webb (1992a), simplified from Joossens and Geboers (1981).

only one-third were due to these causes. Deaths from cancer have increased two and a half times since 1950, as have deaths from heart disease.

These changes in Japanese mortality patterns are consistent with current views on the causal influences of diet upon particular chronic diseases. Note, however, that Japanese boys are now more than 20% taller than they were in 1950, and the total life expectancy of the Japanese has also shown large increases over the same time scale, making them now one of the most long-lived populations in the world. Fujita (1992) reviewed these changes in the Japanese

diet since World War II and concluded that 'they have greatly improved the nutritional status of the Japanese people'. He further suggested that the rise in crude mortality rates from the diseases of industrialization is 'mainly due to increases in the actual numbers of elderly people, rather than to westernising effects in dietary habits'. Of course, this does not mean that the current Japanese diet is optimal, and it certainly does not mean that further 'Westernization' of the Japanese diet will lead to further benefits, but the example does illustrate the potential hazards of focusing too narrowly upon individual causes of mortality or morbidity.

Migration studies

One possible explanation of the differences between the morbidity and mortality patterns of different nations is that they are due to genetic differences between the populations. Studies on migrants are a useful means of distinguishing between environmental and genetic influences on disease frequency. People who migrate are immediately exposed to the new environmental conditions and they also tend to acculturate, i.e. they progressively adopt aspects of the dietary and lifestyle habits of the indigenous population. However, unless widespread intermarriage occurs, migrants retain the genetic characteristics of their country of origin. Migrant populations generally also tend to progressively acquire the morbidity and mortality characteristics of their new homeland (Barker *et al.*, 1998). These trends indicate that, although there are undoubtedly variations in genetic susceptibility to different diseases both within and between populations, environmental factors have a predominant influence upon disease rates in populations. The differences in the mortality and morbidity patterns in different nations are predominantly

- Temporal changes in disease frequency can be associated with corresponding changes in diet in genetically stable populations.
- These changes may take place rapidly if there is rapid industrialization in a country.
- Many of the problems of cross-cultural studies also apply to the analysis of time trends.
- Increases in the proportion of deaths due to chronic diseases are inevitable when the life expectancy of a population rises.

- Migration is accompanied by immediate changes in the environment and usually more gradual changes in diet as migrants acculturate to their new surroundings.
- Unless there is extensive intermarriage, migrant populations remain genetically similar to those in their country of origin.
- Migration is usually accompanied by a gradual shift in disease frequencies away from that in the country of origin and towards that typical of the new country.
- Migration studies suggest that many of the differences in disease frequencies seen between populations are due to environmental and lifestyle factors rather than to genetic differences between races.

due to differences in environment, lifestyle and diet, rather than to the genetic differences between populations. In some instances where the change in risk exposure occurs at the time of migration, the time lag between migration and change in disease frequency may suggest the length of the latent period between the initiating event and the appearance of symptoms.

The mortality rate for stroke has for a long time been much higher in Japan than in the USA, whereas mortality from coronary heart disease has been much higher in the USA than in Japan. In Japanese migrants to the USA and their descendants, mortality rates for these diseases move over time towards those typical of other Americans. This would indicate that these differences in mortality rates in the two countries are due to diet or other lifestyle factors and thus that they are potentially alterable by intervention.

Cohort studies

In a **cohort study**, a sample of people (or cohort) is selected either from within the same population or, less commonly, from several populations. Details of the individual diets, lifestyles and other characteristics of all members of the cohort are then recorded and their subsequent health and mortality are monitored for a period of years or even decades. The sample is selected and the data are collected with the aim of testing an hypothesis about the cause of a disease. The investigators then look for associations between the measured characteristics and the subsequent risks of mortality or morbidity. They

look for dietary, lifestyle or other characteristics that are predictive of increased risk of a particular disease. To use a simple non-health example, a university could record the entry qualifications of its new students and see if these predict the eventual level of achievement of students at the end of their course. Is there a significant association between level of entry qualification and final level of attainment?

Cohort studies usually require large sample groups and long follow-up times in order to obtain enough cases of a disease for meaningful statistical analysis, e.g. a sample of 100 000 middle-aged northern Europeans would be required in order to obtain 150 cases of colon cancer within a 5-year follow-up period. Cohort studies inevitably require major commitments of time, effort and money. They have, in the past, been considered impractical for all but the most common of diseases because, for less common diseases, impossibly long follow-up times or impossibly large sample groups are required in order to gather enough cases for statistical analysis. In recent decades, however, several vast cohort studies have been instigated using tens or even hundreds of thousands of subjects. The European Prospective Investigation of Cancer is using a cohort of 400 000 people from nine countries across Europe (see Bingham, 1996). Such a study is not limited to investigating any one diet-disease relationship; it can potentially be used to provide data on dozens of such relationships. For example, data from the American Nurses Study have been used to investigate several relationships between environmental factors and diseases. Some of these are discussed in this book and listed in the references (see examples below):

- obesity and coronary heart disease (Manson *et al.*, 1990)
- vitamin E and coronary heart disease (Stampfer *et al.*, 1993)
- meat, fat, fibre and colon or breast cancer (Willett *et al.*, 1990)
- *trans* fatty acids and coronary heart disease (Willett *et al.*, 1993).

Many problems of such studies are common to other areas of epidemiology, such as:

- the accuracy of the end-point diagnosis
- the reliability, range and validity of the original measurements (note that several long-running, sophisticated cohort studies make repeat measurements on surviving members of the cohort at regular intervals)
- the problem of confounding variables.

Morris *et al.* (1980) asked 18 000 middle-aged British civil servants to complete a questionnaire on a Monday morning detailing, amongst other things, a 5 minute by 5 minute record of how they spent the previous Friday and Saturday. They found that, over the next 8 years, those who had reported engaging in vigorous leisure time activity had only half the incidence of coronary heart disease of those who reported no such vigorous activity (even allowing for differences in smoking of the two groups). This finding is consistent with the hypothesis that exercise reduces the risk of coronary heart disease. It does not prove that the differences in heart disease rates were caused by differences in exercise levels. There are several other possible explanations, such as:

- Perhaps those best equipped for vigorous physical activity and thus more inclined to participate are also less prone to coronary heart disease.
- Perhaps a reluctance to participate in vigorous leisure time activity is an early indication of existing but undiagnosed coronary heart disease.
- Perhaps participation in leisure time activity is a marker for those who are most health conscious and who eat healthier diets and generally lead healthier lifestyles.

When cohort studies are used to assess the relationships between exposure to an environmental factor and disease, the term **relative risk** is widely used:

$$\text{Relative risk} = \frac{\text{incidence in exposed group}}{\text{incidence in unexposed group}}$$

In the example of Morris *et al.* (1980), the relative risk of coronary heart disease (CHD) in the inactive smokers could be compared to that of the active non-smokers:

$$\text{Relative risk} = \frac{\text{CHD incidence in inactive smokers}}{\text{CHD incidence in active non-smokers}}$$

In many dietary studies, the population is divided up into fifths (or **quintiles**: see Centile in Glossary) according to their level of consumption of a particular nutrient, i.e. the fifth with the lowest consumption (or exposure) through to the fifth with the highest consumption. The relative risk in each quintile can then be expressed in relation to that of the lowest quintile, for example:

- Cohort studies are expensive and time consuming because they require large samples of subjects to be followed for long periods of time.
- Cohort studies are impractical for investigating the causes of uncommon diseases, but can be used to investigate the causes of several diseases simultaneously.
- Deaths in the first few years of the study may represent those with existing disease at the start of the study and may need to be excluded from the analysis.
- Relative risk is the ratio of the incidence of disease in the exposed group to that in the unexposed group.
- Relative risk can be calculated in quintiles of the population, divided according to their level of exposure to the independent variable (e.g. the dietary factor). The incidence in each quintile is expressed as a ratio of that in the lowest quintile of exposure.

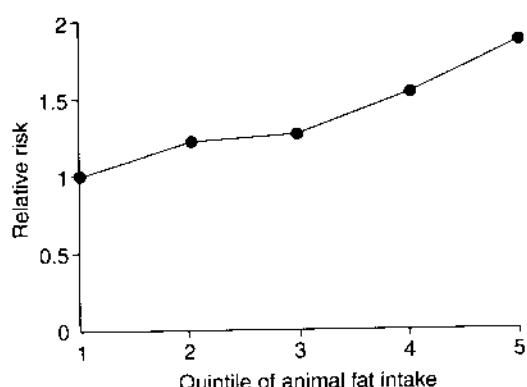


Figure 3.10 Relative risk of colon cancer according to quintile of animal fat intake. A cohort of almost 90 000 American nurses was followed for 6 years and there were 150 recorded cases. Simplified from Willett *et al.* (1990).

$$\text{Relative risk} = \frac{\text{incidence of disease in any other quintile}}{\text{incidence in lowest quintile}}$$

Figure 3.10 shows the relative risk of developing colon cancer in differing quintiles of a cohort of almost 90 000 American nurses divided according to their consumption of animal fat (Willett *et al.*, 1990). The results indicate that there is a progressive increase in the risk of colon cancer associated with increased animal fat consumption; the effect was statistically significant.

Positive results from an apparently rigorous cohort study are usually given particular weight when conclusions are made about relationships between lifestyle factors and disease.

Case-control studies

In **case-control studies**, investigators try to identify differences in the diet, lifestyle or other characteristics of matched groups of disease sufferers and controls. Often, such studies are retrospective. A group of those suffering from or dying of a particular disease is matched with a group who can act as controls. The investigators then try retrospectively to identify differences in the past diets, behaviours or occupations of the two groups that might account for the presence of the disease in the case group. Returning to the example of predicting final degree results from entry qualifications: a case-control study could be used to identify the characteristics of students who

have attained first-class degrees. One could take a group of students attaining first-class degrees (cases) and match them with some obtaining lower grades (controls) and see if the entry or other characteristics of the two groups were different.

Matching of cases and controls is clearly a critical factor in such investigations. Sample sizes are usually small, which means that testing and correction for the effects of confounding variables cannot be undertaken in the final analysis of the results. Ideally, one would match only for confounding variables, i.e. those that are independently linked to both the suspected cause and the disease under investigation. Matching for factors that are linked to the cause but not independently to the disease will actually tend to obscure any relationship between the suspected cause and the disease. Barker *et al.* (1998) use the example of a suspected link between taking oral contraceptives and deep-vein thrombosis. Religious belief may be linked to the use of contraceptives, but is unlikely to be causally linked to the disease. Barker *et al.* make the point that 'perfect matching of influences which determine exposure to the suspected cause will result in the frequency of exposure in cases and controls becoming identical'.

In addition to the problem of control selection, it may be extremely difficult to obtain reliable information about past behaviour, especially past dietary practices. Present behaviour may be an unreliable guide to past behaviour because the disease or awareness of it may modify current behaviour; differences between controls and cases may be as a result of the disease rather than its cause. Some diseases, such as cancer, are probably initiated many years before overt clinical symptoms appear. This may make it desirable to compare the behaviours of the two groups many years previously. If one tries to assess things like past smoking habits or the position in which a baby was laid down to sleep, one could expect reasonably accurate responses. If one wants to assess habitual salt intake 5 years ago, this will present a much greater challenge.

Case-control studies helped to highlight the apparently causal link between prone (front) sleeping and risk of cot death. More than 20 case-control studies were published over a 20-year period from the early 1970s. Each of these found that babies who had died from cot death were more likely to have slept on their fronts than those who did not die (see Department of Health, 1993; Webb, 1995).

Cramer *et al.* (1989) tried to test the hypothesis that high dietary galactose consumption (from lactose or milk sugar) was implicated in the aetiology of ovarian cancer. Approximately equal numbers of sufferers and non-sufferers from the disease were asked to fill in a food frequency questionnaire, focusing upon the dietary patterns in the previous 5 years but asking the 'cases' to ignore changes in food preferences since their diagnosis of cancer. Lactose consumption was not significantly different in the cases and controls, and there were no significant differences when several other specific dairy foods were considered separately. However, eating yoghurt at least once a month was associated with significantly increased relative risk of ovarian cancer. The implication was that yoghurt consumption is causally linked to the development of ovarian cancer, but there are many other explanations for these results, such as those below.

- Yoghurt is a food with a 'healthy image' and ovarian cancer is sometimes called the silent cancer because the disease has often progressed to an advanced stage before its clinical recognition. Perhaps minor subclinical effects of the disease persuaded the cases to include more healthy foods in their diets prior to the onset of clinically overt symptoms.
- The more associations one tests, the more likely it is that one of them will be statistically significant simply by chance. Lactose consumption *per se* was not associated with risk of ovarian cancer, and so these authors then looked for a link with several individual dairy foods.
- The sufferers also differed from the controls in several other respects. The cases were more likely to be Jewish, college educated, never to have been married, never to have had children and never to have used oral contraceptives. It is not unreasonable to suggest that women with such characteristics might also be more likely to eat yoghurt. The differences in yoghurt consumption may be due to some other characteristic of women who develop ovarian cancer, rather than yoghurt consumption being directly, causally linked to the development of ovarian cancer.

Case-control studies may be prospective. Clinical records or stored samples may be used to yield information that is more reliable or at least

less subject to bias than asking questions about past behaviour. Occasionally, such information may be recorded or stored with the specific intention of referring back to it when cases have become apparent. Wald *et al.* (1980) investigated the possibility of a link between vitamin A status and risk of cancer. Serum samples were collected and frozen from 16 000 men over a 3-year period. Eighty-six men from this large sample subsequently developed cancer and were matched with 172 men who did not develop cancer. They were matched for age, smoking habits and time at which the samples were collected. The serum retinol (vitamin A) concentrations were then measured in the stored samples from both groups of men. Serum retinol concentrations were found to be lower in the cases than in the controls – consistent with the hypothesis that poor dietary vitamin A status increases cancer risk. There are several reasons (see below) why these results and those of other similar studies need to be interpreted cautiously.

- The time between sample collection and clinical diagnosis was short (i.e. 1–4 years). It is quite likely that an existing, undiagnosed cancerous condition may have resulted in reduced serum retinol concentration. The difference in serum retinol concentrations of cases and controls may be a product of early cancer rather than its cause.
- Serum retinol concentration is an insensitive measure of dietary status for vitamin A in individuals. Serum retinol concentrations only start to fall when liver stores are seriously depleted. Other factors, such as infection, may influence plasma retinol concentrations.

In cohort studies, relative risk could be calculated directly, but in case-control studies, an indirect measure of relative risk, the **odds ratio**, has to be used. This is defined as the odds of exposure to the suspected cause amongst the cases divided by the odds of exposure amongst the controls:

$$\text{Odds ratio} = \frac{\text{odds of case exposure}}{\text{odds of control exposure}}$$

In most case-control studies this is calculated using the equation:

$$\text{Odds ratio} = \frac{\text{number of exposed cases} \times \text{number of unexposed controls}}{\text{number of unexposed cases} \times \text{number of exposed controls}}$$

If the experiment has been designed using matched pairs of cases and controls, then:

$$\text{Odds ratio} = \frac{\text{number of pairs with case exposed but control unexposed}}{\text{number of pairs with case unexposed but control exposed}}$$

As with relative risk in cohort studies, the odds ratio can be calculated at each of several different levels of exposure, e.g. with quintiles of exposure, as in Figure 3.10.

Case-control studies of some type account for many of the putative links between diet or lifestyle and disease that fleetingly attract the attention of the health correspondents of the media. They can be conducted cheaply and quickly and so are available even to those with very limited research facilities. This type of study is clearly an important tool for the epidemiologist seeking to identify causes of many diseases. Many of these studies relating to diet have major methodological difficulties and are open to several interpretations. Often, these studies do no more than point to the possibility of a causative link that is worthy of further investigation.

'Experiments' of nature

Observations made upon victims of some congenital or acquired disorder may provide useful evidence about the relationships between diet and disease. For example, the observation that people with **familial hypercholesterolaemia** (an inherited tendency to very high plasma cholesterol concentration) are very prone to premature coronary heart disease supports the belief that plasma cholesterol levels and coronary heart disease risk are positively associated. The Nobel Prize-winning work of Brown and Goldstein (1984) indicates that the high plasma cholesterol in these people is a primary consequence of their genetic defect, and this supports the proposition that the association is causal.

Animal experiments

This topic has previously been the subject of two critical reviews by the author (see Webb, 1990, 1992b). Animal experiments allow researchers to perform high-precision, well-controlled and, if necessary, long-term experiments to directly test hypotheses. Technical, ethical or financial constraints would often make it difficult or impossible to undertake such experiments using human subjects. The reliability of well-designed and competently executed animal experiments should be high. The researcher can thus have confidence that a statistically significant difference between, say, an experimental group and a

- In case-control studies, the matching of cases to controls is of paramount importance. Ideally, they need to be matched only for likely confounding variables that are independently linked to both the suspected cause and the outcome measure.
- Ideally, one should compare diets of cases and controls at the time of disease initiation.
- It may be very difficult to obtain reliable measures of past diet. The current diets of disease sufferers may be a product of the disease rather than a cause.
- Stored blood or other samples may be used to give more objective measures of past nutritional status.
- Odds ratio is the measure of relative risk used in case-control studies. In most studies, it is the product of the number of exposed cases and unexposed controls divided by the product of the unexposed cases and the exposed controls.

control group is due to the treatment. It should be repeatable if the experimental circumstances are replicated. Epidemiologists, on the other hand, have to rely upon finding statistical associations between naturalistic observations of diseases or disease markers and potential causes. These are not truly experiments because the investigator does not impose any constraint or intervention upon the subjects to test a hypothesis. These observational human studies cannot be properly controlled. Confounding variables may result in a false assumption that two associated variables are causally linked and, conversely, other influences upon the dependent variable (the disease) may obscure a causative association.

Animal experiments are thus likely to be much more reliable than human epidemiology and even many human experiments. The problem with animal experiments is their questionable **validity**. It may be difficult to decide to what extent hypotheses generated or supported solely by the results of animal experiments may be applied to people. Strictly speaking, animal experiments can only be used to generate hypotheses about humans. The hypothesis should then be tested by direct experiment with humans or, if that is not possible, the extent to which human observations are consistent with the animal-generated hypothesis can be assessed. In practice, however, results from animal experiments may be extrapolated to people with little experimental confirmation in people and sometimes with little consideration of biological factors which may make such projections unsafe. It is a measure of the robustness of animal experimentation that often, despite such lack of rigour, essentially correct conclusions about humans are drawn from these experiments.

There is abundant evidence, for those willing to be convinced, that animal experiments have played a key role in progressing our knowledge of human biology. They have also, therefore, made an important contribution to producing the practical medical, surgical and nutritional benefits that have resulted from the application of that knowledge. Only around 20% of the Nobel Prize winners in Physiology and Medicine have used human subjects for their research, and only around 10% have not made use of some non-human species (Webb, 1992b). These Nobel prizes represent a yardstick to assess the contribution of animal experiments because they acknowledge landmark achievements in advancing physiological and medical knowledge. Almost all

prize winners have made use of non-human species in their research.

There is also clear evidence that animal experiments have sometimes misled those doing research in human biology and medicine, and two examples are given below.

- Dietary cholesterol was once widely regarded as the dominant influence upon plasma cholesterol concentration. Frantz and Moore (1969), in a review of 'the sterol hypothesis of atherogenesis', concluded that, although saturated fat was an important influence upon plasma cholesterol, its effect was probably to potentiate the effects of dietary cholesterol. Most current nutrition education advice suggests that dietary cholesterol is normally a relatively minor influence upon plasma cholesterol concentration. According to Frantz and Moore, experiments with herbivorous, laboratory rabbits were the major evidence directly supporting their view of dietary cholesterol as the dominant influence upon blood cholesterol concentration.
- In Chapter 10 it is argued that experiments with laboratory rats were an important factor in exaggerating the protein requirements of children, and thus in falsely indicating a huge deficit in world protein supplies.

Animal experiments are a vital research tool in human nutrition and the other biomedical sciences. In the past, they have made major contributions to advancing understanding of human biology and will continue to be vital for the foreseeable future. Animal experiments also have a considerable potential to mislead the human nutritionist, especially if results generated in experiments with small laboratory animals are extrapolated to people, with little consideration of the factors that might invalidate such projections and with little attempt to corroborate the animal-generated hypothesis with studies in people.

The following are several problems that may make it difficult to extrapolate the results of nutrition experiments with laboratory animals to people.

SPECIES DIFFERENCES IN NUTRIENT REQUIREMENTS

There may be both qualitative and quantitative differences in the nutritional requirements of different species. Most mammals, with the exception of primates and guinea-pigs, make their own vitamin C

and therefore do not require a dietary supply. All mammals require an exogenous source of protein, but the requirements of children and other young primates are low compared to other mammals because of their slow growth rate. The milk of most mammals has 20–25% of the energy as protein, whereas in humans and chimpanzees, this figure is only around 6% (see Chapter 10).

SPECIES DIFFERENCES IN FEEDING PATTERN AND DIET CHARACTER

Animals may be exposed to a type of diet or a mode of feeding that is totally different from their natural one. Their response in such an experiment may not be a good guide to the response of people, especially if the experimental diet or feeding pattern is much more akin to the natural one in humans. For example, rabbits are herbivores and would not normally encounter cholesterol in their diets because cholesterol is found exclusively in foods of animal origin. Rabbits may be metabolically ill-equipped to handle large dietary loads of cholesterol. They certainly seem to be much more sensitive to the atherogenic effects of cholesterol than many other laboratory species. Omnivorous and carnivorous species should be better adapted to handling dietary loads of cholesterol and thus less susceptible to its atherogenic effects. With the benefit of hindsight, the assumption that the rabbit's response to dietary cholesterol would predict the human response seems unsafe. This partly explains why dietary cholesterol was regarded as a much greater influence on plasma cholesterol and atherosclerosis than it is now.

Some animals spend long periods of time nibbling or grazing, whereas other species consume large amounts of food rapidly in short-duration and relatively infrequent meals. The idea that nibbling regimes might be useful in the treatment of human obesity was largely based upon experiments that monitored the effect of meal feeding upon naturally nibbling species, like rats and mice.

CONTROLLED EXPERIMENTAL CONDITIONS VERSUS REAL LIFE

Animal experiments are usually performed under highly controlled conditions. Experimental and control groups are matched and treated identically except for imposed differences in the experimental variables.

Strictly speaking, one is then only able to assume that the test treatment produces the results obtained under the specified conditions of the experiment. In practice, one tends to assume that essentially similar effects would result from the treatment even if some of the background conditions were changed (e.g. in free-living, genetically diverse, wild animals of the same species). Changes in background conditions might sometimes substantially alter the response of even the experimental species to the experimental treatment, i.e. controlled experiments with laboratory animals might sometimes not even predict the response of free-living wild animals of the same species, let alone the response of human beings.

Take as an example the key role that animal experiments play in the toxicological testing of food additives. Such tests often involve exposing experimental animals to huge doses of single additives over their whole life span under laboratory conditions. If additives show no adverse effects under these test conditions, it is then assumed that they are generally non-toxic to that species in those doses. Under different conditions, however, the additive might have harmful effects. There might be interactions between the additive and some other factor. The additive might, for example, become harmful in one of the following circumstance:

- in the presence of other additives, drugs or chemicals
- when there is a particular nutrient inadequacy or the diet is low in antioxidants
- when there is exposure to cigarette smoke or alcohol
- if animals with differing genotypes are exposed.

There is, for example, a well-known interaction between certain antidepressant drugs (mono amine oxidase (MAO) inhibitors) and certain foods containing tyramine (e.g. cheese). The tyramine causes a dangerous rise in blood pressure in people taking these drugs. There are several other well-documented interactions between drugs and diet.

DIFFERENCES IN SPECIES' SIZES

One of the major advantages of using laboratory animals for experiments is their small size, which greatly reduces the costs. More than 90% of all animals used for experiments in the UK are rodents or rabbits (Webb, 1992b). This small size also leaves

experimenters with the difficulty of how to allow for differences in size when projecting results from animal experiments to humans. It is usual to scale dosages and requirements according to body weight, but this is more because of convention and convenience than because there is convincing evidence of its validity. Earlier in the chapter, it was explained how four-fold differences could be obtained when projecting the rat's rate of vitamin C synthesis to human size, depending upon whether relative body weight or relative metabolic rate was used as the basis for the scaling.

Let us assume that one might wish to model, in a mouse experiment, the sugar consumption of two human populations, one consuming around 40 kg per person per year and one consuming around half this amount. These amounts represent somewhere near 20% and 10% of total energy intake respectively for the human populations. If one scales this human consumption according to relative body weight, to set the experimental consumption in the mice, sugar would only represent around 2.5% and 1.3% of total calories in the two groups of mice.

As discussed earlier in this chapter, the relative metabolic rate (i.e. energy expenditure per unit weight) declines with increasing body size in mammals. Small animals such as rats and mice have much more rapid metabolic rates than larger species such as humans. This accounts for the large differences produced when the scaling is by relative body weight or relative metabolic rate in the two examples above.

Rodents are widely used in studies of energy balance that are directed towards increasing the understanding of human energy balance regulation (see Chapter 7). The size difference in relative metabolic rate is just one of several factors that greatly complicates the human nutritionist's interpretation of energy balance studies conducted with small mammals. It is sometimes difficult to project even qualitatively the results from energy balance studies on rodents to humans. Two examples are given below.

1. The relative energy costs of locomotion (i.e. energy required to move a specific mass a specific distance) declines with increasing body size.

- The use of laboratory animals for experiments extends the range of experiments that are ethically and technically feasible.
- Well-designed animal experiments should yield reliable and repeatable results.
- The results from experiments with laboratory animals may not validly predict the responses of free-living people.
- Animal experiments have made a major contribution to the understanding of human biology and to the practical benefits that have resulted from the application of that knowledge.
- Animal experiments also have the potential to mislead human biologists, particularly if the results are simply extrapolated to humans.
- There are species differences in the quantitative requirements for nutrients and even some differences in the range of nutrients that are essential.
- Different species have different types of diets and different patterns of feeding. The response of a species to a foreign diet or feeding pattern may be a poor guide to the response of a species for which it is normal.
- The effects of a dietary alteration under a set of controlled conditions may not predict what would happen under a different set of background conditions and thus those of free-living animals or people.
- It is very difficult to scale doses or nutrient requirements between species of vastly different sizes. Scaling according to body weight may be misleading.
- It can also be difficult to scale the energy costs of exercise and homeothermy between large and small species. This is a particular problem if small animals like mice are used as models for humans in energy balance studies.

There is also the added complication that, in humans, bipedal locomotion seems to double the energy cost of walking compared to that of a similarly sized quadruped. This makes it difficult to assess the importance of activity level to overall human energy expenditure from small-animal studies.

2. Small animals rely upon heat generation as their principal physiological response to cold, whereas large animals and people rely much more upon heat conservation to maintain body temperature in the cold. When mice are used, the situation may be still further complicated because mice have been shown to become torpid when fasted, i.e. they allow substantial and prolonged falls in body temperature in order to conserve energy (e.g. Webb *et al.*, 1982). Most investigators in the past have assumed that mice are always homeothermic. Such factors would clearly make it difficult to predict the effects of environmental temperature upon human energy needs from experiments with mice. They may also complicate more fundamental decisions about the relative importance of variations in energy intake or expenditure to overall energy balance and thus to the development of obesity (see Chapter 7).

Human experimental studies – an overview

Human experimental studies have been divided up into three categories for ease of discussion – short-term experiments, clinical trials and intervention trials – but these divisions are rather arbitrary. The overall design strategy is similar in the three types of study; the differences are to do with the scale and duration of the study, the nature of the experimental constraints and whether or not the subject is expected to gain therapeutic benefit from the experimental treatment.

Short-term human experiments

Most of the hypotheses relating diet to the risk of chronic disease suggest that prolonged exposure to the dietary risk factor eventually leads to an increased risk of developing clinical disease. Clearly, such hypotheses are not directly testable in acute experiments. However, these diet–disease hypotheses often

involve hypothetical chains of apparently reversible steps between exposure to the initial dietary risk factor and the onset of clinical disease. The diet–heart hypothesis, for example, envisages changes in plasma cholesterol resulting from high intake of dietary saturated fats leading to lipid deposition in artery walls, atherosclerosis and ultimately an increased risk of coronary heart disease (see Chapter 11 for details). It may be possible to test individual steps in such an hypothesis in short-term experiments without exposing the experimental subjects to any significant long-term risks. Keys *et al.* (1959) began a classical series of studies that have had a tremendous influence upon thinking about the relationship between diet and heart disease. They were able to demonstrate that changing the ratio of polyunsaturated to saturated fats in the diet had predictable acute effects upon plasma cholesterol concentration. When subjects switched from diets high in saturated fat and low in polyunsaturated fat to low-saturated-fat, high-polyunsaturated-fat diets (i.e. from low to high P:S ratio), their plasma cholesterol concentration fell (Figure 3.11). All of the subjects showed some decline

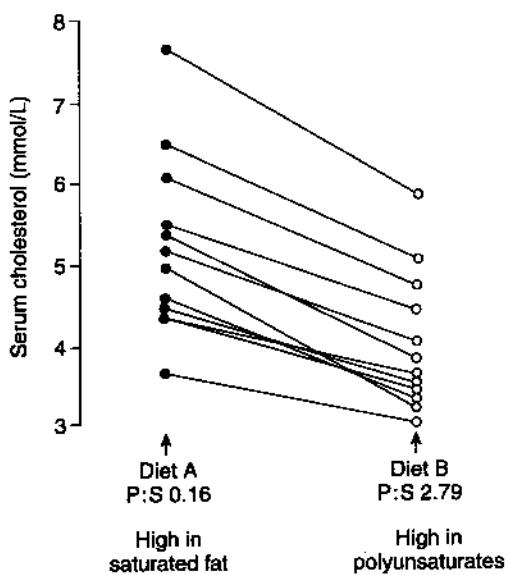


Figure 3.11 Effects of different types of fat on serum cholesterol in 12 men. Total fat in the two diets was matched. Data from Keys *et al.* (1959); reproduced from Webb (1992a). P:S ratio = ratio of polyunsaturated to saturated fatty acids.

in plasma cholesterol, even though the degree of responsiveness varied. The total fat content of the two diets was matched. There are two particular problems with this type of experiment:

- The results of such acute experiments may not necessarily predict the chronic response. The subjects may adapt to any change in the longer term.
- Support for one element of any hypothesis does not confirm the whole hypothesis. In the example above, even if high polyunsaturated-fat, low saturated-fat diets reduce plasma cholesterol, it does not necessarily follow that this will lead to improved health and longevity or even necessarily to a reduced risk of coronary heart disease.

Further consideration of the data in Figure 3.11 may help to explain the apparent contradiction between the results of between-population and within-population studies of the relationship between saturated fat intake and plasma cholesterol concentration. Saturated-fat intake appears to be a major determinant of average population plasma cholesterol, despite the fact that, within populations, there appears to be no significant association between serum cholesterol concentration and saturated-fat intake. In this experiment, the dietary change in P:S ratio has had a cholesterol-lowering effect in these individuals, and the appropriate statistical test (a paired 't' test) would show that this effect was highly statistically significant. This dietary change appears to have shifted the whole distribution of individual serum cholesterol concentrations downwards. If the sample was large enough, one would expect this would significantly reduce average plasma cholesterol of the sample population. However, the

range of cholesterol concentrations of subjects consuming a matched diet, say diet A, is huge and the average reduction in cholesterol resulting from the dietary change to diet B is relatively small in comparison. The difference between the two means is less than a third of the difference between the highest and lowest values on diet A, despite the fact that the magnitude of this experimental dietary change was enormous: it is probably towards the extreme ends of the range of practical feasibility with regard to P:S ratio on a high-fat diet. Although the ratio of polyunsaturated to saturated fat clearly affects individual plasma cholesterol levels, and therefore average population levels, it is not the primary factor determining where any individual lies within the population range of plasma cholesterol concentrations.

Clinical trials

Clinical trials are usually used to test the effectiveness of a treatment on the course or severity of a disease or symptom. The aim of such trials is to isolate and quantify the effect of the treatment under test. It is therefore necessary to eliminate, minimize or allow for other influences on the disease progression or severity, such as those listed below.

- There may be variation in disease progression or severity that is unrelated to the intervention. This requires that there be a matched control group or control period.
- There may be considerable variation in the initial disease severity and this could be a source of bias if

- A dietary factor may contribute to the cause of a disease by initiating a reversible chain of events that eventually results in clinical disease.
- It may be possible to test elements of this chain of events in short-term, human experiments without exposing the subjects to significant risk.
- In the diet heart hypothesis, high saturated-fat intake is postulated to raise plasma cholesterol concentration, which ultimately can lead to increased risk of coronary heart disease; this element of the hypothesis can be tested experimentally.
- The experimental confirmation of one step in the hypothetical chain does not prove the whole hypothesis.
- Acute responses to dietary change may not always be the same as chronic responses.

the control and experimental groups are not initially well matched. If the sample is large enough, simply allocating subjects randomly to control and experimental groups will make it likely that groups will be initially well matched. In some studies, particularly small-scale studies, there may need to be more formal matching for disease severity.

- A **placebo** effect of treatment may occur, i.e. a psychologically based improvement that results from the patient's expectation that treatment will yield benefits. If the outcome measure involves any subjective grading of symptom severity, this placebo effect is likely to be particularly important. Ideally, the control subjects would be given a dummy treatment which they cannot distinguish from the real treatment – a placebo.
- There may be unconscious bias on the part of the operator who measures outcomes and who administers the real and dummy treatments. Ideally, the experimenter should also be unaware of which patients are receiving real and dummy treatments until after the data collection has been completed. This is particularly important if the outcome measure involves any subjective judgement to be made by the experimenter.

When dealing with clinical trials of dietary interventions, there may be particular problems, such as those listed below, that may make it impossible to achieve the design aims listed above.

- A dietary intervention aimed at varying the intake of a particular nutrient will usually involve consequential changes in the intakes of other nutrients.
- If the intervention is a special diet, it may be difficult or impossible to allow for the placebo effect and to eliminate operator bias.
- If the intervention is a special diet, there are likely to be variations in the degree of compliance and it may be difficult or impossible to objectively assess the level of compliance.

One ideal model of a clinical trial is the double-blind, random-crossover trial. Patients are given indistinguishable real or dummy treatments during two consecutive trial periods and some measure of treatment success is made at the end of both treatment periods. Patients would be randomly assigned to receive either real or dummy treatment first. Neither patient nor experimenter knows whether the real or dummy treatment is being given at any particular

time. It may even be possible to ensure uniform compliance by administering the treatment (e.g. a tablet) in the presence of the experimenter. Below is an example to illustrate this type of experimental design.

MacGregor *et al.* (1982) prescribed patients with mild to moderate hypertension a diet designed to be low in sodium. All subjects were required to remain on this experimental diet throughout the 10 weeks of the experiment. After a 2-week adjustment period, the subjects then received either slow-release sodium tablets or placebos during two consecutive monthly periods. The sodium tablets raised salt intake and negated the effects of the low-salt diet. The real and dummy tablets were identical and were administered using the double-blind, random-crossover model (see Figure 3.12 for a summary of the experimental protocol). Blood pressure was found to be significantly lower at the end of the placebo period compared to the salt-supplementation period. This

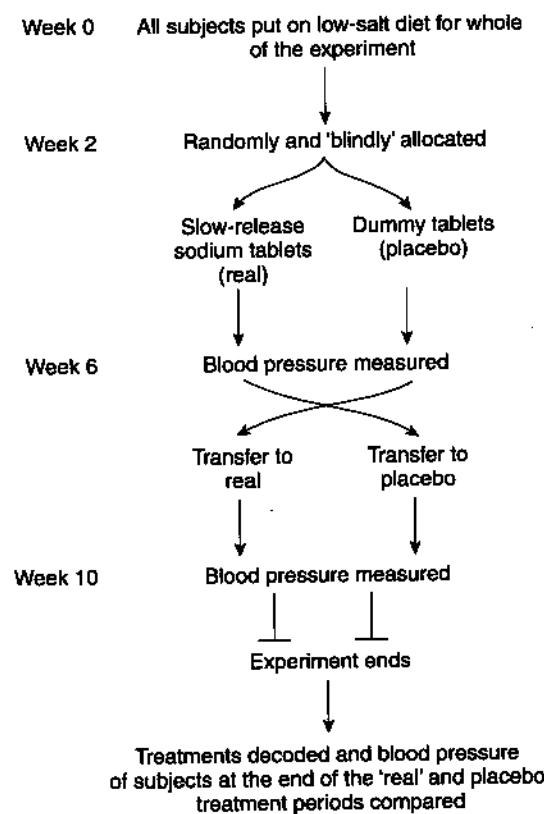


Figure 3.12 A plan of the experimental protocol of Macgregor *et al.* (1982) to illustrate a double-blind, random, cross-over trial.

- Clinical trials aim to isolate and quantify the effects of a treatment upon the course or severity of a disease.
- With large samples, it may be sufficient to randomly allocate subjects to control and test groups, but in small studies there often needs to be more direct matching of subjects for initial severity.
- The use of a dummy treatment (placebo) in the control group can allow for any psychological effects of treatment.
- In the double-blind design, neither patient nor researchers know who has been allocated to the treatment and control groups until the end of the study.
- Clinical trials of dietary interventions have particular difficulties, such as:
 - it may be difficult to produce isolated changes in diet
 - it may be difficult to design a placebo treatment that is not easily recognizable by either subjects or researchers
 - there may be variation in compliance with the prescribed diet that may be difficult to quantify.

supports the hypothesis that moderate salt restriction can have at least some short-term effect of lowering blood pressure in this category of patient.

In Chapter 14, a large, double-blind clinical trial is described which was set up to test whether folic acid supplements given to high-risk pregnant women reduced the risk of babies being born with **neural tube defects**. The subjects were almost 2000 women, from seven countries and 33 centres, who were at high risk of having babies with neural tube defects (because of a previously affected pregnancy). They were randomly assigned to receive one of four treatments starting when they planned to become pregnant again:

- folic acid supplements
- other vitamin supplements
- both of these
- neither of these, just a placebo.

There were many less-affected babies in the two folic acid-supplemented groups compared to the two groups not receiving folic acid supplements (MRC, 1991).

Intervention trials

These are essentially long-term, population-level experiments. The design aim is to make the intended variable or variables the only difference between the matched control and experimental groups. The scale and duration of most intervention trials and perhaps

also the nature of the intervention itself may make it impossible to impose the level of rigour in design that can be accomplished with small-scale, short-term experiments or clinical trials.

Some of the problems that may be encountered when designing intervention trials are listed below.

SELECTION OF SUBJECTS

In some intervention trials, a 'high-risk' group may be used as subjects. Even if an intervention is convincingly shown to benefit such subjects, this does not necessarily mean that the intervention will yield net benefit to the population as a whole.

ETHICS OF WITHHOLDING TREATMENT OR INFORMATION FROM THE CONTROL GROUP

When such high-risk subjects are used, there may be an ethical dilemma about how to treat the control group. If they are made aware of their high-risk status, control groups may significantly alter their behaviour during the course of the experiment.

MULTIPLE INTERVENTION

Most of the chronic diseases of industrialization are assumed to have a multifactorial aetiology. Multiple risk factor interventions may be used to maximize the likely benefits of intervention. In some cases, the experimental aspects of the trial may be given low priority

because the primary purpose is to promote health within a population rather than to test the efficacy of the intervention. Using such multiple interventions may make it impossible to quantify the contribution of any particular intervention to the overall benefits (or harm) demonstrated for the experimental group. Factorial design of experiments can aid in this respect, but may not be practical in many trials. Note that the previously discussed clinical trial of folic acid supplementation in pregnant women is an example of a factorial design. With such a design, it is possible to statistically test the effects of several interventions and also to test for any possible interaction between interventions.

MEASURE OF OUTCOME

An intervention trial may be designed to test the effect of the intervention upon a particular disease risk marker or upon morbidity or mortality from a particular disease. Even if the intervention produces the expected beneficial effect in such cases, the ultimate holistic test of the benefit of the intervention is whether such narrow benefits result in a reduction in total morbidity or mortality. Statistically significant reductions in these holistic measures may require impossibly large subject groups and long periods of study.

Passmore and Eastwood (1986) summarized studies on the effects of water fluoridation upon the dental health of UK children. These come closer to an 'ideal' design than most intervention trials. In 1956, three experimental areas of the UK started to fluoridate their water supply, whilst three neighbouring areas were used as controls. The prevalence of caries in the teeth of the children in the control areas was similar to that of the experimental areas prior to fluoridation, but by 1961 the experimental areas had shown a considerable relative improvement. One of the experimental areas later discontinued fluoridation and, within 5 years, caries prevalence had risen back up to the level in the control areas. This trial provided very convincing evidence of the benefit of fluoridation to the dental health of children. The controversy that has prevented widespread fluoridation of water supplies in the UK revolves around its long-term safety rather than its efficacy in preventing dental caries. Absolute, long-term safety is, of course, much more difficult to demonstrate convincingly.

In the Multiple Risk Factor Intervention Trial (1982), a third of a million American men were screened to identify 13 000 who were classified as at 'high risk' of coronary heart disease (often referred to as the MR FIT study). These high-risk men were then randomly assigned to experimental and control groups. The experimental group received intensive dietary counselling aimed at the normalization of both plasma cholesterol concentration and body weight. They received aggressive treatment for hypertension and intensive counselling to help them reduce their use of tobacco. The control group received no counselling but they, and their personal physicians, were advised of their high-risk status and the results of annual physical examinations sent to their physicians. They were classified as 'usual care'. The intervention was apparently very successful in modifying behaviour and in producing measurable reductions in the objective risk markers (e.g. plasma cholesterol concentration). However, the study failed to show any beneficial effects of intervention, i.e. there was no difference in either total mortality or even coronary heart disease mortality between the two groups after 7 years of follow-up. This failure has been partly explained by the behaviour of the control group, who also appeared to modify their behaviour once they became aware of their high-risk status. It has also been suggested that harmful effects of the antihypertensive therapy may have cancelled out and obscured the beneficial effects of other aspects of the intervention.

Both of these trials could be described as primary intervention trials – they aimed to prevent the onset of disease in asymptomatic subjects. Rarely have such trials produced significant reductions in total mortality when dietary interventions have been used. Other trials are described as secondary intervention trials because they use subjects who have already experienced a disease event, e.g. have already had a myocardial infarction ('heart attack'). These subjects are an extreme example of a high-risk group and one would need to be particularly cautious about assuming that any benefit demonstrated in such a group would have net beneficial effects on the population as a whole. Burr *et al.* (1991) tested the effects of three dietary interventions upon total 2-year mortality in 2000 men who had recovered from a myocardial infarction. The three dietary interventions were advice to:

- increase cereal fibre consumption
- reduce total fat intake
- eat oily fish or take fish-oil capsules twice weekly.

The subjects were randomly allocated to receive or not to receive advice on each of these three interventions, and the allocations were made independently of allocations for the other two interventions. Thus, there were eight possible combinations of interventions or non-interventions, including a group who received no intervention at all. The fish-oil intervention almost uniquely, for

a dietary intervention, did produce a significant reduction in total all-cause mortality. One suggestion as to why this trial alone has produced a statistically significant fall in total mortality is because there appeared to be almost no tendency for non-fish-advised men to increase their fish consumption. At the time, the potential therapeutic effects of fish oils were not widely known in the UK. In many other intervention trials, control subjects have also tended to modify their behaviour in the direction of that counselled for the intervention group.

- Intervention trials are long-term, population-level experiments.
- The scale and duration of these studies and the nature of the interventions may pose considerable problems in designing these studies.
- If high-risk subjects are used, it cannot be assumed that any net benefit demonstrated in the study will translate to a net benefit for the general population.
- It may pose ethical problems if high-risk control subjects are denied the intervention that is expected to be beneficial.
- It will undermine the study if control subjects alter their behaviour in the direction of that suggested for the experimental groups.
- If multiple interventions are used, it may be difficult to decide which intervention has produced any study benefits.
- A reduction in mortality for a particular disease or disease risk factor does not necessarily mean that the intervention has produced a net benefit.
- Statistically significant reductions in total mortality may require impossibly large groups and long periods of study. Primary dietary interventions have rarely produced statistically significant reductions in total mortality.
- A secondary intervention trial uses subjects who have already experienced a disease event.

Dietary guidelines and recommendations

The range of 'expert reports' and their consistency	115	Alcohol	122
Variations in the presentation of guidelines and recommendations	117	How do current UK diets compare with 'ideal' intakes?	124
'Food' recommendations	118	Other nutrients	125
Energy and body weight	119	Willingness to change	126
Recommendations for fats, carbohydrates, protein and salt	120	Some barriers to dietary change	128
		Aids to food selection	129
		Concluding remarks	134

THE RANGE OF 'EXPERT REPORTS' AND THEIR CONSISTENCY

Over the past couple of decades there have been dozens of expert reports published in the industrialized countries that have produced sets of dietary guidelines or recommended changes in national diets. These guidelines and recommendations have been aimed at reducing morbidity and mortality from the diseases of industrialization and so ultimately at promoting health and longevity. Prior to this, most published dietary recommendations and guidelines were geared towards assuring adequate intakes of the essential nutrients and preventing the consequences of deficiency. The first fairly comprehensive standards of adequacy were published as early as 1943 in the USA (NRC, 1943). For those readers interested in international comparisons, Truswell (1999) gives a review of dietary guidelines around the world.

Some of these sets of dietary guidelines have focused upon one aspect of diet and health. For example, in the UK, COMA (1994a) focused upon '*Nutritional aspects of cardiovascular disease*', whereas COMA (1998) was concerned with '*Nutritional aspects of the development of cancer*', as were the guidelines of the American Cancer Society (1997).

Other reports have attempted to synthesize current ideas on the relationship between diet and individual diseases and produced more general and comprehensive guidelines aimed at 'reducing chronic disease risk' and promoting health, e.g. the National Research Council (NRC, 1989b) in the USA and COMA (1991) in the UK.

These many reports from around the world inevitably vary somewhat in the way the recommendations are framed, the scope of their recommendations, the precise quantitative targets and the priority attached to the various recommendations. Nevertheless, in general qualitative terms, there is a striking level of agreement and consistency between almost all of these reports; examples of fundamental disagreement are rare. This certainly adds a cumulative weight to these recommendations and guidelines. Most of these reports directly or indirectly recommend the changes listed below, or at least make recommendations that are consistent with them.

- Maintain **body mass index (BMI)** within the ideal range (i.e. avoid excessive weight gain), either by restricting energy intake and/or increasing energy expenditure (exercise).
- Eat a variety of foods, with ample amounts of starchy, fibre-rich foods and plenty of fruits and vegetables.

- Reduce the proportion of fat, especially saturated fat, in the diet and perhaps reduce or do not increase cholesterol intake.
- Reduce salt consumption.
- Reduce or do not increase the consumption of added sugars (i.e. those not naturally present in fruit, vegetables and milk).
- Limit the consumption of alcohol.

In the UK, the report of the National Advisory Committee on Nutrition Education (NACNE, 1983) had a great influence on nutrition education. This committee, for the first time in the UK, attempted to offer a comprehensive set of quantitative dietary targets for the UK population. Its publication generated considerable debate and controversy in the UK and this debate helped focus attention upon the relationships between diet and chronic degenerative disease. NACNE suggested that its full recommendations could be implemented within 15 years (by 1998) and that a third of the recommended changes could be achieved within 5 years (by around 1988). In retrospect, several of the NACNE targets now look unrealistic to try to achieve within a 15-year time-scale, for example, their aim of reducing total fat intake by almost a quarter and almost halving intakes of saturated fat and added sugars. NACNE's short-term target for total dietary fat (i.e. to reduce total fat to no more than about 35% of food energy by 1988) appeared in the Department of Health report on *The health of the nation* as a target for the year 2005! (Department of Health, 1992). The COMA reports on *Dietary reference values*, *Nutritional aspects of cardiovascular disease* and *Nutritional aspects of the development of cancer* (COMA, 1991, 1994a and 1998, respectively) are currently (early 2001) regarded as the authoritative sources of UK dietary recommendations and guidelines. In this chapter, I use a synthesis of the recommendations in these three reports as the current UK position.

In the USA, '*Dietary goals for the United States*', published in 1977, was the first report to focus specifically upon reducing the risk of chronic degenerative disease. It also aroused considerable critical comment and discussion at the time of publication. According to Truswell (1999), one collection of commentaries upon this report contains almost 900 pages. Some of the criticisms of this report seem very familiar and many of them are still commonly made about all reports of this type. Some examples of these criticisms are listed below (after Truswell, 1999).

- Technical: some of the recommendations are wrong, too extreme or some important issues are not adequately covered. In this particular report, there were criticisms that the target for salt was too ambitious and that there were no targets for obesity.
- The report is premature and based upon inadequate scientific evidence.
- The recommendations might not be appropriate for the whole population. In particular, the guidelines might be more appropriate for overweight, middle-aged males than for active, rapidly growing children and perhaps even for women.
- The report is unnecessary because health and life expectancy are already improving.
- The recommended changes might have serious negative impact upon important industries such as the sugar industry and the dairy industry.
- There may be a general resentment that governments and their experts are trying to tell people what to eat.

The National Research Council report of 1989 (NRC, 1989b) offered a comprehensive set of dietary recommendations and goals that are still the basis for health promotion in the USA.

The World Health Organization (WHO, 1990) has also published a set of nutritional goals in the form of suggested minimum–maximum ranges for selected nutrients. The minimum suggested intake of the nutrient should be sufficient to prevent deficiency, whilst the maximum should be consistent with the aim of reducing the risk of chronic degenerative disease. These goals were intended to have worldwide application and are summarized below.

- Total fat should make up between 15% and 30% of energy intake and countries consuming high amounts of fat should ultimately work towards the lower end of this range.
- Saturated fatty acids should not exceed 10% of total energy, with no lower limit.
- Polyunsaturated fatty acids should make up between 3% and 7% of energy intake.
- Cholesterol intake should be below 300 mg/day, with no lower limit.
- Carbohydrate should provide 55–75% of dietary energy and complex carbohydrate (starch) 50–75%.
- Dietary fibre intake should be 27–40 g/day (which corresponds to approximately 20–25 g of non-starch polysaccharide).

- In recent decades, the focus of nutrition education has been on preventing chronic disease rather than simply ensuring dietary adequacy.
- Many sets of recommendations and guidelines have been produced in the last 25 years and there is a remarkable consistency about the general qualitative changes that they recommend for industrialized countries.
- Consumers should eat more starch, fibre, fruits and vegetables, but less fat, saturated fat, added sugar and salt. Excessive alcohol consumption should be avoided.
- Criticisms of these reports, particularly the earlier ones have been on the following grounds:
 - a recommendation is wrong or impractical
 - a recommendation is premature and there is insufficient evidence to justify it
 - a recommendation may be inappropriate for some sectors of the population
 - implementing the recommendation(s) will have adverse economic consequences
 - governments should not try to control what people eat.
- The WHO has published a set of nutritional goals that are intended to have worldwide application.

- Added sugars should contribute no more than 10% of dietary energy, with no lower limit.
- Protein from mixed sources should provide 10–15% of dietary energy.
- Salt intake should be less than 6 g/day, with no lower limit specified (although some small amount of salt intake is essential).

carbohydrates and proteins), these recommendations are usually framed as a percentage of the dietary energy, as seen in some of the WHO recommendations listed above. This method of presentation allows the amount of the nutrient to be expressed in a single number that does not have to be adjusted as energy intake varies or qualified by a statement of the energy intake at which it applies. It effectively represents the concentration of the nutrient in the diet. Further discussion of this procedure and the calculations involved can be found in Chapter 6. A recommendation that carbohydrates should be 50% of total energy means that carbohydrates should provide 50% of the total energy consumed. Sometimes, recommendations are expressed as a percentage of the food energy rather than total energy to eliminate the distorting effects of alcohol. Recommendations to Britons to reduce fat to 33% of their total average energy intake is approximately equivalent to 35% of their average food energy, excluding alcohol. Alcohol normally has a diluting effect upon the proportion of energy derived from fat (and carbohydrate and protein).

In the UK, dietary recommendations have usually been framed as targets for average population consumption, for example:

To reduce the average percentage of food energy from about 40% in 1990 to no more than 35% by 2005.

(Department of Health, 1992)

VARIATIONS IN THE PRESENTATION OF GUIDELINES AND RECOMMENDATIONS

Some reports on diet and health frame their advice in the form of simple, qualitative or semi-quantitative recommendations that are easy to understand by the general public, e.g. 'eat five portions of fruits and vegetables each day', 'reduce consumption of fatty foods, especially those rich in saturated fat'. Other, more technical reports aimed at scientists and health professionals usually frame their recommendations in a more quantitative way, like the WHO goals listed earlier. It is expected that health professionals and health promoters will interpret these quantitative ideals and translate them into practical dietary recommendations for their clients.

Some quantitative recommendations, such as those for salt and fibre, are given simply as grams per day. However, for the energy-yielding nutrients (fats,

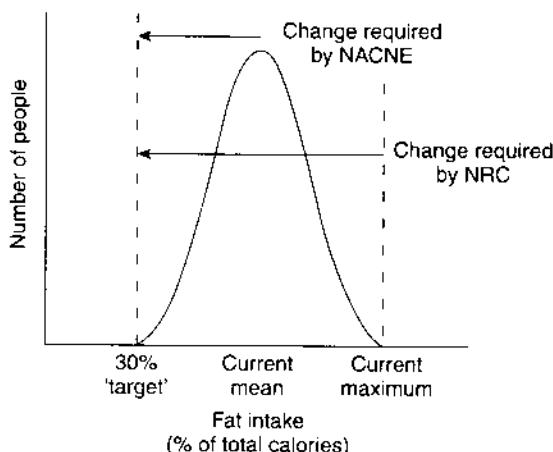


Figure 4.1 The theoretical changes in the current 'normal' distribution of population intakes required to meet the recommended fat intakes of NACNE (1983) and NRC (1989b). NACNE recommended a new population average of 30% of energy from fat, whereas NRC recommended 30% as an individual target.

In the USA, the NRC (1989b) gave their recommendations in the form of individual minimum or maximum intakes:

Fat should make up no more than 30% of an individual's energy intake.

This difference in approach means that, even if a British and NRC recommendation seems to correspond, the NRC target requires greater change than the British one. This is illustrated in Figure 4.1 by

reference to the apparently common 30% target for proportion of total energy from fat in the current NRC guidelines and those of the NACNE committee back in 1983. If individuals are all to meet this target as recommended by NRC, the average population intake must be some way below 30% because of individual variation in fat intakes.

Some recommendations may be targeted at specific groups (e.g. guidelines for pregnant women) or they may exclude particular groups (e.g. children under the age of 5). In this chapter, I generally only consider recommendations that are intended for all healthy people over the age of 5 years. Recommendations for specific groups are dealt with in Part 4.

'FOOD' RECOMMENDATIONS

Consumers in Britain, America and elsewhere are encouraged to eat a variety of foods each day. Different foods contain different profiles of essential nutrients and different profiles of potentially toxic substances. Eating a variety of foods makes it more likely that adequate amounts of all the essential nutrients will be consumed and less likely that any toxins in food will be consumed in hazardous amounts. Consumers in most countries are encouraged to eat foods from different categories or food groups to ensure the adequacy of their diets (see later). Japanese consumers are advised to eat at least 30 foods each day!

- Some reports frame their recommendations in simple qualitative terms that can be interpreted by the general public.
- Other reports offer quantitative targets that health professionals and health promoters need to translate into practical terms for their clients.
- Quantitative fat, carbohydrate and protein targets are usually expressed as a percentage of total energy or food energy (excluding alcohol).
- This method of expression means that the numerical recommendation does not need to be qualified by the energy intake at which it applies and effectively represents the concentration of fat, carbohydrate or protein in the diet.
- In Britain, targets are usually expressed as population averages, whereas in the USA they are given as individual maxima and minima. Even if a British and American target apparently corresponds, this means that the American target will still require substantially greater change.

In both Britain and the USA, consumers are being advised to increase their consumption of fruits and vegetables. More specifically, there are active campaigns in both countries that urge consumers to eat a minimum of five portions of fruit and vegetable each day, with only one of this five being fruit juice. Diets low in fruits and vegetables have been associated in numerous epidemiological studies with elevated rates of cancer and cardiovascular disease. There are several possible reasons why fruits and vegetables might exert a protective effect against these diseases, such as those listed below.

- They are major sources of dietary fibre, including the soluble components of dietary fibre.
- They are low in energy and so they make the diet bulky, which may help in weight control.
- They increase the proportion of dietary energy that is derived from carbohydrate and so reduce the proportion from fat.
- They are the major sources of vitamin C and other antioxidants, e.g. β -carotene and the other carotenoids and non-nutrient antioxidants. These antioxidants inhibit the oxidative damage to cells caused by free radicals that may be important in the aetiology of both cancer and cardiovascular disease.
- They are important sources of potassium, and high potassium intake may help in the prevention of high blood pressure.

A relatively high proportion of nutritionists might be sceptical about the putative health benefits of any one of these compositional changes, but almost all would support the recommendation to increase consumption of fruits and vegetables. The cumulative evidence supporting this food change is overwhelming.

It is recommended in both Britain and the USA that starchy foods like cereals and potatoes should be more prominent in the diet. This would increase the amount of dietary fibre and the proportion of dietary energy coming from starch. This would almost inevitably lead to a reduction in the proportion being derived from fat and/or sugar because there is not much variation in protein intake.

Britons and Americans are encouraged to increase their consumption of oily fish (e.g. mackerel, herring, salmon, trout and sardines) to around two portions per week. These oily fish are the only rich sources of the long-chain n-3 (ω -3) polyunsaturated fatty acids. These fatty acids may have a number of effects that would be regarded as beneficial for many people, such as:

- an anti-inflammatory effect
- an anti-aggregating effect upon blood platelets that reduces the tendency of blood to clot and increases bleeding time
- beneficial effects upon blood lipoprotein profiles.

There is a more detailed discussion of the actions of fish oils in Chapter 11. Concentrated fish-oil supplements are not advised for the general population, although they may be useful for some individuals.

ENERGY AND BODY WEIGHT

Most dietary guidelines emphasize the importance of maintaining an ideal body weight (BMI in the range 20–25) by matching energy intake to expenditure. More than half of British and American adults are either overweight or obese, and these figures have

- People should eat a variety of foods selected from each of the major food groups.
- Dietary variety helps to ensure adequacy and reduces the risk of toxins being consumed in hazardous amounts.
- Five daily portions of fruit and vegetables are recommended in the UK and USA.
- Fruits and vegetables are good sources of dietary fibre, potassium and antioxidants. High fruit and vegetable diets tend to be bulky and low in fat, which may aid weight control.
- Starchy cereals and potatoes should be more prominent in UK and US diets.
- An increased consumption of oily fish is recommended.

been rising alarmingly during the past two decades. Obesity and being overweight are associated with reduced life expectancy, increased illness and disability as well as having social and psychological consequences that further reduce the quality of life. In particular, obesity is associated with an increased risk of type 2 diabetes, hypertension, cardiovascular diseases and several types of cancer. (See Chapter 8 for details on the prevalence and consequences of obesity.)

In both Britain and the USA, increased physical activity is recognized as a key factor in the maintenance of an ideal body weight. In a recent government White Paper, Britons have been advised to be more physically active and to take 30 minutes of activity at least five times a week. There are also specific targets for increasing physical activity for Americans in *Healthy people 2000* (DHHS, 1992), and these are discussed in Chapter 16. Both the US and UK populations have become increasingly sedentary in recent decades. There is clear evidence in Britain that, as a consequence of our increasingly sedentary lifestyle, average energy intakes have fallen substantially in the last few decades (e.g. Prentice and Jebb, 1995). Over the same period, average BMI has increased and the numbers of overweight and obese adults have also increased sharply. A national survey of activity and physical fitness of British adults found very high levels of inactivity and low levels of aerobic fitness amongst the majority of the adult population (Allied Dunbar, 1992; see Chapter 16 for further details). In Chapter 8, it will be argued that inactivity is a major cause of the epidemic of obesity that is sweeping through most industrialized and many developing countries. Increased activity should ultimately lead to a reduction in average BMI and some moderation in the numbers of obese and overweight people in these countries. Concentrating on reducing energy intake still further does not

seem to be the solution to this still-increasing problem of obesity. Increased activity and fitness have many other health benefits, which are discussed in Chapter 16.

RECOMMENDATIONS FOR FATS, CARBOHYDRATES, PROTEIN AND SALT

Targets

- The contribution of fat to total energy should be reduced. In the UK the aim is to get fat down to a population average of 35% of food energy (33% of total energy). The US target is that fat should provide 30% or less of total energy.
- The contribution of saturated fatty acids should be reduced to an average of 10% of total energy (in the USA to a maximum 10% of total energy).
- Polyunsaturated n-6 (ω -6) fatty acid intake should not increase any further. COMA (1991) suggested a population average of 6% of total energy from these fatty acids, with no individual's intake exceeding 10%. Similarly, Americans are advised not to increase their intake of these fatty acids beyond current levels (7% of energy), and a 10% individual maximum is again recommended.
- Cholesterol intakes in the UK should not increase. (In the USA, an intake of 300 mg/day or less is recommended.) Cholesterol is generally regarded, at least in the UK, as a relatively minor influence upon plasma cholesterol. If saturated fat intake is reduced, this may well also cause a drop in cholesterol intake as some foods are important contributors of both cholesterol and saturated fat to the diet.

- Body mass index should be kept within the ideal 20–25 kg/m² range.
- Increased activity is important in the control of body weight, and inactivity is an important cause of obesity.
- Britons eat much less now than 30 years ago, but are now much fatter.
- Increased activity and fitness would have numerous other health benefits.

- COMA (1994a) recommended increasing the intake of long-chain n-3 (ω -3) polyunsaturated fatty acids from an average of 0.1 g/day to 0.2 g/day. This is effectively a recommendation to eat twice as much oily fish.
- COMA (1991) recommended that carbohydrates should provide an average of 47% of total energy (50% of food energy). As American recommendations for total fat are lower than those in Britain, it means that Americans are advised to obtain at least 5% more of their energy from carbohydrates than Britons.
- In Britain it is recommended that added sugars or **non-milk extrinsic sugars** should make up less than 10% of total energy intake. The term non-milk extrinsic sugars covers all of the sugars that are not present in milk or within the cell walls of fruits and vegetables – it is almost synonymous with added sugars, except that it would include the sugar in extracted fruit juices. The intrinsic sugars are those that are found naturally in fruits and vegetables, and Britons are encouraged to eat more of these, along with the complex carbohydrates. NRC (1989b) in the USA merely counselled against any further increase in intakes of added sugars.
- If Britons are to obtain 47% of their total energy from carbohydrate and only 10% from added sugars, it follows that 37% of their energy should come from starches and intrinsic sugars combined.
- COMA (1991) recommended a 50% increase in the intakes of dietary fibre or non-starch polysaccharide, i.e. to aim for a daily intake of about 18 g of non-starch polysaccharide (equivalent to around 25 g of dietary fibre).
- Current protein intakes should be maintained. Both COMA (1991) and NRC (1989b) specifically counsel against taking in more than double the Reference Nutrient Intake (RNI) (Recommended Dietary Allowance, RDA) of protein.
- British and American consumers are advised to make substantial reductions in their intake of salt, with an upper limit of 6 g/day being suggested. This amounts to a reduction of somewhere around a third for Britons.

In Table 4.1, these varied recommendations and targets have been synthesized into a British diet of 'ideal' composition.

Table 4.1 *The composition of a diet that would meet current UK guidelines*

Nutrient	'Ideal' intake
Total fat	33
Saturated fatty acids	10
Polyunsaturated fatty acids	7
Total carbohydrate	47
Added sugars	10
Alcohol	5
Protein	15
Non-starch polysaccharide (g/day)	>18
Salt (g/day)	<6

Values are percentage of total energy, unless otherwise stated.
These recommendations would be consistent with American guidelines
except that there would be a lower fat target of 30% and a corresponding
increase in the proportion of energy derived from carbohydrate.

Rationale

Overall, these changes would be expected to make the diet more bulky and to reduce the amount of energy derived from each gram of food. Reducing the content of fat and sugar and replacing it with foods rich in starch, fibre and also probably with a higher water content would reduce the **energy density** of the diet (amount of energy per gram of food). This bulky and less energy-dense diet should aid weight control and lessen the risks of excessive weight gain and obesity. Such a diet would also have more nutrients per unit of energy (higher nutrient density) because fats and sugars (and alcohol) add a lot of energy but few nutrients.

The changes in fat consumption would be expected to lead to the following benefits.

- Better weight control (fat is palatable, a concentrated source of energy, less satiating than carbohydrate and very efficiently converted to body storage fat).
- Reduced plasma cholesterol concentrations and a corresponding reduction in atherosclerosis and risk of cardiovascular diseases. High intakes of saturated fat raise plasma cholesterol levels, and replacing saturated fat with unsaturated fat reduces plasma cholesterol levels. An elevated plasma cholesterol level is causally linked to an increased risk of coronary heart disease. These changes may also reduce the tendency of blood to clot and form thromboses.

- Reduced risk of some cancers, such as bowel cancer, which are associated with a high-fat, low-carbohydrate diet.

A reduced consumption of added sugars should lead to improved dental health. Both dental decay (caries) and gum disease are strongly linked to high sugar consumption, especially frequent consumption of sugary snacks and drinks between main meals. The priority should therefore be to concentrate on reducing this between-meal sugar and reducing the frequency of consumption of sugary drinks and snacks.

Increased intakes of complex carbohydrate are an inevitable requirement if fat and sugar intakes are to be reduced. Non-starch polysaccharide should additionally lead to improved bowel function, less constipation and a reduced incidence of minor bowel conditions such as haemorrhoids (piles) and diverticulosis. High-fibre diets are also associated with a reduced risk of bowel cancer, and some forms of fibre may have a plasma cholesterol-lowering effect.

A reduction in salt intake is expected to lead to a reduction in average population blood pressure and a

reduced rate of rise of blood pressure with age. This should lead to a reduced incidence of hypertension and ultimately to a reduced incidence of strokes, coronary heart disease and renal failure. High salt intake may also be a factor in the aetiology of stomach cancer and so a reduction in salt intake should further decrease the prevalence of this form of cancer.

ALCOHOL

There is a practically unanimous belief that excessive alcohol intake is damaging to health and that alcohol should therefore be used only in moderation or not at all. What constitutes an acceptable and healthy alcohol intake is a rather more contentious issue. There is persuasive evidence that small amounts of alcohol may have some beneficial effects upon health, but that excessive alcohol consumption has severe health and social consequences. High alcohol consumption leads to an increased risk of liver disease, strokes, fatal accidents, birth defects, hypertension, and several forms of cancer. COMA (1998)

- Dietary targets (population average in the UK; individual targets in the USA):
 - fat: no more than 33% of total energy (30% in USA)
 - saturated fat: no more than 10% of total energy (10% USA)
 - cholesterol intakes should not increase (maximum 300 mg/day in USA)
 - long-chain n-3 (ω -3) polyunsaturated fatty acid intake doubled
 - no increase in n-6 (ω -6) polyunsaturated fatty acids, with an individual maximum of 10% of energy in USA and UK
 - carbohydrates should make up at least 50% of food energy (55% in USA)
 - added sugars should be less than 10% of total energy (USA – no increase)
 - dietary fibre to increase by 50% (USA – increase)
 - salt intake should be cut to 6 g/day (USA – 6 g/day).
- These changes should:
 - produce a diet that is less energy dense but more nutrient dense
 - improve body weight control
 - reduce plasma cholesterol concentrations and so reduce the risk of cardiovascular diseases
 - reduce the risk of bowel and some other cancers
 - improve dental health
 - reduce constipation and other bowel disorders such as haemorrhoids and diverticulosis
 - reduce average blood pressure and the prevalence of hypertension and so lead to reduced incidence of strokes, renal failure and coronary heart disease.

suggested an annual social cost of excessive alcohol consumption in Britain of £2.5 billion.

Many studies have reported higher mortality in those who abstain from alcohol completely as, compared to those consuming moderate amounts of alcohol. High alcohol intakes are, on the other hand, associated with sharply increasing mortality rates, the so-called J-curve (or U-curve) of mortality. It has been argued that this J-curve of mortality is merely an artefact due to the inclusion of reformed alcoholics in the abstaining group. However, there are many reports that support the idea of a protective association between moderate alcohol consumption and, in particular, coronary heart disease and that this remains even after correcting for this effect (e.g. Rimm *et al.*, 1991).

A 10–12-year cohort study conducted in Denmark (Gronbaek *et al.*, 1994) confirmed the J-shaped relationship between total mortality and alcohol consumption, with lowest relative mortality risk in those consuming 1–6 units of alcohol per week. This study also reported that only with weekly alcohol consumption in excess of 42 units per week was total mortality significantly higher than in those consuming low amounts of alcohol. It found that age did not change the relationship between mortality risk and alcohol consumption and that the apparent benefits of small amounts of alcohol applied to all adults. Some other studies in the USA (e.g. Klatsky *et al.*, 1992) have reported that the relative risk of dying as a result of alcohol consumption is higher in younger people than in older people because of the alcohol-related increase in deaths from violence and traffic accidents in younger adults. Gronbaek *et al.* (1994) suggest that this difference in findings between the American and Danish studies may be because accidents and violence are less prominent as causes of death amongst young Danes than amongst young Americans.

With alcohol, the priority is to moderate the intakes of high consumers rather than to aim for a more universal reduction. Despite a substantial body of evidence that alcohol in moderation may have some beneficial effects, expert committees have generally shied away from recommending people to drink alcohol. This reluctance can be explained by the fact that many people find it difficult to gauge and to regulate their intake of alcohol. Back in 1983, NACNE recommended a reduction in alcohol consumption from an average 6% of total energy to 4%

of total energy. They suggested that this should be achieved by curtailing the consumption of high consumers rather than shifting the whole distribution of intakes, as with their other recommendations. The NRC (1989b) recommended that, if people did drink alcohol, they should limit their consumption to no more than about 2 units of alcohol in any one day. A unit of alcohol is defined as 8 g of pure alcohol, which is equivalent to half a pint of ordinary beer (about 250 mL), a small glass of wine (75 mL) or a single measure of spirits. In Britain, the recommended maximum intake of alcohol was 21 units per week for men and 14 units for women. In 1995, a new Department of Health statement rather controversially suggested raising these safe drinking limits to 28 units per week for men and 21 units for women. This alcohol should be spread fairly evenly throughout the week and high consumption on particular days (binge drinking) should be avoided. This Department of Health statement also contained some positive encouragement for older people who do not drink to consider taking 1 or 2 units per day. It is generally recommended that pregnant women should avoid alcohol. Alcohol impairs judgement and co-ordination and slows down reflexes, so should be avoided when driving or participating in any other activity for which such impairments could be hazardous, e.g. operating machinery.

Table 4.2 shows the distribution of alcohol intakes in English adults.

Table 4.2 *The distribution of alcohol intakes in English adults aged over 16 years.*

Amount of alcohol (units per week)	Men (%)	Women (%)
0–1	15	31
1–10 (low)	33	37
10–21	22	17
21–35	15	8
>35	15	5
		2

Values are the percentage of people with alcohol intakes within the weekly range

From Prescott-Clarke and Primatesta (1998).

Mechanisms by which alcohol may reduce the risk of coronary heart disease

Low to moderate alcohol intake seems to be associated with reduced total mortality, largely as a result of reduced mortality from coronary heart disease. Even though heavy drinkers have higher total mortality, this is largely due to causes other than coronary heart disease. Below are listed some suggested mechanisms by which alcohol may protect against coronary heart disease (see COMA, 1998, for primary sources).

- Alcohol intake is associated with increased levels of **high-density lipoproteins (HDLs)** in plasma. This is the cholesterol-containing lipoprotein fraction in plasma that removes excess cholesterol to the liver and affords protection against atherosclerosis and coronary heart disease. (See Chapter 11 for further details on HDLs.)
- Some alcoholic drinks (especially red wine) may contain substances that act as antioxidants, which help to prevent the free-radical damage that is implicated in the aetiology of atherosclerosis.
- Alcohol may reduce the tendency of blood to clot and for thromboses to form. Alcohol consumption is associated with reduced plasma fibrinogen levels and reduced activity of platelets (fibrinogen

is a key protein in blood clotting, and platelet activation is also a key step in blood coagulation).

HOW DO CURRENT UK DIETS COMPARE WITH 'IDEAL' INTAKES?

In 1990, *The dietary and nutritional survey of British adults* was published (Gregory *et al.*, 1990). This survey used a large, representative sample of British adults and directly assessed their food and nutrient intakes using a 7-day weighed inventory. This survey provides the most direct assessment of the 'typical British diet' and the results are summarized in Table 4.3. In order to comply with the ideal British values suggested in Table 4.1, there would need to be substantial reductions in the prominence of fat and saturated fat in the diet from 38% and 16% of energy, respectively, down to 33% and 10%, respectively. Total carbohydrate consumption would need to rise to compensate for the reduction in fat, and within the carbohydrates there would need to be a substantial switch from added sugars to starch and intrinsic sugars. There would need to be a 50% increase in starch consumption. Non-starch polysaccharide intake would need to increase by 50% and salt intake to be reduced by a third. Gregory *et al.* (1990) found that only 12% of men and 15% of women met the

- Alcohol should be consumed in moderation or not at all. Alcohol should be avoided during pregnancy and when driving.
- High alcohol consumption has serious health and social consequences and is associated with steeply increasing mortality. Low alcohol consumption is associated with lower mortality than complete abstention and seems to afford some protection against coronary heart disease.
- The relationship between alcohol consumption and mortality risk is a J-shaped curve.
- Alcohol consumption raises the HDL-cholesterol concentration in blood and reduces the tendency of blood to clot.
- Some alcoholic drinks contain antioxidants that may afford protection against damage by free radicals.
- Recommendations for alcohol vary from 2–3 units per day (14–21 per week) in women and from 2–4 units per day (14–28 per week) in men.
- A unit of alcohol is a small glass of wine, a half-pint of ordinary beer or a standard measure of spirits (about 8 g of pure alcohol).
- About 30% of English men say that they drink more than 21 units per week and 15% of women drink more than 14 units per week.

Table 4.3 The composition of the typical British diet (in 1987)

	Typical British diet
Fat	38
Saturated fat	16
Polyunsaturated fat	6
Protein	15
Total carbohydrate	42
Sugars	18
Alcohol	5
Salt (g/day)	9
Non-starch polysaccharide (g/day)	12

Values are percentage of total energy unless otherwise stated. Data from Gregory *et al.* (1990).

target of 35% of food energy from fat. Only 6% of men and 8% of women had both no more than 35% of food energy from fat and 15% from saturated fat (the 1988 target of NACNE, 1983).

Unfortunately, the fieldwork for this survey was conducted in 1986–7 and so the results summarized in Table 4.3 do not reflect changes that have occurred since that time. The annual *National Food Survey* gives up-to-date information on the composition of household food purchases in Britain, but the numerical data are not directly comparable to those of Gregory *et al.* (1990). Nevertheless, the National Food Survey does suggest the direction and general magnitude of dietary changes that have taken place in Britain since 1986. The 1996 National Food Survey (MAFF, 1997) gives an analysis of trends in fat and

saturated-fat consumption over the 10-year period 1986–96. Over this period there was a decrease in the percentage of household food energy (excluding alcohol, soft drinks and confectionery) derived from fat from 42.6% to 39.7%, and a decline in saturated fat from 17.7% in 1986 to 15.4% in 1996. The decline in saturated fat has been fairly evenly spread over the 10 years, but most of the decline in total fat occurred after 1992. Over this period carbohydrates have increased roughly in line with the reduction in fat, from 44% of household food energy in 1986 to 46% in 1996. There was also a 10% decrease in the energy yield of household food over the 10-year period. There seems to have been no significant increase in the contribution of starch to total energy intake, and intakes of dietary fibre (as non-starch polysaccharide) were practically unchanged.

OTHER NUTRIENTS

The importance of obtaining intakes of all the essential nutrients in line with current dietary standards is generally acknowledged, i.e. of ensuring that the diet contains adequate amounts of energy and all of the essential nutrients.

Calcium and iron

Milk and milk products are the most important sources of calcium in British and American diets.

- In 1986–7, the typical British diet, when compared to current UK guidelines, had:
 - 5% too much of dietary energy as fat
 - 6% too much as saturated fat
 - 5% too little as carbohydrate
 - too much added sugar (about 5% of energy)
 - 3 g per day too much salt
 - only two-thirds of the recommended amount of dietary fibre.
- Since then, the National Food Survey indicates that there has been a continuing steady decline in the proportion of energy as saturated fat and since 1992 a decrease in the proportion from all fats.
- Since 1987, there has been some increase in the proportion of energy from carbohydrate, but no significant increases in starch or dietary fibre.

Meat is a major source of dietary iron in its most readily absorbed form as haem. The NRC (1989b) were concerned that calcium and iron intakes should be maintained despite cuts in saturated-fat consumption. They felt these dual objectives could be achieved if low-fat dairy products, fish, lean meat and poultry were substituted for fatty meats, full-fat dairy products and other fried and fatty foods. Back in 1983, NACNE also recognized the importance of milk in maintaining adequate calcium intakes. They suggested that the wider availability of lower fat milk would allow calcium intakes to be maintained despite the recommended cuts in saturated-fat intake. Low-fat milk has, indeed, become more readily available in the UK, and semi-skimmed milk now accounts for the majority of milk drunk in Britain. The issue of calcium and its importance in bone health is discussed at length in Chapter 13.

Fluoride

The NRC (1989b) specifically recommended an optimal intake of fluoride, especially in children whose teeth are developing. They recommended the use of fluoridated water or dietary supplements where this was not available. COMA (1991) also endorsed the fluoridation of public water supplies up to a level of 1 ppm and set a safe intake of fluoride for infants of 0.05 mg/kg body weight per day. The effects of fluoride on dental health are discussed in Chapter 9.

Potassium

Increased consumption of fruits and vegetables would lead to an increase in potassium intakes.

COMA (1994a) specifically recommended an increase in average potassium intake to about 3.5 g/day in adults. There is evidence that high potassium intake protects against the hypertensive effects of salt. Gregory *et al.* (1990) estimated that British men were about 0.5 g/day short of this target and women almost 1 g/day below it.

WILLINGNESS TO CHANGE

It is sometimes argued that health promotion and nutrition education in particular are doomed to be ineffective because consumers are unwilling to make changes to their diets on health grounds. In a short article that is largely critical of health promotion, Watts (1998) suggests that 'no one has yet fathomed out how to persuade anyone to live more healthily'. To support his argument, he uses the failure of several major intervention trials to demonstrate significant benefits for the intervention group. It is quite clear, however, that health advice and health promotion can produce major shifts in national behaviour. In the 1970s and early 1980s, many parents in Britain and elsewhere were persuaded to use a front sleeping position for their babies rather than the traditional back sleeping position. In the early 1990s, many were then persuaded to revert to the traditional back position to reduce the risk of cot death (see Chapter 1, and Webb, 1995). Smoking is much less prevalent now than it was when the links between smoking and lung cancer were first established and publicized. In several sectors of British society, smoking has gone from being accepted as the social norm to a minority and antisocial activity.

There is also ample evidence that Britons and Americans made huge changes to their diets in the

- It is important that adequate amounts of all essential nutrients are consumed.
- Calcium and iron intakes can be maintained despite reductions in saturated fat if lean meat, poultry and low-fat milk replace fatty meats, full-fat dairy products and other fatty foods.
- In both Britain and the USA, minimum intakes of fluoride are recommended, especially in young children, in order to protect against dental decay.
- High potassium intakes may help to prevent high blood pressure. Increased fruit and vegetable consumption would lead to increased potassium intakes and British adults have been recommended to increase potassium intake to 3.5 g/day.

last quarter of the twentieth century. These changes have been largely driven by the desire to control body weight and the publicity about the links between diet and health. Table 4.4 shows a comparison of fat purchases for home consumption in the UK in 1975 and 1996, and some other major changes are listed below.

- Low-fat milks only became readily available in the early 1980s in Britain and so, before this, almost all of the fresh liquid milk sold was whole-fat milk. Between 1986 and 1996, low-fat milk increased from 19% of liquid milk sales to around 60%.
- In 1986, about 7% of the soft drinks purchased for home consumption were low-calorie versions and by 1996 this had risen to 20%.
- In the late 1960s, sales of sugar amounted to around 0.5 kg per person per week, but by 1996 this had fallen to 144 g per week. (Note that total added sugar consumption has not dropped by anything like this amount because it has been largely offset by greater consumption of sugar in manufactured products.)
- In 1976, sales of wholemeal bread made up a negligible proportion of total bread sales. Improvements in the texture of wholemeal bread coupled with health messages about dietary fibre caused them to rise to about 18% of total bread sales in 1986, although this has since fallen back to about 13% in 1996.

Table 4.4 A comparison of fats purchased for home consumption in the UK in 1975 and 1996

<i>Yellow fats</i> (as % of yellow fat sales)		
Butter	68	25
Soft margarine	14	21+
Other margarine	18	<3
Low-fat spreads	—	51
<i>Cooking fats</i> (as % of cooking fat sales)		
Vegetable oils	22	77
Other fats (mainly lard)	78	23
<i>Fats as percentage of food energy</i>		
Total fat	41.3	39.7
Saturated fatty acids	20.3	15.4
Polyunsaturated fatty acids	4.6	7.2
<i>P : S ratio</i>	0.19	0.46

P : S ratio = ratio of polyunsaturated to saturated fatty acids.

Data source: *National Food Survey*.

Many of these changes that have occurred in recent decades are the result of simple replacement of a traditional product with another of similar or better utilitarian value, so they have been easy to understand and implement. Butter and hard margarine have been replaced by soft margarine and low-fat spreads in most British homes (Table 4.4). Semi-skimmed (1.7%) and skimmed milk have largely replaced whole milk. Even though the new

- Despite assertions to the contrary, consumers have made substantial dietary changes as a result of health information and health promotion.
- Since 1995, the following changes to the British diet have occurred:
 - a massive switch from the use of butter and hard margarine to soft margarine and low-fat spreads
 - a huge switch from the use of animal-derived cooking fats like lards to the use of vegetable oils for cooking
 - replacement of whole milk with low-fat milks, which now account for the majority of milk sales
 - large increases in the sales of low-calorie soft drinks and wholemeal bread.
- Many of these dietary changes made in Britain involve the simple replacement of a traditional product by an alternative that, although possibly less palatable, has similar utilitarian characteristics and is perceived as healthier.
- Past experience thus suggests that if consumers are offered simple, realistic and relatively painless ways of improving their diets, many will make such changes.

'healthier product' is often considered slightly less palatable than the traditional product, people have been prepared in large numbers to make this relatively small sacrifice for their health or to try to lose weight. Table 4.4 shows that there have been dramatic changes in the types of cooking and spreading fats used in British homes, but relatively small and only fairly recent changes in the overall prominence of fat in British diets. Health promoters and nutrition educators should consider this when framing their dietary advice or devising their campaigns. People need to be offered simple, realistic and relatively painless ways of achieving or moving towards the desired nutritional goals.

SOME BARRIERS TO DIETARY CHANGE

Figure 1.4 in Chapter 1 summarizes the pathway by which a health promotion message is transmitted to the public and successfully results in improvements in health. At each step on this pathway there are potential barriers that may prevent effective change being made. Below are listed a number of factors that may discourage consumers from changing their diets in the direction suggested in this chapter – a number of potential barriers to change.

Palatability

Fat and sugar are major contributors to the palatability of Western diets. Starch and fibre are bland and so high-fibre, high-starch diets are often perceived as deficient in taste, smell and having an unappealing texture. As populations become more affluent, their natural inclination is to replace much of the starchy food that makes up the bulk of peasant diets with foods that are much higher in fat and sugar.

Sugar-fat seesaw

It may prove particularly difficult to reduce simultaneously the contribution of both fats and sugars to the diet. This would inevitably involve a substantial increase in starch consumption. In free-living, affluent populations, there tends to be an inverse relationship between sugar and fat consumption. People who

get a low proportion of their energy from fat tend to get a high proportion from sugar and *vice versa*; this is the so-called 'sugar-fat seesaw'. Survey data quoted by McColl (1988) suggested that people who consumed less than 35% of their calories as fat, derived, on average, more than 25% of their calories from sugars. Gregory *et al.* (1990) found that only 13% of their large representative sample of British adults met the target of no more than 33% of total energy from fat. Only 1% of their sample met both this fat target and that for added sugars (10% of energy). Those men and women who met the fat targets and sugar targets had higher alcohol intakes than those who did not, i.e. alcohol tended to dilute the fat and sugar and so help some individuals to meet sugar or fat targets. This sugar-fat seesaw might, in practice, tend to make the British target of 10% of energy from added sugars and the American target of no more than 30% from fat almost mutually exclusive. The NRC (1989b) recommendation to Americans merely not to increase sugar intakes could thus be interpreted as a logical and pragmatic decision to make fat reduction their nutrition education priority.

Availability of healthy alternatives

We saw in the previous section that many of the most striking changes in the British diet in the last 25 years have resulted from the substitution of a traditional food by one perceived as more healthy. If people are to continue to make these substitutions, the healthy alternatives must be readily available, affordable and acceptable in terms of their palatability. The massive switch from whole to low-fat milk coincided with the much greater availability of low-fat milks in the early 1980s. The increase in sales of wholemeal bread was partly the result of improved texture of these products. Low-fat spreads were unavailable 25 years ago, but now account for half the 'yellow fat' market in Britain. It may be argued that public demand will be sufficient to ensure the increasing availability of these 'healthier' alternatives. However, in order to demonstrate demand, the public must be given reasonable access to a product. Will a perceived future demand always be sufficient incentive for manufacturers and retailers to provide healthy options, particularly for a product likely to yield no increase in or perhaps even lower profit margins? The range of these low-fat, low-sugar and low-salt products is

increasing all the time, but some of them may be relatively expensive or perceived as poor value for money by low-income families trying to satisfy their appetites with strictly limited funds (see examples below).

- Extra lean mincemeat (hamburger) is much more expensive than the standard version.
- Low-fat milk costs the same as whole milk but provides fewer calories; the healthier version provides far fewer calories per penny (this applies to most low-fat and sugar products).
- Fruits and vegetables are an expensive source of energy, especially the more exotic and appealing varieties.
- The cheapest white bread in British supermarkets costs only about 30% as much as a wholemeal loaf.

Skills and knowledge

It is easier for people to reduce their fat intake or increase their fibre intake if they have a reasonable idea of the relative amounts of these in different foods. Have people got the skills and knowledge to prepare meals that are both appetizing and more in tune with current guidelines? Are food, nutrition and food preparation given sufficient weighting in our schools?

Traditional nutritional beliefs

Some consumers may be reluctant to increase their consumption of starchy foods because they still perceive them as fattening. A reduction in the consumption of meat and cheese may be hindered by their perception as valuable sources of first-class protein.

AIDS TO FOOD SELECTION

Selecting food solely according to biological criteria requires considerable knowledge of nutrition and food composition. It takes some skill to routinely translate current views on optimal intakes into meals and diets. Dietitians and nutrition educators need ways of translating complex technical guidelines about diet composition into simple, practical advice that is easy to understand, remember and implement. When adequacy was the nutritional priority, **food groups** were a simple and effective means of helping people to select food on a constituent basis. Foods were grouped together according to the nutrients that they provided – consumers could then be advised that, if they consumed specified minimum numbers of portions from each of the groups, nutritional adequacy would probably be assured.

- A number of factors may act as barriers or hindrances to the implementation of current nutritional guidelines.
- Sugar and fat add palatability to the diet, whereas starch and fibre are bland – this may discourage the replacement of some fat and sugar by starchy foods.
- In affluent diets there is often a sugar–fat seesaw, which makes it difficult simultaneously to reduce the proportions of energy derived from both fat and sugar.
- Substitution of traditional foods by ‘healthier’ alternatives requires that the healthier options are readily available, affordable and palatable.
- Many of the ‘healthier’ options are more expensive than the traditional product, particularly when costed as calories per penny. This may hinder dietary improvement in low-income groups.
- People may lack the knowledge and skills to implement the guidelines.
- Some outdated beliefs may discourage beneficial dietary change, e.g. ‘starchy foods are fattening’, ‘high intakes of meat and cheese are necessary to meet protein needs’. They may simply provide some people who are reluctant to change with an excuse for not changing.

The most widely used of these food group systems was the **four food group plan** (see Food groups in Glossary). Foods were divided up into four basic food groups and specified minimum numbers of servings from each group were recommended. These four food groups and an outline of their essential nutrient profiles are listed below.

- *The milk group* (milk, cheese, yoghurt and other milk products): these provide good amounts of energy, good-quality protein, vitamin A, calcium and riboflavin.
- *The meat group* (meat, fish, eggs and also meat substitutes such as pulses and nuts): these provide protein, vitamin A, B vitamins and iron.
- *The fruit and vegetable group* (fruits and those vegetables not classified as meat substitutes): these provide carotene, vitamin C, folate, riboflavin, potassium and fibre.
- *The bread and cereals group* (bread, rice, pasta, breakfast cereals and products made from flour): whole-grain cereals are good sources of B vitamins, some minerals and fibre. White flour still provides reasonable amounts of fibre and it is often fortified with vitamins and minerals (in the UK, with iron, calcium and some B vitamins).

People were advised to eat a minimum of two portions each day from the meat and milk groups and a minimum of four portions each day from the other two groups. As the foods in these four groups have differing profiles of essential nutrients, then provided consumers ate enough to satisfy their appetite and ate these minimum numbers of portions, nutritional adequacy was practically assured.

A number of foods are not covered by any of these categories, e.g. butter, oil and other fats, sugar, alcoholic drinks etc. This is because they are largely sources of energy and provide few nutrients, so no minimum portions were recommended. They were to be used sparingly because of their low nutrient density.

Several generations of American schoolchildren and college students have been taught the basic four food group plan since its development in 1955. This classification system will still be familiar to many American adults, although much less familiar to most Britons.

Such food-group schemes were originally designed to ensure nutritional adequacy rather than to guide consumers towards a diet that helps to pre-

vent chronic disease. Many foods in the milk and meat groups are now seen as rich sources of dietary fat and cholesterol as well as nutrients, e.g. whole milk, cheese, fatty meat. Many foods in the cereals group are seen as low in fat and sugar and high in fibre and starch in addition to being sources of some nutrients. Many foods in the fruit and vegetables group are very low in energy density and practically fat free in their natural, unprocessed state. In order to comply with the compositional guidelines discussed in this chapter, consumers need to be encouraged to:

- increase their consumption of cereals, vegetables and fruit
- moderate their consumption of meats and dairy foods
- minimize their consumption of fatty and sugary foods that do not fit into any of the four groups
- choose foods within the food groups that are low in fat and added sugars but high in fibre.

For some time, nutrition educators tried to adapt this four food group guide to modern nutritional priorities, for example highlighting the low-fat/low-sugar options within each group. However, a food guide that merely suggested minimum quantities from the food groups was unsuited to modern dietary guidelines, with their emphasis on moderating intakes of some foods and food groups and increasing the prominence of other food groups in the diet.

In 1992, the US Department of Agriculture (USDA, 1992) published a new guide to food selection intended to update and replace the basic four plan – the **food guide pyramid** (Figure 4.2). This food pyramid is a development of the food groups that has been designed to reflect the dual aims of selecting a diet that is not only adequate, but also meets nutritional guidelines aimed at reducing the risk of chronic disease. At the base of the food pyramid are the starchy grain foods that should contribute more servings than any other single group to the ideal diet – 6–11 servings, depending upon total energy intake (in the range 1600–2800 kcal or 6.7–11.8 MJ). On the next tier of the pyramid are the fruit and vegetable groups – a total of five to nine servings from these groups, depending upon energy intake. At the third level on the pyramid are the foods of animal origin (including the vegetarian alternatives to meat and milk such as pulses,

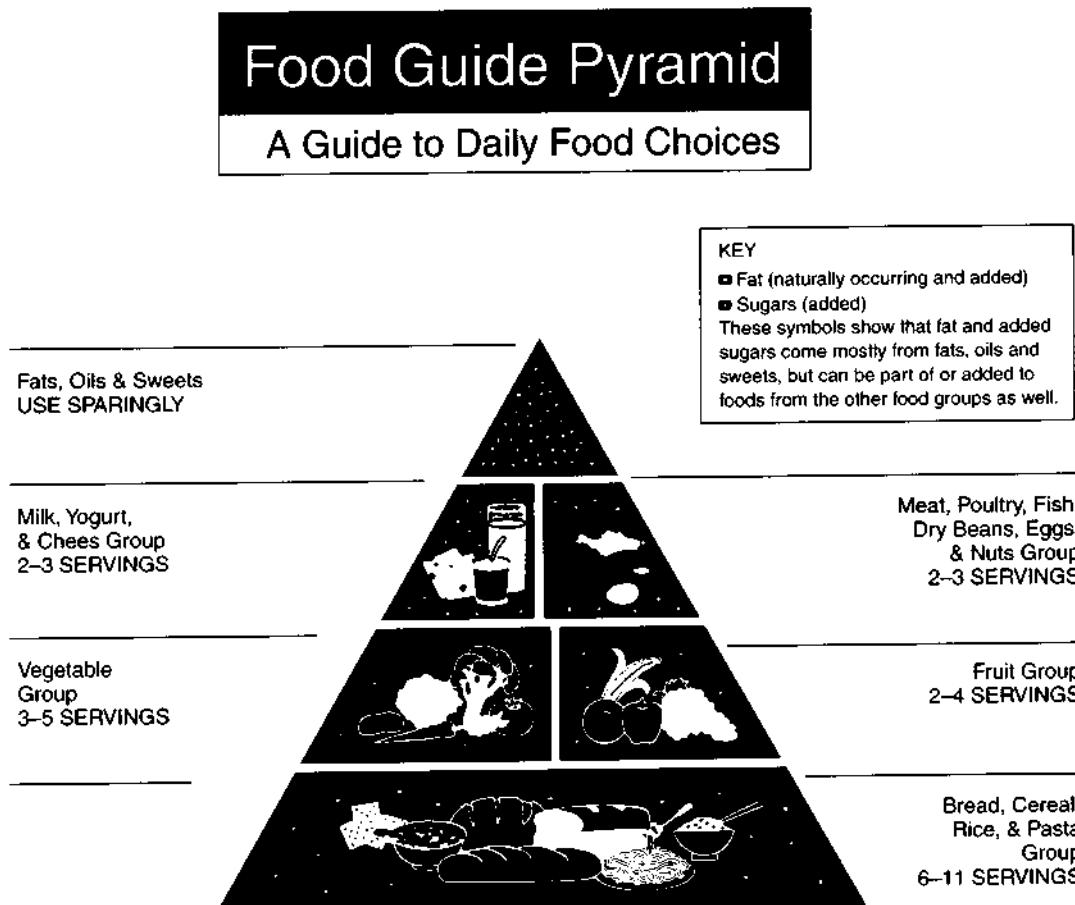


Figure 4.2 The American food guide pyramid. Source: US Department of Agriculture/US Department of Health and Human Services.

nuts, soya milk and modern meat substitutes). There should be two to three servings from each group on this tier, the meat and milk groups. At the top of the pyramid are the fats, oils and sweets – foods such as salad dressings, cream, butter, margarine, soft drinks, sweets (candies), sweet desserts and alcoholic drinks. These foods provide few nutrients but are high in energy, sugars and fats and thus should be used sparingly in the ideal diet, especially for those seeking to lose weight. In Figure 4.2, a triangle symbol is distributed lightly within the cereal, fruit and milk groups to show that some foods from these categories may contain added sugars, e.g. sweetened breakfast cereals, fruit canned in syrup, some flavoured yoghurts, milk shake and ice cream. Also in Figure 4.2, a circle symbol is distributed lightly throughout the cereal and vegetable groups to show that these may contain added fats,

and more densely in the meat and milk groups to show that many of the foods in these groups contain substantial amounts of naturally occurring fat. The top layer of the pyramid contains a high density of both triangles and circles to emphasize its role as a major contributor of fats and added sugar to the diet.

The food pyramid thus indicates a general dietary structure that should ensure adequacy and yet at the same time make it more likely that other nutrition education guidelines aimed at reducing chronic disease will also be met. Highlighting the likely sources of fat and added sugars in the diet with the circle and triangle symbols should aid consumers in making choices that reduce intakes of fat and added sugars.

Figure 4.3 shows an estimate of what Americans really eat laid out in the form of a pyramid. This real

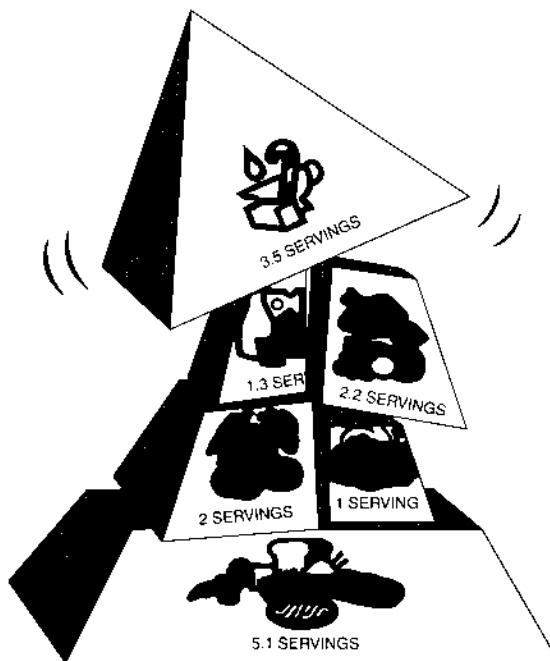


Figure 4.3 The real food guide pyramid – what Americans really eat. Source: National Cattlemen's Beef Association.

pyramid is much more top heavy than the ideal one. Foods from the fats and sugars group provide more servings to Americans than those from the fruits and vegetables groups. This shows just how large the dietary changes required by the US food guide pyramid are.

Following the launch of the food guide pyramid in the USA, a British version of this pyramid appeared in 1993. This was an unofficial attempt to provide Britons with some form of nationally accepted food guide for the first time. It was produced jointly by the Flour Advisory Bureau and the Dunn Nutrition Laboratory in Cambridge. 'The British Healthy Eating Pyramid' was in most of its essentials identical to the US version, except that potatoes were included with the cereals at the base of the pyramid (rather than with the other vegetables). The fruits and vegetables were represented as a single group on the second tier of the pyramid, with a recommended five to nine daily servings.

In early 1994, the Nutrition Task Force had identified the need for a national pictorial food guide as a priority for implementing the dietary and nutritional targets set out in *The health of the nation*. Research on

the best form that this food guide should take was undertaken on behalf of the Health Education Authority. A variety of different formats were tested with both the general public and health professionals. This research showed that such pictorial guides could improve people's understanding of nutritional concepts. A plate shape was found to be more effective at conveying those nutritional concepts than an American-type pyramid. The tilted-plate model proved most appealing to members of the public (Hunt *et al.*, 1995). As a result of these findings, the Health Education Authority in Britain published the first official National Food Guide in mid-1994. It used the image of a tilted plate with five food groups occupying different-sized segments of this plate (Figure 4.4).

Two large segments occupy about 30% of the plate each:

- starchy foods – the bread, cereals and potatoes group
- the fruit and vegetables group.

Two medium-sized segments occupy about 15% of the plate each:

- meat, fish, eggs and pulses
- milk and milk products like cheese and yoghurt.

The smallest sector of the plate represents less than 10% of the area and contains:

- fatty and sugary foods.

The general dietary structure indicated in the tilted-plate model is essentially the same as that suggested by the American pyramid. The diet should contain large amounts of starchy foods and fruit and vegetables, more moderate amounts of foods from the meat and milk groups, and strictly limited amounts of fatty and sugary foods. The form that is used for general display, e.g. in supermarkets and for general guidance to the public, contains no recommended numbers of portions. It is felt that the wide variation in individual energy intakes and thus the wide variation in numbers of recommended portions would make this information difficult for the general lay person to interpret. The tilted plate on general release (Figure 4.4) does contain additional advice for selection within each food group:

- choose low-fat options from the meat and milk groups where possible
- eat a variety of fruits and vegetables

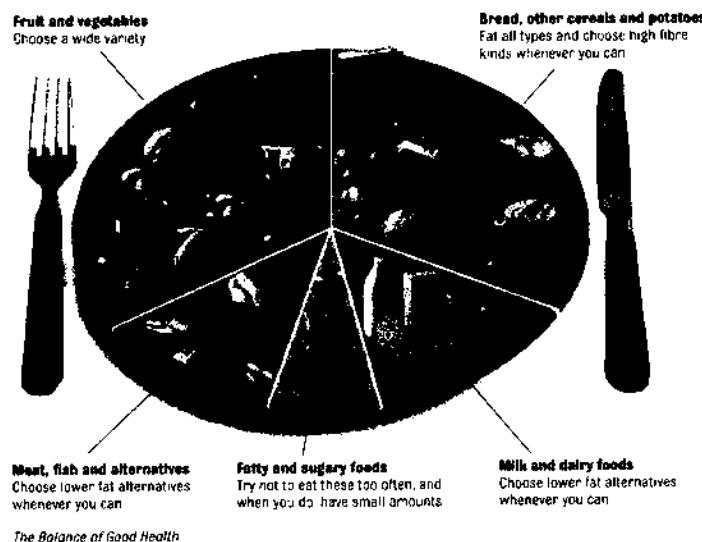


Figure 4.4 The tilted plate model used as the National Food Guide in Britain. Source: Health Education Authority, London.

- Nutrition educators have used a variety of guidance tools to translate technical nutritional guidelines into practical and comprehensible dietary advice.
- Food groups helped to guide consumers towards a nutritionally adequate diet. Foods were grouped into four or more categories with differing nutrient profiles and consumers were advised to eat a minimum number of portions from each category daily.
- Milk, meat, cereals, and fruit and vegetables were the four groups of the basic four-group plan. Sugary and fatty foods were in a separate category outside the main food groups.
- New food guides became necessary when 'preventing chronic disease' became a major nutritional priority and when guidelines often stressed reducing or moderating intakes of certain dietary components.
- The USA now has a food guide pyramid. Foods are arranged on the pyramid in order to indicate their ideal prominence in the diet. The large base section of the pyramid is occupied by the bread and cereals group, on the next level are the fruit and vegetable groups, then the meat and dairy groups. The small section at the top of the pyramid is occupied by fatty and sugary foods, emphasizing the need to restrict intake of these.
- In Britain a tilted-plate model is used. Two large sectors of the plate are occupied by starchy foods (cereals and potatoes) and the fruit and vegetable group, and two moderate-sized sectors by the meat and alternatives group and the milk group. The last, small, section of the plate is occupied by fatty and sugary foods.
- In the American pyramid, the density of circular and triangular symbols indicates the likely sources of fat and added sugar, respectively.
- In the UK, text messages on the plate urge consumers to chose low-fat and high-fibre options from within the appropriate groups.
- The American pyramid has suggested ranges for ideal numbers of daily portions from each of the food groups. Only the version of the British guide intended for use by health professionals has these portion guidelines.

- eat all types of starchy foods and choose high-fibre varieties where possible.

There is a version of this tilted-plate model that does contain recommended numbers of portions of foods. It is intended for use by health professionals in one-to-one counselling with their clients where they can interpret this quantitative information in a way that is appropriate to the individual.

These portion guides are similar to those suggested on the American pyramid guide:

- 5–9 measures of fruits and vegetables
- 5–14 measures of bread cereals and potatoes
- 2–3 measures of meat, fish and alternatives
- 2–3 measures of milk and dairy foods
- 0–4 measures of fatty and sugary foods.

CONCLUDING REMARKS

Although many nutritionists would argue about specific diet/disease issues, few would disagree with the proposition that significant health benefits would be likely to accrue if Americans and Britons were leaner, more active, consumed alcohol moderately, ate less fat, sugar and salt but more fruit, vegetables and cereals. The basic qualitative message transmitted by the American food guide pyramid and the British tilted plate enjoys consensus support of nutritionists and dietitians.

In this chapter, recommendations have only been considered in terms of their effects upon the health of people eating them. Dietary changes will also inevitably have social, economic and environmental consequences. Some dietary changes that might be deemed nutritionally desirable may not be sustainable if everyone adopts them. If those in the wealthy, industrialized countries make dietary changes that soak up more of the world's resources, will health promotion in wealthy countries be at the expense of still greater deprivation in poorer ones? Some examples of these issues are listed below.

- The world's oceans are already suffering depletion of fish stocks due to overfishing. If people are encouraged to eat more fish, will this situation worsen still further? Is fish farming a realistic alternative to wild fish? Is it capable of producing fish on a huge scale without heavy

inputs of energy and other negative ecological repercussions?

- If all Western consumers start eating five portions of fruits and vegetables each day, will many of these foods have to be transported into northern countries with all the resultant fuel costs? What effects will their growth have upon ground-water resources in the producing country? Will they be grown as cash crops for export at the expense of the staple crops needed to feed the local population?
- If Western consumers reduce their consumption of meat and dairy foods and replace them with starchy cereals and roots, this should have a positive impact upon sustainability. It takes 3–10 kg of grain to produce 1 kg of meat or poultry. However, what if these Western consumers are encouraged to waste many of the expensively produced calories in animal² foods for example if they only eat the leaner cuts of meat and skimmed milk and discard most of the butterfat and meat fat?
- The dietary changes brought about by health promotion have already had a noticeable effect upon the nature of farming and the countryside. In Britain, huge yellow swathes of land devoted to the growth of oilseed rape have become a feature of many parts of the countryside. Can the hill land in the more mountainous parts of Britain be used productively for purposes other than grazing animals?
- Some of the recommended dietary changes seem to be encouraging a more artificial and less natural diet, e.g. the use of low-fat spreads, low-calorie drinks and some meat substitutes.
- As an alternative to the American food guide pyramid, Walter Willett and his colleagues at the Harvard School of Public Health have suggested an alternative pyramid based upon the traditional Mediterranean diet and in which most of the dietary fat would come from olive oil. Olive oil is a relatively expensive (and delicious) regional product. Is it practical to recommend such a diet for mass consumption, e.g. in northern Europe and the colder regions of North America?

Detailed discussion of such issues is beyond the scope of a book such as this one and beyond the capabilities of its author, but they are issues that need to be considered by those advocating dietary change.

- Most nutritionists would agree with the broad thrust of current nutritional guidelines, despite questioning particular elements of them.
- The environmental, economic and cultural impacts of recommended dietary changes need to be considered as well as simply their nutritional desirability for residents of the home country.
- Expert committees must ensure that any recommendations they make are sustainable if implemented on a mass scale by many populations.

Cellular energetics

Aim	136	Metabolism of fats	140
Introduction and overview	136	Metabolism of protein	141
Metabolism of carbohydrate	138	The pentose phosphate pathway	142

AIM

This very brief and greatly simplified outline of the energy conversions within the cell is not intended to be an introduction to the study of biochemistry. Only the minimum of factual biochemistry to allow illustration of the broad principles and concepts has been included. The purpose is to give readers with no biochemical background enough perception of the biochemical processes of the cell to help their understanding of nutrition. This short summary may also provide a useful overview for readers who have studied biochemistry. Certain nutritionally important observations should be clarified by this discussion, such as those listed below.

- Fatty acids cannot be converted to glucose, but glucose can be converted to fat.
- The human brain cannot use fatty acids as substrates, but during starvation it obtains more than half of its energy from the metabolism of **ketone bodies** that are derived from fatty acids.
- Many vitamins are essential as precursors of the **coenzymes** that are necessary to allow enzyme-mediated reactions to occur.
- Amino acids can serve as energy sources and can be used to generate glucose during fasting.
- Certain substances can ‘uncouple’ the link between oxidation – energy production – and – **adenosine triphosphate** phosphorylation (ATP) synthesis – and thus cause this energy to be released as heat. This is how **brown fat** in some animals

generates heat for maintaining body temperature in the cold.

INTRODUCTION AND OVERVIEW

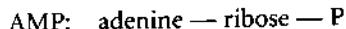
The catabolism (breakdown) of carbohydrates, fats and proteins within cells releases chemical energy. This energy is used to drive all of the energy-requiring processes of the cell, e.g. synthetic processes, muscle contraction, transport of materials across cell membranes, nerve conduction. ATP plays a pivotal role as a short-term energy store within the cell. The chemical energy released during the catabolic metabolism of foodstuffs is stored as high-energy ATP and then the breakdown of ATP is used to drive the other energy-requiring processes. Each cell produces its own ATP; it is not transported between cells.

Although not discussed in the following outline, it should be borne in mind that every cellular reaction is catalysed (speeded up) by a specific enzyme. The reactions would not occur to any significant extent in the absence of the specific enzyme – it is the enzymes within the cell that determine what reactions can occur within the cell and thus determine the nature of the cell. Enzymes are proteins and the genetic code (contained within the DNA molecule) codes for the proteins of the cell. It is the proteins an organism produces that determine its characteristics. Genetic diseases are the result of an error in one of the proteins produced by that individual. Many enzymes

require non-protein moieties known as coenzymes or cofactors in order to function. Several of these coenzymes are derivatives of vitamins and it is their roles as precursors of coenzymes that account for the essentiality of several vitamins. Cofactors that are bound strongly to the enzyme and become an integral part of the enzyme structure are termed **prosthetic groups**.

Nature and functioning of ATP

Adenosine triphosphate is comprised of a purine base (adenine), a pentose sugar (ribose) and three phosphate groups. The hydrolysis of ATP to adenosine diphosphate (ADP), by removal of one of the phosphate groups, is a highly **exergonic reaction**, i.e. a considerable amount of chemical energy is released during the reaction. This energy would be released as heat if the ATP hydrolysis were conducted in a test tube. Similarly, the hydrolysis of ADP to adenosine monophosphate (AMP) is also highly exergonic. Conversely, the conversions of AMP to ADP to ATP are highly **endergonic reactions**, i.e. they absorb large amounts of energy; if these reactions were conducted in a test tube, one would expect to need to provide heat energy to make the thermodynamics of the reactions favourable.



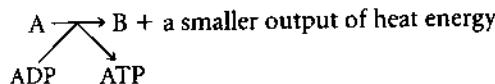
P = phosphate group (PO_4)

Within the cell, highly exergonic reactions in the catabolism of foodstuffs are coupled to ATP synthesis. The combined reaction remains slightly exergonic, but much of the chemical energy of the exergonic, catabolic reaction is stored as ATP rather than being released into the cellular fluids as heat.

For example, take the hypothetical exergonic catabolic reaction:

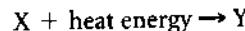


In the cell:

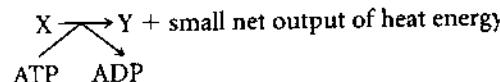


Other reactions within the cell are endergonic (energy consuming) and one would expect that, if these reactions were being carried out in a test tube, one would have to provide heat to make them thermodynamically favourable and allow them to occur. Within the cell, however, such an endergonic reaction can be coupled to ATP hydrolysis to make the combined reaction exergonic and thus thermodynamically favourable at body temperature.

For example, take the hypothetical endergonic reaction:



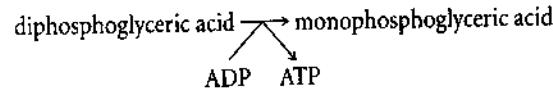
In the cell:



Substrate-level phosphorylation

In catabolic pathways, some highly exergonic steps are directly linked to ATP formation (substrate-level phosphorylation), as illustrated by the theoretical A → B reaction above.

For example, in the glycolysis pathway, one reaction involves the conversion of diphosphoglyceric acid to monophosphoglyceric acid, and this reaction is directly coupled to ATP synthesis:

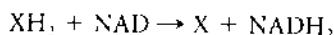


Oxidative phosphorylation

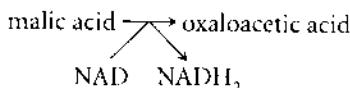
Much of the energy release in the catabolic pathways of metabolism occurs as a result of oxidation. Many oxidative steps involve the removal of hydrogen atoms from substrate molecules and their acceptance

by hydrogen acceptor molecules such as **nicotinamide adenine dinucleotide (NAD)**.

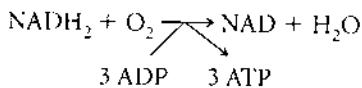
For example, take the hypothetical oxidative reaction:



In this reaction, XH_2 has been oxidized to X, and NAD has been reduced to NADH_2 . As a real example, in the citric acid cycle (discussed below), malic acid is oxidized to oxaloacetic acid by the removal of hydrogen and this reaction is coupled to the reduction of NAD by hydrogen addition:



NAD is derived from the vitamin niacin. Niacin is essential because it is a precursor of NAD and the phosphorylated derivative NADP. Re-oxidation of these reduced hydrogen acceptors in the **mitochondria** of cells results in the production of large quantities of ATP (three molecules of ATP per molecule of NADH_2 re-oxidized):



This mitochondrial process is called **oxidative phosphorylation**. In this process, there are a series of oxidation-reduction reactions, which culminate in the reduction of molecular oxygen to water. Reduced NAD (NADH_2) reduces the next compound on the electron transport chain and is itself re-oxidized; this reduced compound then reduces the next compound and so on to molecular oxygen. This sequence is called the electron transport chain because reduction can be defined as either the addition of electrons or the addition of hydrogen.

During this sequence of oxidation-reduction reactions, there is considerable release of chemical energy. Some of this energy is released as heat, which helps to keep the body warm, but some of it is used to force hydrogen ions or protons (H^+) through the inner mitochondrial membrane. As the inner mitochondrial membrane is generally impermeable to protons, this creates a considerable proton gradient across this membrane. There are pores in the inner

mitochondrial membrane where protons can pass back into the mitochondrion, and this controlled flow of protons provides the energy that drives ATP synthesis in the mitochondrion. This is the chemiosmotic theory of oxidative phosphorylation, which explains how the oxidation of reduced cofactors (e.g. NADH_2) is linked to ATP synthesis.

If the inner mitochondrial membrane become more permeable to protons, this will compromise or 'uncouple' this link between oxidation and ATP synthesis. Uncouplers make the inner mitochondrial membrane more permeable to protons and so they cause oxidation to accelerate, but the energy is released as heat rather than being used to drive ATP synthesis. In a tissue known as brown fat, there is an uncoupling protein (UCP1) which, when activated by sympathetic nerve stimulation, allows protons to leak though the inner mitochondrial membrane and this generates heat. This tissue is capable of producing large amounts of heat to warm the body and brown fat is prominent in small mammals, hibernating animals and human babies. There are other uncoupler proteins in other tissues (UCP2 and UCP3); their significance is discussed in Chapter 7. Some drugs and poisons also uncouple oxidative phosphorylation, e.g. dinitrophenol. These substances increase body heat production and can cause rapid weight loss, but those currently available are too toxic to be used therapeutically for this purpose.

Oxidative phosphorylation normally yields the vast bulk of the energy released in catabolic metabolism. Cells which do not have mitochondria (e.g. red blood cells) or which have insufficient oxygen supply (e.g. a muscle working beyond the capacity of the blood system to supply oxygen) have to rely upon the anaerobic metabolism of glucose to supply their energy needs. One molecule of glucose when metabolized anaerobically to lactic acid gives a net yield of only two molecules of ATP, whereas when metabolized aerobically to carbon dioxide and water, it yields 38 ATP molecules.

METABOLISM OF CARBOHYDRATE

Dietary carbohydrates are digested to their component monosaccharides before being absorbed. Digestion of starch yields glucose and digestion of

disaccharides yields glucose plus other monosaccharides, fructose from sucrose (cane or beet sugar) and galactose from lactose (milk sugar).

Glycolysis

Glycolysis is the initial pathway involved in the metabolism of carbohydrates and is summarized in Figure 5.1. The first three steps in this pathway involve activation of the glucose molecule by the addition of phosphate groups and a molecular rearrangement (isomerization) of glucose phosphate to fructose phosphate. There is consumption of two ATP molecules during these activating steps. The fructose diphosphate thus produced is much more reactive (unstable) than glucose, and the six-carbon (6C) molecule can be split by an enzyme into two three-carbon (3C) molecules. The other major dietary monosaccharides, fructose and galactose, also feed into the early stages of this glycolysis pathway, as does glucose phosphate derived from the breakdown of body glycogen stores.

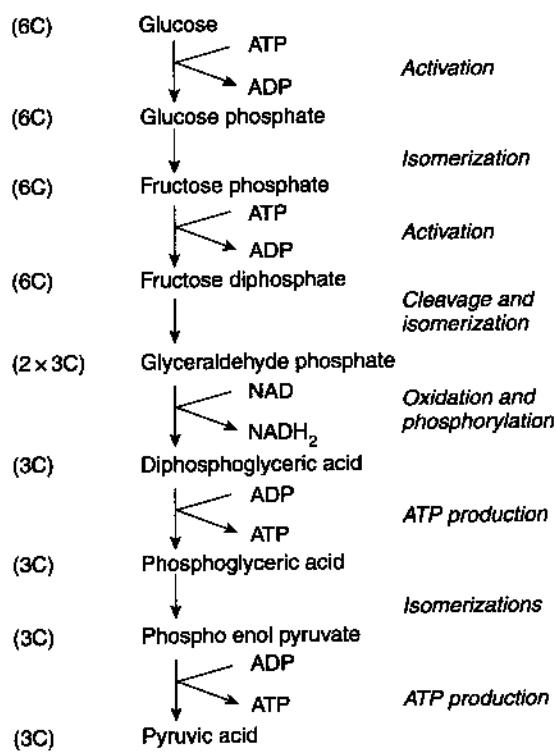
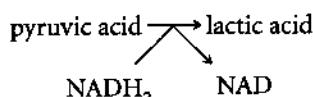


Figure 5.1 Outline of the glycolytic sequence. C = number of carbon atoms.

During the steps that convert glyceraldehyde phosphate into pyruvic acid, there is production of two molecules of reduced NAD and four molecules of ATP from each starting molecule of glucose. Taking into account the two molecules of ATP used to generate fructose diphosphate, this gives a net yield of two ATP molecules at the substrate level for each molecule of glucose. Under aerobic conditions, the reduced NAD will be re-oxidized in the mitochondria using molecular oxygen and will yield a further six molecules of ATP in oxidative phosphorylation. Under anaerobic conditions, this reduced NAD must be re-oxidized by some other means, otherwise the very small quantities of oxidized NAD within the cell would be quickly exhausted and the whole process halted. In mammalian cells, this NAD is regenerated under anaerobic conditions by the reduction of pyruvic acid to lactic acid:

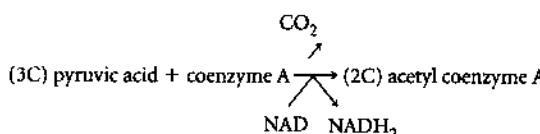


Lactic acid is the end product of anaerobic metabolism in mammalian cells, and anaerobic energy production is only possible from carbohydrate substrate. Red blood cells do not have mitochondria and thus they metabolize glucose only as far as pyruvic acid and lactic acid. During heavy exercise, the oxygen supply to a muscle limits aerobic metabolism and so the muscles will generate some ATP anaerobically and produce lactic acid as a by-product. Accumulation of lactic acid is one factor responsible for the fatigue of exercising muscles. Note also that in thiamin deficiency (beriberi) there is effectively a partial block in the metabolism of carbohydrate beyond pyruvic acid because the conversion of pyruvic acid to acetyl coenzyme A requires a coenzyme, thiamin pyrophosphate, that is derived from thiamin. Lactic acid and pyruvic acid therefore also accumulate in victims of beriberi because of this metabolic block.

The lactic acid produced by anaerobic metabolism is used for the re-synthesis of glucose in the liver (Cori cycle). This re-synthesis of glucose is effectively a reversal of glycolysis and thus it is possible to synthesize glucose from any intermediate of the glycolysis pathway, although the process does consume ATP.

Aerobic metabolism of pyruvic acid

Under aerobic conditions, pyruvic acid will normally be converted to acetyl coenzyme A (effectively activated acetate):



Coenzyme A is a large moiety that is derived from the B vitamin pantothenic acid; addition of the coenzyme A moiety increases the reactivity of acetate. It is released in the next reaction and recycled.

This step cannot be reversed in mammalian cells and thus glucose cannot be synthesized from acetyl coenzyme A.

The acetyl coenzyme A (2C) then enters a sequence known as the **Krebs' cycle** or the citric-acid cycle (Figure 5.2); it combines with oxaloacetic acid (4C) to give citric acid (6C). This citric acid then goes through a sequence of eight reactions, which ultimately result, once again, in the production of oxaloacetic acid, i.e. a cyclic process. During two of the reactions, a molecule of carbon dioxide is produced and in four of them a molecule of reduced coenzyme (e.g. NADH₂) is produced; in only one reaction is there direct substrate-level ATP production. Each starting molecule of glucose yields two molecules of acetyl coenzyme A and thus each molecule of glucose metabolized under aerobic conditions results in two 'turns' of the citric acid cycle.

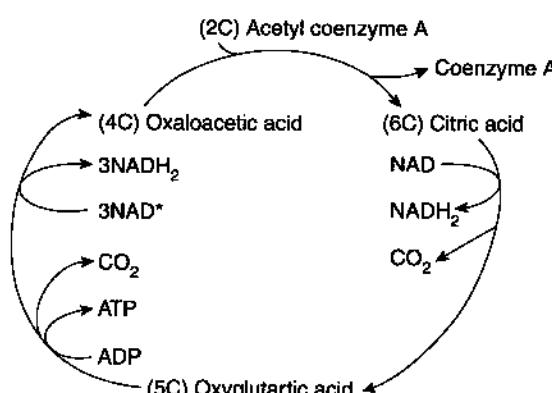


Figure 5.2 Outline of the Krebs cycle. C = number of carbon atoms.

After undergoing the reactions of glycolysis and the citric-acid cycle, all of the six carbon atoms of the original glucose molecule will thus have been evolved as carbon dioxide. When the reduced coenzyme (e.g. NADH₂) is re-oxidized in the mitochondria, this will result in the production of water. Thus, overall, the glucose has been metabolized to carbon dioxide and water.

METABOLISM OF FATS

Dietary fat, or the fat stored in adipose tissue, is largely **triacylglycerol (triglyceride)** and it yields for metabolism one molecule of glycerol and three fatty-acid molecules. The three-carbon (3C) glycerol is converted to glyceraldehyde phosphate and enters glycolysis – it can thus be used directly as a source of energy or it can be used to generate glucose (by reversal of glycolysis). Fatty acids are metabolized to multiple units of acetyl coenzyme A, in a pathway known as the **β -oxidation pathway**. A 16-carbon fatty acid (e.g. palmitic acid) would thus yield eight two-carbon units of acetyl coenzyme A (summarized in Figure 5.3).

Note that, as the conversion of acetyl coenzyme A back to pyruvic acid is not possible in mammalian cells, fatty acids cannot be used to generate glucose.

Brain cells do not have the enzymes necessary for β -oxidation and therefore they cannot directly use fatty acids as an energy source. When carbohydrate is available (e.g. when carbohydrates are eaten regularly), brain cells use glucose as their substrate, but they are not, as was previously thought, totally

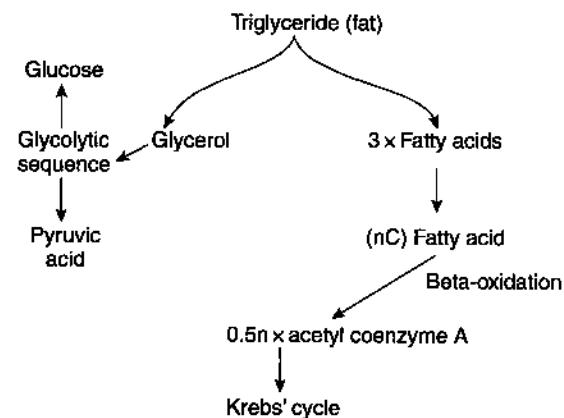


Figure 5.3 Outline of fat metabolism. C = number of carbon atoms.

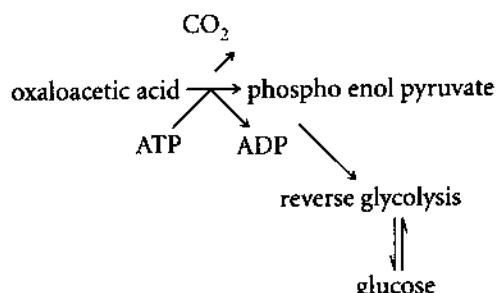
dependent upon carbohydrate as a substrate. During fasting, they can utilize certain ketones or ketone bodies that are made from acetyl coenzyme A. This means that during starvation the brain can indirectly utilize fatty acids that have been converted to these ketones in the liver. (This is discussed more fully under 'Metabolic adaptation to starvation' in Chapter 6.)

Fatty acids are synthesized by a process that is essentially a reversal of β -oxidation, e.g. to synthesize 16C palmitic acid, eight units of two-carbon acetate (as acetyl coenzyme A) are progressively assembled. Thus, fatty acids can be synthesized from carbohydrates via acetyl coenzyme A. Breakdown of fatty acids to acetyl coenzyme A is an oxidative process (hence β -oxidation), thus the synthesis of fatty acids from acetyl coenzyme A is a reductive process. The reduced form of the phosphorylated derivative of NAD, NADPH₂, is used as the source of reducing power in this pathway – NADPH₂ is generated in the pentose phosphate pathway (see later).

METABOLISM OF PROTEIN

Surplus amino acids can be used as an energy source. The nitrogen-containing amino group is removed to leave a moiety, the keto acid, that can be converted to pyruvic acid, acetyl coenzyme A or one of the intermediates of the citric acid cycle. The amino group can be converted to the waste product urea or it can be transferred to another keto acid and thus produce another amino acid – a process called **transamination**.

It is possible to make glucose from protein. If the amino acid (e.g. alanine) yields pyruvic acid, then glucose synthesis merely involves the effective reversal of glycolysis. If the amino acid (e.g. glutamic acid) yields a citric-acid cycle intermediate, then this intermediate will be converted to oxaloacetic acid, which can then be converted to the phospho enol pyruvate of glycolysis:



Note that acetyl coenzyme A (and thus fatty acids) cannot be used to synthesize glucose via this route because the two carbon atoms of the acetate that enter the citric-acid cycle have been lost as carbon dioxide by the time oxaloacetic acid has been regenerated. Utilization of existing citric-acid cycle intermediates to synthesize glucose is theoretically possible, but they would be so rapidly depleted that their contribution to glucose supply would be insignificant. The metabolic routes for the metabolism of the different foodstuffs are summarized in Figure 5.4.

During starvation or carbohydrate deprivation, the only routes available for the maintenance of carbohydrate supplies are by manufacture from amino acids or the glycerol component of fat. This process of generation of glucose from amino acids (**gluconeogenesis**) occurs in the liver, but it is inefficient, energy expensive and will, of course, lead to depletion of the protein in muscle and vital organs. Note, however, that the need for gluconeogenesis is limited during starvation or carbohydrate deprivation by the use of ketone bodies as an alternative to carbohydrate substrate.

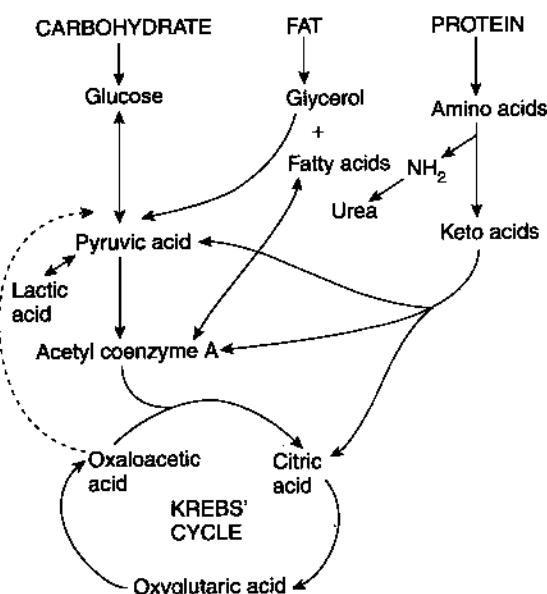
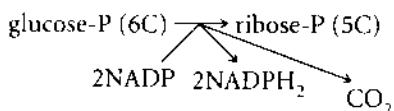


Figure 5.4 Summary of metabolic routes for foodstuffs.

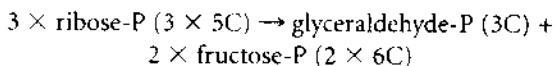
THE PENTOSE PHOSPHATE PATHWAY

This pathway generates reducing power in the form of NADPH₂ (necessary, for example, in fatty-acid biosynthesis) and also generates the pentose sugar, ribose phosphate, essential for nucleotide (e.g. ATP) biosynthesis and nucleic acid (RNA and DNA) synthesis. The first part of this pathway involves the conversion of glucose phosphate to ribose phosphate:



If the demand for NADPH₂ and ribose phosphate is balanced, these two will represent end products of the pathway. If, however, the demand for NADPH₂, say for active lipid synthesis, exceeds the demand for ribose phosphate, the excess ribose is converted, by a complex series of reactions, to three-carbon glyceraldehyde phosphate and six-carbon fructose phosphate.

Overall reaction:



The enzymes used in this series of reactions are called transketolase and transaldolase; transketolase requires thiamin pyrophosphate, derived from vitamin B₁, (thiamin) as a coenzyme. The glyceraldehyde phosphate and fructose phosphate produced by these reactions can both enter glycolysis, either to be metabolized to pyruvic acid or to be used to regenerate glucose phosphate.

If the demand for ribose phosphate exceeds that for NADPH₂, the reverse of the transketolase/transaldolase reaction can generate ribose phosphate from glyceraldehyde phosphate and fructose phosphate.

Wernicke–Korsakoff syndrome is a neuropsychiatric disorder caused by lack of dietary thiamin (see Chapter 12 for details). In industrialized countries, it is usually associated with thiamin deficiency brought on by alcoholism. Some people seem to have a genetic predisposition to this syndrome because their transketolase enzyme has a low affinity for thiamin pyrophosphate, making them extremely susceptible to the harmful effects of thiamin deficiency. The symptoms of this syndrome are probably due to inadequate NADPH₂ synthesis leading to impaired synthesis of **myelin**, the fatty sheath that surrounds many neurones.

Part 2

Energy, energy balance and obesity

Introduction to energy aspects of nutrition

Units of energy	145	Energy density	151
How energy requirements are estimated	146	Nutrient density	152
Variation in average energy requirements – general trends	148	Sources of dietary energy by food groups	154
The energy content of foods	148	Starvation	155
Sources of dietary energy by nutrient	149	Eating disorders	159
		Cancer cachexia	162

Although the range of nutrients required by living organisms varies enormously, all organisms require an external supply of energy. Plants 'trap' energy from sunlight and use it to make sugars in the process of photosynthesis. Photosynthesis is the ultimate source of all the energy in living systems and fossil fuels. In human diets, the food macronutrients fat, carbohydrate and protein (plus alcohol) are the sources of energy.

UNITS OF ENERGY

The Standard International unit of energy is the **joule**, which is defined as the energy expended when a mass of 1 kg is moved through a distance of 1 metre by a force of 1 Newton. For nutritionists the kilojoule – kJ (a thousand joules) – and the megajoule – MJ

(a million joules) – are more practical units. Traditionally, nutritionists have used a unit of heat, the kilocalorie – kcal – as their unit of energy. Although strictly a thousand calories, most people, when dealing with nutrition, tend to use the terms calorie and kilocalorie as if synonymous. A kilocalorie is defined as the heat required to raise the temperature of a litre of water by 1 °C. In practice, the kilocalorie is still widely used both by nutritionists and by non-scientists. The kilocalorie is a convenient unit, both because people with a limited knowledge of physics can understand its definition and because nutritionists may use heat output as their method of measuring the energy yields of foods and the energy expenditure of animals and people.

To interconvert kilocalories and kilojoules:

$$1 \text{ kcal} = 4.2 \text{ kJ}$$

- A joule is the energy expended when a mass of 1 kg is moved through a distance of 1 metre by a force of 1 Newton
- A kilojoule (kJ) = 1000 joules
- A megajoule (MJ) = 1 million joules (1000 kJ)
- A kilocalorie is 1000 calories, but in nutrition the term calorie usually means kilocalorie
- 1 kilocalorie (kcal) = 4.2 kJ.

HOW ENERGY REQUIREMENTS ARE ESTIMATED

Table 6.1 shows some selected UK Estimated Average Requirements (EARs) for energy and their corresponding US equivalents (Recommended Dietary Allowances, RDAs). Remember that the dietary standards for energy represent the best estimate of average requirement, whereas the most commonly used reference values for most nutrients, the Reference Nutrient Intake (RNI) or RDA, are the estimated needs of those with the highest requirement (see Chapter 3 for discussion and rationale).

The basal metabolic rate (BMR) is the minimum rate of energy expenditure in a conscious person or animal. It is determined by measuring the metabolic rate (see Chapter 3) of a subject who has been resting for some time in a warm room and after an overnight fast. BMR represents the energy required to keep the body's internal organs and systems functioning, e.g. for breathing, circulation of blood and brain function. In a sedentary person, BMR accounts for more than two-thirds of total energy expenditure and only in those who are extremely active will it make up less than half of total energy expenditure. If BMR is measured in large samples of individuals, within specified age and sex bands, there is a linear relationship between BMR and body weight. This means that one can derive regression equations that allow the average BMR of a group of people within an age band to be predicted from the average weight of the group.

Table 6.1 Selected current UK EARs for energy and corresponding US RDAs

Age (years)	UK EAR		US RDA	
	kcal	MJ	kcal	MJ
1–3 boys	1230	5.15	1300	5.44
1–3 girls	1165	4.86	1300	5.44
7–10 boys	1970	8.24	2000	8.37
7–10 girls	1740	7.28	2000	8.37
15–18 men	2755	11.51	3000	12.56
15–18 women	2110	8.83	2200	9.21
19–50 men	2550	10.60	2900	12.14
19–50 women	1940	8.10	2200	9.21
75+ men	2100	8.77	2300	9.63
75+ women	1810	7.61	1900	7.96

All values are kcal/MJ per day. Data sources: COMA (1991) and NRC (1989a).

For example, in women aged 18–29 years the following regression equation allows prediction of average BMR from average weight:

$$\text{BMR (kcal/day)} = 14.8 \times \text{weight (kg)} + 487$$

So, if the average weight of women in this age band is 60 kg, then:

$$\text{average BMR} = (14.8 \times 60) + 487 = 1375 \text{ kcal/day}$$

In order to estimate the average daily energy requirement of this group, one must multiply the BMR by a factor that reflects the level of physical activity of the group, this is called the **physical activity level (PAL)**. If this population is assumed to be sedentary in both their occupation and in their leisure time, a PAL value of 1.4 would be appropriate and so the average energy expenditure of the group (and thus their average requirement) would be estimated at:

$$\text{BMR} \times \text{PAL} = 1375 \times 1.4 = 1925 \text{ kcal/day}$$

This is essentially how the UK panel setting the dietary standards (COMA, 1991) arrived at the figure of 1940 kcal/day (8.1 MJ/day) for women aged 19–50 in Table 6.1 (the slight difference in numbers occurs because of interpolation over the larger age band).

In principle, this is the method COMA (1991) used to set the EARs for energy for all adults and older children. They used regression equations like the one above to predict BMR from the average weight of Britons in each age group. For most groups they multiplied the BMR by a PAL of 1.4 to reflect the generally sedentary nature of the British population. For adults aged over 60 years, they used a PAL of 1.5. This higher PAL for elderly people seems surprising given the clear evidence that activity decreases in the elderly (the reasons for this anomaly are outlined in the next section). For younger children, estimates of energy requirements are largely based upon survey measurements of average energy intakes.

The US panel that set the RDAs (NRC, 1989a) used very similar regression equations to predict the average BMR (or the resting energy expenditure, REE) of Americans in the various age bands and multiplied by an activity (PAL) factor. The large differences between the UK and US values in Table 6.1 are mainly due to differences in the activity factor

Table 6.2 A guide to predicting the PAL multiple of an individual or group

Non-occupational activity	Occupational activity					
	Light		Moderate		Moderately heavy	
	M	F	M	F	M	F
Non-active	1.4	1.4	1.6	1.5	1.7	1.5
Moderately active	1.5	1.5	1.7	1.6	1.8	1.6
Very active	1.6	1.6	1.8	1.7	1.9	1.7

After COMA (1991).

used by the two panels. Americans also tend to be a little heavier than Britons. The US panel used PAL multiples of between 1.55 and 1.7 in older children and adults up to 50 years, compared to the UK panel's 1.4. Like the UK panel, NRC used a PAL multiple of 1.5 for older people.

It must be borne in mind when using figures such as those in Table 6.1 that these are estimates of average requirements. They are not intended to be used as an accurate statement of the requirements of individuals. Many different factors will affect the energy expenditure and thus the energy requirements of an individual, such as:

- size – in general, the bigger the body the greater the energy expenditure
- body composition – lean tissue is metabolically more active and uses more energy than adipose tissue
- activity level (as discussed above)
- environmental conditions such as the ambient temperature

- physiological factors such as hormone levels
- rate of growth in children
- individual genetic variability.

It is possible to estimate roughly the energy requirements of an individual. One could use the regression equation appropriate for the person's age to estimate their BMR from their weight (a list of regression equations can be found in COMA, 1991). Alternatively, one can assume that in adults under 50 years, the BMR is 0.9 kcal/hour for a woman (1.0 for a man). Such approximations assume average proportions of fat and lean in the body and so they become increasingly unreliable as one moves towards extremes of fatness or leanness. Using this rule of thumb, the estimated BMR of a 65-kg man would be:

$$65 \times 1.0 \times 24 = 1560 \text{ kcal/day}$$

This estimated BMR must then be multiplied by an appropriate PAL. Table 6.2 was offered by COMA (1991) as a guide to deciding upon the appropriate

- The average BMR of a group can be estimated from their average body weight using regression equations derived from measurement of BMR and body weight in a large sample of people.
- This BMR can be used to estimate average energy requirements by multiplying it by a factor that reflects the activity level of the group – the physical activity level (PAL).
- This PAL can vary from under 1.3 in elderly, housebound people to well over 2 in athletes during training.
- In the UK, the dietary standard for energy is the Estimated Average Requirement (EAR) and in the USA it is the Recommended Dietary Allowance (RDA).
- Both the EAR and the RDA are estimates of average requirements and any individual's requirements will depend upon many factors, such as their size, body composition, activity level, physiological state and environmental conditions.

- Children have higher relative metabolic rates than adults and thus require more energy per kilogram of body weight.
- Men have higher average energy requirements than women because they are bigger and have a higher lean-to-fat ratio in their bodies.
- Basal metabolic rate (BMR) declines in the elderly due to reduced lean body mass and this compounds with declining activity levels to substantially reduce the average energy requirements in the very elderly.

PAL multiple. It uses three categories each for both occupational and non-occupational activity levels. Only those who are extremely active (e.g. athletes in training) or extremely inactive (e.g. elderly, housebound people) will fall outside this range.

It takes about half an hour's brisk walking to raise the PAL by 0.1 on that day.

VARIATION IN AVERAGE ENERGY REQUIREMENTS – GENERAL TRENDS

In Table 6.1, we saw quite large differences between the estimated average energy requirements of Britons and Americans which largely reflect the different assumptions about activity levels and average body weights, as discussed in the previous section. Other trends that are apparent upon close inspection of this table are summarized below.

- The energy requirements of males are generally higher than those of females. This is largely because the average male is bigger than the average female. In older children and adults, females also have a higher proportion of adipose tissue, which is metabolically less active than lean tissue.
- The relative energy requirement (per unit body weight) of growing children is higher than that of adults – a 2-year-old child may be only around a fifth of the weight of an adult but requires more than half of the adult energy intake. In the case of older teenagers, even the absolute energy requirement is higher than in adults. This relatively high energy requirement of growing children partly reflects the energy requirements for growth but, particularly in small children, it is also partly a manifestation of the general inverse relationship between the relative metabolic rate and body size

(see Chapter 3). Younger children also tend to be more active than adults.

- Both the UK and US values in Table 6.1 imply that there is a considerable reduction in energy needs in the elderly. This suggests that older people will need less energy and eat less food than younger adults. As requirements for most other nutrients are not thought to decline with age, this increases the vulnerability of elderly people to nutrient inadequacies (see Chapter 14 for further discussion). The amount of metabolically active lean tissue declines in old age and the proportion of body fat increases. There is also no doubt that activity levels show an accelerating decline in middle age and old age. This decline in activity is reflected in the use of a reduced PAL multiple for elderly people in the US values. However, the UK panel took the surprising decision to use a PAL multiple of 1.5 for older adults, compared to only 1.4 in younger adults. This decision was made because the panel was keen to discourage overconsumption of energy in younger people because of the prevalence of obesity. In elderly people, the panel's priorities were to maintain sufficient energy intake to allow adequate intake of other nutrients and to prevent elderly people becoming underweight. The values in Table 6.1 almost certainly underestimate the real decline in energy intake that occurs in the very elderly. The real PAL of elderly housebound people may be 1.3, or even less.

THE ENERGY CONTENT OF FOODS

The oxidation of foodstuffs in metabolism is often likened to the oxidative processes of combustion or burning. This would suggest that the heat energy released during the burning of a food should be an

- Food yields less energy when metabolized than when burnt because of losses in faeces (undigested material) and in urine (mainly urea).
- Fats yield about 37 kJ/g (9 kcal/g) when metabolized.
- Carbohydrates yield about 16 kJ/g (3.75 kcal/g).
- Proteins yield 17 kJ/g (4 kcal/g).
- The metabolizable energy of foods is determined by measuring the content of fat, carbohydrate and protein and using the energy equivalents of these nutrients.

indicator of its metabolic energy yield. In a **bomb calorimeter**, a small sample of food is placed in a sealed chamber, which is pressurized with oxygen to ensure complete oxidation (burning). The food is then ignited electrically and the heat energy released during combustion of the food sample is measured. In practice, the heat energy released during combustion will significantly overestimate the energy that is metabolically available from most foods. The predictive value of the energy of combustion is worsened if the food is high in indigestible material or high in protein.

Some components of a food may burn and release heat energy but not be digested and absorbed and so this fraction of the energy of combustion will be lost in the faeces, e.g. some components of **dietary fibre**, lignin (wood) and indigestible proteins like those in hair.

In the bomb calorimeter there is complete oxidation of protein to water, carbon dioxide and oxides of nitrogen, whereas in metabolic oxidation most of the nitrogen is excreted in the urine as urea. Urea will burn and release energy and so this means that some energy released from protein when it is burnt is lost in the urine when it is metabolized. The metabolic energy yield from protein is only about 70% of the heat energy released during combustion; the other 30% is lost in the urine.

To determine the **metabolizable energy** using a bomb calorimeter, the heat energy released during combustion has to be corrected for this energy lost in the faeces and urine.

Nowadays, the usual way of determining the energy value of foods is to measure the available carbohydrate, fat, protein and alcohol content of the food and then to use standard conversion values for the metabolizable energy of each of these nutrients:

1 g of carbohydrate yields 16 kJ (3.75 kcal)

1 g of protein yields 17 kJ (4 kcal)

1 g of fat yields 37 kJ (9 kcal)

1 g of alcohol yields 29 kJ (7 kcal)

In the main, these approximations hold irrespective of the type of fat, carbohydrate or protein; the carbohydrate figure is applied to both starches and sugars. Small variations in these conversion factors are of little significance given the many other sources of error in estimating dietary energy intake. Non-starch polysaccharide or dietary fibre has traditionally been assumed to contribute nothing to metabolizable energy, but recent studies suggest that it may yield up to 8 kJ (2 kcal) per gram (via its fermentation by intestinal bacteria to volatile fatty acids and the subsequent absorption and metabolism of those acids).

SOURCES OF DIETARY ENERGY BY NUTRIENT

Dietary recommendations and survey reports often quote carbohydrate, fat, protein and even alcohol intakes as a percentage of total energy. This method of presentation allows meaningful comparison of the diet composition of people with widely differing energy intakes. It can be seen as a way of comparing the 'concentration' of fat, carbohydrate or protein in different foods and diets. Quoting absolute values for intakes of these major nutrients (e.g. grams per day) may be of limited usefulness and may even, on occasion, be misleading. For example, someone eating

4.2 MJ (1000 kcal) per day and 30 g of fat will get the same proportion of their dietary energy from fat (27%) as someone consuming 8.4 MJ (2000 kcal) and 60 g of fat.

As another example of the usefulness of this tool, absolute *per capita* intakes of fat (g/day) in the UK have dropped very sharply over the last couple of decades. However, this reduction is largely accounted for by a reduction in total food intake over this period and only in the last few years has the proportion of energy from fat started to fall.

In order to calculate macronutrient intakes as a percentage of total energy, the energy equivalents of fat, carbohydrate and protein given in the previous section are used in the following formula:

$$\frac{\text{g of nutrient consumed} \times \text{energy equivalent (as above)}}{\text{total energy intake}} \times 100$$

For example, to calculate the proportion of energy from fat in an 11 000-kJ diet containing 100 g of fat:

$$\frac{100 \times 37}{11\,000} \times 100 = \frac{3700 \times 100}{11\,000} = 33.6\% \text{ of energy as fat}$$

Table 6.3 shows estimates of the contribution of the various energy-yielding nutrients to the total energy supplies of adults in the UK. This was measured directly using a weighed inventory of the intakes of a nationally representative sample of British adults (Gregory *et al.*, 1990). At first sight, these

Table 6.3 Sources of energy in the diets of British adults expressed as percentages of total energy

Nutrient	Men	Women
Fat (%)	37.6	39.4
Protein (%)	14.1	15.2
Carbohydrate (%)	41.6	43.1
Sugars (%)	17.6	19.2
Starches (%)	24.0	23.9
Alcohol (%)	7.1	2.3
Total (kcal/day)	2450	1680
(MJ/day)	10.26	7.03

Data source Gregory *et al.* (1990)

figures seem to indicate that there are quite significant differences in the contribution of the major macronutrients to the energy intakes of men and women. However, these differences are largely a consequence of the higher alcohol consumption of men. The energy contribution of the major macronutrients is often expressed as a percentage of food energy, i.e. excluding alcohol. When the results of this survey are expressed in this way then the values for men and women are very similar (Table 6.4).

The data in Tables 6.3 and 6.4 were collected well over 10 years ago, in 1986–7. In Britain, the annual National Food Survey (NFS) (see Chapter 3 for details) has been providing regular information about British diets for 60 years. The NFS can be used to estimate the nutritional composition of household food purchases (Table 6.5). Until 1992, the NFS did not record confectionery, soft drinks or alcoholic drinks and so the results of this survey are now presented both including and excluding these items in order to facilitate historical comparisons.

It is clear from Tables 6.3, 6.4 and 6.5 that fat and carbohydrate together provide the bulk of the dietary energy in Britain, around 85% of the total. This will also be true of most European and American diets and, indeed, it will be true for all but the most unusual or extreme diets, e.g. old-fashioned 'high-protein' reducing diets in which almost the only foods eaten are lean meat, white fish, eggs and skimmed milk. As the bulk of any diet's energy usually comes from fat and carbohydrate, this means that, as the proportion of one goes up, so the proportion of the other tends to go down. An inverse relationship between the proportion of dietary energy derived from fat and carbohydrate is almost inevitable and could be termed a carbohydrate–fat seesaw. This means that in most Western diets, there also tends to be an inverse relationship between dietary fat content and the major classes of carbohydrate, i.e. sugars, starches and non-starch polysac-

Table 6.4 Some of the figures in Table 6.3 expressed as a percentage of food energy, i.e. excluding alcohol

Nutrient	Men	Women
Fat (%)	40.4	40.3
Protein (%)	15.2	15.6
Carbohydrate (%)	44.7	44.2

Data of Gregory *et al.* (1990)

Table 6.5 The percentage of energy obtained from fat, carbohydrate and protein in household food purchases in the UK

	Included	Excluded
Fat (%)	38	40
Protein (%)	13	14
Carbohydrate (%)	47	46
Sugars (%)	21	19
Starches (%)	26	28
Total energy <i>per capita</i>		
(kcal)	1960	1850
(MJ)	8.3	7.8

Data source: MAFF (1997).

charide (NSP)/ dietary fibre. The consequences of the sugar–fat seesaw were discussed in Chapter 4. In Chapter 9, we will also see that the inverse relationship between dietary fat and fibre makes it difficult to determine whether any proposed benefits of high-fibre (NSP) diets are due to fibre *per se* or due to the fact that they also tend to be low in fat.

ENERGY DENSITY

The energy density of a food or diet is the metabolizable energy yield per unit weight of food (e.g. kcal or kJ per 100 g food). If a food is of high energy density, then a little of it provides a lot of energy, whereas one can eat a lot of a low energy density food and yet obtain little energy. This is illustrated in Table 6.6,

where the weights of various foods required to provide 4.2 MJ (1000 kcal) are shown. The range of energy densities in Table 6.6 is very large: just over a tenth of a kilo of vegetable oil will provide 4.2 MJ, but one needs over 9 kg of boiled swede or 7 kg of lettuce. A number of the general trends illustrated by Table 6.6 are summarized below.

- As fat yields more than twice as much energy as either carbohydrate or protein, fat content must be a major factor increasing the energy density of foods or diets. High-fat foods are inevitably of high energy density (e.g. nuts, seeds and oils), whereas low-fat foods are generally of low or moderate energy density.
- Most fruits and vegetables are almost fat free and are high in water and fibre. They thus have very low energy densities (e.g. grapes and lettuce). Drying fruits (e.g. raisins) or adding sugar to them (e.g. strawberries canned in syrup) increases their energy density. Cooking or preparing vegetables with fat or oil greatly increases their energy yield (e.g. French fried potatoes, hummus, and fried tofu).
- When eaten boiled, starchy cereals and roots are low in fat and absorb substantial amounts of water. This means that, despite their importance as major sources of energy in many diets, their energy density is rather modest (e.g. boiled potatoes, boiled rice and corn grits).
- Very lean meat and white fish are low in fat and so their energy density is moderate or low (e.g. roast chicken meat), especially if boiled (e.g. lean boiled ham, poached cod). The meat in our ancestors' diets (and even a few populations today) would

- The macronutrient content of diets and foods is often expressed as the percentage of the total energy derived from that nutrient, e.g. the percentage of energy as fat.
- This is a measure of the 'concentration' of these nutrients in a diet.
- Using a weighed inventory method, 40% of the food energy of the British diet was found to come from fat, 45% from carbohydrate and 15% from protein. Similar estimates have been made using the National Food Survey.
- In most diets, fat and carbohydrate together provide most of the food energy (about 85% in the UK).
- As the proportion of energy derived from fat in a diet goes up, so the proportion from carbohydrate tends to go down and *vice versa* – the so-called carbohydrate–fat seesaw.
- In most Western diets, there also tends to be a sugar–fat seesaw.

have come from wild animal carcasses and this is usually low in fat.

- Many of the modern processed foods in Table 6.6 have a relatively high energy density. Many modern methods of preparing and processing foods tend to remove water and fibre but add fat and/or sugar (e.g. fish fingers, chicken nuggets, fruit/meat pies, cornflakes, cookies and cakes).

The present-day diets of affluent populations is of much higher energy density than the diet of our hunter-gatherer ancestors or in subsistence agricultural communities today. The cross-population changes in diet that are associated with increasing affluence also increase the energy density of the diet, i.e. the partial replacement of vegetables and starchy, fibre-rich foods with foods rich in fats and sugars but low in fibre. These compositional changes produce diets that, as well as being more energy dense, are also more palatable. This combination of high palatability and high energy density may increase the likelihood of energy intake exceeding requirements and encourage weight gain. If gut-fill cues (e.g. stomach distension) play any role at all in the regulation of energy intake, then one would expect high energy density of the diet to reduce the effectiveness of food intake-regulating mechanisms. This may be particularly so if there is also low energy expenditure due to inactivity. It is widely believed that a reduction in the energy density of adult diets in affluent, industrialized countries will help people to control their energy intake better and so lessen the risks of excessive weight gain.

Diets high in fat, and therefore of high energy density, have often been used to induce obesity in rodents.

According to Miller (1979), high energy density, rather than fat content *per se*, is the primary obesifying influence. Varying the energy density and the proportion of fat independently (by using an inert bulking agent) produced a strong correlation between the fatness of the animals and dietary energy density, but not between body fatness and dietary fat content.

At the other end of the spectrum, it is now widely believed that the low energy density of some children's diets, especially weaning diets, may be a major precipitating factor for malnutrition. Weaning diets, particularly in some developing countries, may have energy densities that are so low that, even under optimal conditions, children may have difficulty in consuming a sufficient volume of food to satisfy their energy needs. Children fed these low-energy diets may be in energy deficit, i.e. starved, despite apparently being fed enough to satisfy their demands. These low-energy diets are based upon starchy staples, are often practically fat free and have large amounts of water added to produce a consistency considered suitable for babies and young children. Children fed such low energy density diets may need to consume anything up to eight times the weight of food that children on a typical Western weaning diet would need to obtain the same amount of energy.

NUTRIENT DENSITY

Nutrient density is the amount of nutrient per unit of energy in the food or diet (e.g. μg nutrient per $\text{kcal}/\text{k}\text{g}$). In the diets of the affluent, nutrient density is

- Energy density is the number of kJ/kcal per unit weight in a food or diet.
- Fat is a concentrated source of energy so high-fat foods are energy dense and adding fat (e.g. frying) to foods raises their energy density.
- Foods with high water content such as fruits and vegetables have very low energy density.
- Starchy foods have only modest energy density, despite being the major energy source in most pre-industrial human diets.
- Modern diets of affluent populations tend to be much higher in energy density than those eaten by hunter-gatherers or subsistence farmers.
- High energy density diets probably predispose to obesity.
- Very low-energy density weaning diets may precipitate malnutrition because children are unable to eat enough of the dilute food to satisfy their needs.

Table 6.6 The weights of selected foods that would need to be eaten to obtain 4.2 MJ (1000 kcal)

		<i>Dairy foods</i>	
<i>Fruits</i>			
Orange	5950	Whole milk	1515
Banana	2280	Skimmed milk	3030
Grapes	1667	Double cream	223
Raisins	368	Cheddar cheese	243
Strawberries	3704	Butter	136
Strawberries (canned in syrup)	1538		
Fruit pie (individual purchased)	271	<i>Meat, fish and products</i>	
		Boiled egg	680
<i>Vegetables</i>		Fried egg	560
Lettuce	7143	Lean boiled ham	758
Tomato	5882	Roast chicken meat	606
Green beans	4000	Southern fried chicken	350
Broccoli	4167	Chicken nuggets	338
Boiled swede	9091	Grilled rump steak	
Red kidney beans (boiled)	971	with fat	459
Soya beans (boiled)	709	trimmed	595
Tofu (fried)	383	Cheeseburger	380
Chick peas (boiled)	836	Salami	204
Hummus	535	Sausage roll	210
		Grilled pork sausage	314
<i>Starchy staples and products</i>			
Boiled potatoes	1389	Poached cod	1064
French fries	357	Cod fried in batter	503
Potato crisps (chips)	183	Steamed salmon	508
Boiled yam	752	Boiled prawns	935
Boiled sweet potato	1190	Fish fingers (fried)	429
Boiled rice	725	Fried scampi (breaded)	316
Spaghetti (boiled)	962		
Wheat bread	464	<i>Nuts, seeds and oils</i>	
Corn grits (made with water)	1666	Peanuts (raw)	177
Porridge		Walnuts	145
made with water	2041	Sunflower seeds	167
made with whole milk	862	Margarine	136
Tortilla/chapatti (no added fat)	478	Low-fat spread	256
Corn flakes	278	Vegetable oils	112
Weetabix®	279		
Chocolate biscuits (cookies)	191	<i>Miscellaneous</i>	
Fruit cake	282	Sugar	254
Doughnuts (ring)	252	Chocolate	189
Pizza (cheese and tomato)	400	Quiche	318

Adapted from Webb (1998).

often almost a mirror image of energy density; adding ingredients to foods or diets that contain much energy but few nutrients (e.g. fats, sugar and alcohol) raises the energy density but reduces the overall nutrient density. Diets and foods high in nutrient density help to ensure nutritional adequacy, but those low in nutrient density increase the possibility that energy requirements may be met and appetite satisfied with-

out also fulfilling the requirements for essential nutrients. Those consuming relatively small amounts of energy (e.g. those on reducing diets or elderly, immobile people) or those requiring higher amounts of some nutrients (e.g. pregnant women) may need to take particular care to ensure that their diet is nutrient dense. It is possible to predict the likely adequacy of a combination of foods for any particular nutrient by

Table 6.7 The amount of energy from various foods that would need to be consumed in order to obtain the UK adult RNI for vitamin C (40 mg)

New potatoes (boiled)	175	735	
French fries			
burger bar	2764	11611	
home-made UK	840	3528	
Avocado pear	1266	5317	
Banana	345	1499	
Orange	27	113	
Mandarin oranges (canned in syrup)	139	584	
Green pepper (capsicum)	5	21	
Tomatoes			
grilled	45	189	
fried in oil	228	958	

Compare with the UK EAR for an adult male of 10.6 MJ/day (2550 kcal/day).

multiplying the nutrient density of the mixture by the EAR (or RDA) for energy:

nutrient density × energy EAR(or RDA) = amount of nutrient consumed by subjects meeting their energy needs

If this figure exceeds the RNI (RDA) for that nutrient, the diet is probably adequate for that nutrient.

Similar calculations are also often made for individual foods in order to illustrate their value as a source of particular nutrients; this is illustrated by some examples in Table 6.7.

Clearly, oily French fries from a burger bar are a poor source of vitamin C. Even fresh, home-made

chips (fries) are much worse than boiled new potatoes. Oranges, peppers and grilled tomatoes are clearly good sources of the vitamin. The addition of sugar (canned oranges) or fat (fried tomatoes or potatoes) reduces the nutrient density because it adds energy but no more vitamin C. Avocados have relatively modest amounts of vitamin C and the nutrient density is further depressed because they are relatively high in fat and energy.

SOURCES OF DIETARY ENERGY BY FOOD GROUPS

The same surveys that were used to estimate the contribution of the major macronutrients to energy intake in the UK (see Tables 6.3–6.5) can also be used to estimate the contribution that the major food groups make to energy intake (Table 6.8).

Considering the very different methodologies of the two surveys used to devise Table 6.8, the two sets of results are reassuringly similar. Some particular factors that would cause differences between the two surveys are listed below.

- In the weighed inventory (Gregory *et al.*, 1990), cooking fats and oils will be incorporated into other foods, whereas in the National Food Survey (MAFF, 1997) they are measured as purchased and incorporated into the fats and sugars group.
- Data for eating out are collected and reported separately in the National Food Survey and so the figures given for MAFF (1997) do not include food eaten out.
- There may have been changes in eating habits over the 10 years (1986 versus 1996) separating the data-collection periods of the two surveys.

- Nutrient density is the amount of nutrient per kJ or kcal in a food or diet.
- Adding sugar, alcohol or fat to foods lowers nutrient density because it adds energy but little or no nutrient.
- A nutrient-dense diet helps to ensure nutritional adequacy, but low nutrient density increases the chances that appetite can be satisfied without fulfilling all nutrient requirements.
- People with low energy intakes especially need to ensure that their diet is nutrient dense, e.g. elderly housebound people and those on reducing diets.

Table 6.8 The percentage contribution of the major food groups to total energy intakes in the UK

	Gregory et al. (1990)	MAFF (1997)
Cereals and cereal products	30	33
Milk and milk products (including cheese)	11	13
Meat, fish, eggs and products	20	16
Fats and sugars (including soft and alcoholic drinks and confectionery)	21	23
Fruits and vegetables (about half of this energy comes from potatoes and potato products)	14	14

As measured by weighed inventory (Gregory *et al.*, 1990) and by household food purchases (MAFF, 1997)

- The weighed inventory was confined to adults aged 16–64 years, but the National Food Survey includes all age groups

If one compares Table 6.8 with the 'ideal diets' suggested by the **UK food guide plate** (or the US **food guide pyramid**), the fats and sugars would seem much too prominent; these provide few nutrients but lots of energy, largely as fat and sugar. Cereals plus potatoes are the largest contributors of energy (almost 40%), but a deeper analysis would indicate that highly refined products and those with considerable added fat and/or sugar are much more prominent than would be considered ideal. Fruits and vegetables (excluding potatoes, nuts, etc.) provide only around 5–6% of total energy. This is lower than ideal, but a very low energy yield from these foods is almost inevitable given the very low energy density of most of them.

STARVATION

The immediate causes of starvation

For most of the world population, starvation results from local food shortages or from lack of money to buy food. It was also suggested earlier in the chapter that some children may be unwittingly starved because they are fed foods of such low energy density that they are unable to consume a sufficient volume of the dilute food to meet their energy needs. Inadequate food intake usually results in hunger and, if food is available, in eating. In some circumstances, hunger may not occur or be suppressed and in some circumstances nourishment will not reach the tissues despite hunger and eating. In affluent countries, these causes of inadequate intake may predominate, such as the examples below.

- According to data in the National Food Survey:
 - about 40% of the energy in the UK diet comes from the cereals and potatoes
 - 13% from milk and milk products
 - 17% from the meat group
 - 23% from fats and sugars (including drinks and confectionery)
 - 5–6% from the fruits and vegetables food group.
- These figures are quite similar to those obtained by weighed inventory using a representative sample of British adults.
- Ideally, more of the energy should come from unrefined cereals and potatoes, fruits and vegetables, slightly less from the milk and meat groups and much less from the fats and sugars group.

- Illness, injury or therapy may result in loss of appetite.
- Infirmitiy may reduce the ability to eat or to obtain and prepare food, e.g. in stroke patients or elderly, immobile people.
- Illness may result in hypermetabolism or loss of nutrients, e.g. the increased metabolic rate in those with overactive thyroid glands or the glucose and ketones lost in the urine of diabetics (see Chapter 15).
- Psychological state or psychological illness may result in inadequate food intake despite its ready availability, e.g. anorexia nervosa.
- Some diseases may result in poor digestion and absorption of nutrients, e.g. coeliac disease and cystic fibrosis (see Chapter 15).

Physiological responses and adaptations

The most obvious physiological response to starvation is wasting. If energy intake is not sufficient to cover expenditure, the deficit must be made up by burning body energy reserves. All tissues, with the exception of the brain, will waste during starvation – the most obvious wasting will be in body fat and muscle, but other vital organs will also waste. After prolonged starvation, the heart may be only half its original weight and this wasting of heart muscle can lead to death from circulatory failure. The intestines of starving people will atrophy: they become thin and have greatly reduced digestive and absorptive capacity. This may cause problems in re-feeding people after prolonged starvation.

A healthy and well-nourished, non-obese man should be able to lose more than 25% of his initial body weight before his life is threatened. During such a 25% weight loss, the man would lose around 70% of his stored fat and around 25% of his body protein (Passmore and Eastwood, 1986). The total carbohydrate content of the body is small (about 500 g), even in well-fed people, and so the small loss of carbohydrate during this prolonged period of weight loss would make an insignificant contribution to total losses of body energy. The energy yield of this 25% loss of body weight represents around 6–7 weeks' energy expenditure in a totally starved and sedentary man. This means that well-nourished adults can survive several weeks of total starvation, and this was

strikingly illustrated by the IRA hunger strikers in the Maze Prison in Northern Ireland during the 1980s. It took up to 2 months of total starvation for these healthy young men to reach the point of death (obese people may survive much longer). This capacity of people to survive long periods of food shortage, or even complete starvation, clearly represents a very considerable aid to the survival of the species. If there is a temporary food shortage, there is a good chance that many people will survive until food supplies are restored, e.g. a new harvest. Children are more vulnerable to starvation than adults: the survival times of completely starved babies or children may be measured in days rather than weeks, depending upon age.

Many cells can utilize either glucose or fatty acids as an energy source. As body stores of carbohydrate are minimal, during a fast such cells will increasingly switch to using fatty acids as an energy source. The cells of some tissues, including the brain and red blood cells, cannot use fatty acids directly as an alternative energy source and most tissues have a limited capacity to use fatty acids. Glucose can be manufactured in the liver from the glycerol component of stored fat and from amino acids released from the protein in muscles and vital organs, but *glucose cannot be synthesized from fatty acids* (see gluconeogenesis in Chapter 5). This process of gluconeogenesis is energy expensive (i.e. wasteful) and supplies of amino acids are limited because they are being taken from protein in muscles and essential organs. A maximum of around 5 kg of protein is available for gluconeogenesis in an adult man.

It was once thought that the brain relied exclusively upon glucose as its substrate for energy production. As there are negligible carbohydrate stores, this would mean that the brain would have to rely entirely upon glucose produced from gluconeogenesis during prolonged fasting. The brain normally uses the equivalent of around 100 g of glucose per day and, unless some adaptation occurred, the use of protein in gluconeogenesis would lead to rapid depletion of body protein reserves, and the size of these protein reserves would limit the duration of survival during fasting. Starving people would be likely to die from depletion of body protein reserves before body fat stores were exhausted. If one monitors nitrogen excretion (i.e. protein breakdown) during fasting, one finds that it declines as the fast is extended and after a week or so is down to about

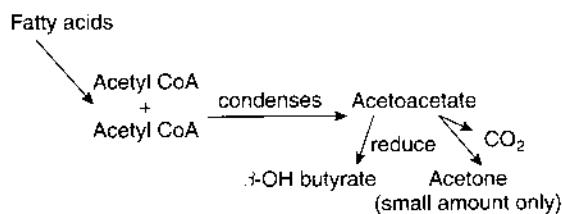


Figure 6.1 The origins of ketone bodies. CoA, coenzyme A.

30% of the initial rate. An important adaptation to starvation occurs that preserves lean tissue during prolonged fasting. During the course of fasting, the blood levels of substances called ketone bodies rise as they are produced from fatty acids in the liver. After a few days of fasting, these ketone bodies become the most abundant metabolic fuel in the blood. It is now clear that they act as alternative substrates for the brain and other tissues during prolonged fasting. These ketones are synthesized in the liver from the acetyl Coenzyme A produced during the β -oxidation of fatty acids (Figure 6.1; see Chapter 5).

After prolonged fasting, as much as three-quarters of muscle energy production and more than half of the brain's energy supply will be from β -hydroxybutyrate (β -OH butyrate) and acetoacetate (acetone is produced in small amounts as a by-product of the process but is not metabolized). The brain is thus indirectly using fatty acids as an energy source. During a prolonged fast, the brain's requirement for glucose will fall from 100 g/day to 40 g/day and most of this will come from the glycerol component of stored fat. This use of ketones as a substrate thus represents a substantial brake upon the depletion of lean tissue during starvation.

For many years, these ketones were thought of solely as abnormal metabolic products produced during uncontrolled diabetes. In severe diabetes, they are produced to excess and their concentration builds up to the point at which they become highly toxic. They are toxic principally because they are acidic and produce a metabolic acidosis. Acetone is responsible for the 'pear drops' smell on the breath of people in diabetic (hyperglycaemic) coma. People on severe reducing diets or diets in which carbohydrate intake is severely restricted will experience mild ketosis.

A third important adaptation to starvation is a reduced rate of energy expenditure. This slows the rate at which reserves are depleted and extends the

survival time. Below are listed some of the very obvious reasons why energy expenditure decreases during fasting (or severe dieting).

- With little or no eating, there is less food to digest, absorb and assimilate. These processes each have an energy cost and eating raises resting metabolic rate.
- Most starving people curtail their physical activity – many anorexics are an exception to this generalization; they exercise in order to lose yet more weight.
- As the body wastes, it gets smaller and so the energy costs of maintenance decrease.
- The energy costs of any activity are also reduced if the body is smaller.
- There is an increase in metabolic efficiency, perhaps related to changes in body temperature regulation.

This means that, during starvation, the rate of energy expenditure decreases and the rate at which energy stores are used up also decreases. During deliberate weight loss, the amount of energy needed to maintain body weight also declines. Many dieters find that a diet initially succeeds in achieving weight loss, but then their weight stabilizes and they need to reduce their intake still further to get continued weight loss. Prentice *et al.* (1991) estimated that the crude reduction in energy expenditure during dieting might exceed 25% but, when corrected for changes in lean body mass, it varies between 5% and 15%, depending upon the severity of the intake restriction.

Some adverse consequences of starvation

Some of the consequences of starvation are summarized in Table 6.9. Those suffering from anorexia nervosa (see later in the chapter) vividly illustrate these consequences in developed countries.

Endocrine function becomes deranged during starvation. Pituitary gonadotrophin secretion is greatly depressed in starvation and this manifests itself outwardly in a cessation of menstruation in starving women. Puberty is delayed by malnutrition. In industrialized countries, the age of onset of menstruation has become markedly lower over the last century as children have become better nourished

Table 6.9 Some consequences of starvation including the starvation associated with anorexia nervosa

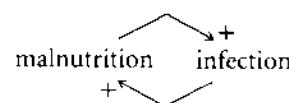
Wasting of lean tissue and vital organs
Shrinkage of heart muscle, irregular heart rhythm and risk of heart failure
Reduced digestive and absorption capability in the gut, often leading to diarrhoea
Reduced growth and development in children and young people, perhaps intellectual as well as physical development
Depressed immune function and increased susceptibility to infections
Delayed puberty, lack of menstruation, infertility and loss of sex drive
Reduced bone density (osteoporosis) and increased risk of fractures, especially stress fractures in anorexic athletes
Slow healing of wounds
Reduced strength and physical work capacity and impaired performance in dancers and athletes
Hypothermia
Insomnia, changes in personality and brain function

After Webb (1998).

and grow faster. Women need a minimum threshold amount of fat to maintain their fertility and proper menstrual cycling. Underweight women may have difficulty in conceiving a child and they are more likely to have an underweight baby. The reduced secretion of sex hormones also leads to loss of bone mass. This makes fractures more likely and may, for example, predispose women who have suffered from

anorexia nervosa to osteoporosis fractures later in life.

Infection is often associated with starvation and there is said to be a cycle of infection and malnutrition:



Malnutrition predisposes to infection and infection increases the severity of malnutrition, or may precipitate malnutrition in areas where there is not overt food shortage. Sick children (or adults) may have reduced appetite and increased requirement for nutrients. They may also be starved or fed a very restricted diet in the belief that this is the correct way to treat sick children. Malnutrition has very profound, specific and deleterious effects upon the functioning of the immune system; these are discussed at some length in Chapter 15. Cohort studies in malnourished children in which nutritional status has been related to subsequent death rate showed that the risk of dying increases exponentially as one moves from the better nourished to the mildly, moderately and severely malnourished (Waterlow, 1979).

There may also be very pronounced psychological effects and personality changes produced by starvation. During the 1940s, in Minnesota, a group of young male volunteers were subjected to a period of partial starvation that resulted in a 25% body weight loss (Keys *et al.*, 1950; see Gilbert, 1986, for sum-

- Worldwide, starvation is usually the result of famine or poverty, but it can also be a consequence of illness or infirmity.
- During starvation, the body uses stored fat and tissue protein to make up its energy deficit.
- Healthy, non-obese adults can survive up to 2 months of total starvation and for much longer on restricted rations.
- During starvation, ketone bodies (β -OH butyrate and acetoacetate) produced from fat become the major body substrates.
- Ketone use reduces the need to produce glucose and spares body protein.
- More than half of the brain's energy needs are met by ketones during prolonged starvation.
- Starvation leads to a large fall in energy expenditure (less digestion and assimilation, reduced activity, reduced tissue mass and increased metabolic efficiency).
- Adverse consequences of starvation include tissue wasting, reduced growth and healing, depressed immune function, reduced physical performance, endocrine disturbances and reproductive impairment, and psychological disturbances.

mary). The feeding behaviour of these men changed markedly during their period of deprivation, including a dramatic decline in their speed of eating. They became preoccupied with food, it dominated their conversations, thoughts and reading. They became introverted, selfish and made irrational purchases of unwanted items. Personality tests conducted during the period of starvation showed a considerable movement towards the neurotic end of the scale as starvation occurred. Towards the end of the period of weight loss, the results from personality tests with these starved men became comparable with those of people with frank psychiatric disease.

EATING DISORDERS

The term **eating disorders** is used to describe the diseases **anorexia nervosa** and **bulimia nervosa**. The term could also encompass the conditions of many other people who have some symptoms of these eating disorders but who do not meet the formal diagnostic criteria for either of them.

Characteristics and consequences

Anorexics have very low energy intakes and are severely underweight; occasionally, they may starve themselves to death. Anorexia nervosa is characterized by an obsessive desire to be thin and an intense fear of being fat, even though the victim may be severely emaciated. The disease has significant mortality, estimated at between 2% and 16%. Many sufferers appear to make only a partial recovery; they continue to eat very restricted diets, to maintain very low body weights and fail to start menstruating. The consequences of this self-starvation are essentially the same as for starvation generally (summarized in Table 6.9 earlier in the chapter). Some symptoms are particularly associated with anorexics, such as excessive growth of dry, brittle facial and body hair ('lanugo hair'), constipation and peripheral oedema. Those who take purgatives or resort to induced vomiting are also likely to suffer from adverse consequences of these behaviours, including:

- electrolyte imbalances, which may lead to muscle weakness, fatigue and increased susceptibility to cardiac arrhythmias

- irritation of the throat and damage to the teeth caused by the corrosive effects of vomitus.

The diagnostic characteristics of anorexia nervosa are:

- low body weight – at least 15% below expected minimum for age and height
- intense fear of gaining weight or becoming fat, even though underweight
- distorted body image – seeing themselves as fat even though they are alarmingly emaciated to the outside observer
- lack of menstrual periods in females for 3 consecutive months.

Many anorexics also frequently exhibit the following behaviours:

- frequent use of purgatives, induced vomiting or emetics
- high levels of physical activity, in contrast to most other victims of starvation.

Victims of the disease may also go to great lengths to hide their restricted food intake and their physical emaciation from relatives and friends. They may wear loose clothing, avoid social eating situations, eat very slowly, choose only foods very low in calories, conceal food that they have pretended to eat and then dispose of it, or induce vomiting after they have eaten. It is, of course, important to rule out other physical and psychological causes of undernutrition and weight loss before concluding that an underweight person is suffering from anorexia nervosa.

Bulimia nervosa is characterized by recurrent bouts of binge eating. Periods of restricted eating are interspersed with sometimes huge binges in which massive quantities and bizarre mixtures of foods may be consumed in very short periods of time. The voracious eating may continue until abdominal pain, sleep or interruption trigger its end. Often the binge is followed by self-induced vomiting or purging. The morbid fear of fatness and distortion of body image seen in anorexia are also seen in this condition. A diagnosis of bulimia would be made if the following features are present in people who are not very underweight:

- Recurrent bouts of binge eating in which large amounts of food are consumed and in which the person loses control of their eating.

- Recurrent bouts of inappropriate compensatory behaviour to prevent weight gain after binges, e.g. induced vomiting, purging, excessive exercise or fasting.
- These bouts of bingeing and compensation should occur at an average rate of twice a week for 3 months to qualify for a diagnosis of bulimia.
- The person's self image is unduly dependent upon his or her body shape and weight.

This condition can also have serious repercussions upon the health of sufferers, including:

- damage to the throat, teeth, oesophagus (gullet) and stomach caused by overloading or the induced vomiting, with consequent risk of infection
- possible harmful side-effects of any laxatives or emetics used.

Incidence

Some surveys amongst college students and amongst more random samples of women in the UK and USA have suggested disturbingly high proportions of subjects exhibiting bulimic behaviour (as many as 20% of sampled women in some surveys). Gilbert (1986) concluded 'that at any one time up to 2% of women up to the age of forty may be experiencing problems with controlling their eating'.

Anorexia nervosa is the best known of the eating disorders; the condition was first described more than a century ago, but there is general agreement that it has become more prevalent in recent decades. It affects less than 0.5% of women. Anorexia nervosa affects mainly females who typically are:

- white
- aged 15–25 years (although it can occur at any time between 12 and 44 years)
- from middle-class and upper-class families, well-educated and with a good knowledge of nutrition.

Cases do occur in men (less than a tenth of the frequency in women), and amongst other social and racial groups. As it becomes more common, so anorexia is radiating out from its traditional risk group and is being reported increasingly in Asian and Afro-Caribbean women and in women of other social classes.

Bulimia nervosa occurs in people whose weight is within the normal range and, as they are guilty and secretive about their behaviour, this makes it difficult to diagnose and difficult to make estimates of its incidence – hence the wide variability in the estimated frequency of eating disorders noted earlier. Perhaps the best estimates of frequency are about 1% of adolescent girls and young women and about a tenth of this frequency amongst the corresponding male population.

Causes

There seems to be a general acceptance that these eating disorders have become more prevalent in recent decades. It is widely believed that they are triggered by dieting and social pressures upon young people to be thin, pressures that may be particularly acutely felt by young women. The majority of adolescent girls have tried dieting, and preoccupation with body weight and dieting has been reported in children as young as 10 years old. Many bulimic patients have a history of cycles of weight loss and regain, or '**weight cycling**'. The average woman (and man) in most industrialized countries is getting fatter, as judged by surveys of weight and height, but images of the ideal woman are apparently getting thinner. Many of the models used in classical painting of the female form would be overweight by present standards, and studies on American winners of beauty contests and models for men's magazines indicated that the 'ideal' body shape got thinner during the 1960s and 1970s. Almost all of the models used by fashion photographers are lean and many are frankly underweight. Young women aspire to a body shape that is unhealthy and unattainable except by the most extreme restrictive diets. These disorders do seem to occur more often in girls and women for whom thinness is a career requirement, e.g. models, ballet dancers and airline stewardesses. In male athletes and ballet dancers, eating disorders are almost as common as in women, despite their general rarity in men.

Over the years, many suggestions have been made as to the individual causes of anorexia nervosa and related eating disorders – social conditions may favour increasing prevalence of the disease, but what individual factors make some women develop the disease? It is likely that the cause of these diseases is

multifactorial and that there is no one simple cause that applies to all sufferers.

The effects of starvation upon the men in the experiment of Keys and his colleagues in Minnesota were discussed earlier, and may be of some relevance in understanding the development of anorexic behaviour. The slowness of eating seen in these starving men is typical of anorexics. The men's unsociability, their obsessive interest in food and their curtailment of other activities are all recognized as traditional symptoms of anorexia nervosa. These men also developed personality profiles that were characteristic of people with frank psychiatric disease as they starved. During the period of re-feeding and rehabilitation of these men, many of them showed a marked tendency to binge eat, behaviour in many ways similar to that found in bulimia. Some months after the re-feeding period, a high proportion of the men had put themselves on a reducing diet; men who had previously had a normal interest in food became preoccupied with dieting after this bout of starvation.

These observations on starving individuals have encouraged the suggestion that anorexia and bulimia may be triggered by social and cultural conditions that require young women, in particular, to be thin. The

dieting itself may, in genetically susceptible individuals, cause some of the grossly aberrant and apparently irrational behaviour characteristic of anorexia nervosa and bulimia. Weight loss and leanness are generally admired and praised in our culture and thus the initial positive reactions to the results of anorexic behaviour may serve as powerful reinforcement or reward and thus encourage learning of anorexic behaviour. In someone whose self-esteem is low and who feels that, because of their family environment, they lack control over their own lives, this initially praised ability to totally control some aspect of their lives may also act as positive reinforcement and encourage continuation and intensification of the anorexic behaviour.

There are numerous other theories as to the origins of anorexia nervosa. Some have suggested that there is some defect in the hypothalamic mechanisms regulating intake, but this idea is largely discounted these days and the frequent occurrence of binge-eating suggests suppression of appetite rather than lack of appetite.

Some adherents to Freudian psychological theories have suggested that the subconscious goal of these women and girls is to avoid adult responsibility and sexuality by starving themselves. The cessation of menstruation, inhibition of sexual development

- The term **eating disorders** encompasses the two well-defined conditions anorexia nervosa, and bulimia nervosa, but could also include many other people with some features of these diseases who fail to meet the formal diagnostic criteria.
- People with anorexia nervosa starve themselves, sometimes to the point of death, because of an obsessive preoccupation with their weight and a pathological fear of fatness, even when they are grossly emaciated.
- People with bulimia nervosa are not seriously underweight, but engage in recurrent bouts of bingeing followed by inappropriate compensatory behaviour to prevent weight gain, e.g. induced vomiting, purging, fasting or excessive exercise.
- The consequences of anorexia nervosa are similar to those of starvation. In addition, people with eating disorders may suffer adverse consequences associated with induced vomiting, use of purgatives or emetics, and overloading of the stomach during binges.
- Eating disorders are predominantly found in adolescent girls and young women.
- Estimates of frequency vary widely, but in young women, estimates of 0.5% for anorexia and 1% for bulimia may be reasonable; the frequency in men is less than a tenth of that in women.
- There are many theories about the causes of these conditions, but there is a widespread belief that they may be triggered in susceptible people by dieting and preoccupation with being slim.
- Eating disorders are more common in men and women for whom thinness is a career requirement.

and lack of adult female form found in anorexics are seen as the aims and rewards of anorexic behaviour. This theory has attracted much attention over the years and is probably the one theory about anorexia that non-specialists will have heard of. It would be fair to say that this theory has gone out of fashion, as has, to some extent, the Freudian view of psychology.

Some have suggested that anorexia is a symptom of a frank affective psychiatric disorder such as clinical depression. This is supported, to some extent, by claims of the successful treatment of anorexics with antidepressant drugs. One major problem of trying to investigate the psychological and physiological characteristics that bring about the disease is that, by the time the disease is recognized, it will be impossible to decide which abnormalities are due to starvation and which are causes of the self-starvation. A detailed discussion of either the causes or treatment of eating disorders is beyond the scope of this book – the aim of this brief discussion has been to increase the awareness of non-specialists and to help them to recognize potential sufferers. The diseases require specialist management; the non-specialist's role can only be to be able to recognize likely sufferers, appreciate that this is a 'real' and potentially life-threatening illness, and direct sufferers towards specialist help.

The discussion may also give those in contact with anorexics some insight into the psychological consequences of anorexia/starvation and thus help them to deal more sympathetically with difficult and incomprehensible behaviour. A detailed and well-referenced account of eating disorders may be found in Gilbert (1986) and a shorter, more recent, summary may be found in Huse and Lucas (1999).

CANCER CACHEXIA

Severe emaciation and malnutrition are frequently associated with terminal malignant disease, and starvation may, in many cases, be the immediate cause of death. This is termed **cancer cachexia**. There may be

several readily identifiable causes for the loss of weight seen in these cancer sufferers, such as:

- when malignancy affects the alimentary tract
- where the disease makes eating difficult or painful
- when appetite is depressed in patients who are in great pain or extremely distressed
- the anti-cancer therapies may themselves induce nausea and anorexia.

There are, nonetheless, occasions when patients with malignant disease may lose weight initially for no clearly apparent reason. Sometimes weight loss may be the symptom that triggers the investigation leading to the diagnosis of malignancy. Hypermetabolism (increased energy expenditure) seems to be partly responsible for such weight loss and this may be a reflection of the metabolic activity associated with rapidly dividing malignant cells. There is additionally thought to be an anorexic effect of tumour growth. The metabolic disturbance or metabolic by-products of tumour growth have a depressing effect upon appetite. In simple starvation, around three-quarters of weight loss is adipose tissue, but in cancer cachexia a much higher proportion of the weight loss is lean tissue. This accelerated loss of lean tissue hastens death in the cachectic patient. Increasing food intake does not always reverse the cachectic process, and some studies have found that even intravenous feeding fails to stabilize weight in the long term. In the short term, intravenous feeding does seem to increase body fat, but does not reverse the loss of muscle and lean tissue.

Cytokines are proteins that are secreted as part of the immune response and several cytokines are known to be produced as a response to tumour growth, e.g. **tumour necrosis factor**. These cytokines may be partly responsible for the anorexia and metabolic changes seen in malignant disease. All of the cytokines produced in cancer cause anorexia and weight loss. Once cancer is suspected or confirmed, the psychological reactions to the disease, the effects of any treatment, the pain and discomfort caused by the disease or treatment as well as the direct pathological effects of the disease will all interact to reduce appetite and hasten weight loss. A short, referenced review of cancer cachexia may be found in Tisdale (1997).

- Large weight losses due to both reduced appetite and hypermetabolism are a usual feature of malignant disease – cancer cachexia.
- In cancer cachexia, there is greater loss of lean tissue than in simple starvation.
- Many factors contribute to the weight loss in cancer, but cytokines, such as tumour necrosis factor, produced as part of the immune response depress appetite and play a part in this response to malignancy.
- Neither raised food intake nor intravenous feeding seems able to correct the excess loss of lean tissue seen in cancer, even if body weight is increased or maintained.

Energy balance and its regulation

The concept of energy balance	164	External influences that affect food intake	169
Is there physiological regulation of energy balance?	166	Control of energy intake – early work with experimental animals	169
Set point theory	166		
Is energy expenditure regulated?	167	Control of food intake – the discovery of leptin	173

THE CONCEPT OF ENERGY BALANCE

When we digest and metabolize food, the oxidation of fats, carbohydrates, protein and alcohol releases chemical energy that can be used for body functions. The first law of thermodynamics states that 'energy can neither be created nor destroyed but only changed from one form to another'. This means that all of the energy released during the oxidation of the macronutrients in food (the metabolizable energy) must be accounted for. It must be used to keep the body's internal systems functioning (this is ultimately lost from the body as heat energy), to do external work (e.g. move something), or be stored within the body (e.g. as fat in adipose tissue).

For adults, if their energy intake (the metabolizable energy in food) exactly equals their energy expenditure (as internal and external work), they are in **energy balance** and their total body energy content (and thus their fat stores) will remain constant (Figure 7.1).

The 'energy in' is the sum of the metabolizable energy yields of all of the food and drink consumed; it will normally only be influenced by the nature and amount of food and drink consumed. The 'energy out' can be directly measured by determining heat output in a calorimeter or, as is more usual, by predicting the heat output from measurements of oxygen consumption and carbon dioxide evolution (see Chapter 3). This output figure will be influenced by

a number of factors (see list below) that raise heat output above the resting level, i.e. factors that have a thermogenic (heat-producing) effect.

- *The thermogenic effect of exercise:* metabolic rate rises as exercise load is increased. The sweating and rise in skin temperature induced by exertion are obvious outward signs of this thermogenesis. Some of this extra energy output may also be used to do external work.
- *Thermoregulatory thermogenesis:* the increase in metabolic rate, and therefore heat output, associated with shivering or non-shivering mechanisms used for heat generation in a cold environment. Energy expenditure also increases in a very hot environment.
- *Drug-induced thermogenesis:* the increase in heat output brought about by certain drugs (e.g. caffeine) that have a stimulating effect upon metabolic rate.
- *The thermic effect of feeding* or postprandial thermogenesis: the increase in metabolic rate and heat output that follows feeding and is the energy expended in digesting, absorbing and assimilating the food (e.g. in the synthesis of fat and glycogen).
- *Adaptive thermogenesis:* an increase in metabolic rate as a consequence of overfeeding that is claimed to help in the maintenance of energy balance (see later in the chapter).

For a healthy adult whose weight is stable, over a period of time, the energy intake and output

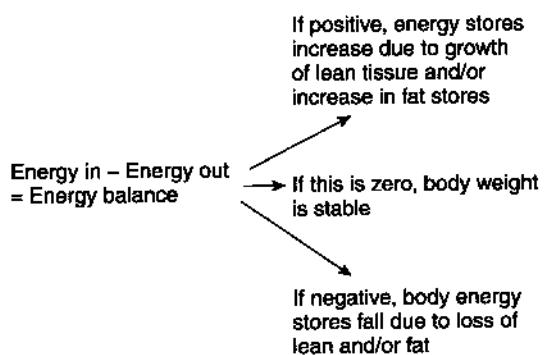


Figure 7.1 The energy balance equation.

must be matched. There is a zero balance, i.e. the energy content of the body remains constant. There is very wide variation in the energy intakes that different individuals require to maintain balance (see under 'Energy requirements' in Chapter 6).

If energy intake exceeds output, that individual is said to be in **positive energy balance** (see Balance in Glossary). Under these circumstances, body energy content must be increasing because 'energy cannot be destroyed'. The surplus energy is converted from chemical energy of food to chemical energy of body tissues. An increase in body energy content means either an increase in lean body mass (growth) or an increase in body fat deposition, or both. Children,

pregnant women and those regaining weight after a period of illness or starvation would all properly be expected to show a positive balance. For most of the rest of the population, positive balance means increasing fat stores which, if unchecked, will ultimately lead to obesity.

If energy output exceeds intake, that individual is said to be in **negative energy balance** (see Balance in Glossary) and, because 'energy cannot be created', the shortfall must be made up from body energy stores. Individuals in negative energy balance must be losing body energy and this will almost certainly be reflected in weight loss. It is possible that body energy content can fall without weight loss if fat loss is compensated for by increases in lean tissue and/or water content, e.g. in starvation there may be excess accumulation of tissue fluid – **oedema**. Loss of body energy may be due to loss of lean tissue (protein) or loss of fat or, as is most likely, both. People who are starving or successfully dieting will be in negative energy balance, as will those in the examples below:

- many sick or seriously injured people
- those losing large amounts of nutrients, e.g. uncontrolled diabetics
- those whose expenditure is particularly high, such as people with hyperactive thyroid glands or those undertaking intense physical activity for prolonged periods.

- Energy can neither be created nor destroyed, only changed from one form to another.
- If energy intake exactly equals expenditure, a person is in energy balance and his or her weight should remain stable.
- Energy intake is the metabolizable energy content of food and drink.
- Energy expenditure is comprised of basal metabolic rate (BMR) plus extra energy expended as a result of exercise, thermoregulation, drug action, feeding etc.
- If energy intake exceeds expenditure (positive energy balance), there must be either growth of lean tissue or extra fat accumulation.
- Growing children, pregnant women and those recuperating after illness or starvation would all be expected to show a positive energy balance.
- If expenditure exceeds intake (negative energy balance), there must be a corresponding decrease in body energy stores.
- Negative energy balance occurs in starvation, serious illness, and in those undertaking intense physical activity. Reducing diets are intended to achieve a negative energy balance.

IS THERE PHYSIOLOGICAL REGULATION OF ENERGY BALANCE?

Many adults maintain their body weight within a very narrow range over periods of years or even decades. Let us consider a normal 65-kg man who consistently takes in 420 kJ (100 kcal) per day more than he uses. This small surplus represents only 3–4% of his estimated requirements and would be equivalent in food terms to not much more than the generous layer of butter or margarine on a slice of bread. If one assumes that 1 kg of adipose tissue yields about 30 MJ (7000 kcal), then, if this surplus were maintained for a year, it would be energetically equivalent to 5 kg of adipose tissue and in 10 years to 50 kg of adipose tissue! If we consider the opposite scenario, if the man had a similar constant daily deficit, then he would lose adipose tissue at this rate and his fat stores would be totally exhausted in 1–2 years.

Such calculations imply that, unless there is extraordinarily precise long-term matching of energy intake to expenditure, over a period of decades, massive changes in fat stores will occur. Of course, in reality these changes would tend to be self-limiting. For the man taking in the small daily surplus, his energy requirements would tend to rise in response to the constant surplus intake, for reasons that are listed below.

- Even adipose tissue uses some energy, so as he gets bigger, it costs more energy to maintain his bigger body – BMR increases with size.
- The energy costs of any activity are increased as he gets bigger.
- It costs energy to convert surplus food energy into stored fat, especially if the surplus is in the form of carbohydrate or protein.

These changes would tend to bring the man back into balance after a modest weight gain unless he increased intake or curtailed his activity to maintain the imbalance. We saw in the previous chapter a corresponding list of changes that occur when there is an energy deficit – changes which tend to reduce energy requirements and weight loss.

SET POINT THEORY

According to the '**set point theory**', body weight control mechanisms operate to maintain a fixed level of body fat stores. If volunteers are weighed accurately every day for a period of months, one finds that, even though their weight may be unchanged over the period, nevertheless their day-to-day weight does tend to fluctuate and seems to oscillate around a set point. This is analogous to the oscillations in temperature that occur in a room where the temperature is controlled by a thermostat. During periods of illness or food restriction, weight decreases. During times of feasting, weight tends to increase. However, when normal conditions are restored, the subject's weight tends to return to the set point. Similar observations have been made in animals. Even rats that have been made obese by damage to the brain centres that control energy balance seem to control their body weight to a set point, albeit a higher set point than in normal animals. Unfortunately (as we will see in Chapter 8), over the course of adulthood the set point of many people seems to drift upwards, so that the majority of middle-aged Europeans and Americans become overweight or obese.

- Long-term weight stability requires very accurate matching of intake and expenditure.
- Even tiny sustained imbalances can lead to massive changes in body fat stores over years or decades.
- This implies that there is regulation of energy balance.
- Weight losses or gains do tend to be self-limiting, because overconsumption increases expenditure and underconsumption decreases expenditure.

- Physiological control mechanisms may operate to maintain a constant level of body fat stores – the set point theory.
- Body weight appears to oscillate around this set point, although in the long term there is often an upward drift in body fat stores with age.

IS ENERGY EXPENDITURE REGULATED?

Regulation of energy balance could be achieved by regulation of energy intake or expenditure or by a combination of the two. The sensations of hunger and satiation are universally experienced manifestations of some energy intake-regulating mechanism. Traditionally, therefore, studies on the regulation of energy balance have focused upon the mechanisms regulating food intake, and these are discussed at length later in this chapter. The regulation of energy expenditure as a mechanism for controlling energy balance has been a more controversial topic. We have already seen that overfeeding increases energy expenditure and that energy restriction reduces expenditure and these changes tend to restore balance. There is little doubt that an increase in 'metabolic efficiency' occurs during fasting and food restriction: BMR drops by up to 15%, even when one allows for any decrease in lean body mass (Prentice *et al.*, 1991). The point of controversy is whether any mechanism exists whose primary purpose is to compensate for overeating by 'burning off' surplus calories.

In 1902, the German scientist Neumann coined the term **luxoskonsumption** to describe an adaptive increase in energy expenditure in response to overfeeding. The term **adaptive thermogenesis** is often used today. Essentially, it is envisaged that some energy-wasting metabolic process is activated when people overeat and that this 'burns off' the surplus calories in order to restore energy balance or to reduce the imbalance.

Rats can be persuaded to overeat and become obese by the provision of a variety of varied, palatable and energy-dense human foods (**cafeteria feeding**). This mode of feeding in caged animals is designed to mimic the affluent human lifestyle. Rothwell and Stock (1979) reported that some rats fed in this way stored less of the surplus food energy than they predicted and appeared to burn off some of this extra

energy. Sims *et al.* (1973) did similar experiments with young men who were persuaded to consume an extra 3000 kcal (13 MJ) per day for a period of 8 months. At the end of this period, only about 15% of the surplus energy that had been eaten could be accounted for in the increased fat stores. This implied that some adaptive thermogenesis had occurred. Many subsequent shorter and less extreme overeating experiments have failed to find evidence of significant adaptive thermogenesis in human subjects. If there is any significant capacity for adaptive thermogenesis, one would expect that, under conditions of abundant food supply, people might eat more than their bare minimum requirements and burn off some energy. The apparent increase in metabolic efficiency seen during fasting might represent a switching off of this adaptive thermogenesis and there might be limited further capacity to increase thermogenesis under conditions of experimental overfeeding.

Rothwell and Stock (1979) proposed that, in small mammals, a tissue known as **brown adipose tissue (BAT)** or **brown fat** is the site of adaptive thermogenesis. Brown fat is the major site of cold-induced heat production (non-shivering thermogenesis) in small mammals and is particularly prominent in hibernators. The sympathetic stimulation of brown adipose tissue results in an uncoupling of oxidative phosphorylation, i.e. oxidation accelerates but does not result in ATP production. The chemical energy produced during oxidation is released as heat and this warms the body during cold exposure. This mechanism is well established for thermoregulatory thermogenesis in small mammals (and human babies), and Rothwell and Stock (1979) suggested that this also occurs during overfeeding. The heat generated during overfeeding is then dissipated from the relatively large surface of the small animal. This theory was given added credence by evidence of abnormal thermoregulation in several animal models with inherited obesity. The most widely used of

these models is the ***ob/ob mouse***, which is homozygous for a mutation at a gene locus referred to as *obese*, or *ob*. These mice have low body temperature, reduced heat production and die during cold exposure. They were said to have a 'thermogenic defect' which made them prone to a metabolic obesity as they would have little capacity for adaptive thermogenesis. A similar thermogenic defect might also predispose people to obesity.

It would be unwise to be too ready to extrapolate these findings in laboratory rodents to people. Adult human beings have very limited capacity for non-shivering thermogenesis, and brown adipose tissue has traditionally been regarded as vestigial in adult humans. People keep warm largely by heat conservation supplemented by shivering rather than non-shivering thermogenesis. A large animal with a relatively small surface area to volume ratio is less able to dissipate large amounts of surplus heat generation; this might also make adaptive thermogenesis a less advantageous strategy for the maintenance of energy balance in people. Accurate long-term measurements of metabolic rate in obese people provide no evidence for any thermogenic defect in the obese and, indeed, the BMRs and total energy expenditure of obese people are increased because of their higher body mass (Prentice *et al.*, 1989). Drugs that are highly specific activators of the β -3 adrenoreceptor on brown fat (β -3 agonists) cause large increases in metabolic rate in rodents but have very little effect in people.

Even in the *ob/ob* mouse, the abnormal thermoregulation probably represents an adaptation rather than a defect (Webb, 1992b). If the genetic defect of these mice makes them unable to detect their large fat stores, they would respond as if being continually starved (see discussion of **leptin** later in the chapter). One of the strategies that mice employ to conserve energy during fasting is to reduce their heat output and lower their body temperature. Fasting mice can enter a state of torpor in which they allow their body temperature to fall by as much as 15 °C for several hours at a time (Webb *et al.*, 1982). The low body temperature of *ob/ob* mice could represent a permanent semi-torpid state as an adaptation to perceived starvation, and their poor cold tolerance might represent a form of disuse atrophy of their brown fat. Torpor is not a usual human response to starvation. Children have now been identified who have an analogous gene defect

to that of the *ob/ob* mouse, but these very obese children have normal body temperature and thermoregulation (Montague *et al.*, 1997). (See Webb, 1990 and 1992b, for critical discussion of the applicability to people of these studies on small-animal models of obesity.)

The publication of this brown fat theory by Rothwell and Stock in 1979 generated a flurry of interest and excitement in the early 1980s. It raised the prospect that adaptive thermogenesis could represent a major component of human energy expenditure and that differences in thermogenic capacity might explain the apparent variation in the susceptibility of people to obesity. It also raised the prospect of thermogenic drugs that would specifically stimulate adaptive thermogenesis. These might offer an attractive alternative to appetite suppressants for the pharmacological treatment of obesity. β -3 adrenoreceptor agonists that specifically stimulate brown fat thermogenesis have shown promising results in rats but not in people. This is probably because of the low capacity for brown fat thermogenesis in people. It would be fair to say that the initial enthusiasm for this theory has cooled considerably over the last 20 years. There is little evidence that adaptive thermogenesis makes a major contribution to human energy balance control.

Some more recent work has further undermined the theory that adaptive thermogenesis is a major factor in controlling body weight. Enerback *et al.* (1997) have produced transgenic mice that lack the key protein required for the uncoupling of oxidation and phosphorylation in brown fat (UCP1). This protein is the key initiator of heat generation in brown fat. These mice are cold sensitive but they are not obese and their ability to regulate their body weight is normal. This work seemed to be a decisive blow against the brown fat theory of obesity. There are, however, now known to be other uncoupling proteins called UCP2 and UCP3. UCP2 is expressed not only in both white fat and brown fat but also in other tissues (Fleury *et al.*, 1997); UCP3 is found mainly in skeletal muscle. (See Hirsch, 1997, for a short, referenced summary of these developments.) The role of these new uncoupling proteins (UCP2 and UCP3) in the control of human metabolic rate is as yet unclear, but Schrauwen *et al.* (1999) suggest that they may be important. Uncoupled oxidative phosphorylation (proton leakage) may account for 20% of the resting metabolic rate in people.

- It seems self-evident that food intake is regulated, but do animals and people have the capacity to burn off surplus energy and thus limit weight gain as an adaptive response to overfeeding (adaptive thermogenesis)?
- Rothwell and Stock (1979) proposed that overfeeding increases sympathetic stimulation to brown fat, which uncouples oxidative phosphorylation in this tissue and so increases heat generation.
- This mechanism is well established as a means of heat generation in small mammals and human babies during cold exposure (non-shivering thermogenesis).
- Rothwell and Stock also proposed that genetically obese mice (*ob/ob*) and humans might become fat because of some thermogenic defect that impaired their capacity for adaptive thermogenesis.
- Adaptive thermogenesis is not now thought to be a major component of human energy expenditure.
- There is no evidence of any thermogenic defect in most obese people and, indeed, the BMR and total energy expenditure of obese people tend to be higher than those of lean people.
- Transgenic mice that lack a key protein required for brown fat thermogenesis (UCP1) are hypothermic but not obese.
- Even the well-documented hypothermia of some obese mice probably represents an inappropriate energy-conserving response in animals whose brains falsely perceive that their fat stores are depleted.
- The discovery of two new uncoupler proteins (UCP2 and UCP3) that are expressed in tissues outside brown fat may yet mean that the regulated uncoupling of oxidative phosphorylation may play an important role in controlling human metabolic rate.
- Up to 20% of resting metabolic rate may be due to uncoupled oxidative phosphorylation (proton leakage).

EXTERNAL INFLUENCES THAT AFFECT FOOD INTAKE

The rest of this chapter considers the physiological regulation of energy intake. However, it is not only 'internal' physiological factors that influence when, what, where and how much we eat. Eating is also a conscious, voluntary act that is affected by a host of 'external' influences:

- The palatability of food and the pleasure derived from eating; we do not have to be hungry to eat and indeed we may be persuaded to eat some delicious item even when feeling distinctly full.
- Habit: for example always eating a particular snack when reading, driving or watching a favourite television programme.
- Our psychological state: eating to relieve psychological states other than hunger, e.g. to relieve anxiety, boredom or unhappiness.

- Eating for social rather than physiological reasons: for example to participate in a social 'feast' or to please someone who has offered us food.

It is perhaps only in newborn babies that eating is regulated largely by internal, physiological hunger-control mechanisms, uncluttered by these external influences. We may, on occasion, ignore or override our physiological mechanisms that control food intake and this may be one reason why we are so prone to obesity.

CONTROL OF ENERGY INTAKE – EARLY WORK WITH EXPERIMENTAL ANIMALS

Over the last 100 years, there have been several reports in the medical literature of patients becoming very obese after damage to their hypothalamus caused by trauma or tumours. This pointed to a

- Even though physiological control mechanisms affect feeding, eating is a voluntary activity that is also influenced by a host of social, cultural and psychological factors.
- These external influences on feeding may reduce the effectiveness of physiological control mechanisms.

role for the hypothalamus in weight control and it is now well established that the hypothalamus is involved in other physiological regulatory systems, such as:

- body temperature control
- control of water balance
- control of the female reproductive cycle.

Around 1940, the development of the stereotaxic apparatus allowed experimenters to make discrete and accurately located lesions deep within the brains of experimental animals. Over the next couple of decades, this lesioning technique showed that damage to certain areas of the hypothalamus could profoundly affect a rat's eating and drinking behaviour. Damage to the ventromedial region of the hypothalamus in many mammalian species, ranging from mice to monkeys, produces a period of massive overeating and weight gain, which leads to severe and permanent obesity. The weight of rats with such lesions stabilizes at a much higher level than that of unoperated rats, and their food intake returns to a level that maintains this 'new set point'. If these rats are slimmed down, they will go through a second bout of overeating and they appear to defend their new set point. It was also found that lesions in the lateral part of the hypothalamus completely abolished eating and drinking behaviour in rats and they died unless tube-fed.

As a result of such observations, a simple theory of food-intake regulation was proposed, the **dual-centre hypothesis** (Figure 7.2). A spontaneously active feeding centre that initiates food seeking and eating behaviour was envisaged as being located in the lateral hypothalamus. Its destruction by lateral hypothalamic lesions results in a cessation of eating behaviour. This feeding centre was said to be periodically inhibited by a satiety centre located in the ventromedial region of the hypothalamus. This satiety centre would become active in response to cer-

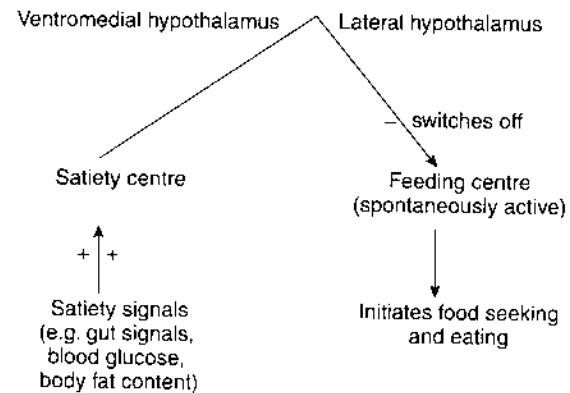


Figure 7.2 The dual-centre hypothesis.

tain **satiety signals**, which strengthen after feeding; it would then inhibit the feeding centre and produce satiation. As time after feeding elapses, so one would envisage these satiety signals dying away, the satiety centre would become quiescent, allowing the feeding centre to become active again and producing hunger. Destruction of this satiety centre would reduce satiation and result in excessive food consumption and obesity. The hypothalamus is thus envisaged as acting like a meter or '*appetstat*' that adjusts the feeding drive in response to a variety of physiological signals that reflect the short-term and long-term energy status of the body – energy intake is thus matched to output and energy balance is maintained.

The dual-centre hypothesis has had enormous influence on thinking about the control of energy balance and body weight. It was prominent in many textbooks published in the 1960s and 1970s and it may still be found in some modern texts. It is now regarded as a rather simplistic model of a complex process. Since these early studies, much detailed work has been done attempting to identify hypothalamic and extra-hypothalamic brain areas that are involved in the control of feeding and body weight.

Detailed discussion of the anatomical locations and interconnections of neural centres controlling energy balance is beyond the scope of this book, but several hypothalamic areas are known to play a role. It has been confirmed that the ventromedial region of the hypothalamus (the satiety area of the dual-centre hypothesis) plays a pivotal role in the control of feeding and perhaps energy expenditure. Animals with lateral hypothalamic lesions were said to recover spontaneous feeding behaviour if kept alive by tube-feeding. This led some workers to question the essential role of the lateral 'feeding centre' in causing spontaneous feeding (see Kennedy, 1966). I will use the term *appetstat* (cf. *thermostat*) as a term to collectively describe the centres in the hypothalamus that control energy balance. References to the early experimental work on the hypothalamic mechanisms regulating food intake have been excluded. Interested readers will find detailed and well-referenced accounts of this work in editions of many texts published during the 1960s and 1970s or by reference to Mayer (1956, 1968), Kennedy (1966) or Brobeck (1974), three of the most prolific and influential researchers in this field.

Despite nowadays being regarded as simplistic, the dual-centre hypothesis does have the three basic elements that any system for controlling food intake must have (Figure 7.3):

1. a series of physiological inputs to the appetstat that provide information about the feeding status and the level of body energy stores – so-called satiety signals
2. areas within the brain that integrate information about feeding status and level of body energy stores – the appetstat

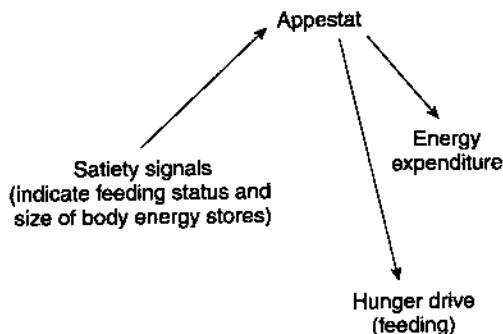


Figure 7.3 The three major elements of the physiological control of energy balance.

3. a hunger drive that can be modified by the appetstat in response to the information received from the satiety inputs (also mechanisms for controlling energy expenditure).

Various satiety signals have been proposed and investigated over the years. These fall into the three major categories below.

1. Signals emanating from the alimentary tract and transmitted via sensory nerves or hormones released from the gut. These would provide information about the amount and the nature of food in the gut.
2. Blood substrate levels. After eating, the blood concentration of substrates such as glucose and amino acids rises as they are absorbed from the gut. These concentrations fall in the post-absorptive state. If we eat three meals per day and they each take around 4 hours to digest and absorb, for most of our waking hours we are absorbing food. The notion that blood glucose concentration or the rate of glucose utilization is a major satiety signal has historically been very popular – the **glucostat theory**.
3. Signals that emanate from adipose tissue and indicate the level of body fat stores – the **lipostat theory**.

Gut-fill cues

We all experience the feeling of stomach fullness that occurs after a meal and the feelings of stomach emptiness and hunger contractions that occur after fasting. Feeding normally finishes before there have been detectable changes in blood substrate levels. These observations would suggest that signals from the alimentary tract play some part in signalling satiation and have some role in short-term appetite regulation. This could be mediated either through sensory nerves in the gut directly affecting the appetstat or perhaps through the effect of hormones released from the gut by the presence of food.

One can greatly alter the energy density of the diets of laboratory rats by the use of an inert bulking agent (e.g. methylcellulose) or by the addition of energy-rich fat. Rats fed diluted diets are able to maintain their body weights, and compensate for the dilution by increasing the volume of food they

eat. However, long-term feeding with diets of high energy density causes obesity in rats, which suggests that overconcentration of the diet may reduce the effectiveness of feeding control mechanisms. It is widely believed that reducing the energy density of human diets will help to prevent excessive weight, which implies widespread acceptance of the belief that dietary bulk has some modulating effect upon appetite. One could, for example, envisage alimentary signals playing an important role in signalling satiety at the end of a meal but, when food is very dilute, being overridden or damped by other physiological signals indicating the need for energy intake. The mechanisms regulating energy balance may be more effective in preventing underconsumption rather than overconsumption (see Chapter 8).

The hormone **cholecystokinin (CCK)** is released from the intestine when food is present and it induces satiety and reduces feeding. It is said to act by activating sensory branches of the vagus nerve in the gut and these relay the information to the appesstat (Smith *et al.*, 1981).

The glucostat theory

In this theory, it is envisaged that there are glucose receptors or sensors in the appesstat, and perhaps elsewhere, that respond to blood glucose concentration (Mayer, 1956). These glucose receptors induce satiety when blood glucose concentration is high and hunger when it is low. The occurrence of both high blood glucose and hunger in uncontrolled diabetes was explained by the suggestion that the glucoreceptor cells needed insulin for glucose entry in the same way as, say, resting muscle cells do. In the absence of insulin (i.e. in diabetes), despite a high blood glucose concentration, the glucoreceptor cells are deprived of glucose, resulting in hunger. As the brain was regarded as being exclusively dependent upon a supply of glucose in the blood, hypothalamic glucoreceptors sensitive to fluctuations in blood glucose concentration seemed a logical homeostatic device.

This theory was given an important boost in the 1950s by work that showed that a glucose derivative, **gold thioglucose**, produced both obesity and lesions in the ventromedial region of the hypothalamus of mice. No other gold thio compounds produced this effect; the glucose moiety

appeared to be essential. This supported the proposal that there are glucoreceptors in this region that bind the gold thioglucose and this leads to accumulation of neurotoxic concentrations of gold in their vicinity (reviewed by Debons *et al.*, 1977). Later versions of this theory suggested that a high rate of glucose metabolism might produce satiation rather than high blood concentration *per se*. Satiation correlates very well with the difference between arterial and venous blood glucose concentration, which is a measure of the rate of glucose utilization.

When the rate of glucose metabolism is high, fat use is low. In healthy people, this high glucose-low fat usage occurs after eating carbohydrate and produces satiation. When glucose metabolism is low, that of fat is high and this results in hunger. High-fat-low-glucose utilization occurs:

- during fasting
- in the absence of insulin (uncontrolled diabetes)
- if meals with little or no carbohydrate are eaten.

Insulin, when released after a meal, causes glucose to be taken up and metabolized by the cells and this leads to satiation. The use of fat in metabolism is associated with hunger. Carbohydrate intake increases the use of carbohydrate and reduces the use of fat, but fat intake has little effect on either fat or carbohydrate metabolism (Schutz *et al.*, 1989). As fat is absorbed, much of it is shunted directly into our fat stores. This may help explain why high-fat diets may be less satiating than high-carbohydrate diets and why high-fat diets seem to predispose to weight gain and obesity (see Chapter 8).

The evidence that some glucostat mechanism plays an important role in the short-term control of appetite is very persuasive.

The lipostat theory

This theory was used to explain how body energy stores (i.e. fat) could be regulated in the long term by regulation of apparently short-term phenomena like hunger and satiety. It is difficult to see how short-term satiety signals such as gut-fill, blood glucose concentration or even the rate of glucose utilization could produce long-term weight stability, i.e. over a period of months, years or decades. It was first proposed by Gordon Kennedy in the 1950s (see

Kennedy, 1966). It is envisaged that some appetite-regulating chemical (a satiety factor or satiety hormone) is released from adipose tissue so that its concentration in blood indicates to the appestat the total amount of stored fat. This would allow the total amount of body fat to have a modulating effect upon appetite and would offer a relatively simple mechanism for the long-term control of energy balance. This lipostat theory would fit well with the notion of a set point for body weight, discussed earlier in the chapter.

Some very persuasive but indirect evidence for the existence of such a satiety hormone (now called leptin) came from the work of Coleman (1978), using two types of genetically obese mice. *Obese* (*ob/ob*) and *diabetes* (*db/db*) are two forms of

Table 7.1 The results of parabiotic pairings of lean and obese rodents

Pairing - A normal animal paired with one that has had its ventromedial hypothalamus (part of the appestat) destroyed

Result The animal with the damaged hypothalamus overeats and gains weight as expected; the normal animal eats less and loses weight

Explanation - The animal with its hypothalamus damaged gets fat because the area where leptin acts in the brain has been destroyed and so it cannot respond to leptin. As this lesioned animal gets fatter, so its adipose tissue produces large amounts of leptin. This extra leptin enters the blood of the normal animal and so this animal's appestat responds by reducing food intake

Pairing A normal and an *ob/ob* mouse

Result The *ob/ob* mouse eats less and loses weight

Explanation - The *ob/ob* mouse is unable to produce its own leptin but can respond to leptin produced by its lean partner, and so eats less and loses weight

Pairing – A normal and a *db/db* mouse

Result – The normal mouse stops eating and starves

Explanation - The *db/db* mouse cannot respond to leptin and so gets fat and produces excess leptin. This leptin enters the blood of the normal mouse and acts upon this animal's appestat to suppress food intake

Pairing – An *ob/ob* and a *db/db* mouse

Result – The *ob/ob* mouse eats less and loses weight

Explanation – The *ob/ob* mouse is responding to the excess leptin produced by its *db/db* partner

genetic obesity in mice. Both *ob/ob* and *db/db* mice inherit a very severe obesity associated with diabetes. On a similar genetic background, the two syndromes are practically identical and yet are caused by mutations at different gene loci. Coleman proposed that in the *ob/ob* mouse the genetic defect is in the leptin (satiety hormone) gene, leaving them unable to produce leptin and making their appestat unable to detect any fat stores. As a consequence of this, they overeat and get very fat. He suggested that in the *db/db* mouse, the abnormal protein is one that allows the appestat to respond to leptin (the leptin receptor). *db/db* mice would thus produce lots of leptin because they have lots of fat in their adipose tissue. However, without any leptin receptor, their appestat cannot detect the leptin and so, once again, they cannot sense any fat stores and so they overeat. At this time, leptin had not been discovered and was a purely hypothetical hormone. These suggestions were the result of an ingenious series of parabiosis experiments; these are experiments in which pairs of animals are surgically linked together so that hormones and other substances can pass between their blood systems. The results obtained from four of these parabiotic pairings are explained in Table 7.1.

CONTROL OF FOOD INTAKE – THE DISCOVERY OF LEPTIN

It is a bit of a cliché that scientific books may become out of date before they are published. Nonetheless, a paper was published just after the manuscript of the first edition of this book was finished which has radically changed scientific thinking about the physiological mechanisms that control energy balance. In 1994, Zhang *et al.* identified the *obese* (*ob*) gene and the protein that is produced by it in healthy, lean mice. The *obese* gene produces a secretory protein of 167 amino acids that seemed to be produced exclusively in adipose tissue and has been named leptin (note that later studies have shown that it is also produced in the placenta and stomach). It seems to have all the characteristics of the hypothetical satiety factor from adipose tissue that was proposed in the lipostat theory of weight control. In normal mice, leptin production falls during starvation and increases in environmentally induced obesity. Leptin production

- The hypothalamus plays a major role in the control of feeding, and experimental lesions in the hypothalamus profoundly affect feeding behaviour.
- The hypothalamic 'appetite' receives input of satiety signals that indicate the feeding state and level of body fat stores.
- The hypothalamus alters the feeding drive and rate of energy expenditure in response to these satiety inputs.
- Satiety signals relay information to the appetite about the amount and nature of food in the gut, the level or rate of usage of blood substrates and the amount of fat stored in adipose tissue.
- Sensory nerves in the gut and hormones released from the gut indicate the nature and amount of gut contents.
- In the glucostat theory, it is envisaged that information from glucose sensors in the hypothalamus and elsewhere has a major role in the short-term control of feeding.
- In the lipostat theory, Kennedy proposed that some satiety factor (leptin) is released from adipose tissue in proportion to the amount of stored fat. This would enable the appetite to regulate feeding and energy expenditure to maintain a constant level of these stores.
- Indirect evidence suggested that obese mice with mutations at the *obese* locus (*ob/ob*) do not produce this satiety factor, whereas mice with mutations at the *diabetes* locus (*db/db*) do not respond to it.

seems to be proportional to the mass of adipose tissue. In genetically obese (*ob/ob*) mice a mutation in the *ob* gene results in the production of a shorter, inactive protein that is not secreted. This leaves the appetite of *ob/ob* mice with no input from adipose tissue. The appetite perceives this as absence of fat stores and so initiates appropriate starvation responses.

Leptin administration to *ob/ob* mice causes weight loss, and normalization of body composition and corrects all of the other metabolic and hormonal abnormalities of these mice. Leptin also causes loss of fat but not lean when it is administered to normal mice. Leptin administration has no effect in the *db/db* mouse, an observation that is entirely consistent with the earlier suggestion of Coleman (1978) that the genetic defect in these mice is in a gene that codes for the leptin-sensing system (the leptin receptor). In 1995, Tartaglia *et al.* were able to identify and characterize the leptin receptor. They confirmed that the mutation in *db/db* mice leads to abnormalities in the leptin receptor. The leptin receptor can be detected in all the areas of the hypothalamus known to be involved in the control of energy balance and in the vascular region where leptin is conducted across the blood-brain barrier (the choroid plexus).

Leptin seems to play a central role in the long-term regulation of energy balance (Figure 7.4). During star-

vation and weight loss, leptin production from adipose tissue decreases. This decrease in leptin is detected by the hypothalamic appetite, which then triggers the normal responses to starvation that are seen at their most extreme in the *ob/ob* mouse, such as:

- increased food intake if food is available
- reduced activity and resting energy expenditure, including reduced body temperature and the permanent semi-torpid state of *ob/ob* mice
- reduced production of hormones in the pituitary gland (gonadotrophins) that leads to delayed puberty and infertility
- other changes in the secretion of hormones from the pituitary, which in turn regulate most of the other endocrine glands. The output of pituitary hormones is controlled by the hypothalamus.

During weight gain, as the fat content of adipose tissue increases, so increasing leptin production should act via the hypothalamus to reduce food intake and increase energy expenditure. Obesity would occur under either of the circumstances listed below.

- Complete failure of leptin production (as in *ob/ob* mice) or inadequate leptin production, i.e. a failure to produce sufficient leptin to reflect the level of fat stores to the hypothalamus.

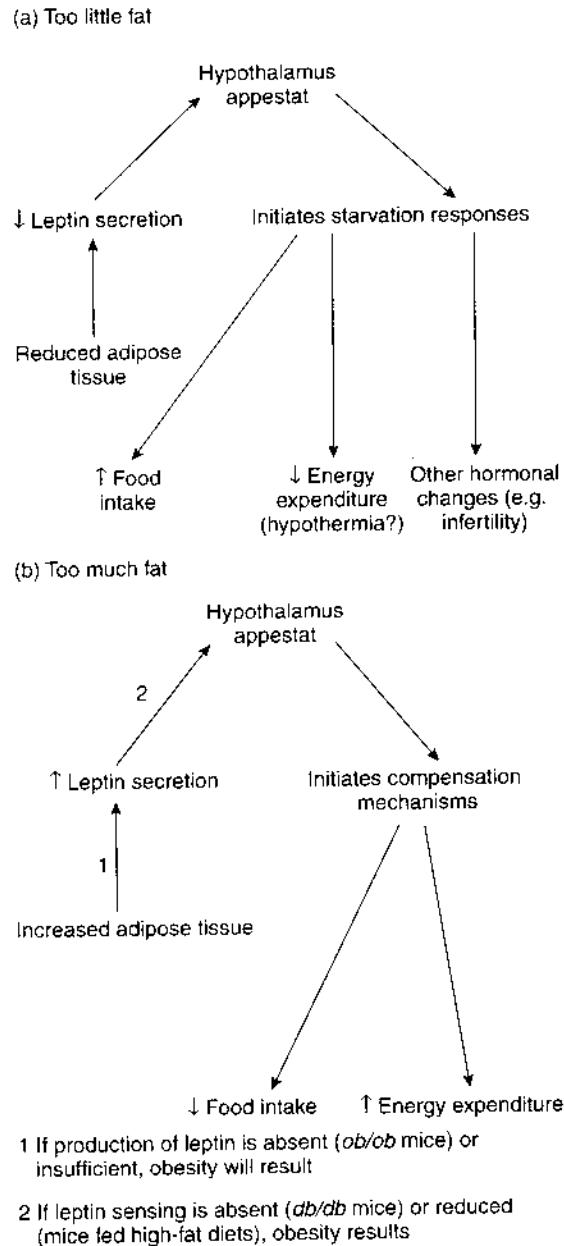


Figure 7.4 The role of leptin in responses of mice to excess or inadequate fat stores.

- A complete lack of response to leptin (as in *db/db* mice) or a reduced response to leptin.

The observation that leptin levels are high in most induced animal obesities suggests that reduced leptin sensitivity is the likely scenario in most obesity. Increasing the fatness of animals by feeding them a

high-fat diet does, indeed, lead to a decrease in their sensitivity to leptin. There is thus a complete spectrum of leptin sensitivity in mice, from the high sensitivity of *ob/ob* mice through to the complete insensitivity of *db/db* mice (Figure 7.5). Normal mice lie between these two extremes, but inducing obesity by dietary means reduces leptin sensitivity. There is a clear precedent for diet and lifestyle causing increased production and declining sensitivity to a hormone. The commonest and mildest form of diabetes begins in middle or old age and does not normally require insulin injections (type 2 **diabetes**). In type 2 diabetes there is no primary failure of the pancreas to produce insulin, as there is in the severe form of the disease that starts in childhood and requires regular insulin injections for survival (type 1 diabetes). In type 2 diabetes, an affluent Western lifestyle (inactivity, high-fat diet and excessive weight gain) triggers a progressive decrease in the sensitivity to insulin, probably as a response to high insulin secretion. In genetically susceptible people, this eventually leads to symptoms of diabetes.

Leptin appears to play a key role in long-term energy balance control, even though other factors such as blood glucose and gut-fill cues seem to control meal spacing and the short-term control of hunger and satiety. Leptin levels do not rise noticeably after a meal and it is not thought to immediately induce satiety during feeding and signal the end of a meal. However, the observation that some leptin is produced in the stomachs of rats and that it is released during feeding may herald some role for leptin in short-term hunger and satiety control. Administration of the hormone CCK also causes

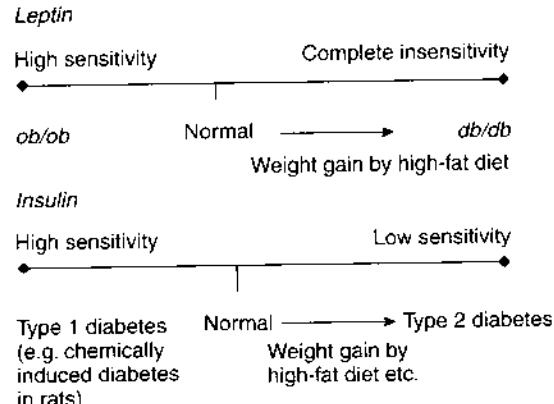


Figure 7.5 Variations in leptin and insulin sensitivity.

- In 1994, the protein produced from the *obese* gene in healthy mice and people was identified and characterized.
- The *obese* gene product is a secreted protein called leptin, which is mainly produced in adipose tissue.
- Leptin seems to be the hypothetical satiety factor from adipose tissue that was predicted in the lipostat theory of weight control.
- In general, leptin concentration in blood is proportional to the amount of stored fat.
- Obesity could result from inadequate leptin production or inadequate response to leptin as this would cause the appetstat to 'underestimate' the fat content of adipose tissue.
- Obese mice with mutations at the *obese* locus (*ob/ob*) do not produce functional leptin, and mice with mutations at the *diabetes* locus (*db/db*) do not respond to leptin because they have a defective receptor protein.
- A few very obese children have been identified with mutations at the leptin gene locus (analogous to *ob/ob* mice) and at the leptin receptor locus (analogous to *db/db* mice).
- Only a small proportion of obese people have low leptin levels; most have high leptin levels that reflect their large amounts of stored fat.
- Most human obesity is probably due to an environmentally induced reduction in sensitivity or response to leptin. By analogy, most human diabetes is due to reduced insulin sensitivity.
- In experimental animals, leptin sensitivity declines if they are made obese by environmental means, e.g. feeding a high-fat diet.

release of leptin from the stomach. CCK is released from the gut during feeding and is known to induce satiety (Bado *et al.*, 1998).

So far, this discussion has been largely confined to studies in experimental animals. What is the relevance of this work to people and, more especially, what future role can leptin play in human obesity therapy? In their original leptin paper, Zhang *et al.* (1994) identified a human version of leptin produced in human adipose tissue. Measurements of serum leptin concentrations in people with varying levels of fatness suggest that leptin concentration rises in proportion to the adipose tissue mass. This means that, in general, leptin levels are higher in obese people than in lean people. For example, in a large sample of Polynesian men and women, Zimmet *et al.* (1996) found a very high correlation ($r = 0.8$) between serum leptin concentration and body mass index (i.e. body fat content).

In 1997, two British children of Pakistani origin were found to have mutations in their leptin genes, i.e. they are the apparent human equivalents of *ob/ob* mice. These children were extremely obese: at 8 years the girl weighed 86 kg, while the boy weighed 29 kg at 2 years old (Montague *et al.*, 1997). There are tech-

nical reasons why one would expect that this mutation (analogous to the *ob* mutation seen in *ob/ob* mice) would be expected to be extremely rare in humans (Sorensen *et al.*, 1996). According to Friedman and Halaas (1998), about 5–10% of obese humans may have relatively low leptin levels. The individuals of one group who are particularly prone to obesity, the Pima Indians, tend to have low leptin levels, which may explain their extreme propensity to obesity. Three members of a French immigrant family have been identified who have a mutation in their leptin receptor gene, the human equivalent of *db/db* mice (Clement *et al.*, 1998). When humans with leptin or leptin receptor mutations reach the normal age for puberty, they show similar reproductive abnormalities to *ob/ob* and *db/db* mice: they do not go through normal pubertal changes and the females do not menstruate. Studies such as these provide confirmation that leptin plays a similar role in the regulation of energy balance in mice and humans (see Figure 7.4).

When the discovery of leptin was first announced, a US pharmaceutical company paid millions of dollars for the patent on the leptin gene. Leptin seemed to offer the prospect of a major advance in the treat-

ment of obesity, even though most obesity seems to be associated with leptin resistance rather than reduced leptin production. Leptin is a protein that needs to be administered by injection, a major drawback to its potential therapeutic usage. Early, preliminary trials with leptin in obese people suggest that it does produce a small average weight loss over a month of use. However, the individual response to leptin was variable. One might expect it to be most effective in the minority of obese people with low

leptin secretion rather than in those who are leptin resistant. In type 2 diabetes, despite the insulin resistance, if enough insulin is given, it is an effective therapy. By analogy, one might expect a similar effect of leptin in leptin resistance obesity. At present, technical problems (e.g. local reactions at the injection site) limit the amount of leptin that can be administered to patients in trials. For a fully referenced review of these leptin developments, see Friedman and Halaas (1998).

Obesity

Defining obesity	178	Inactivity as a cause of obesity	194
Prevalence of obesity	179	Prevention and treatment of obesity in populations	196
The medical and social consequences of obesity	182	Obesity treatment in individuals	199
The causes of obesity	188	More 'aggressive' treatments for obesity	202
Is fat more fattening than carbohydrate?	193		

DEFINING OBESITY

Obesity strictly refers to an excess accumulation of body fat but, as we saw in Chapter 3, body fat content is difficult to measure quickly and accurately in people. For this reason, the **body mass index (BMI)** is now very widely used to determine whether or not someone is excessively fat. (See Chapter 3 for a critical discussion of the use of BMI for this purpose.) The BMI classification system in Table 8.1 is now used internationally and the **World Health Organization (WHO)** defines clinical obesity as a BMI of over 30. This classification is a simple one because it uses:

- a small number of ranges
- whole numbers
- the same cut-off points for men and women.

This simplicity must have a cost in terms of accuracy and sensitivity, and so these ranges should be

Table 8.1 *The internationally recognized classification of people according to their body mass index (BMI)*

BMI range	Classification
<20	Underweight
20–25	Ideal range
25–30	Overweight
>30	Obese
>40	Severely obese

BMI = weight (kg) ÷ height² (m)

used as a general guide rather than absolute cut-off points. A small difference in BMI is still small even if one value is above and one below a barrier. For example, values of *slightly* less than 20 in young women are fairly common and not usually a cause for concern. In the USA, another classification system has been widely used. This had more categories, different cut-off points for men and women, and did not use round, or even whole, numbers. In this US

- Body mass index (BMI) is the standard way of classifying people according to fatness.
- $\text{BMI} = \text{weight (kg)} \div \text{Height}^2 (\text{m})$.
- $20\text{--}25 \text{ kg/m}^2$ is taken as the ideal range for BMI in adults.
- People below the ideal range are classified as underweight and those above it as overweight.
- A BMI of over 30 is taken as the definition of clinical obesity.

system, BMIs of 27.8 for men and 27.3 for women were taken as the crucial boundaries for being obese. In this book, overweight always refers to a BMI in excess of 25 and obese to a BMI of over 30, unless otherwise specified.

PREVALENCE OF OBESITY

There is little doubt that overweight and obesity are more common in the UK and USA than in previous generations and that the prevalence is still rising. The 1980s and 1990s seem to have been a time of particularly rapid increases in prevalence. Well over half of English adults are currently overweight or obese (i.e. BMI 25+) and over 16% are clinically obese (i.e. BMI 30+). Figure 8.1 shows the proportion of English adults in various BMI categories. In the USA, the situation is even worse, with over 60% of all adults having BMIs that are above 25 and in 1988–1991 a third had BMIs above the thresholds then used for obesity in the USA (27.3 for women and 27.8 for men).

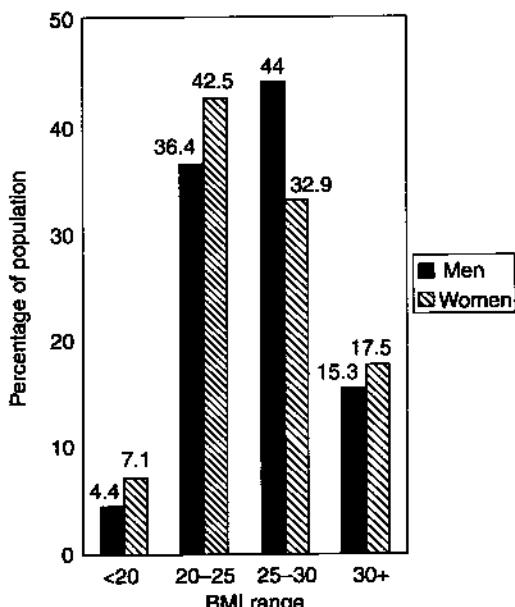


Figure 8.1 Distribution (%) of English adults aged over 16 years according to body mass index (BMI) category. Data source: Prescott-Clarke and Primatesta (1998).

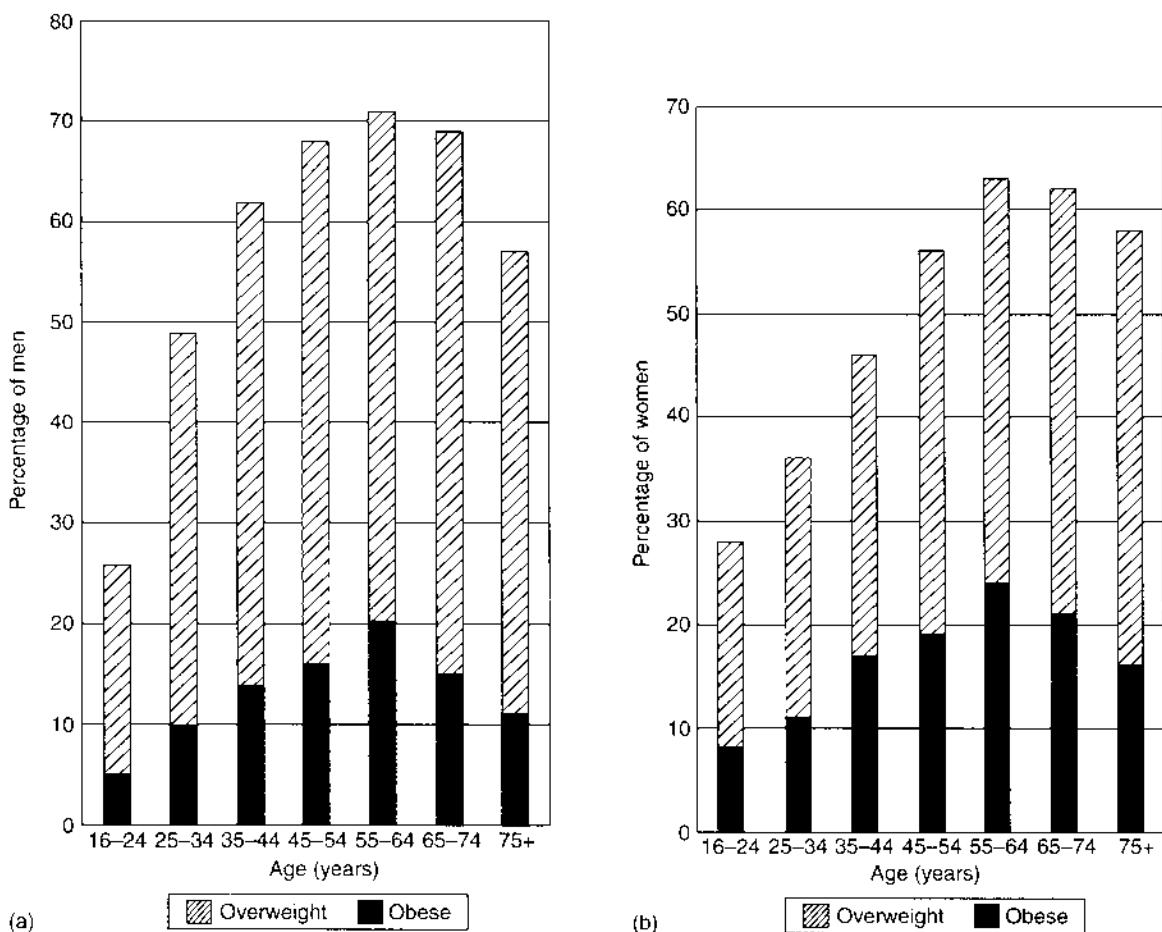
Even the high frequency of obesity indicated in Figure 8.1 somewhat understates the problem in middle-aged people. Figure 8.2 shows how the proportion of people who are overweight or obese changes with age in English adults. The proportion of people who are overweight and obese rises steeply throughout early adulthood and middle age; only in the over-60s do the proportions start to fall back slightly. In the 45–64-year-old age band, more than two-thirds are either overweight or obese. Essentially, similar age-related changes in prevalence are seen in the USA (Kuczmarski *et al.*, 1994).

Figure 8.3 shows recent changes in the estimated prevalence of obesity in England. Obesity prevalence has more than doubled since 1980. In 1992, the British government set a target of reducing clinical obesity prevalence in England to no more than 6% of men and 8% of women by the year 2005 (Department of Health, 1992). At first sight, these seem to be totally unrealistic figures, and yet these targets represent the estimated prevalence of obesity in 1980. If current trends continue, it is likely that the real prevalence in 2005 will be around three times the target values.

Kuczmarski *et al.* (1994) have reported similar rapid increases in obesity prevalence in the USA during the 1980s. A nationally representative survey conducted in 1988–91 found that just over 33% of all American adults (20–74 years) had BMIs that exceeded 27.8 for men and 27.3 for women. A similar survey conducted a decade earlier found only 25% of adults above these threshold values. The stated aim of the US government in 1992 was that, by the year 2000, no more than 20% of Americans should have BMIs that exceed 27.8 for men and 27.3 for women (DHHS, 1992). Based upon current trends, the number of people exceeding these thresholds is likely to be around twice this 20% target value.

Trends around the world

The prevalence of obesity is rising not just in the UK and the USA, but throughout the industrialized world. Table 8.2 shows recent changes in the estimated prevalence of adult obesity in the USA and several European countries. Differences in the time intervals and in the cut-off points used to define



Figures 8.2 The effects of age on the percentage of (a) men and (b) women who are overweight or obese. Data source: Prescott-Clarke and Primatesta (1998).

obesity hinder direct comparison; nevertheless, it is clear that rates are rising in all of these countries, although the prevalence and rate of increase vary amongst countries. In Sweden, for example, there seems to be a relatively low prevalence of obesity and a relatively slow rate of increase. There are other signs that a continuing upward trend is not inevitable; there seems to have been a decline in obesity rates amongst Danish women in recent years, and rates of obesity in Holland appear to have stabilized (information supplied by the Obesity Research Information Centre, London).

Obesity is not confined to the industrialized countries; it is an increasing problem in many of the developing countries as they adopt 'Western' dietary and lifestyle patterns. As an extreme example, in urban areas of Western Samoa, 60% of men and

about 75% of women are obese and less than 10% are not overweight, i.e. BMI below 25 (see Table 8.2). Hodge *et al.* (1994) reported large increases in obesity rates in all regions of Western Samoa over the period 1978–1991, but increases in the rural areas were particularly dramatic, e.g. in men in the rural Tuasivi region, obesity rates quadrupled in 13 years. In 1978, there was a large urban–rural difference in obesity rates in Western Samoa. Obesity rates were much higher in the urban areas, where many people were employed in sedentary jobs and had ready access to imported Western foods, but less prevalent in the rural areas, where the diet and lifestyle were more traditional. With increasing development and industrialization, the rural populations are rapidly becoming as obese as their urban neighbours.

Table 8.2 Changes in the prevalence of obesity in selected countries

Country	Time span	Percentage obese
England	1980→95	
Men		6→15.3
Women		8→17.5
US	1976/80→1988/91	
Men (BMI > 27.3)		24.1→31.7
Women (BMI > 27.8)		26.5→34.9
Sweden	1980/81→1988/89	
Men		4.9→5.3
Women (BMI > 28.6)		8.7→9.1
Netherlands	1981→1993	
Men		3.9→8
Women		6.2→10
Germany	1985→1990	
Men		15.1→17.2
Women		16.5→19.3
Western Samoa	1978→1991	
Apia (urban) men		38.7→58.5
Women		61.2→77.7
Tuasivi (rural) men		9.9→39.3
Women		26.6→57.2

Obesity is defined as a BMI of over 30 unless indicated otherwise in brackets
 Main data source: Obesity Research Information Centre, London. Data for Western Samoa from Hodge *et al.* (1991).

Effects of ethnicity and social status upon obesity prevalence

Within the UK and USA, there are marked differences in the prevalence of obesity among different socioeconomic and ethnic groups. In the USA, obesity is much more prevalent amongst native Americans, Mexican Americans and black women (Table 8.3). Rates of obesity and Type 2 diabetes are extremely high in some tribes of native Americans such as the Pima of Arizona. In Britain, obesity is also more common in black women and people of South Asian origin; this latter group is particularly prone to abdominal obesity. An English survey (Prescott-Clarke and Primatesta, 1998) found that 45% of middle-aged black women were obese, compared to 28% and 21%, respectively, of middle-aged Indian and white women (unpublished statement by P. Primatesta).

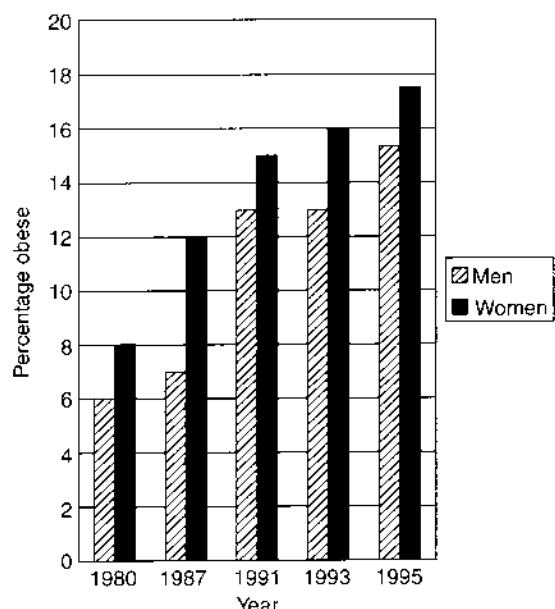
In the industrialized countries, obesity is more prevalent in the lower socioeconomic groups and amongst the less well educated. There is a pronounced increase in

Table 8.3 Ethnic differences in obesity prevalence in the USA

USA	Percentage obese
Women (BMI 27.3+)	
White	32.1
Black	48.5
Mexican American	47.2
Men (BMI 27.8+)	
White	31.6
Black	31.2
Mexican American	39.1

(Data from Kuczmarski *et al.* 1994).

the prevalence of obesity in English women as one goes down the socioeconomic scale (Prescott-Clarke and Primatesta, 1998). The prevalence in the lowest social class (V) is twice that in the highest social class (I). This class-related pattern of obesity is much less clear in English men. There is a strong inverse correlation between level of educational attainment and obesity: those with higher levels of education are less likely to become obese than those with more limited education. Sobal and Stunkard (1989) reviewed the influence of socioeconomic status on obesity around the world. They concluded that, in industrialized countries, obesity tends

**Figure 8.3** Changes in the estimated prevalence (%) of obesity in England since 1980.

- Obesity rates are rising rapidly in most industrialized countries and in many developing countries.
- In England in 1996, 16% of men and 18% of women were obese and a further 45% of men and 34% of women were overweight.
- More than 60% of American adults have a BMI of over 25.
- In 1980 it was estimated that only 6% of English men and 8% of women were obese.
- The proportion of people who are overweight or obese rises steeply throughout early adulthood and middle age so that, by middle-age, two-thirds of English people have a BMI of over 25.
- In the USA, obesity is much more prevalent amongst native Americans, Mexican Americans and black women than it is in white Americans.
- In Britain, South Asians and Afro-Caribbean women are more likely to be obese than whites.
- In industrialized countries, obesity is more prevalent in the lower socioeconomic groups and less well-educated sectors of the population. In England, this trend is particularly pronounced in women.
- In developing countries, obesity may be more prevalent in the upper social classes; it is often only these groups that have the opportunity to get fat.

to be more prevalent amongst the lower socioeconomic groups, whereas in poorer countries it is more common in the upper social groups because only the relatively affluent have the opportunity to get fat.

THE MEDICAL AND SOCIAL CONSEQUENCES OF OBESITY

The relationship between BMI and life expectancy

Studies conducted by life insurance companies since early in the twentieth century have found that being overweight or obese is associated with excess mortality and that mortality rises steeply when BMI is over 30; mortality ratio is doubled at a BMI of around 35. In a study of 200 morbidly obese American men, death rates were 12 times higher than normal in those aged 25–34 and still three times higher than normal in the 45–54-year age group when the total number of deaths is much higher (Drenick *et al.*, 1980).

According to life insurance studies, the health effects of obesity seem to be greater in men than in women. Such studies usually show some increase in mortality in those who are underweight as well as in those who are overweight, i.e. there is a so-called

J-curve when mortality is plotted against BMI. The higher death rate at low BMI has usually been explained by suggesting that being underweight is often an indicator of poor health, or the result of smoking or of alcohol or drug abuse. Smoking, for example, is known to depress body weight and is also a cause of much ill-health and many premature deaths. Figure 8.4 uses data from a study of 115 000 American female nurses and shows a typical J-curve of mortality against BMI. However, this figure also shows the results from this study when the following 'corrections' were made:

- Those who died within the first 4 years were excluded (i.e. already in poor health when the study started?).
- When only those whose weight was stable at the start of the study were included (recent unintentional weight loss may be an indicator of existing ill-health).
- When only women who had never smoked were included (smoking damages health and lowers body weight).

The J-curve disappears when these women are excluded. The increase in death rate with increasing BMI is also more pronounced. These data suggest that raised mortality amongst smokers and women whose weight is depressed by poor health tends to

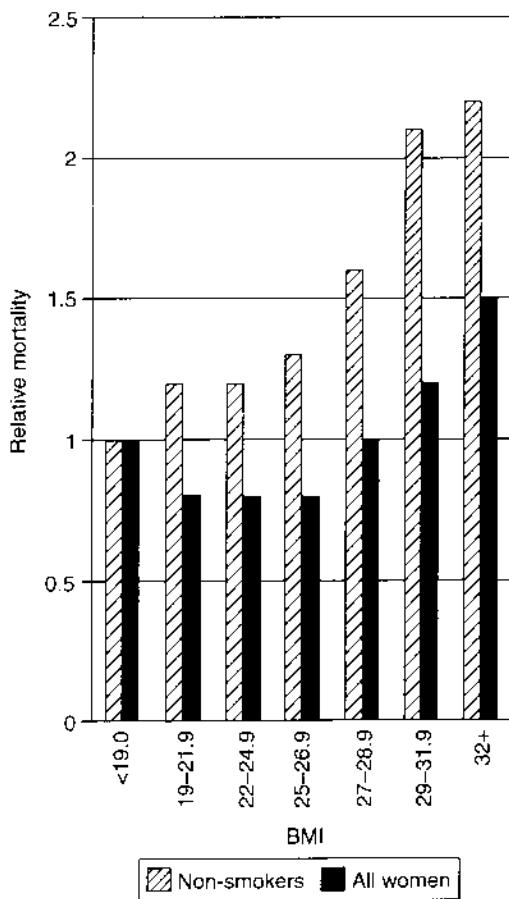


Figure 8.4 Body mass index (BMI) and mortality in women; the effects of smoking and ill-health. See text for details. Data from Manson *et al.* (1995). Relative mortality (RM) is the number of times higher the death rate is than for the corresponding group when the BMI is less than 19, e.g. if the RM for a group of smokers is 2, this means that the death rate is twice that of non-smokers with BMIs of less than 19.

obscure the strength and steepness of the association between BMI and mortality.

The mortality risk associated with being overweight or obese seems to be greater in young people. Andres *et al.* (1985) used life insurance data to calculate the BMI associated with the lowest risk of dying during the follow-up period in various age bands. This analysis is summarized in Figure 8.5 and it seems to indicate that the 'ideal BMI' (the one associated with lowest mortality risk) seems to rise progressively with age. This apparent age dependence in the risks of obesity is discussed in the section on the elderly in Chapter 14, where several studies are

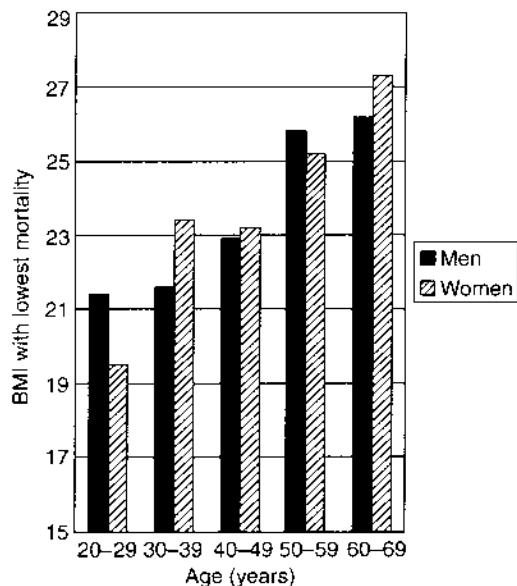


Figure 8.5 The effects of age on the apparent 'ideal' Body mass index (BMI), i.e. the BMI at which mortality rate is lowest. Data source: Andres *et al.* (1985).

reviewed which indicate an apparent protective effect of overweight in the very elderly.

Table 8.4 shows a list of conditions that are associated with high BMI and that are probably caused or worsened by obesity. Increased numbers of premature deaths from some of these conditions is largely responsible for the increased mortality and reduced life expectancy of obese people. Death rates from strokes, coronary heart disease, diabetes, diseases of the liver and digestive system and certain cancers in women are particularly high in the obese.

Obesity and the quality of life

Some of the conditions listed in Table 8.4 may cause ill-health and disability, even if they are unlikely to be a primary cause of death, e.g. gout and arthritis. Type 2 diabetes increases the risk of cardiovascular disease and strokes and also leads to a number of problems that are disabling even though not necessarily acutely fatal, e.g. blindness, peripheral neuropathy, gangrene and progressive renal failure (see Chapter 15).

One Finnish study (Rissanen *et al.*, 1990) found that overweight people were much more likely to claim a disability pension than normal-weight

Table 8.4 Some conditions that are thought to be caused by obesity or worsened by it

<i>Heart and circulatory system</i>	
Coronary heart disease	
Strokes	
Hypertension	
Angina pectoris	
Sudden death due to ventricular arrhythmia (abnormal heart rhythm)	
Congestive heart failure	
Varicose veins, haemorrhoids, swollen ankles, venous thrombosis	
<i>Joints</i>	
Osteoarthritis – knees	
<i>Endocrine/metabolic</i>	
Type 2 diabetes (maturity-onset diabetes)	
Gout	
Gallstones	
<i>Cancer</i>	
In women, increased risk of cancer of the ovary, cervix (neck of the womb), breast and endometrium (lining of the womb)	
In men, increased risk of prostate cancer and cancers of the colon and rectum	
<i>Other</i>	
Increased risk of pregnancy complications	
Reduced fertility	
Menstrual irregularities	
Increased risk during anaesthesia and surgery	
Reduced mobility, agility and increased risk of accidents	
Adverse psychological, social and economic consequences (see later in the chapter)	
<i>Risk factors</i>	
Raised blood lipids, including blood cholesterol	
Raised blood uric acid levels	
Elevation of blood pressure	
Insulin resistance	
Increased plasma fibrinogen concentration	

people. The authors concluded that being overweight is a major cause of disability in Finland. This effect of overweight on disability seemed to start at relatively low BMIs, where the effect upon mortality is normally small or non-existent (25+ for women and 27+ for men). The effect of overweight on disability also seemed to be more pronounced in women than in men. Being overweight or obese not only shortens life, but also reduces the proportion of life that is spent free of major illness and disability.

Also listed in Table 8.4 are several **risk factors** for heart disease that are elevated in the obese. Risk factors are parameters that predict the risk of developing or dying of particular diseases and in many cases are assumed to contribute to the cause of the disease. It is increases in these risk factors that probably lead to much of the excess mortality of obese people. In a large survey of English adults, Bennett *et al.* (1995) found the following examples of associations between BMI and cardiovascular risk factors.

- Overweight and obese people have high blood pressure. In this survey, people with a BMI of over 30 had double the risk of **hypertension** as compared to people with a BMI in the 20–25 range. Blood pressure rose with increasing BMI in all age and sex groups.
- Overweight and obese people have reduced insulin sensitivity and high average blood glucose concentrations (early indications of the onset of type 2 diabetes). **Glycosylated haemoglobin** concentration in blood tends to rise with average blood glucose concentration and so it is used as a marker for high blood glucose concentration. In this survey, obese people had the highest levels of glycosylated haemoglobin, and those with a BMI of less than 20 had the lowest levels.
- People who are overweight or obese tend to have the highest blood cholesterol levels. In this survey, there was a clear tendency for blood cholesterol to rise with increasing BMI.
- Obese people in this survey tended to have increased plasma uric acid levels.
- Plasma fibrinogen concentration is used as an indicator of the blood's tendency to clot and form thromboses. Plasma fibrinogen concentration tended to rise with increasing BMI in this survey.

In general, these risk factors tend to rise during weight gain and to fall during weight reduction, and this indicates that weight gain *per se* or the behaviours that lead to weight gain are the direct cause of these elevated risk factor levels. Type 2 diabetes seems to be triggered by a decline in target tissue sensitivity to insulin (insulin resistance) rather than by a primary failure of insulin secretion. In experimental studies, deliberate overfeeding and weight gain reduce insulin sensitivity in animals and people; insulin sensitivity increases during weight loss (e.g. Sims *et al.*, 1973). Weight loss is known to lessen the symptoms of type 2 diabetes. For men in their

forties, being moderately obese (BMI 31) increases the risk of developing type 2 diabetes by tenfold, and in those with a BMI of over 35 the risk is 77 times higher (Chan *et al.*, 1994). Weight loss is also an accepted part of the treatment for hypertension and gout. Gout is the painful joint condition caused by crystals of uric acid forming in the joints when blood uric acid levels are elevated.

In addition to higher morbidity and mortality, there are major social and economic disadvantages to being overweight. In some cultures, obesity may be admired as a symbol of wealth and success, and fatness may be regarded as physically attractive. In most Western countries, however, the obese have long been poorly regarded and discriminated against. In his book on obesity, Mayer (1968) largely devotes a whole chapter to literary examples of hostility towards the obese, and Gilbert (1986) also includes a short, referenced review of negative attitudes towards the obese.

Several studies have been conducted in which children have been asked to rate the likeability of other children from pictures and to assign various character traits to them. They consistently rate the obese children in these pictures as less likeable than the lean children and, in some studies, as even less likeable than children with physical deformities such as missing limbs or facial disfigurement. They attribute unpleasant characteristics to the obese children in the pictures – with comments such as they are 'lazy', 'dirty', 'stupid', 'ugly' and that they are 'cheats' and 'liars'. Basically similar attitudes are also widespread amongst adults of all social classes. Even amongst those professionally involved in the treatment of obesity, the obese may be seen as self-indulgent people who lack willpower and overeat to compensate for their emotional problems (these early studies are reviewed and referenced in Wadden and Stunkard, 1985).

Obese people are subject to practical discrimination as well: they are less likely to be accepted for university education and for employment. Gortmaker *et al.* (1993) did a 7-year follow-up study of 10 000 young people in Boston who were aged 16–24 years at the start of the study. They found that, particularly amongst women, there were substantial social and economic disadvantages to being obese. After 7 years, those women who were obese at the start of the study:

- had completed fewer years at school

- had lower household incomes and were more likely to be below the poverty line
- were less likely to be married (another study has reported that obese lower-class women are much less likely to marry men of higher social class than lean lower-class women).

Similar trends were seen in the men, but the magnitude of the disadvantages attributed to being obese was substantially smaller. The investigators concluded that prejudice and discrimination were the main causes of these socioeconomic disadvantages of being obese. They were not due to differences in the socioeconomic status of the obese subjects at the start of the study or to differences in levels of ability as measured by initial performance of the young people on intelligence and aptitude tests. This conclusion was also supported by their finding that there was no evidence of similar reduced socioeconomic prospects for those subjects who at the start of the study were suffering from a chronic physical condition such as asthma, diabetes or a musculo-skeletal problem. Such conditions could well have reduced physical performance, but being obese was more of a socioeconomic handicap over this 7-year period than suffering from one of these chronic conditions.

These negative attitudes to the obese may well be rooted in the historically assumed association of overweight, excessive food intake and inactivity, i.e. obesity is widely assumed to be due to greed and laziness. The obese are seen as responsible for their own condition due to the presence of these two undesirable personality traits.

Is all body fat equally bad?

Studies in the early 1980s indicated that the **waist-to-hip ratio** was an important determinant of the health risks associated with obesity. This measure is the circumference of the body at the waist divided by that around the hips and it is used as a simple indicator of body-fat distribution. The health risks seem to be greater if the obesity is associated with a high waist-to-hip ratio. Such people are sometimes called 'apple-shaped' and this is the typically male (ovoid) pattern of fat distribution. There is much less risk if the obesity is associated with a low waist-to-hip ratio. This is termed 'pear-shaped' or gynoid and is

the typically female pattern (reviewed by Seidell, 1992). Inactivity, smoking and high alcohol consumption seem to be associated with high waist-to-hip ratio and **hormone replacement therapy** in post-menopausal women seems to reduce central fat deposition (see Samaras and Campbell, 1997).

As one would expect, average waist-to-hip ratio is higher in men than in women and it tends to rise with age in both sexes, i.e. people tend to thicken around the waist as they get older – 'middle-aged spread' (Figure 8.6). Ideally, this ratio should be below 0.8 in women and below 0.9 in men. Simple measures of fat distribution such as waist-to-hip ratio are being increasingly used to gauge the magnitude of risk associated with a particular level of overweight or obesity and thus the priority that should be given to weight reduction (see Ashwell, 1994, for a review of these measures). A BMI of 27 in a person with a high waist-to-hip ratio seems to carry more

health risk than a BMI of 30 in someone with a low waist-to-hip ratio.

This research on waist-to-hip ratio and related measures suggests that fat deposited in the abdominal cavity may be largely responsible for the detrimental health consequences of being overweight or obese. Large amounts of abdominal fat seem to predispose people to diabetes and heart disease. The increase in heart disease may be largely a consequence of the predisposition to diabetes. South Asians seem to be particularly prone to diabetes and heart disease and have higher waist-to-hip ratios than Caucasians at any given BMI (see Chapter 16).

Weight cycling

In line with the high prevalence of obesity, there is also a high prevalence of weight dissatisfaction and dieting, particularly amongst women. At any one time, perhaps a third of American women are dieting, and studies in England and Ireland have found that most adolescent girls have tried to lose weight (see Ryan *et al.*, 1998). Many dieters fail to lose significant amounts of weight, but many others do achieve substantial short-term weight losses, only to regain this weight once their diet is relaxed. This cycle of weight loss and regain has been termed **weight cycling** or yo-yo dieting.

It is a common perception that people find it increasingly difficult to lose weight after previous cycles of weight loss and regain. One obvious explanation for this observation is simply that it becomes increasingly difficult for people to summon up the effort and willpower to stick to a weight-reduction programme after being dispirited by previous experience of rapid weight regain soon wiping out any hard-won weight losses. It has also been suggested that cycles of weight loss and regain might lower metabolic rate and so predispose people to weight gain because of the replacement of lean tissue by fat during these cycles of loss and regain. Those studies that have attempted to test this hypothesis with animals or people have generally not found evidence to support it (e.g. Jebb *et al.*, 1991).

Weight fluctuation is also associated with increased mortality risk, leading to speculation that yo-yo dieting might have detrimental health effects *per se*. One large US study did, indeed, find that men who had experienced weight loss or weight

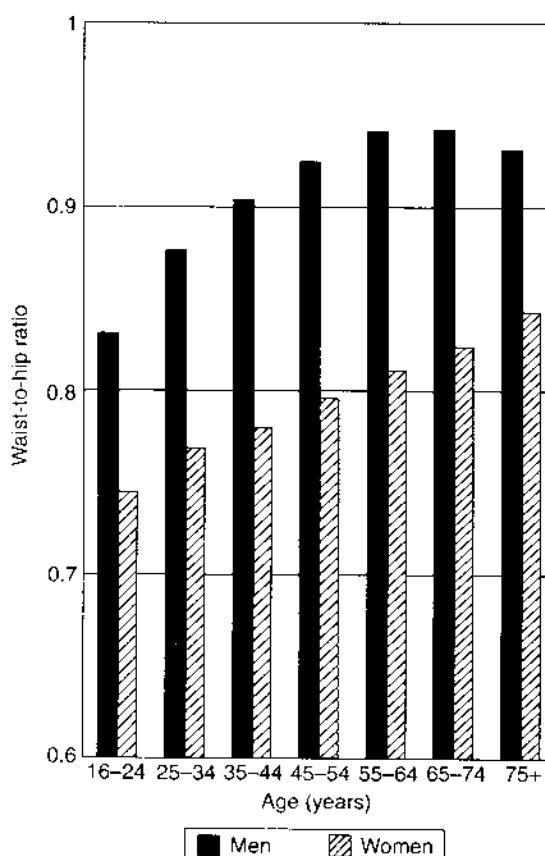


Figure 8.6 The rise in average waist-to-hip ratio with age.
Data source: Bennett *et al.* (1995).

fluctuation had an elevated mortality risk. However, when smokers and those with existing ill-health were excluded from the analysis, this association between weight fluctuation and elevated mortality risk disappeared. This led the authors to conclude that any association between weight fluctuation and elevated mortality is due to the effects of pre-existing illness and smoking and that weight cycling is not, in itself, directly harmful to health (Iribarren *et al.*, 1995). Note, however, that many patients with eating disorders do have a history of yo-yo dieting.

Does high BMI directly cause an increase in mortality?

The earlier observation that risk factors (e.g. high blood pressure and insulin resistance) tend to rise and fall with induced changes in body weight provides persuasive support for the belief that the association between BMI and mortality is a cause-and-effect relationship. There are two probable scenarios:

- either the presence of excessive amounts of body fat directly hastens death via development of the diseases in Table 8.4
- or the behaviours that cause people to gain weight also increase the risk of ill-health.

Such a subtle distinction might seem to be pedantic because, in either case, changes in behaviour that prevent or reduce obesity should lead to corresponding improvements in health and longevity. Some work by Lee *et al.* (1998) suggests that this may be a significant distinction. They related measured fitness to all-cause mortality in a large sample of men divided into strata on the basis of their BMI. They found, as one might expect, that in each BMI stratum, mortality was inversely related to fitness (i.e. the fit were less likely to die). More surprisingly, they also found that men in the high-fitness and moderate-fitness categories had similar death rates irrespective of BMI category (even those in the BMI 30+ category), and that men in the unfit category had high death rates, again irrespective of their BMI category. This suggests that men with high BMI have high mortality because they are much more likely to be inactive and unfit – the inactivity and lack of fitness are the common cause of both their high BMI and high mortality risk. More encouragingly, this would also suggest

that increasing activity and improving fitness could substantially reduce the health consequence of overweight and obesity, even if it does not lead to substantial weight loss. This lends support to the suggestion in an editorial in the influential *New England Journal of Medicine* (Byers, 1995) that it might be time to shift the emphasis of research and health education away from body weight *per se* and to focus more upon the behaviours that lead to excessive weight gain. Rather than focusing on body weight reduction, we should concentrate more on encouraging healthier food choices and increased physical activity and upon reducing the behavioural and cultural barriers that prevent people from adopting this lifestyle pattern.

This is an important point for those working in health education and health promotion to take on board. Huge numbers of people go on diets and spend lots of money on a variety of books and products in order to try, usually unsuccessfully, to achieve an 'ideal' body shape. The severe social and psychological consequences of being obese mean that very few people are content to be obese, and many people are desperate to become or remain lean and are willing to adopt desperate measures to achieve this. Despite this, obesity rates continue to increase. The work of Harris (1983) exemplifies this almost universal wish to be lean. She asked Australian psychology students how they would feel about being obese themselves. All of the students gave negative responses to this question and 60% chose the highest level of negative response offered. Studies with teenage girls in Britain and Ireland have found a high level of dissatisfaction with their body weight. For example, Ryan *et al.* (1998) found that 60% of a sample of 15-year-old Dublin schoolgirls wanted to be slimmer and 68% had tried to lose weight. Most worryingly, many normal-weight and even underweight girls also expressed this wish to be thinner. Many of these girls reported using a variety of unhealthy practices to try to lose weight, including purging, smoking, periodic fasting and avoidance of staple foods. Health promotion campaigns that focus upon body fat *per se* are, at best, probably pointless, because most people already desperately want to become or remain lean. At worst, they may increase the prejudice and discrimination against obese people and push some impressionable young people into trying to lose weight when it is inappropriate and into using methods that are harmful.

- An elevated BMI is associated with reduced life expectancy, increased risk of illness and disability, and increased levels of cardiovascular risk factors.
- The health consequences of obesity are greater in the young and probably greater in men.
- Low BMI is also associated with reduced life expectancy, but this is probably because of an excess of smokers and those with existing ill-health in this category.
- Obesity has adverse social and economic consequences for the individual; obese people, especially women, are the subject of widespread prejudice and discrimination.
- Obesity that is associated with a high waist-to-hip ratio is more harmful than if the waist-to-hip ratio is low; excess fat stored in the abdominal cavity is largely responsible for the adverse health consequences of obesity.
- There is no convincing evidence that cycles of weight loss and regain (yo-yo dieting) are directly harmful to health or have physiological consequences that predispose to obesity. Any excess mortality attributed to yo-yo dieting (weight fluctuation) is probably due to the weight instability of many smokers and those with existing ill-health.
- Cardiovascular risk factor levels (e.g. insulin resistance and high blood pressure) tend to rise and fall with weight gain and loss. This suggests that the relationships are probably causal, i.e. either weight gain *per se* or the behaviours that cause weight gain are responsible for the elevated risk factor levels.
- Some evidence suggests that the inactivity and low physical fitness of many obese people may be largely responsible for both their obesity and their reduced health prospects.
- Health promotion should concentrate upon encouraging and enabling people to increase their activity and fitness and to eat a healthy, low-fat diet.
- Most people already desperately want to become or remain lean, so health promotion campaigns that focus upon body weight *per se* may achieve little. They may simply:
 - increase prejudice and discrimination against the obese
 - increase the dissatisfaction of overweight people
 - encourage some young people to aim for an inappropriately low body weight and to adopt harmful strategies to achieve it.

THE CAUSES OF OBESITY

During the course of human evolution, intermittent food shortages and an active lifestyle would have been the norm. The assured availability of a varied, highly palatable and energy-dense diet coupled with a negligible requirement for physical work would not have been a general experience. The reality would often have been a general state of food shortage or insecurity interspersed with occasional gluts of food. Those best equipped to cope with limited rations and to survive prolonged bouts of severe food shortage would be the most likely to survive, reproduce and pass on their genes. Some deposition of extra fat reserves during periods of abundance, to act as a buffer during the almost inevitable periods of food

shortage, would have been a considerable aid to survival. Permanent and severe obesity is unlikely under these circumstances because, as we saw in Chapter 7, energy expenditure tends to rise under conditions of energy surplus and this tends to restore energy balance. The monotony, low palatability and low energy density of primitive diets would have tended to limit excess consumption. Overweight and obese people may also have been at a disadvantage in the competition for food. Only where people have an unlimited supply of palatable, varied and energy-dense food and where they can curtail their activity as they gain weight would obesity be likely to be prevalent, progressive and often severe. The struggle to ensure energy adequacy and security would have been a major selection pressure shaping our mechanisms of energy-balance control. The need to prevent excess

fat accumulation would have been a more minor influence.

Given these selection pressures, one would expect our regulatory mechanisms to be more efficient at preventing energy deficit rather than at preventing energy surplus. Those best equipped to cope with periods of famine and shortage would be more likely to survive and reproduce, whilst those best equipped for conditions of continual plenty and inactivity would tend to be selected out during times of famine. It is therefore not too surprising that, under conditions of affluence and plenty, many people store more body fat than is desirable for their long-term health prospects. Some populations seem to be even more prone to obesity than Europeans and their descendants, e.g. the Polynesian people of Western Samoa and some native American tribes such as the Pima Indians of southern Arizona in which the majority of the adult populations are clinically obese. Often populations that have only recently adopted Western dietary and lifestyle habits and are historically accustomed to a harsher energy-supply environment seem to be particularly prone to obesity and diabetes. They may have evolved a 'thrifty' genotype that makes them better able to survive such conditions. There is also some evidence that a period of food shortage in early life (or other adverse circumstances such as cold exposure) may programme people to become obese in later life if environmental conditions allow it.

Nature or nurture?

Both genetic and environmental influences must be involved in the genesis of obesity; only the relative contribution of nature and nurture is disputed. As we saw in Chapter 7, certain relatively uncommon genetic defects can cause profound obesity in animals and people. It is also possible to selectively breed animals for leanness or fatness. Dietary and environmental changes can also markedly affect the degree of fatness of laboratory or farm animals, e.g. offering rats a varied, palatable and energy-rich diet when confined in cages (cafeteria feeding) causes them to get fat. The situation of the cafeteria-fed rat seems to mimic the situation of many affluent humans.

There is a strong correlation between the fatness levels of parents and their children and between the fatness levels of siblings, but it is difficult to dissect

out how much of this is due to common environmental factors and how much to shared genes. Several lines of enquiry point towards a considerable genetic component in this tendency of obesity to run in families:

- There is a much higher correlation between the fatness of monozygotic (identical) twins than between dizygotic (non-identical) twins. When they share the same environment, the fatness levels of the monozygotic twins are closer because all their genes are common.
- When monozygotic twins are adopted and reared apart, their fatness levels are still very similar. Despite their differing environments, their identical genes mean that they show similar levels of fatness.
- Large studies on adopted children in Denmark suggest that correlations between the BMI of adopted children and their natural first-degree relatives (parents and siblings) are similar to those in natural families – suggesting that family resemblances in adiposity are mostly due to genetic factors.

Of course, the environmental and dietary variation between Danish families may be relatively small and this will tend to exaggerate the prominence of genetic factors in determining obesity levels. See Sorensen (1992) for a referenced review of these genetic aspects of obesity.

Clearly, certain environmental conditions are necessary for the expression of any genetic tendency to obesity, but, equally, some individuals and some populations, such as the Western Samoans and Pima Indians, do seem to be more susceptible than others. The widespread use of animal models with some inherited or induced lesion that produces extreme obesity may have encouraged the view that obese and lean are two distinct categories. In reality, there will be a continuum of genetic susceptibility to obesity in people, ranging from very resistant to very susceptible. As the conditions in affluent countries have increasingly favoured the development of obesity, so more and more people's 'threshold of susceptibility' has been reached, and so the number of overweight and obese people has risen. The recent large rises in the prevalence of obesity in many populations cannot be due to genetic change – they must be caused by lifestyle changes. The increase in obesity as one goes down the socioeconomic classes is unlikely to be

due to genetics, particularly as it is much less obvious in men than in women. The relatively low rates of obesity in some affluent European countries and the low rates of increase suggest that very high rates of obesity and rapid increases in its prevalence are not an inevitable consequence of affluence. The fact that obesity rates in England in 1980 were well under half current levels confirms that this change must be the result of dietary and/or lifestyle changes. If English people reversed some of the lifestyle and dietary changes of the last 20 years, there would be corresponding falls in the levels of obesity.

Studies on the Pima Indians provide striking evidence of the need for environmental triggers to produce widespread obesity, even in a population considered genetically susceptible to obesity. Several centuries ago, one group of Pima Indians settled in southern Arizona whilst another group settled in the mountains of Mexico. These two groups remain genetically very similar. Present-day Mexican Pimas do not suffer from widespread obesity; they are subsistence farmers who spend many hours each week in hard physical labour and still eat a fairly traditional low-fat, high-carbohydrate diet. In contrast, the Arizona Pimas have ceased to be subsistence farmers; they are now a sedentary population, and they eat a typical high-fat American diet. Consequently, they have one of the highest rates of obesity in the world (see Gibbs, 1996).

A weakening link between hunger and eating? The internal/external hypothesis and behaviour therapy

Hunger is the drive or need to eat food that is initiated by the appetstat in response to internal physiological signals such as the absence of food in the gut, low blood glucose concentration or depleted fat stores. As noted in Chapter 7, many 'external' factors in addition to physiological need determine when, where, what and how much we eat. **Appetite**, the desire to eat palatable food, may persist even after hunger has been quenched. Social convention, emotional state or just the hedonistic pleasure of eating may cause us to eat in the absence of hunger. Hunger will usually drive us to eat, but the absence of hunger will not necessarily prevent us from eating. By analogy, we do not have to be thirsty to make a cup of tea or to accept the offer of a social drink. Thirst will

cause us to drink, but we may drink for pleasure or to be sociable when we are not thirsty. We can easily excrete excess fluid, but we are less well equipped to burn off surplus calories, notwithstanding the discussion of adaptive thermogenesis in Chapter 7. Perhaps modern diets and lifestyles weaken the link between hunger and eating and make us more prone to excess consumption and weight gain. For the majority of people in affluent populations:

- food is more consistently abundant than it has ever been
- there is less effort required to prepare food and more encouragement to consume it (e.g. advertising) than ever before
- food is more varied, palatable and energy dense than it has ever been
- the requirement for food has never been lower because of our inactive lifestyles and the high energy density of our diets.

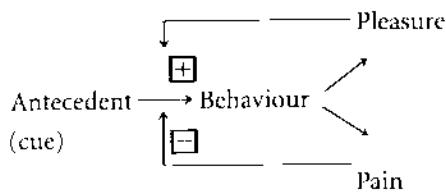
Schachter (1968) proposed that lean people regulate their food intake primarily in response to their internal physiological hunger mechanisms, whereas obese people are much more responsive to external, non-physiological influences upon their feeding behaviour. Schachter and Rodin (1974) identified certain behavioural characteristics of obese rats and then, in an ingenious series of experiments, demonstrated very similar behavioural characteristics in obese people (see the examples below).

- Obese rats and people consumed more food than lean ones when it was readily available, but less when they were required to work for the food. Obese rats generally ate more than lean ones but, if they were required to run through a maze to obtain food, the lean ate more than the obese. In another experiment, lean and fat people were offered nuts to eat in a disguised situation. If these nuts were opened and easy to eat, the obese subjects ate many more than the lean ones, but when the subjects had to crack the nuts open for themselves, the obese ate far fewer nuts than the lean!
- Obese rats and people consumed more than lean ones when the food was palatable, but less when it was made unpalatable, e.g. by the addition of quinine – they were more finicky.
- The feeding behaviour of obese rats and people was more affected by their anxiety state than was that of the lean ones.

Schachter and Rodin (1974) concluded that the feeding behaviour of obese rats and obese people was motivated primarily by appetite (the learned, hedonistic desire to eat food) rather than by hunger (the physiological need to eat food). The lean appeared to be driven to eat by hunger, but the obese were persuaded to eat by ease of access, high palatability and for emotional comfort.

This theory has lost favour in recent years – external orientation of food regulation is not confined to the obese. Nevertheless, it has had a great practical influence on the treatment of obesity. Much of the behavioural treatment of obesity is based upon this concept of external orientation of food-intake regulation in obese people.

Behaviour is envisaged as being initiated by certain cues or antecedents and then, if the behaviour results in pleasure that acts as a reward or reinforcement, that behaviour is learned. If the behaviour has unpleasant consequences, this acts as negative reinforcement and discourages repetition of the behaviour.



Thus, obese people are triggered into inappropriate eating behaviour by certain circumstances or cues and their inappropriate behaviour is rewarded, and thus learned, because of the pleasure or relief of anxiety that results from the inappropriate eating behaviour. **Behaviour therapy** seeks to identify the cues or antecedents to inappropriate behaviour (e.g. eating high-energy foods). It then involves devising ways of avoiding these cues and ways of reinforcing or rewarding appropriate behaviours (e.g. taking exercise or achieving weight loss targets) so that they are 'learned'.

In order to identify the antecedents or cues that trigger inappropriate eating behaviour, patients are often asked to keep a diary of when and what they eat, and in what circumstances they eat it. They are then encouraged to modify their habits and environment to reduce these cues, using techniques like those below.

- They may be encouraged to avoid the temptation to buy high-energy foods on impulse by shopping strictly from a list.
- They may reduce casual, impulse consumption of high-energy snacks by making them less accessible, e.g. keeping them on a high shelf or in a locked cabinet so that this snacking has to be a premeditated act.
- They may reduce the temptation to pick at food they are preparing by encouraging other family members to prepare their own snacks.
- They can reduce the temptation to eat leftovers by throwing them away immediately.

They may be encouraged to modify the process of eating using techniques such as:

- always using utensils and never eating with fingers
- slowing down the eating process by chewing food thoroughly, using a smaller spoon, pausing between mouthfuls or introducing a short gap in the meal
- never eating alone, but always with other family members
- leaving some food on their plate at the end of a meal.

They are also encouraged to increase the cues to appropriate behaviour as in the examples below:

- keeping walking shoes and equipment for active pursuits in easily accessible and visible locations
- keeping low energy snacks ready prepared and in a prominent place in the refrigerator.

Most therapists today avoid the use of active 'punishment' in their schemes. Clients are encouraged to eat with others and ask them to praise and encourage appropriate behaviour but to ignore inappropriate behaviour and simply withhold praise. Scolding is seen as attention giving and thus a form of reward, whereas withholding of attention is seen as punishment. In the past, some behaviour therapists have used punishment as an active part of their programme – so-called aversion therapy. Patients were shown images of problem foods followed by painful electric shocks or some other unpleasant image in the hope that they would learn to associate these foods with unpleasant sensations and so avoid them! A thorough and referenced review of the behavioural methods employed to treat obesity may be found in Gilbert (1986).

Variety of food and sensory-specific satiety

Rolls *et al.* (1982) reviewed a series of experiments which would suggest that increasing food variety may encourage increased consumption. Volunteers were asked to rate the pleasantness of different foods before and after eating. Previous consumption of a food reduced its pleasantness rating and the desire to consume more of it, but had little effect on that of other foods – they termed this phenomenon **sensory-specific satiety**. When people ate sausages, their rating for sausages was reduced, but the pleasantness rating for other foods, e.g. fruit or cookies, was largely unaffected by eating sausages. Subjects offered successive plates of sandwiches ate more if the sandwiches contained different fillings than if they were all the same. More surprisingly, subjects offered successive plates of pasta ate more if the pasta shape was

varied than if the shape was the same. Variation in taste or appearance increased the amount eaten. If a diet is monotonous, we are less likely to overeat than if it is varied. This is not too surprising a conclusion – most of us can usually be persuaded to accept a little dessert even when satiated by earlier courses of the meal. Even laboratory rats eat much more when they are offered a variety of foods (cafeteria feeding).

Sensory-specific satiety should increase the range of foods that people eat and limit the consumption of any single food. It should therefore decrease the likelihood of nutritional inadequacies because they are unlikely if the diet is varied. It would also reduce the likelihood that any of the toxins found in small amounts in many foods would be consumed in sufficient quantities to be hazardous. In an affluent population, nutrient deficiencies are unlikely; food is always plentiful and its variety appears almost limitless. Under such conditions, sensory-specific satiety

- Evolutionary pressures may have caused the mechanisms that regulate energy balance to be relatively ineffective at preventing excess energy storage.
- Some populations (e.g. the Pima Indians and Western Samoans) are extremely prone to obesity; a harsh environment in the past may have caused them to evolve with a 'thrifty' genotype.
- Studies with twins and adopted children suggest that much of the family resemblance in fatness levels is due to genetic factors.
- Clearly, both environmental and genetic influences determine an individual's or a population's fatness.
- Certain environmental conditions are necessary before any inherited predisposition to obesity is expressed.
- Recent changes in obesity prevalence must be due to lifestyle changes.
- Even in those groups with an extreme propensity to obesity (such as the Pima Indians), it only becomes prevalent when environmental conditions are conducive to obesity.
- Hunger will drive us to eat during fasting or fat-store depletion, but absence of hunger will not always prevent us from eating.
- 'External' influences such as pleasure of eating, habit, social pressures or anxiety may cause us to eat in the absence of hunger.
- The feeding of obese people was said to be more influenced by external factors than that of lean people, who were thought to be more responsive to their internal regulatory mechanisms.
- Behavioural therapy is based upon the notion of external orientation of feeding control in obese people.
- In behaviour therapy, obese people are encouraged to identify and avoid the external cues that trigger inappropriate eating and to facilitate and reward appropriate behaviour (e.g. taking exercise).
- A wide variety of food encourages higher consumption. As one food is eaten, so satiation for that food develops, but the desire to eat other foods is much less affected – sensory-specific satiety.

probably encourages overeating and excessive weight gain.

IS FAT MORE FATTENING THAN CARBOHYDRATE?

This topic is reviewed in Livingstone (1996) and most of the sources not listed here are referenced in this review. For many years, it was very widely held that carbohydrates, and especially sugar, were the most fattening components of the diet. This notion was reinforced by a string of popular diet books that promoted a variety of low-carbohydrate diets as the best diets for weight loss. These popular diet books included one written by an eminent British professor of nutrition (John Yudkin – *This slimming business*). In his book, Yudkin made the point that severe restriction of dietary carbohydrate would not only affect the carbohydrate foods that provided a high proportion of the dietary energy, but would also indirectly reduce the fat that is normally consumed with these foods. He expected that, on his diet, fat consumption would either drop or stay the same. Thus, bread, sugary foods and drinks, breakfast cereals, rice, pasta, potatoes, cakes, biscuits etc. are deliberately restricted, but so, indirectly, are, for example, butter/margarine that is spread on bread, fat used to fry potatoes, rice or other carbohydrate foods, the fat in battered or breadcrumbed fish or meat, the fat in pies and milk on breakfast cereals. Under these circumstances, it is difficult to make up the lost carbohydrate calories by replacing them with fat because concentrated fat is nauseating. The net result would often be a reduction in total energy intake and this accounts for the frequent short-term successes of some of these low-carbohydrate diets.

Current wisdom and advice are the opposite to these earlier views: fat is now regarded as more fattening than carbohydrate and the current consensus amongst nutritionists is that diets that are rich in starch and fibre but low in fat are most likely to achieve long-term weight control. There are, of course, other important reasons for moderating fat intake. In cross-sectional studies of populations, dietary and anthropometric data are collected from large samples of people. It is often found in such studies that there is a positive correlation between the proportion of energy that is derived from fat and

BMI: those who get the highest proportion of their energy from fat are the most likely to be fat. Given the carbohydrate–fat and sugar–fat seesaws discussed in Chapter 6, it almost inevitably follows that, as the proportion of energy from carbohydrate or even sugar increases, so BMI tends to fall: those who get more of their energy from carbohydrate or sugar tend to be leaner. In one study of more than 11 500 Scots, Bolton-Smith and Woodward (1994) found that the prevalence of overweight and obesity decreased with an increase in the proportion of energy derived from carbohydrate or from sugar, but increased with increases in the proportion of energy that came from fat. In the 20% of the men in this Scottish population who had the highest fat-to-sugar ratios, rates of obesity were 3.5 times higher than in the 20% with the lowest ratios (double in women).

There is evidence that very low-fat diets (say, under 20% of energy from fat) are associated with weight loss even when total energy intake is not restricted. High-fat diets, on the other hand, lead to overconsumption of energy (positive energy balance) in studies using free-living subjects. There may well be an interaction between activity and dietary fat content in their effects upon energy balance control. In one study, the proportion of energy from fat was varied (20%, 40% or 60% of energy from fat) in active and sedentary men. The active men were in negative energy balance on the 20% fat diet and still slightly at 40%, but had a small positive balance when fat was raised to 60% of energy. The inactive men, however, only achieved balance on the 20% fat diet and were in substantial positive balance on the other two diets. These observations suggest that, in sedentary populations, diets with fat contents typically found in Western industrialized societies are likely to promote positive energy balance and weight gain. As we become more sedentary, so our high-fat diets will make us increasingly susceptible to obesity.

Listed below are several theoretical reasons why one might expect fat to be more fattening than carbohydrate.

- Fat is energy dense. It yields more than twice as much energy as the same weight of carbohydrate or protein. This means that high-fat diets are inevitably energy dense. This may be the predominant reason why high-fat diets encourage over-

consumption in free-living situations (see the discussion of energy density in Chapter 6).

- Fat increases the palatability of food and this will tend to encourage overconsumption, especially if it replaces bland starch. Overweight and obese people seem to have an increased preference for high-fat foods.
- The conversion of excess dietary fat to body fat is a very efficient process and only about 4% of the energy is wasted in the conversion process. In contrast, the conversion of dietary carbohydrate to fat leads to the loss of about 25% of the energy in the conversion process. The conversion of dietary protein into fat is even more wasteful. In practice, on normal, high-fat, Western diets, there is very little conversion of dietary carbohydrate to fat and most carbohydrate is either used directly as an energy source or converted to glycogen.
- Calorie for calorie, fat seems to be less satiating than carbohydrate. Numerous experiments have been performed in which subjects are given iso-energetic preloads of varying composition and the effect upon their subsequent hunger ratings and the amount consumed at a subsequent meal are compared. These studies consistently suggest that fat is slightly less satiating than carbohydrate (usually sugar) and that this difference is more pronounced in obese than in lean subjects.
- Consumption of carbohydrate leads to an accelerated use of glucose in metabolism and a reduction in fat oxidation – conditions that are associated

with satiation. Consumption of fat has little effect on the metabolism of fat or carbohydrate.

INACTIVITY AS A CAUSE OF OBESITY

The decline in physical activity is strongly implicated as an important cause of the rapid increases in the number of overweight and obese people in Europe, the USA and elsewhere (Prentice and Jebb, 1995). Car ownership, together with automation in the home, garden and workplace have all combined to substantially reduce our mandatory energy expenditure – the energy we must expend each day to accomplish the tasks of everyday life. Few people in the USA and western Europe now have jobs that are physically demanding. Some people have increased their leisure-time activities to compensate, e.g. recreational walking, jogging, exercise classes, participation in sports and active games. However, for most people, any increase in recreational activity does not compensate for the decline in mandatory activity.

There is persuasive evidence that *per capita* energy intakes have dropped substantially in Britain in recent decades. According to Durnin (1992), average energy intakes of teenagers were up to a third higher in the 1930s than they are now. According to the National Food Survey, *per capita* household food consumption has dropped by more than a quarter since 1950; a similar trend is seen using wider measures of energy consumption (Figure 8.7). This fall in energy intake has corresponded with the very large

- Obesity was at one time widely blamed upon high-carbohydrate, and especially high-sugar, diets.
- Low-carbohydrate diets were widely recommended for weight reduction and often achieved short-term success because they indirectly reduced fat and total energy intake.
- In cross-population studies, people who get the highest proportion of their energy from fat tend to have the highest BMI, whereas those who obtain the highest proportion from carbohydrate or even sugar tend to have the lowest BMI.
- Very low-fat diets are associated with weight loss even when total intake is not restricted, whereas high-fat diets encourage weight gain.
- Active people may be able to tolerate medium and high-fat diets without weight gain, whereas in inactive people they lead to weight gain.
- High fat intake encourages weight gain because fat is energy dense, palatable, less satiating than carbohydrate and more efficiently converted to storage fat.

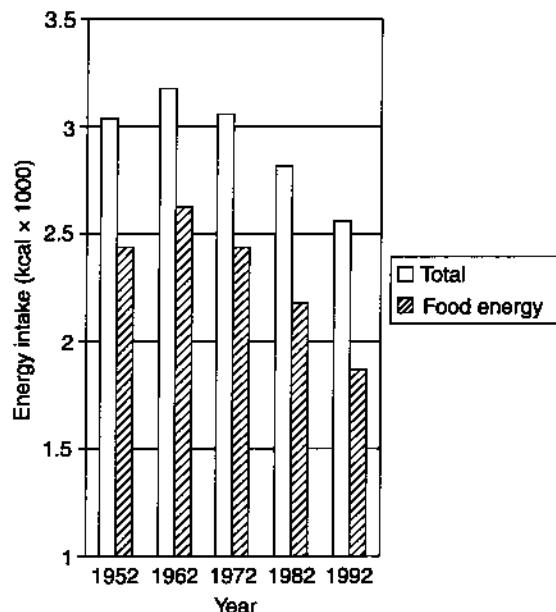


Figure 8.7 The apparent decline in average energy intake in the UK. Food energy data are derived from the National Food Survey and total energy intake from food balance sheets. Most data taken from COMA (1984).

increases in the number of people who are overweight and obese (as detailed earlier in the chapter), i.e. we are eating less but getting fatter. The only obvious explanation for these opposing trends is that activity levels have declined sharply. The fall in intake has lagged behind the fall in expenditure, resulting in an increasingly positive population energy balance that has led to a fattening of the population.

Prentice and Jebb (1995) have reported that increases in obesity prevalence in Britain since 1950 have tended to mirror increases in car ownership and the number of hours that Britons spend watching television.

Large cross-sectional surveys in both the USA and Britain have found that those who report being the least active are the most likely to be overweight or obese. Bennett *et al.* (1995) recorded BMI and assessed activity by questionnaire in a large representative sample of English adults. Figure 8.8 shows the variation in obesity in three sections of this English population when they were stratified according to activity level. In both men and women, obesity was more than twice as common in the lowest activity band compared to the highest.

In women, there is a very pronounced class variation in obesity prevalence. Prentice and Jebb (1995)

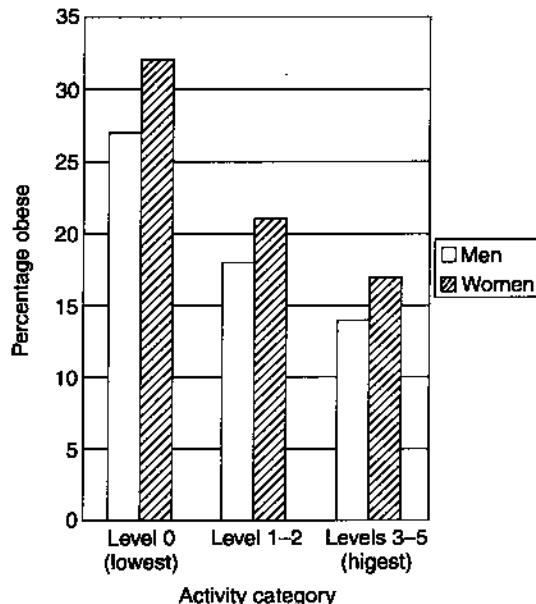


Figure 8.8 Obesity prevalence in different activity categories in middle-aged English people. Data source: Bennett *et al.* (1995).

showed that hours of television viewing, number of people reporting being inactive and obesity rates all rise as one goes down the social classes. Gregory *et al.* (1990) reported that BMI increased with declining social class, but that recorded energy intake was also lower in the lower social classes. Women in the lower classes were fatter but reported eating less, again an indication of reduced activity in the most overweight categories. One note of caution must be sounded about all of these studies. Both food consumption and activity level are self-reported in most studies and there is persuasive evidence that food intake tends to be under-recorded and activity level overestimated in such studies. This tendency to be overoptimistic about activity levels and energy intakes is probably more pronounced in the obese.

Television viewing is now probably the most popular recreational activity in most industrialized countries. The average Briton now spends twice as much time watching television as he or she did in 1960. Many people now spend more time watching television each week than they do at work or at school. Dietz and Gortmaker (1985) analysed data on television viewing in American children when they were aged 6–11 years and again when they were aged 12–17 years. When they divided the children up

according to the number of hours of television watched, they found that at both times the prevalence of obesity rose with increased viewing hours. This trend was most marked in adolescence, and at this stage there was a 2% rise in obesity prevalence for every extra hour of television viewing. They also found that the number of hours of television watching in the 6–11 year olds was significantly correlated with the prevalence of obesity 6 years later. Those 6–11 year olds who watched the most television were the most likely to be obese when they reached adolescence – a strong indication that the relationship is a cause-and-effect one.

Of course, even strong and consistent associations between obesity and inactivity in a variety of studies do not prove that inactivity causes obesity. Apart from the difficulty of measuring activity level, it is also likely that excess weight gain reduces fitness and exercise tolerance and this discourages activity – i.e. an effect-and-cause relationship. There is almost certainly a large element of truth in both of these sug-

gestions, and so a dangerous inactivity–obesity cycle is created (Figure 8.9).

PREVENTION AND TREATMENT OF OBESITY IN POPULATIONS

Adopting a ‘low-risk’ lifestyle

For effective, long-term weight control, one needs to identify dietary and lifestyle patterns that are associated with a low risk of obesity without the need for continual restraint on food consumption and calorie counting. From the earlier discussion on the causes of obesity, there would seem to be two changes that would lessen an individual’s risk of becoming overweight:

1. an increase in activity and fitness levels of the population (or perhaps just a reduction in inactivity)
2. a reduction in the proportion of dietary energy that is derived from fat.

If these changes were generally adopted, this would lead to a levelling off and eventually a fall in the population prevalence of overweight and obesity. Both of these changes would also be likely to have substantial health benefits over and above their effects upon body weight. High-fat diets are implicated in the genesis of cardiovascular disease, cancer and perhaps diabetes. The many health benefits of regular exercise are detailed in Chapter 16.

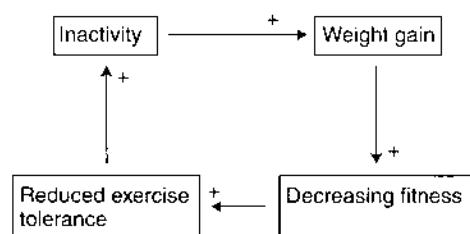


Figure 8.9 The inactivity–obesity cycle.

- Reduced requirement for activity to travel and to accomplish household and occupational tasks has led to a general decline in energy expenditure.
- Average energy intake in Britain has fallen substantially in the last 50 years and yet levels of fatness have increased markedly.
- Energy intake has fallen more slowly than energy expenditure, leading to increased levels of fat stores in the population.
- Indirect measures of population inactivity in Britain, such as amount of television viewing and car ownership, have mirrored increases in obesity rates.
- There is persuasive evidence in children that the amount of television viewing is positively linked to the risk of obesity.
- Inactivity is a major factor in recent increases in the prevalence of overweight and obesity.
- Inactivity encourages weight gain and weight gain probably discourages activity, leading to a cycle of inactivity and weight gain.

The most obvious way of increasing activity and fitness is by participation in active games, sports or formal exercise sessions and, of course, these need to be encouraged and facilitated. However, the major reason that affluent populations are now so inactive is because mandatory activity has been almost eliminated from everyday life for many people. Therefore, in order to bring about a substantial increase in the weekly energy expenditure of most people, we also need to find ways of incorporating more activity into everyday living, such as by the examples below.

- Making more journeys on foot or bicycle rather than by car. In Britain, journeys of less than a mile account for a third of all car journeys and these have been growing rapidly as a proportion of the total. More of these short journeys need to be made on foot or bicycle.
- Making more use of stairs rather than always taking lifts (elevators).
- Becoming less reliant on labour-saving gadgets and doing some tasks by hand.
- Engaging in more leisure-time pursuits that, although not formally exercise, do have a physical activity element, e.g. gardening, home improvement/maintenance, playing with children or walking pets.
- Limiting the time that is spent in totally passive pursuits such as watching television or playing computer games. Such pursuits demand physical inactivity and studies with obese children have found that simply reducing access to these sedentary activities was more effective in long-term weight control than programmes involving strenuous exercise sessions. Evidence was given earlier that children who watch the most television have the highest risk of becoming obese.

Health promotion campaigns that are aiming to improve population weight control need to concentrate on the following.

- Making it clear that the priorities for improving weight control are increased activity and a reduction in dietary fat.
- Educating people about the ways in which they can achieve these objectives, e.g. the sources of fat in the diet and dietary choices that would reduce fat consumption (see Chapter 10), how more

activity can be incorporated into everyday living, as well as encouraging participation in formal exercise sessions.

- Identifying and then looking for ways to minimize the barriers that make it difficult for people to be active and consume less fat.

It must be said that personal choice is a major reason why people are inactive and eat high-fat diets. High fat and sugar content tends to make food more palatable and so affluent populations tend to consume diets that are high in fat and sugar. Many affluent people have no necessity to be active and, indeed, their sedentary occupations may force them to be inactive during long working and commuting hours. To fill their leisure time, they have access to pleasurable and addictive pursuits that are also totally inactive. It is therefore not surprising that many people choose to be very inactive under these circumstances. Health promoters can only try to encourage people to go against their natural inclinations and adopt healthier lifestyle patterns. However, there may also be circumstances that encourage unhealthy lifestyles or barriers that prevent people adopting healthier lifestyles, such as:

- A lack of access to sports facilities or just to safe, attractive, open spaces, e.g. no local facilities, inconvenient opening hours, high cost of entry or membership fees, lack of money to buy sports clothing or equipment.
- Real or perceived danger when walking or cycling, e.g. vulnerability to injury by motor vehicles, fear of mugging or molestation, the noise and pollution caused by motor vehicles.
- Town planning decisions that assume that the car is the only mode of transport and make no provision for pedestrians or cyclists.
- Long hours spent in the workplace, in sedentary jobs, in an environment that discourages physical activity and with no opportunity to do anything active.

Some barriers to reducing fat consumption are listed below.

- Low-fat diets tend to be less palatable than high-fat diets.
- Lack of availability of low-fat foods. In fact, the availability of low-fat or reduced-fat foods has increased in line with consumer demand for these

products. Low-fat spreads, reduced-fat cheddar cheese, low-fat milk and yoghurt, reduced-fat crisps (chips) and savoury snacks, extra lean meat, reduced-fat French fries, low-fat cakes and biscuits (cookies) and reduced-fat ready meals are all readily available now. Note that some of the products listed still contain substantial amounts of fat and/or sugar even though their fat content is lower than the standard product, e.g. reduced-fat crisps and savoury snacks, reduced-fat cheddar cheese. The food provided in restaurants and some fast-food outlets often seems to have much more limited low-fat options.

- The increased cost of a diet that is both lower in fat and still palatable. Starchy foods are cheap but rather bland. Many low-fat or reduced-fat products carry a price premium over the standard product and almost all do if one expresses cost as calories (or joules) per penny. Fruits and vegetables, lean meat and white fish are all relatively expensive using this criterion. These extra costs may be significant for those on low incomes.
- Lack of knowledge about the sources of dietary fat.

Suggesting measures to reduce these barriers takes us firmly into the political arena and so I am not going to attempt to suggest detailed solutions. An increased risk of obesity is just one of many health consequences of poverty and poor educational attainment. If we are substantially to increase activity levels in the population as a whole, then the safety and comfort of pedestrians and cyclists need to be given a higher priority by politicians, town planners, property developers and law enforcement agencies. This may have to include active measures to discourage car usage, especially for short, urban journeys.

Targeting anti-obesity measures or campaigns

Is it possible to identify an age groups in which intervention to prevent or reduce obesity is likely to be most effective? Intuitively, childhood would seem to be a crucial time, because this is when good dietary and lifestyle patterns can be set for the rest of life, and this would seem like a better option than trying to change bad habits later in life. Most obese adults were not obese as children, but two-thirds of obese

children do become obese adults and it may well be bad habits learned earlier in life that are responsible for much adult-onset obesity. Some possible measures that could be directed towards children are listed below.

- Giving proper emphasis to food and nutrition, including food preparation skills, in the school curriculum. If people are advised to reduce their fat consumption, they must have a clear idea of where our dietary fat comes from and what practical measures would reduce fat consumption. They should have the skills and confidence to prepare appealing, healthy dishes.
- Recognition of the importance of physical education to the long-term well-being of children. It should not be sacrificed to make room for more academic work in the school timetable.
- Games' teaching should seek to involve all children, not just the talented minority. There should also be more emphasis on those activities that are more likely to be continued into adulthood. We would think poorly of a school that concentrated all of its academic efforts onto the brightest pupils and on subjects that were unlikely to be used in later life.
- Trying to limit the time spent watching television and playing computer games. Ideally, this should be by persuasion rather than coercion, e.g. by offering children more attention and more opportunities to spend some of their leisure time in enjoyable alternative activities.
- Every effort should be made to facilitate activity in children, e.g. by ensuring that they can walk or cycle to school safely. One can try to change the **subjective norm** by making activity more fashionable, e.g. by emphasizing the environmental benefits of walking to school.

At one time, the **fat-cell theory** (Hirsch and Han, 1969) gave a formal theoretical basis for this belief that childhood was the best time to act. Hirsch and Han suggested that early overfeeding permanently increased the number of fat cells and thus predisposed children to obesity in later life. This theory was based upon some dubious extrapolation of experiments with rats to people. In the first few weeks of life, Hirsch and Han suggested that rat adipose tissue grew both by cell division increasing the number of fat cells and by increases in fat-cell size. According to this theory, at about 6 weeks old, the number of fat cells in the rat

- Obesity prevalence would probably decrease if there was a general increase in activity and fitness and a decrease in the proportion of fat in the diet.
- Increased activity and a lower-fat diet would have substantial health benefits over and above any effects upon body weight.
- Increased participation in active leisure-time pursuits is highly desirable, but there is also a need to increase the activity associated with the tasks of everyday life.
- Health promotion campaigners must encourage these changes, educate people about how to make them and seek to minimize the barriers that prevent change.
- Reducing some of the barriers to change may involve measures that are politically controversial.
- Fat children tend to become fat adults, but as obesity rates rise steeply with age, most obesity is adult onset.
- The dietary and lifestyle patterns that predispose to obesity are probably laid down in childhood (Dietz, 1999).
- Nutrition, food preparation and all forms of physical education must be given due emphasis in the school curriculum.
- Inactivity in children needs to be moderated (television viewing and computer) and activity needs to be encouraged and facilitated.
- Serious doubts have been cast upon the notion that early overfeeding of children increases fat-cell number and permanently predisposes them to obesity – the fat-cell theory.

was fixed and any increases in adipose tissue in adult rats could only occur by increases in the size of existing fat cells. Overfeeding before 6 weeks of age permanently increased the rats' numbers of fat cells and so might have permanently increased their susceptibility to obesity. Pond (1987) provided very persuasive arguments for not applying these ideas to people. Pond had found that fat-cell numbers could be increased throughout life in primates and that the capacity of primate fat cells to increase by expansion alone was rather limited compared to the rat. She suggested that the key premise of this fat-cell theory was false. It was also suggested earlier that a period of deprivation in early life may programme people to become obese in later life if conditions allow it.

and a reasonably active lifestyle may be effective in preventing much excess weight gain. However, to lose fat that has already accumulated within a reasonable time-scale will probably require a period of deliberate and uncomfortable food restriction (dieting). Most treatments for obesity represent different strategies for making this dieting easier, more effective and more acceptable. This applies even to most anti-obesity drugs and surgical treatments for obesity.

If one assumes that 1 kg of adipose tissue yields about 29 MJ (7000 kcal), then even a daily deficit of 4.2 MJ (1000 kcal) will only result in the loss of 1 kg of adipose tissue per week. This is probably the most that can be hoped for from conventional dieting and, for small women or inactive people, perhaps as little as half of this may be a more realistic target. During the first week of a diet or in severe dieting, weight loss may be much faster than this because of the loss of glycogen reserves and protein (lean tissue). Each kilogram of body glycogen or protein is associated with 3 kg or more of water, so to lose 1 kg via these routes requires an energy deficit of only about 4.2 MJ (1000 kcal). However, body glycogen reserves are very limited and loss of lean tissue is not the aim of dieting. Any weight lost via these routes can be rapidly

OBESITY TREATMENT IN INDIVIDUALS

Realistic rates of weight loss

In order to lose substantial amounts of weight, energy expenditure must exceed energy intake for a considerable length of time. A moderately low-fat diet

regained once the diet is relaxed. It also takes significantly more than 29 MJ (7000 kcal) to synthesize 1 kg of adipose tissue and so, on a more positive note, an occasional lapse in an otherwise successful diet – say, a 1-MJ (250-kcal) chocolate bar – is not going to undo weeks of hard-won losses.

The aim of dieting is to lose fat but to minimize losses of muscle and other metabolically active lean tissue. It is inevitable and proper that some of the weight lost during slimming is lean tissue, because obese people have more lean tissue as well as more fat, and even adipose tissue has a cellular, protein, component. A ratio of about three-quarters fat to one-quarter lean may be a reasonable rule of thumb for planned weight loss. Extreme dieting and very rapid weight loss increase the proportion of lean that is lost, whereas exercise seems to protect against lean-tissue loss during dieting (reviewed by Prentice *et al.*, 1991).

As we saw in Chapter 6, starvation reduces resting metabolic rate (RMR) and this conservation of energy during starvation is an aid to survival during food shortage. It will also tend to thwart the efforts of the dieter. This decline in RMR remains even after correcting for changes in lean body mass and is rapid. It can be detected within 24 hours of the start of dieting, and a rise in RMR occurs within 24 hours of the return to normal eating (Prentice *et al.*, 1991). The magnitude of this protective response to dieting is not so large as to present an insurmountable barrier to successful weight loss. According to Prentice *et al.* (1991), even in severe dieting (say, under 3 MJ or 700 kcal per day), the decline in RMR per unit of lean body mass rarely exceeds 15%, and in more moderate dieting (over 5 MJ or 1200 kcal per day), it is only about 5%.

The reducing diet

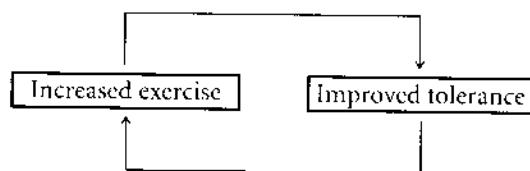
The diet should aim to produce a daily energy deficit of 2.1–4.2 MJ (500–1000 kcal). It should achieve this by selectively restricting fat, added sugars and alcohol. Such measures will drastically reduce the energy density of the diet, but will actually increase the nutrient density, which makes nutrient deficiencies unlikely despite the reduced energy intake. As we saw in Chapter 7, most fruits and vegetables have extremely low energy yields and so can be eaten in large quantities without adding much energy to the diet. Starchy foods such as potatoes and cereals are relatively low in energy density and wholegrain cereals have the added

advantage of high amounts of dietary fibre (**non-starch polysaccharide**), which may increase their satiation value (see Chapter 9). Low-fat milk, very lean meat, white fish, pulses and the vegetarian options in the meat group will also be essential components of a healthy reducing diet. Foods from the fats and sugars group of the food plate (or pyramid), high-fat meats and dairy products, fried foods and other foods with added fats and sugars are the prime candidates for reduction in a healthy reducing diet.

The role of exercise

The evidence that inactivity is strongly implicated as a cause of obesity seems very persuasive. This would also mean that regular exercise should be of major importance in preventing excess weight gain. However, evidence that exercise has a major impact on rates of weight loss during dieting is less persuasive. Most studies that have looked at the contribution of exercise to weight-loss programmes have concluded that the exercise has led to only a modest increase in weight loss as compared to dieting alone. However, when the long-term outcome has been assessed, the evidence is that an exercise component increases the chances of long-term success, i.e. it helps to prevent rapid regain of lost weight.

It takes a long time to walk off 2.1–4.2 MJ (500–1000 kcal), say, 2–6+ hours, depending on speed and the person's size. Very few people have this much time to devote to formal exercise. Many very overweight people will have a low exercise tolerance and so the amount of energy that can be consumed in a formal exercise session will be small. These observations highlight the need to incorporate more activity into everyday living as well as in formal exercise sessions. Earlier in the chapter, it was suggested that inactivity and weight gain become part of a vicious cycle. Increased activity should break this cycle and create a virtuous cycle – exercise improves tolerance and so enables more exercise to be taken:



The energy expenditure and energy requirements of an individual are determined by the person's basal metabolic rate (BMR) and physical activity level (PAL) (see Chapter 6). The PAL is the factor that BMR is multiplied by to allow for the increase in energy expenditure caused by all activities:

$$\text{Energy expenditure} = \text{BMR} \times \text{PAL}$$

Table 8.5 shows how the energy a woman needs to achieve balance and to produce a 2.1–4.2-MJ (500–1000 kcal) deficit increases with the PAL multiple.

A sedentary, largely housebound woman with a PAL multiple of 1.3 can eat less than 3.4 MJ (800 kcal) per day if she is aiming for a 4.2 MJ (1000 kcal) daily energy deficit. As the PAL multiple rises, so the amount that can be eaten while still achieving this deficit rises, so a woman with a PAL multiple of 1.7 could eat over 5.5 MJ (1300 kcal) and still achieve this deficit. A higher PAL multiple increases the amount that can be eaten whilst maintaining an energy deficit, and this also reduces the likelihood that dieting will lead to inadequate intakes of other nutrients. About 4 hours of walking per week increases the PAL multiple by 0.1.

These calculations suggest that achieving and maintaining long-term energy deficits of the required

magnitude may be much easier for people who are active, and may be impractical for people who are very inactive. It should also be borne in mind that exercise may protect against lean tissue loss during dieting, and so increase the quality of weight loss.

Are the obese less 'vigilant'?

In 1992, Garrow suggested that one major reason why some people remain relatively lean whilst others get very fat is that the lean are more vigilant. Lean people monitor their body weight more closely and take steps to correct any weight gain (i.e. eating less and exercising more) before it has become a major problem. He argued that the obese, on the other hand, are less vigilant and therefore do not initiate corrective steps until the condition has progressed to the point where the person cannot cope with the magnitude of weight loss that is required to restore normal body weight. This has been used to explain the social class differences in obesity prevalence; perhaps those in the lower socioeconomic groups tend to be less vigilant in monitoring their body weight and this is why they are more prone to becoming obese. This class difference in body-weight vigilance would be consistent with the general trend for the upper social groups to be more health aware. They

Table 8.5 The energy needs of a 65-kg woman at various physical activity levels (PAL) to achieve energy balance and a 500 kcal (2.1 MJ) or 1000 kcal (4.2 MJ) deficit

Energy needed	PAL multiple					
	1.0	1.3	1.5	1.7	2.0	2.2
For balance						
kcal	1375	1788	2063	2338	2750	3025
MJ	5.78	7.51	8.66	9.82	11.55	12.71
500 kcal (2.1 MJ) deficit						
kcal	875	1288	1563	1838	2250	2525
MJ	3.68	5.41	6.56	7.72	9.45	10.61
1000 kcal (4.2 MJ) deficit						
kcal	375	788	1063	1338	1750	2025
MJ	1.58	3.31	4.46	5.62	7.35	8.51

1.3 represents an extremely inactive person, e.g. a largely housebound, elderly person.

1.5 represents someone such as a teacher who walks around quite a bit while working.

1.7 represents a person who walks around quite a bit at work and also takes part in some regular leisure-time physical activity – such as jogs several times a week.

2.0 represents someone whose job involves a lot of heavy manual work and who probably also takes part in regular, vigorous leisure-time pursuits.

2.2 (or higher) represents a serious athlete during training.

Note that a brisk 30–40-minute walk would increase the PAL by around 0.1 on that day.

Modified from Webb (1998).

are, for example, more likely to choose foods with 'healthy images'. Similar lack of vigilance, it has been argued, allows regain of weight after successful slimming – the weight creeps back on unnoticed by the individual until it is not readily correctable. Once the weight loss or re-loss required is high, the person feels unable to control the situation, becomes disheartened and no longer makes any sustained effort to reduce or even to maintain body weight.

Garrow (1992) reported the use of an adjustable weight cord to increase the vigilance of his patients – the patient is only able to adjust the length of this cord by an amount that allows for normal, day-to-day fluctuations in waist size. Any increase in waist size above the adjustable range would make the cord uncomfortable to the patient and immediately signal the need for corrective measures. He has used such a cord to reduce

the risk of weight regain during and after slimming. The cord is shortened by the physician as slimming proceeds. It could also be used as a preventative measure. A cherished or expensive item of clothing that becomes uncomfortably tight may serve as a similar early-warning signal for many of us who struggle to keep our weight within certain tolerance limits.

MORE 'AGGRESSIVE' TREATMENTS FOR OBESITY

Anti-obesity drugs

Listed below are a number of potential approaches to the drug treatment of obesity.

- Substantial weight loss will require a sustained period of dieting and negative energy balance.
- A daily deficit of 4.2 MJ (1000 kcal) is equivalent to a weekly loss of 1 kg of adipose tissue; this is around the realistic maximum for conventional dieting.
- It is inevitable that both fat and lean are lost during slimming, and a ratio of three-quarters fat to one-quarter lean is a realistic guideline figure.
- Extreme energy restriction increases the proportion of lean that is lost, whereas exercise reduces lean-tissue losses.
- Resting metabolic rate (RMR) falls during dieting and rises when the diet is relaxed. In severe dieting, the RMR per unit of lean body mass may fall by up to 15%, but only by about 5% in moderate dieting.
- The reducing diet should selectively restrict fat, added sugars and alcohol, which will reduce the energy density of the diet but increase its nutrient density.
- Exercise has only a modest direct effect on rates of weight loss in most studies, but programmes that include an exercise element seem to have better long-term success rates.
- It takes a long time to expend 2.1–4.2 MJ (500–1000 kcal) in a formal exercise session. This highlights the need to incorporate more activity into everyday life.
- Many overweight and obese people will have limited exercise tolerance, but increased activity should increase tolerance and create a virtuous cycle.
- An active lifestyle increases the amount of food that can be consumed whilst still achieving the target energy deficit. It may be impractical for inactive people to aim for an energy deficit of 4.2 MJ (1000 kcal) with a conventional diet.
- A moderate diet increases the quality of weight gain (more of it is fat) and causes a smaller decrease in RMR.
- Obese people may be less vigilant in monitoring their body weight. Lean people are thus able to initiate corrective measures early, whereas the obese may not start them until the magnitude of weight loss required is large and very difficult to achieve. A waist cord has been suggested as one way of increasing weight vigilance.

- Appetite suppressants: drugs that affect appetite by affecting neurotransmission in the brain centres that regulate feeding, e.g. the amphetamines, which have a noradrenaline-like (norepinephrine) effect, and drugs such as fenfluramine and dexfenfluramine, which increase 5HT (serotonin) activity. The latter potentiate the effects of serotonin by inhibiting its re-uptake into nerve cells, which is the normal route for serotonin inactivation. A drug called sibutramine has combined noradrenaline and serotonin effects.
- Drugs that reduce digestion or absorption in the gut: e.g. **orlistat** blocks pancreatic lipase and this results in up to 30% of dietary fat not being digested and passing through the gut into the faeces. This drug has recently been approved for use in the UK and several other countries.
- Drugs that mimic or potentiate the effects of cholecystokinin (CCK): CCK is a gut hormone released during feeding that causes satiety and it is also a transmitter in the brain, where it also seems to depress feeding. A compound called butabindide has been shown to prevent the breakdown of CCK in the brain and this results in reduced food intake and weight loss in experimental animals.
- Drugs that increase energy expenditure: e.g. drugs that stimulate the β_3 -adrenergic receptor and lead to uncoupling of oxidative phosphorylation in brown fat.
- Leptin: early trials in obese people have been a little disappointing and, because it is a protein, it needs to be administered frequently by injection. In the long term, perhaps orally active agents may be found that mimic or potentiate the actions of leptin.

Many potential drugs in these categories are still in the experimental or development stages, but amphetamines, drugs that increase serotonin activity and, more recently, orlistat are or have been widely used.

When noradrenaline (norepinephrine) is injected into the brains of experimental animals, it reduces hunger and food intake and also stimulates the sympathetic nervous system. Amphetamine has a noradrenaline-like effect and was the first appetite suppressant to be widely used. Amphetamine itself also increases the activity of the nerve transmitter dopamine and this produces a potentially addictive euphoric effect. Modified amphetamines maintain the noradrenaline-like effect, but the addictive

dopamine-like effect is minimized. They have been widely used in the USA because, until 1996, they were the only appetite suppressants approved by the Food and Drug Administration (FDA). In controlled trials, they produce only modest extra weight loss as compared to placebos, and then only in the first few weeks of treatment.

In Europe and the UK, fenfluramine and dexfenfluramine have, until recently, been the most widely used anti-obesity drugs. When 5HT (serotonin) is injected into the brains of animals, it also decreases feeding. These drugs mimic the actions of serotonin but do not have the addictive dopamine-like effects of the amphetamines. They act, not by reducing the desire to eat, but by causing the patient to feel full or sated earlier. Fenfluramine has been in use for about 30 years and tens of millions of Europeans have taken these drugs over that period (over 60 000 Britons in 1996 alone). In 1996, they were finally approved by the FDA for use in the USA. Long-term studies of over a year have reported significantly greater weight loss in people treated with drug and diet as compared to diet and placebo. Rare, but potentially fatal, side-effects of these drugs have been known for some time (pulmonary hypertension) and there have been suggestions from animal experiments that they may damage serotonin-producing brain cells. However, in 1997, new evidence was produced in the USA suggesting that they might cause damage to heart valves. This led to both drugs being withdrawn from the market in both Europe and the USA.

Surgical treatment for obesity

Surgery is a last-resort treatment for severe and intractable obesity. All surgical procedures carry some risks and these are much higher in obese people. Surgical treatment does seem to offer some help in losing weight to severely obese people and it may even be the most effective treatment for this extreme group, judged purely on the basis of weight lost.

A number of the surgical treatments that have been used in the treatment of obesity are summarized below.

- **Gastric stapling:** this is the most common procedure used in Europe for the surgical treatment of obesity. It involves reducing the capacity of the

stomach by surgically stapling off a large part of it. This means that the patients can only eat small meals; if they eat too much, this may cause them to vomit.

- **Jaw wiring:** this involves wiring the jaws together so that the intake of solid food is restricted whereas the intake of liquids is relatively unhindered. This procedure is often effective in producing short-term weight loss whilst the jaws are actually wired, but patients usually regain the weight once the wires are removed. It may be useful for achieving short-term weight loss prior to surgery.
- **Fat removal:** this can be achieved by cutting or by **liposuction** (literally, sucking out areas of adipose tissue). This unpleasant procedure may be judged by some people to be worthwhile for its short-term, localized, cosmetic effects.
- **Intestinal bypass:** this procedure involves surgically bypassing a large section of the small intestine and shunting food past most of it. This hinders digestion and absorption. Patients often suffer from chronic diarrhoea and nutrient deficiencies and may also suffer from liver and kidney damage. This risky procedure has now been largely superseded by gastric stapling.

Very low-energy diets

Very Low Energy Diets (VLEDs) are commercial products that are formulated to contain all of the essential vitamins and minerals but provide only 1.6–3.2 MJ (400–800 kcal) per day. The product is consumed as a flavoured drink and is used to replace normal food. It is very difficult, using normal food, to provide all of the essential nutrients with such a low energy intake. Sometimes, a cheaper ‘milk diet’ is used as an alternative to these commercial products, i.e. skimmed or semi-skimmed milk plus a multivitamin and mineral supplement. The attraction for dieters is that they are absolved of all responsibility for food choice – the regime is very clearly prescribed, significant mistakes are not possible and weight loss must occur if one sticks to the regime. For people who comply with these regimes, they must produce substantial and rapid weight loss. Some of the disadvantages of VLEDs are listed below.

- Individuals get no experience of eating a real diet that is capable of maintaining any short-term

weight losses induced by this ‘controlled starvation’ – they are likely to regain any weight as soon as control of their eating is returned to them. It has been argued that this could be viewed as an advantage of these products because it minimizes contact with food and breaks bad eating habits.

- The rapid weight loss leads to a high loss of lean tissue. It is difficult to measure this change in body composition accurately and some advocates of these products claim that this problem has been exaggerated.
- The energy deficit is so severe that **ketosis** occurs.
- Might this ‘controlled starvation’ risk precipitating an eating disorder?
- They are potentially dangerous, especially if used for prolonged periods without proper supervision.

Despite these problems, some nutritionists do consider that these products may have a role in the supervised treatment of people with moderate or severe obesity (BMI well in excess of 30) which poses a larger threat to health.

In a similar vein, there are many ‘meal replacement’ products on the market that are designed to replace one or more meals each day with a drink or snack. These products have a clearly defined energy yield and so are supposed to make it easy for dieters to control their energy intake. Often these products are nutrient enriched, but in some cases have similar energy yields to much cheaper conventional equivalents.

Use of these more extreme treatments

All drugs have potential side-effects, all surgical procedures entail some risk, and prolonged use of VLEDs may also have adverse consequences. They are relatively extreme measures that are inappropriate for people who are only moderately overweight and even more inappropriate for normal-weight people who wish to be thinner to enhance their self-esteem. The commercial motivation may persuade some unscrupulous ‘counsellors’ and even physicians to recommend them to people who are not even overweight. For people who are obese, the health and other consequences of their obesity may justify the use of these measures if there is some reasonable prospect that they are likely to be effective. This must be a matter for individual clinical judgement.

These more extreme treatments are not alternatives to diet and exercise. In most cases, their aim is to make it easier for people to comply with their weight-loss programmes. Appetite suppressants may reduce hunger during dieting, psychotherapy may help people to control their hunger, and dietary changes may maximize the satiating value of each food calorie. Despite this, dieting will inevitably involve a considerable period of discomfort and the

dieter will have to exercise a considerable degree of sustained willpower to achieve substantial weight loss. A VLED will not produce weight loss if it is supplemented by other eating, and an appetite suppressant will only produce weight loss if food intake is actually reduced. Some patients who have had their stomach capacity reduced by 90% by surgical stapling still manage to consume enough to prevent weight loss.

- Drugs can combat obesity by interfering with the digestion and absorption of food, by decreasing appetite or by increasing energy expenditure.
- The appetite suppressants fenfluramine and dexfenfluramine have been the most widely used anti-obesity drugs in Europe. Fenfluramine has been available for 30 years, but both drugs were withdrawn from sale in 1997 after new doubts were raised about their safety.
- Orlistat interferes with fat digestion by blocking pancreatic lipase and has been recently approved for use in Britain.
- Jaw wiring, gastric stapling, intestinal bypass and lipid reduction are some of the surgical treatments that have been used to treat obesity.
- Gastric stapling is currently the favoured surgical treatment for obesity. It involves reducing the stomach capacity by stapling off a large part of it and thus severely limiting the size of meal the patient can consume.
- Very Low Energy Diets (VLEDs) are liquid formulae designed to provide all of the essential nutrients but only a very low energy yield – 1.6–3.2 MJ (400–800 kcal) per day. They are used to replace all solid food or perhaps just some meals.
- These more severe measures inevitably carry some risks, which makes them suitable only as treatments of last resort for people for whom the risks of their obesity are likely to outweigh any risks of therapy.
- These treatments are adjuncts to conventional dieting rather than substitutes for it, i.e. they make it easier for people to maintain an energy deficit and only work if it is maintained.

Part 3

The nutrients

Carbohydrates

Introduction	209	Non-starch polysaccharide	221
Nature and classification of carbohydrates	210	Resistant starch	223
Dietary sources of carbohydrates	211	Does dietary fibre/NSP protect against bowel cancer and heart disease?	225
Sugars	212		
Artificial sweeteners	214	Possible mechanisms by which diet may influence the risk of bowel cancer and heart disease	228
Diet and dental health	215		
Starches	219		

INTRODUCTION

Carbohydrates have traditionally supplied the bulk of the energy in most human diets. They still provide the bulk of the energy intake for the majority of the world population whose diets are based upon cereals or starchy roots. Dietary carbohydrates are almost exclusively derived from the vegetable kingdom and the carbohydrate content of most foods of animal origin is dietetically insignificant. The only major exception is lactose or milk sugar, which provides around 40% of the energy in human milk and 30%

of the energy in cow milk. It is usually only in populations consuming large amounts meat, oily fish and dairy produce that fat challenges carbohydrates as the principal source of dietary energy. Some affluent vegetarian populations may derive a high proportion of their dietary energy from fat in the form of extracted vegetable oils and margarine.

In some Third World countries, over 75% of dietary energy is derived from carbohydrate, which is predominantly starch. In the affluent countries of Western Europe, Australasia and North America, carbohydrates are likely to provide only around 45% of dietary energy, and close to half of this is in the form

- Historically, carbohydrates have provided most of the energy in human diets.
- Carbohydrates, principally starch, still provide as much as 75% of the energy in many Third World diets.
- In the UK and many other developed countries, less than 45% of dietary energy comes from carbohydrate and close to half of this may be in the form of simple sugars.
- Almost all dietary carbohydrate is from plant foods, except for the lactose in milk.
- In populations consuming large amounts of animal foods and/or extracted vegetable oils, fat will threaten carbohydrate as the major energy source.
- Western consumers are advised to take 50–60% of their energy as carbohydrate, but, ideally, with less than 10% coming from added sugars.

of simple sugars in some instances. These affluent populations are currently being advised to increase the proportion of their energy derived from carbohydrates to 50–60% of the total and yet, at the same time, to substantially reduce their consumption of simple sugars, perhaps limiting 'extracted' sugars to no more than 10% of the total energy intake.

NATURE AND CLASSIFICATION OF CARBOHYDRATES

Carbohydrates are monomers or polymers of simple sugar units or **saccharides**. Carbohydrates may be described according to the number of saccharide units they contain, i.e. the **monosaccharides** (one), the **disaccharides** (two), the **oligosaccharides** (a few), and the **polysaccharides** (many). Although there is a range of monosaccharides present in human diets, three of them make up the bulk of the monosaccharide that is

released for absorption and utilization during the digestion of most diets, namely glucose, galactose and fructose. These three monosaccharides each contain six carbon atoms and are therefore termed **hexoses**. The two five-carbon sugars (pentoses), ribose and deoxyribose, will be released from the digestion of the nucleic acids DNA and RNA in food.

Carbohydrates must be broken down (digested) to their component monosaccharide units before they can be absorbed in the small intestine. Saliva and pancreatic juice contain an enzyme known as **alpha(α)-amylase** that breaks up digestible polysaccharides (starches) into oligosaccharides and disaccharides. The enzymes maltase and isomaltase, which are localized on the absorptive surface of the small intestine, complete starch digestion by cleaving the products of the partial digestion by α-amylase into glucose. The enzymes lactase and sucrase, which are also located on the absorptive surface of the small intestine, digest any of the disaccharides lactose and sucrose present in the diet.

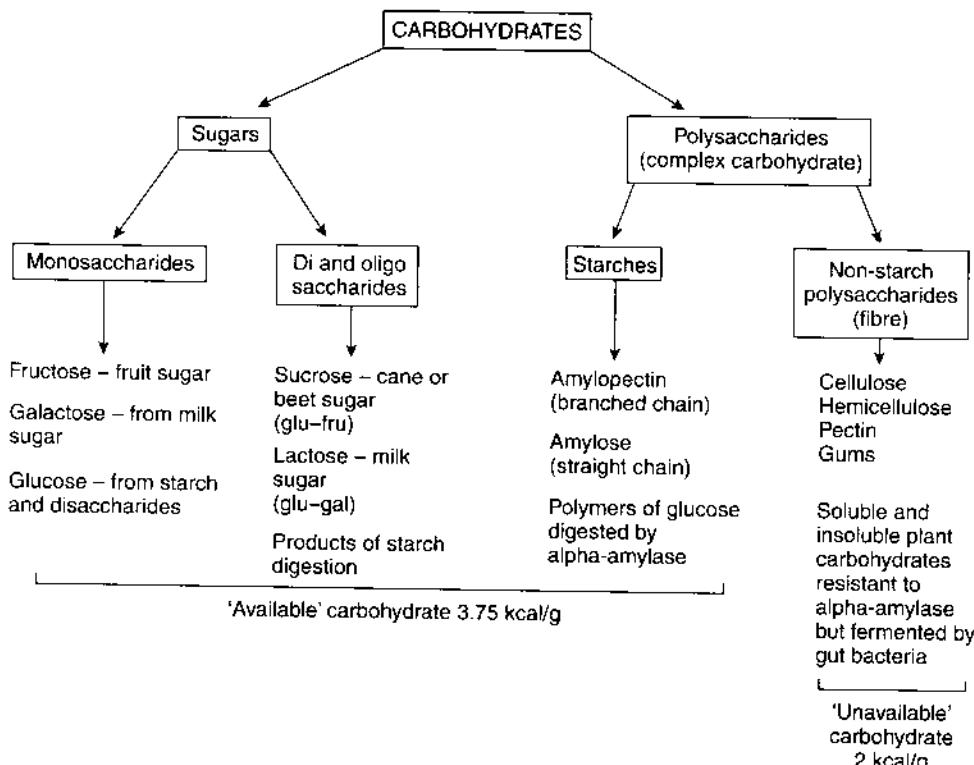


Figure 9.1 A summary of carbohydrate classification.

- Carbohydrates may be monosaccharides (one sugar units), disaccharides (two), oligosaccharides (a few) or polysaccharides (many)
- Only monosaccharides can be absorbed from the gut.
- α -amylase in saliva and pancreatic juice cleaves polysaccharides to disaccharides and oligosaccharides. Other enzymes on the surface of the small intestine complete the digestion of carbohydrates to monosaccharides.
- Sugars, starches and non-starch polysaccharides are the major subdivisions of the carbohydrates.
- Sugars and starches are digested and absorbed in the small intestine.
- Non-starch polysaccharides are indigestible and enter the colon undigested, but may be fermented by bacteria in the colon.

Dietary carbohydrates are usually classified into three major subdivisions: the sugars, the starches and the non-starch polysaccharides (NSPs), summarized in Figure 9.1. Sugars and starches are readily digested and absorbed in the human small intestine; they thus clearly serve as a source of dietary energy and are sometimes termed the available carbohydrates; they all yield around 3.75 kcal (16 kJ) per gram. The NSPs are resistant to digestion by human gut enzymes and are thus referred to as unavailable carbohydrate, although they may yield up to 2 kcal/g (8 kJ/g) if they are fermented by bacteria in the large intestine. This fermentation yields volatile fatty acids (e.g. propionate, butyrate and acetate), which can be absorbed and metabolized.

DIETARY SOURCES OF CARBOHYDRATES

According to the National Food Survey (MAFF, 1997) carbohydrates provide around 47% of the total energy in food and drink purchased for home consumption in Britain, with 21% coming from sugars and 26% from starch. Using a weighed inventory of food as eaten, The Dietary and Nutritional Survey of British Adults (Gregory *et al.*, 1990) reported that 42% of total energy (including alcohol) came from carbohydrates, with 18% from sugars and 24% from starch. According to the National Food Survey, between 1952 and 1996, the carbohydrate content of household food fell from 324 g/day to 228 g/day, i.e. from 53% of food energy in 1952 to 46% in 1996.

The figures for both years exclude alcohol, soft drinks and confectionery because these were not recorded in 1952.

The National Food Survey (MAFF, 1997) indicates that 74% of the starch in British household food comes from cereals and cereal products and 20% from vegetables (mostly from potatoes). About 4% of our starch comes from cereals added to meat and fish products, e.g. breadcrumb coatings. This survey also shows that about 46% of the NSP in British diets comes from cereal foods and 49% comes from fruit and vegetables.

Table 9.1 shows the sources of sugars in the British diet. Clearly, foods in the sugars and fats group of the **food guide plate or pyramid** are the major sources of added sugars in the British and other Western diets. Most of the sugar in the milk group and in fruits and vegetables is sugar that is naturally present in the food.

Table 9.1 *The contribution of food groups to the total sugars in household food and drink purchases in Britain*

Milk and milk products, including cheese	16
Sugar, preserves, confectionery, soft and alcoholic drinks	39
Vegetables	6
Fruit, including fruit juices	14
All cereals	18
Cakes, pastries and biscuits	10

Data source MAFF (1997).

- Carbohydrates comprise about 42% of total energy intake in Britain and around 55% of this carbohydrate is starch.
- The carbohydrate content of British diets has fallen since the 1950s, both in absolute terms and as a proportion of food energy.
- Around 90% of the starch in British diets comes from cereal foods and potatoes.
- The NSP in British diets is derived roughly equally from cereal foods and from fruits and vegetables.
- About 40% of the sugar in British diets comes from sweetened drinks, sweets, preserves and sugar itself and most of this is added sugar.
- Sweetened cereal products, including cakes and biscuits, are another major source of added sugars.

SUGARS

Sugars may be monosaccharides, disaccharides or oligosaccharides; they are characteristically soluble in water and have a sweet taste. The monosaccharides glucose and fructose are present in some fruits and vegetables and in honey. Fructose is three times as sweet as glucose and is largely responsible for the sweet taste of fruits and some vegetables. Sucrose (relative sweetness 100) is used as the standard against which the sweetness of other sugars and also of artificial sweeteners is measured. The relative sweetness of the other major dietary sugars is: lactose 30, glucose 50 and fructose 170.

Lactose or milk sugar

Lactose is a disaccharide found exclusively in milk and is the least sweet of the major dietary sugars. It is comprised of one unit of glucose and one of galactose. Lactose is the only significant source of galactose in the diet. Galactose is essential in several synthetic pathways but can be readily synthesized from glucose and so it is not an essential nutrient. On rare occasions, babies may inherit an enzyme deficiency that renders them unable to metabolize dietary galactose. If such babies are fed with breast-milk or lactose-containing infant formula, galactose accumulates in their blood (**galactosaemia**). This galactosaemia results in a range of symptoms (e.g. vomiting, weight loss and jaundice) and if untreated, it may result in death or permanent disability (e.g. cataracts, retarded physical and mental develop-

ment). Infants with this condition must be fed on a lactose-free formula and will need to continue to avoid milk and milk-containing foods after weaning.

Some babies are born deficient in the enzyme **lactase**, which cleaves lactose into its component monosaccharides in the small intestine (congenital lactase deficiency). This deficiency results in a failure to digest and absorb dietary lactose. Such infants will fail to thrive. They will suffer from diarrhoea, because of the osmotic effect of undigested lactose causing retention of water within the gut, and bloating, because of fermentation of lactose by intestinal bacteria causing gas production. These infants also need to be fed on a lactose-free formula.

In most human populations (and other mammals), lactase activity in the small intestine declines with age after about 4 years of age. This reduces the capacity of the gut to digest and absorb lactose in older children and adults (primary lactase non-persistence). This decline in lactase activity is genetically determined; it is not prevented by continued consumption of lactose and is exhibited to some degree by up to two-thirds of the world population. Lactase activity does not usually decline with age in the white populations of North America and northern Europe. People who exhibit primary lactase non-persistence are termed **lactose intolerant** and they experience symptoms of diarrhoea and bloating when challenged with a high dose of lactose. Lactose intolerance may also occur as a secondary consequence of intestinal infection, inflammatory gut diseases and parasitic infestations of the gut (secondary lactase deficiency). The high prevalence of lactose intolerance has led to suggestions that milk and milk products are inappropriate for use in nutritional

support programmes in many countries. In a recent review, Rosado (1997) confirms that many adults and older children in developing countries are indeed intolerant to the 50 g of lactose (equivalent to 1.5 L of milk) used in standard lactose tolerance tests. However, he suggests that only about 30% of these people are intolerant of the smaller amounts of lactose that are likely to be consumed, e.g. in a glass of milk. Even this 30%, who he terms lactose maldigesters, can tolerate smaller amounts of milk without symptoms, and their colon metabolism will adapt to regular milk intake to eliminate symptoms. He suggests that the problem of primary lactase non-persistence does not warrant the elimination or severe restriction of milk intake in Third World countries.

Sucrose

Sucrose is a disaccharide composed of one unit of glucose and a unit of fructose and it is digested by the enzyme sucrase, which is located on the absorptive surface of the small intestine. Sucrose is found in several fruits and vegetables and is present in large quantities in sugar beet and sugar cane, from which it is extracted and purified on a vast industrial scale. Sucrose is readily available in highly purified or partly purified forms (e.g. white sugar, brown sugars, treacle or syrup). It is also very widely used by food processors to sweeten, preserve and texture a variety of foods. The terms sucrose and sugar are often used as if they are synonymous.

High sucrose consumption has, at one time or another, been blamed for many of the illnesses that afflict affluent populations. The poor health image of sucrose was encapsulated in the phrase 'Pure, white and deadly'. Its poor health image has led many food manufacturers to use other sugars, or sugar-containing extracts, in their products so that they can imply in advertising claims that their product is 'healthier' because it has reduced levels of sucrose. The term **extrinsic non-milk sugar** has been used to cover all of the dietary sugars (still principally sucrose) that are not present as natural components of the fruit, vegetables and milk in the diet. Conversely, sugars that are naturally present within the cell walls of plants are termed **intrinsic sugars**. Note that the sugars in fruit juice are classified as part of the extrinsic non-milk sugars.

Sucrose is a natural plant sugar. It happens to be present in large enough amounts in readily cultivatable plants to make it convenient for extraction in industrial quantities. It seems reasonable to suppose that, if sucrose is indeed innately harmful, as it is used in affluent countries, then other sugars used in the same ways and quantities are also likely to be harmful. Substitution of sucrose by other extracted sugars that have not yet acquired the poor health image of sucrose does not appear to offer a high probability of dietary 'improvement'. Extrinsic non-milk sugars comprise between about 10% and 20% of the total energy intake in most affluent populations.

Sucrose and other sugars have a very high, almost addictive, sensory appeal and this might be expected to encourage overconsumption of food energy and thus to predispose to obesity. Providing sugared drinking water to rats and mice has been used as an experimental means of inducing obesity. In these experiments with rodents, increasing sugar consumption depresses the intake of solid food (and therefore also nutrient intake), but not by enough to compensate fully for the extra energy in the sugar (e.g. Kanarek and Hirsch, 1977). However, as we saw in Chapter 8, most of the epidemiological evidence suggests that those people who take in a high proportion of their energy as sugar are less likely to be obese than those who consume a lower sugar diet. The proportion of energy derived from fat and sugar tends to be inversely correlated in Western diets (the sugar-fat seesaw) and there are good reasons to expect fat to be more fattening than carbohydrates, including sugar. There is thus no convincing evidence to link high sugar intake to increased risk of obesity.

There is little convincing evidence that, at levels typical of UK and US diets, added sugars are directly implicated in the aetiology of cardiovascular diseases, diabetes or hypertension. There is convincing evidence to link some dietary fats to these conditions and the sugar-fat seesaw means that high sugar consumers will often consume less of these fats. At very high intakes, 30% or more of total calories, they may elevate serum cholesterol levels and lead to insulin resistance, which is important in the aetiology of the maturity-onset form of diabetes (COMA, 1991). In an invited review, Bruckdorfer (1992) reviews evidence against the current consensus view that sugar is a relatively benign factor in the

- Sugars have a sweet taste and are soluble in water.
- The relative sweetness of sugars and artificial sweeteners is graded in relation to sucrose (relative sweetness 100). Fructose is the sweetest dietary sugar (170) and lactose the least sweet (30).
- Lactose is a disaccharide of glucose and galactose that is found exclusively in milk.
- Babies with congenital galactosaemia are unable to metabolize galactose and those with congenital lactase deficiency are unable to digest and absorb lactose. These infants need a lactose-free formula.
- In most non-European populations, lactose activity in the gut declines after early childhood, causing them to become lactose intolerant (primary lactase non-persistence).
- Most lactose-intolerant people can tolerate moderate amounts of milk and so lactose intolerance does not warrant the elimination of milk from Third World diets.
- Lactose intolerance may also occur as a secondary consequence of gut disease.
- Sucrose is a disaccharide of glucose and fructose and is the major added sugar in the diet.
- The term non-milk extrinsic sugars covers all sugars (largely sucrose) that are not present naturally in the fruit, vegetables and milk that we eat; the term includes the sugar in fruit juice.
- Sugars contained within the cell walls of plants are termed intrinsic sugars.
- There is little evidence that high sugar intake is causally linked to diabetes, heart disease or obesity.
- The proportion of energy derived from fats and sugars tends to be inversely correlated – the sugar–fat seesaw.
- Low-fat diets, which, because of the sugar–fat seesaw, tend to be high in sugar, are associated with reduced risk of obesity.
- High sugar consumption does reduce the nutrient density of diets and this may be particularly important in groups, such as the very elderly, with marginal intakes of essential nutrients.

aetiology of obesity, cardiovascular disease and type 2 diabetes.

When extracted sugars are added to foods, they add energy but no nutrients. They reduce the nutrient density of the diet, and the phrase 'empty calories' has been coined. The significance of this reduction in nutrient density will depend upon the general level of nutrient adequacy. In populations or individuals with high nutrient intakes, it will be of less significance than in those whose nutrient intakes are more marginal. Some nutrient-rich foods may only become palatable if sugar is added, e.g. some sour fruits. In groups such as the elderly for whom high nutrient density is considered a nutritional priority (COMA, 1992), moderation of added sugar to no more than 10% of total energy should help to achieve high nutrient density. However, when appetite is poor, any reduction in dietary palatability resulting from reduction in sugar intakes may simply depress both energy and nutrient intake.

ARTIFICIAL SWEETENERS

Artificial sweeteners have been developed and manufactured to satisfy our craving for sweetness without the necessity to consume sugars (e.g. saccharin, aspartame and acesulfame K). These substances are not sugars or carbohydrates and are essentially calorie free. They are also intensely sweet; these examples are typically 200–300 times as sweet as sucrose and so very small quantities are needed to replace the sweetness normally provided by sugar. Their use is particularly prominent in drinks. Low-calorie soft drinks, containing artificial sweeteners, account for over a third of all soft drink sales in Britain (MAFF, 1997). Saccharin was the first of these artificial sweeteners and has been in use for much of the twentieth century. Saccharin is absorbed and excreted unchanged in the urine. Very large intakes have been reported to increase the incidence of blad-

der tumours in rats; despite this, it is still a permitted sweetener in most countries, including the USA and UK, because, at practical intake levels, it is not thought to represent a significant hazard. Saccharin has a bitter after-taste and is destroyed by heating, which means that it cannot be used as a substitute for sugar in cooking; this limits its usefulness to food manufacturers. Aspartame is a more recent addition to the sweetener range; it is a dipeptide made up of the amino acids aspartic acid and phenylalanine. Despite being digested to its constituent amino acids and absorbed, it has a negligible calorific yield because so little is used. In the rare inherited condition **phenylketonuria (PKU)**, there is an inability to metabolize dietary phenylalanine and so it can accumulate in the blood, causing severe brain damage and mental retardation in affected children (see Chapter 15). Obviously, PKU sufferers should avoid aspartame, but there have been some, largely unsubstantiated, suggestions that consuming large amounts of phenylalanine in the form of aspartame might also cause brain damage in non-PKU sufferers. Although these intense sweeteners mimic the sweetness of sugar, they do not have the preservative or textural effects of sugar.

These artificial sweeteners would seem to offer great potential benefits, especially to those people who are trying to lose weight. If a glass of regular cola is replaced by a 'diet' version sweetened with one of these sweeteners, this would reduce energy intake by 75 kcal (300 kJ). If a sweetener is used to replace two spoonfuls of sugar in a cup of tea or coffee, this saves 40 kcal (150 kJ) each time. Consumption of these artificial sweeteners is equivalent to billions of sugar calories in Europe and America each year and this saved energy is equivalent to several pounds of theoretically prevented weight gain for every European and American. However, the massive growth in the use of these products has not been accompanied by corresponding reductions in sugar use and it has coincided with a large rise in the prevalence of obesity.

These intense sweeteners do not provide energy, but neither do they have the satiating effect of sugar. This means that, unless users are consciously restricting their energy intake, i.e. dieting, they will tend to replace any lost sugar calories with more of their usual mixed food. The net effect is that there tends to be some replacement of carbohydrate calories (sugar) by fat calories and so the proportion of energy derived from fat tends to rise. Given the arguments

about the greater fattening effects of fat compared to carbohydrate in Chapter 8, this would not be a desirable outcome for improving weight control. Data from the USA show that there has been no significant decline in sugar use to parallel the growth in sales of intense sweeteners – are they really substituting for sugar or largely an addition to the usual diet? Of course, these sweeteners may be helpful to those who are consciously dieting but who do not want give up sweet drinks and other sweet products. They may also have benefits for dental health and nutrient density if they really reduce sugar consumption. The effects of sugar substitutes on food choice and dietary composition have been reviewed by Mela (1997).

Sugar replacers

More recently, there has been a growth in the use of bulk sweeteners or **sugar replacers** that have a similar sweetness to sucrose on a weight-for-weight basis (typically 40–90% of the sweetness of sucrose). These compounds are mostly sugar alcohols (e.g. xylitol, sorbitol and isomalt) that yield fewer calories than sugar because they are incompletely absorbed or metabolized; they typically yield 40–60% of the energy of an equivalent weight of sucrose. They can be used in similar amounts to sugar in some food products and they add not just sweetness but also the textural and mouthfeel properties of sugar. They may be used in combination with intense artificial sweeteners to boost their sweetness. These sugar replacers do not promote dental caries and, because they are only slowly and partially absorbed, they do not cause the same large rises in blood glucose and insulin that sugar does, which may be particularly useful for diabetics. However, if they are eaten in large amounts, because of their limited absorption, large amounts of carbohydrate may enter the large bowel. This can have an osmotic effect and increase bacterial fermentation, leading to diarrhoea and flatulence. Sugar replacers have been reviewed by McNutt (1998).

DIET AND DENTAL HEALTH

COMA (1989a) and Binns (1992) give referenced reviews of the role of sugars in the aetiology of

- Artificial sweeteners such as saccharin and aspartame are intensely sweet and so only need to be used in minute amounts; they are essentially calorie free.
- Aspartame is a source of phenylalanine and so must be avoided by those with phenylketonuria.
- These sweeteners do not have the preservative or textural effects of sugar and this limits their usefulness in cooking and food processing.
- Artificial sweeteners can cut total energy intake if they replace sugar in a calorie-controlled diet.
- Sweeteners do not have the satiating effect of sugar and so, in unregulated diets, any 'saved' sugar calories are likely to be replaced by calories from other foods.
- If sweeteners really replace sugar, they will have benefits for dental health and should improve the nutrient density of the diet.
- Sugar replacers are mostly sugar alcohols, like sorbitol, that have similar sweetness to sugar but yield less energy because they are incompletely absorbed and metabolized.
- Sugar replacers provide similar textural effects to sugar in food processing.
- Sugar replacers do not promote dental caries and may be useful for diabetics because their slower absorption means they cause smaller rises in blood sugar and insulin release than sugar.
- In large doses, sugar replacers spill into the large bowel and can cause flatulence and diarrhoea.

dental caries. Moynihan (1995) gives a more up-to-date review of the relationship between all aspects of diet and dental health. Most of the uncited references used in this section are listed in these reviews.

The notion that sugar is implicated in the aetiology of dental caries (tooth decay) stretches back to the Ancient Greeks. The evidence that sugars, and sucrose in particular, have some effect on the development of dental caries is overwhelming; only the relative importance of sugar as compared to other factors, such as fluoride intake and individual susceptibility, is really disputed. There is a strong correlation across populations between average sugar consumption and the number of **decayed, missing and filled (DMF)** teeth in children. In populations with low sugar consumption (less than 10% of energy intake), dental caries are uncommon. Observations with island populations and other isolated communities have shown a strong correlation between changes in the sugar supply and changes in the rates of tooth decay. In Britain during World War II, rates of tooth decay fell when sugar was rationed. Low rates of tooth decay have been found in diabetic children living in institutions whose sugar intake would have been kept very low as part of their diabetic management. Studies of other groups of people with habitually low sugar intakes have also found low levels of dental caries, e.g. children living in child-

ren's homes, Seventh-Day Adventists and children of dentists. Conversely, groups with high habitual sugar intakes have high rates of caries, e.g. sugar cane cutters and those working in factories producing confectionery.

In a study conducted more than 40 years ago in Swedish mental hospitals, the Vipeholm study, groups of patients were exposed not only to differing levels of sugar intake but also to differences in the form and frequency of that sugar intake. Not only was the amount of sugar important in determining rates of tooth decay, but form and frequency had an even greater impact:

- sugar taken between meals was found to be the most cariogenic
- frequent consumption of small amounts of sugar was more harmful than the same amount consumed in large, infrequent doses
- the most cariogenic foods were those in which the sugar was in a sticky form that adhered to the teeth, e.g. toffees.

Studies with animals have also found that:

- the presence of sugar in the mouth is required for caries to occur
- all common dietary sugars are cariogenic to some extent

- increased frequency and duration of eating sugar-containing foods increases the risk of caries production.

The key event in the development of dental caries is the production of acid by the bacteria *Streptococcus mutans* in dental plaque – plaque being a sticky mixture of food residue, bacteria and bacterial polysaccharides that adheres to teeth. These bacteria produce acid by fermenting carbohydrates, particularly sugars, in foods. When the mouth becomes acidic enough (i.e. the pH drops below 5.2), this causes demineralization of the teeth. The demineralization of teeth induced by the acid can lead to holes in the hard outer enamel layer of the tooth, exposing the softer under layers of dentine and thus enabling decay to proceed rapidly throughout the tooth. This is not a one-way process: between periods of demineralization, there will be phases of remineralization or repair. It is the balance between these two processes that will determine susceptibility to decay.

The key role of oral bacteria in the development of caries is highlighted by experiments with germ-free animals. Rats born by caesarean section into a sterile environment and maintained germ-free throughout their lives do not develop dental caries, even when their diets are high in sucrose. Tooth decay could be considered a bacterial disease that is promoted by dietary sugar. Sugars are ideal substrates for the acid-producing bacteria of the plaque. When sugars are consumed as part of a meal, other constituents of the food can act as buffers that 'soak up' the acid and limit the effects on mouth pH of any acid produced. High saliva production also reduces acid attack by buffering and diluting the acid. The more frequently sugar-containing snacks or drinks are consumed, and the longer that sugar remains in the mouth, the longer is the period during which the overall pH of the mouth is below the level at which demineralization occurs. In the UK, there is a target for reducing extrinsic non-milk sugars to around half of current intakes. Reducing between meal sugar in drinks and sugary snacks is regarded as the priority in the prevention of dental caries. Prolonged sucking of sugary drinks through a teat by infants and young children is particularly damaging to teeth, and manufacturers of such drinks recommend that they be consumed from a cup. Some foods contain

factors that protect against caries, particularly milk and cheese. Despite containing large amounts of lactose, milk is regarded as non-cariogenic. Chocolate may also contain factors that protect against caries, but this is counterbalanced by its high sugar content.

In a review of the role of sugars in caries induction from the sugar industry viewpoint, Binns (1992) emphasized the importance of other factors in determining caries frequency. Binns gave evidence of a substantial reduction in caries frequency in industrialized countries between 1967 and 1983 (Table 9.2). During the period covered by Table 9.2, sugar consumption did not change substantially and Binns suggests that consumption of sugary snacks between meals may even have increased. Moynihan (1995) summarizes more recent UK data which suggest that, between 1983 and 1993, the percentage of 15-year old children who were free of caries rose from 7% to 37% and the average number of decayed, missing and filled teeth fell from 6 to 2.5. There have been similar improvements in the dental health of adults: between 1968 and 1988, the proportion of Britons aged 65–74 years who had no natural teeth dropped from three-quarters to less than half.

The decline in the prevalence of caries in children in Britain and in other industrialized countries may be attributed in large part to the effect of increased fluoride intake. In some places, water supplies have been fluoridated, some parents give children fluoride supplements and, most significantly, almost all toothpaste is now fluoridated. The evidence that fluoride has a protective effect against

Table 9.2 Changes in the average number of decayed, missing and filled (DMF) teeth in 12-year-old children over the period 1967 to 1983

Country	1967	1983
Australia	4.8	2.8
Denmark	6.3	4.7
Finland	7.2	4.1
Netherlands	7.5	3.9
New Zealand	9.0	3.3
Norway	10.1	4.4
Sweden	4.8	3.4
UK	4.7	3.0
USA	3.8	2.6

Data source Binns (1992).

the development of dental caries is very strong. There is an inverse epidemiological association between the fluoride content of drinking water and rates of DMF teeth in the children. These epidemiological comparisons suggest that an optimal level of fluoride in drinking water is around 1 part per million (1 mg/L). At this level, there seems to be good protection against tooth decay without the risk of the mottling of teeth that is increasingly associated with levels much above this. Intervention studies in which the water supply has been fluoridated (see example in Chapter 3) and studies using fluoridated milk and salt have also confirmed the protective effects of fluoride.

The major effect of fluoride in preventing tooth decay is thought to be by its incorporation into the mineral matter of the tooth enamel; incorporation of fluoride into the calcium/phosphate compound

hydroxyapatite to give **fluorapatite** renders the tooth enamel more resistant to demineralization. This means that fluoride is most effective in preventing dental caries when administered to young children (under 5s) during tooth formation. Fluoride in relatively high doses also inhibits bacterial acid production and bacterial growth; it may concentrate in plaque to produce these effects even at normal intake levels (Binns, 1992). Fluoride in the mouth assists in the remineralization process: the more soluble hydroxyapatite dissolves when the mouth is acid and is replaced by the more acid-resistant fluorapatite. Several expert committees have made specific recommendations to ensure adequate intakes of fluoride, especially in young children (e.g. COMA, 1991; NRC, 1989b). COMA (1991) give a 'safe intake' of 0.05 mg/kg body weight for fluoride in infants only; NRC (1989a) give a range of safe intakes for all age groups.

- Cross-cultural comparisons, time trends and observations upon groups with high or low sugar intakes all suggest that high sugar intake promotes dental caries.
- Sugar is most damaging to teeth when it is taken frequently, between meals and in a form that adheres to the teeth.
- Bacteria in the mouth ferment carbohydrates in food, especially sugars, to produce acid and this acidity leads to demineralization of teeth making them prone to decay.
- Germ-free animals do not get dental caries.
- When sugar is taken as part of a meal, other components of the meal may buffer any acid produced. Saliva also buffers and dilutes mouth acid.
- Some foods contain factors that protect against dental caries, including milk, cheese and chocolate.
- A reduction in the consumption of added sugar with a particular emphasis on reducing sugary snacks and drinks between meals is recommended to improve dental health.
- Infants should not be allowed to suck sugary drinks slowly through a teat.
- Rates of dental caries in children have fallen in recent years in many countries and this has been largely attributed to higher fluoride intakes, particularly the use of fluoridated toothpaste.
- In areas with a natural fluoride content in drinking water of 1 mg/L, rates of child dental disease are low and there are no indications of any adverse effects.
- Fluoride is incorporated into the tooth enamel in young children and makes teeth more resistant to acid demineralization. It may also protect the teeth in other ways.
- Expert committees in both the USA and UK have recommended that adequate intakes of fluoride should be ensured, especially in infants and young children.
- Some areas have artificially fluoridated their water up to a level of 1 mg/L, but there has been considerable resistance to widespread fluoridation, largely because of unsubstantiated concerns about the long-term safety of this level of fluoride intake.
- In adults, gum disease is the most frequent cause of tooth loss.
- Gum disease is also promoted by sugar, but improved dental hygiene may be the most effective way of reducing tooth loss by this route.

Despite the very strong evidence for a beneficial effect of fluoride upon dental health, there has been limited fluoridation of water supplies in the UK, in contrast to the USA, where water fluoridation is widespread. Only around 12% of England has optimally fluoridated water. This is because of strong opposition from some pressure groups. Opposition to fluoridation has been partly on the grounds that the addition of fluoride to the public water supply amounts to mass medication without individual consent. There have also been persistent concerns expressed about the long-term safety of this level of fluoride consumption; it is very difficult, for example, to prove beyond any doubt that there is no increased cancer risk associated with lifetime consumption. The consensus is that fluoride in drinking water at optimal levels for dental health is not associated with any increase in risk of cancer or other diseases. The difference between the therapeutic levels proposed and the dose associated with early signs of fluoride overload (mottling of the teeth) is also relatively small (twofold to threefold). There may also be a relatively small number of people who are fluoride sensitive.

Dental caries is not the only cause of tooth loss. In adults, gum disease (periodontal disease) is the most frequent cause of tooth loss. **Gingivitis** is an inflammation of the gums caused by the presence of plaque. It can lead to receding of the gums, which exposes the roots of the teeth to decay and eventually may lead to loosening and loss of teeth. As with caries, sugar is a major causative factor for gum disease, but Moynihan (1995) suggests that practically attainable reductions in sugar consumption are unlikely to have much impact upon this condition. Improved oral hygiene may be the most effective way of controlling this condition – regular brushing, flossing and attention from a professional dental hygienist. There is some speculation about whether the antioxidant vitamins, e.g. vitamins C and E, might protect against gum disease. Certainly, deficiencies of vitamins A, C, E and folic acid have detrimental effects upon gum health.

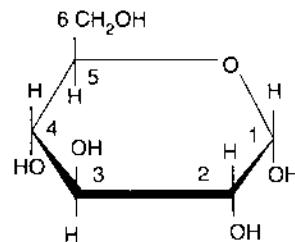
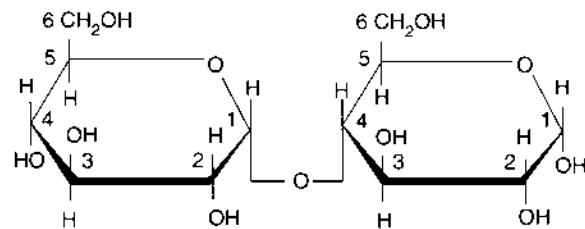
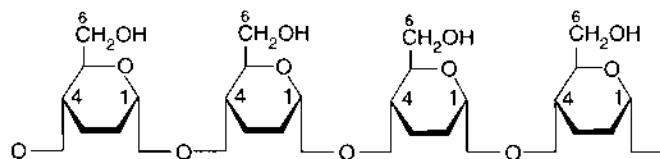
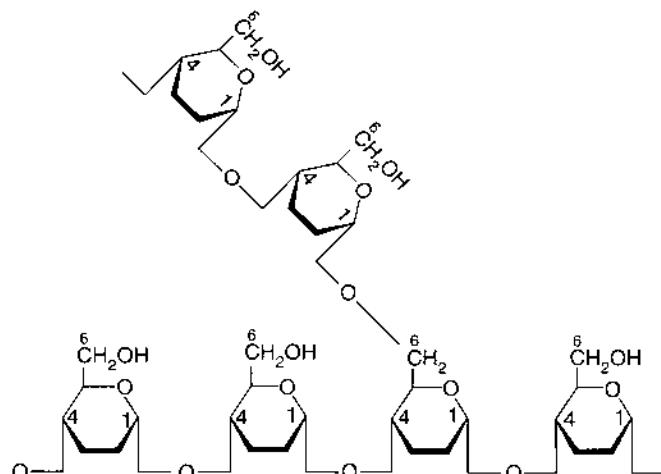
STARCHES

Plants use starches as a store of energy in their roots and seeds. Animals also manufacture and store a

form of starch, **glycogen**, in their muscles and livers, but the amounts present in most flesh foods are dietetically insignificant. The amount of energy that animals and people store as glycogen is very small. Total glycogen stores of people amount to only a few hours' energy expenditure. Glycogen is an inefficient bulk energy store because it yields only around 4 kcal/g (16 kJ/g) and because it holds with it around four times its own weight of water. Fat, on the other hand, yields 9 kcal/g (37 kJ/g) and does not hold water. Glycogen serves as a store of carbohydrate that can maintain blood glucose concentration during normal intermeal intervals.

Starches are polymers of glucose; the glucose may be arranged in predominantly straight chains (**amylose**) or, more frequently, in highly branched chains (**amylopectin**). The glucose residues in starches are joined together by linkages that can be broken (hydrolysed) by human digestive enzymes. These linkages are mainly α 1–4 linkages, with some α 1–6 linkages that cause the branching of the glucose chain; amylose has few α 1–6 linkages, whereas amylopectin has many and the latter's extensively branched structure has been likened to a bunch of grapes (Figure 9.2). The dimers and trimers of glucose that are produced as a result of digestion by α -amylase are then finally cleaved by enzymes on the mucosal surface of the small intestine to yield glucose. The enzyme maltase breaks α 1–4 linkages in these dimers and trimers and the enzyme isomaltase breaks α 1–6 linkages.

Starches provide approximately the same energy yield as sugars, i.e. 3.75 kcal/g (16 kJ/g). They have low solubility in water and they do not have a sweet taste. As starches are bland, starchy foods also tend to be bland. This is one reason why affluent populations take the opportunity to select foods with more sensory appeal and tend to reduce the proportion of their dietary energy that is supplied by starch and replace it with fats and sugars. If both sugar and fat intakes are to be reduced in line with the recommendations discussed in Chapter 4, this must inevitably involve a large increase in starch consumption, perhaps a doubling of current intakes. Something like 90% of the starch in the British diet comes from cereals and potatoes and so any major increase in starch consumption must mean eating more potatoes, bread and other cereals.

Glucose**Maltose**Glucose joined by α 1-4 link**Amylose**Long chains of glucose residues joined by α 1-4 links**Amylopectin**A branch point in amylopectin due to an α 1-6 link**Figure 9.2** Some carbohydrate structures.

- Animals store a form of starch (glycogen) in their livers and muscles, but the starch content of meat is insignificant.
- Glycogen is an inefficient energy store for animals because it has a lower energy yield than fat and also holds large amounts of water.
- Starches are polymers of glucose.
- Amylose is starch in which the glucose residues are joined in long, straight chains by α 1–4 links.
- In amylopectin there is considerable branching of the glucose chain caused by the presence of α 1–6 links.
- The enzyme α -amylase in saliva and pancreatic juice cleaves starches to dimers and trimers of glucose by hydrolysing α 1–4 links.
- The enzymes maltase and isomaltase, located on the surface of the small intestine, complete the digestion of starch to glucose.
- Starches are insoluble in water and are bland, so affluent populations tend to replace much of the starch that predominates in peasant diets with fats and sugars.
- Reducing fat and sugar intakes to recommended levels will involve large increases in starch consumption, which means eating more potatoes, bread and other cereals.

NON-STARCH POLYSACCHARIDE

Halliday and Ashwell (1991a) have written a concise, referenced review of this topic; many of the uncited references used for this discussion are listed in it. NSPs are structural components of plant cell walls (cellulose and hemicellulose) and viscous soluble substances found in cell sap (pectin and gums); they are resistant to digestion by human gut enzymes. Animal-derived foods can be regarded for dietetic purposes as containing no NSP. NSP contains numerous different sugar residues, but it passes through the small intestine and into the large bowel undigested. The term non-starch polysaccharide is used interchangeably with the term **dietary fibre**, although the latter term would also include **lignin** and **resistant starch** (see later). Lignin is an insoluble, chemically complex material that is not carbohydrate; it is found in wood and is the woody component of cell walls. Dietary lignin comes from mature root vegetables, wheat bran and the seeds of fruits such as strawberries. More details of the chemical components of NSP can be found in Englyst and Kingman (1993) and Groff *et al.* (1995).

Americans and Britons are currently being advised to substantially increase their consumption of NSP or fibre – current recommendations suggest increases over current intakes of 50–100%. NSP and

starches are usually found within the same foods, particularly in cereals and starchy vegetables such as potatoes. This means that increasing NSP and starch intakes are likely to go hand in hand. The milling of cereals so that all or part of the indigestible outer layers of the cereal grains (**bran**) is removed substantially reduces the NSP and micronutrient content of the end product, but without removing significant amounts of starch. Whole-grain cereals, therefore, contain substantially more NSP than refined cereals. Some examples of NSP content as grams/100 g are:

Wholemeal bread	5.8
White bread	1.5
Brown rice	0.8
White rice	0.2.

Current average intakes of NSP in the UK are around 12 g/day, with a wide individual variation (5–25 g/day). NSP is frequently categorized into two major fractions: the soluble substances that form gums and gels when mixed with water (the gums and pectins plus some hemicellulose), and the water-insoluble fraction (cellulose and some hemicellulose). The NSP in the UK diet is made up of similar proportions of **soluble** and **insoluble** NSP. The insoluble forms of NSP predominate in the bran of wheat, rice and maize, but oats, rye and barley contain higher proportions of soluble NSP. The

Table 9.3 The soluble non-starch polysaccharide (NSP) content of some common foods

Food	Total NSP (g/100 g)	Percentage soluble NSP
Wholemeal bread	5.8	17
Brown rice	0.8	12
Oatmeal (raw)	7.0	60
'All-bran'	24.5	17
Cornflakes	0.9	44
Potatoes	1.2	58
Apples	1.6	38
Oranges	2.1	66
Bananas	1.1	64
Carrots	2.5	56
Peas	2.9	28
Sprouts	4.8	52
Cabbage	3.2	50
Baked beans	3.5	60
Roasted peanuts	6.2	31

Data source Halliday and Ashwell (1991a).

proportions of soluble and insoluble NSP in fruit and vegetables vary, but they generally contain a substantially higher proportion of soluble NSP than wheat fibre, the principal cereal NSP in the British diet. Table 9.3 shows the total NSP content of a number of common foods and the proportion of the NSP that is soluble.

Although NSP is resistant to digestion by human gut enzymes and passes through the small intestine largely intact, in the large intestine, most of the components of NSP are fermented to variable extents by gut bacteria. The pectins are very readily fermentable, but cellulose remains largely unfermented in the human colon and, in general, soluble NSP is more fermented than insoluble NSP. NSP thus acts as a substrate for bacterial growth within the large intestine; the by-products of this fermentation are intestinal gases and short-chain fatty acids (acetate, butyrate and propionate). The short-chain fatty acids thus produced can be absorbed from the gut and can serve as a human energy source, so that the typical mix of NSP in British and American diets may contribute as much as 2 kcal/g (8 kJ/g) to dietary energy.

(Note that in ruminating animals like sheep and cattle, the short-chain fatty acids produced by the fermentation of starches and NSP by microbes in the rumen are the animal's principal energy source. Even

in pigs, whose gastrointestinal physiology is much more like that of humans, such fermentation in the large intestine, and the resulting short-chain fatty acids, can make up a substantial proportion of the total dietary energy.)

Increasing levels of NSP in the diet can be convincingly shown to have a number of physiological effects in the short term, and these are listed below.

1. High NSP content of foods slows down the rate of glucose absorption from the gut and thus reduces the postprandial peak concentrations of glucose and insulin in blood. This effect is attributed to the soluble components of NSP and it is thought to be partly due to the mechanical effect of the viscous NSP components in reducing the contact of dissolved glucose with the absorptive surface of the intestinal epithelium. Soluble NSP may also slow down the mixing of gut contents with digestive enzymes and so slow the digestive process. Guar gum (a soluble NSP that forms a viscous solution when mixed with water) has been shown to reduce the outflow of dissolved glucose even in *in vitro* experiments using a dialysis bag made from synthetic materials. The NSP can only produce this effect if taken with glucose or digestible carbohydrate – it cannot work if taken a few hours before or after the ingested glucose. Increased intakes of dietary fibre have been shown to improve glucose tolerance; diabetics, in particular, may benefit from this improvement. Dietary guidelines for diabetics now include a recommendation to ensure good intakes of high-NSP foods.
2. Increased intakes of NSP increase the volume of stools produced and reduce the **transit time** (i.e. the time it takes for an ingested marker to pass out in the faeces). The fermentable components of NSP lead to increased proliferation of intestinal bacteria and thus increase the bacterial mass in the stools. The short-chain fatty acids produced during fermentation also tend to acidify the colonic contents and act as substrates for the intestinal cells, and this may increase gut motility. The unfermented components of NSP hold water and thus they also increase the stool bulk and make the stools softer and easier to pass. The increased bulk and softness of stools, together with more direct effects on gut motility, produced by the products of fermentation combine to accelerate the passage of faeces through the colon.

These effects of fibre/NSP on the gut are often presented as being only relatively recently recognized, but they do, in fact, have a very long history. The following are quotes from a booklet published by the Kellogg Company in 1934:

It is imperative that proper bulk be included in the foods that are served. Otherwise, the system fails to eliminate regularly. Common constipation, with its attendant evils, sets in.

This bulk is the cellulose or fibre content in certain natural foods.

Within the body, it (bran) absorbs moisture and forms a soft mass which gently helps to clear the intestines of wastes.

As long ago as the 1880s, Dr Thomas Allinson was championing the cause of wholemeal bread. He suggested many benefits of wholemeal over white bread (i.e. of bran) that are widely accepted today, e.g. the prevention of constipation and of minor bowel disorders such haemorrhoids (piles).

3. NSP and some of the substances associated with it (e.g. phytate) tend to bind or chelate minerals, and probably also vitamins, and thus may hinder their absorption. In compensation, foods rich in NSP tend also to have higher levels of micronutrients than refined foods. If the NSP is fermented in the colon, this releases bound minerals. At intakes of NSP likely to be consumed by most Britons or Americans, this is thought unlikely to pose a significant threat to micronutrient supply.
4. In recent years, there has been considerable interest in the possible effects that increased intakes of soluble NSP may have in reducing the serum cholesterol concentration. A number of controlled studies have indicated that the consumption of relatively large quantities of oats (which are rich in soluble NSP) can have a significant cholesterol-lowering effect in subjects with either elevated or normal cholesterol levels. The most widely promulgated explanation for this effect is that the soluble NSP binds or chelates both dietary cholesterol and bile acids (produced from cholesterol) and this reduces their absorption or re-absorption in the small intestine, and thus increases the rate of faecal cholesterol loss. Most of the considerable quantity of bile acids and cholesterol secreted in the bile each day is reabsorbed and recycled (see entero-hepatic recycling of bile

acids in Chapter 11); high NSP intake may interfere with this process.

5. High-fibre diets tend to be high in starch and low in fat and thus the energy density of such diets tends to be low. There is a widespread belief that such dietary bulking may contribute to maintaining energy balance and thus to reducing overweight and obesity. Blundell and Burley (1987) reviewed evidence that NSP may also have a satiating effect and thus may be a useful aid to regulating energy balance in those prone to overweight and obesity (another effect of bran proposed by Allinson in the 1880s). Some possible reasons why high-NSP foods could be more satiating are:

They reduce eating speed because they require increased chewing (this also stimulates saliva production and may be beneficial to dental health).

Soluble NSP slows gastric emptying and may contribute to feelings of fullness.

NSP slows down the absorption of glucose and other nutrients and reduces the insulin response to absorbed nutrients. A large insulin response to absorbed nutrients may cause the blood glucose concentration to fall below resting levels (the rebound effect), triggering feelings of hunger.

RESISTANT STARCH

Many starchy foods, such as bread, cornflakes and boiled potatoes, contain starch that is resistant to digestion by α -amylase *in vitro*. If resistance to α -amylase digestion is used in the chemical determination of NSP in food, this **resistant starch** would be classified as part of the NSP. Studies on patients with ileostomies (i.e. whose small intestine has been surgically modified so that it drains externally into a bag rather than into the large intestine) show that this resistant starch also resists digestion by gut enzymes *in vivo*. Like NSP, resistant starch enters the large intestine undigested. Resistant starch acts as a substrate for bacteria in the large bowel and this affects bowel function. There is some debate about whether resistant starch should be regarded as part of the 'dietary fibre', with similar health benefits to NSP.

- Non-starch polysaccharides (NSPs) are carbohydrates such as cellulose, hemicellulose, pectin and some gums that are all resistant to digestion by α -amylase.
- NSP passes into the large bowel undigested.
- The term dietary fibre includes not only NSP but also lignin, a complex, woody material that is not carbohydrate, and starch that resists digestion by α -amylase.
- UK and US consumers should increase their fibre/NSP intakes by 50–100%.
- Most of the NSP in cereal grains is in the bran layer and so milling cereals to remove the bran removes much of their NSP.
- Whole-grain cereals contain much more fibre/NSP than refined cereals.
- Some components of NSP are soluble and form gels when mixed with water, whereas other components are insoluble.
- Insoluble NSP predominates in wheat, rice and maize bran, but other cereals (e.g. oats) and most fruits and vegetables contain higher proportions of soluble NSP.
- In the bowel, most components of NSP are fermented to some extent by bacteria to yield short-chain fatty acids.
- These short-chain fatty acids can be absorbed and may yield significant amounts of metabolizable energy.
- Soluble NSP slows glucose absorption and reduces the postprandial peaks of glucose and insulin.
- NSP increases the volume of stools and reduces transit time. Stools have an increased bacterial mass and higher water content.
- NSP may bind vitamins and minerals and hinder their absorption. However, many high-NSP foods are relatively rich in micronutrients and any bound nutrients will be released if the NSP is fermented.
- Soluble fibre has a lowering effect upon plasma cholesterol. One proposed mechanism is that it hinders the recycling of bile acids and so increases faecal losses of cholesterol.
- High-fibre foods may help in the regulation of energy intake because:
 - they reduce eating speed
 - NSP slows gastric emptying
 - NSP slows glucose absorption and may help to prevent rebound hypoglycaemia caused by a large insulin response to a rapid absorption of glucose.

Starch that is within starch granules is in a partial crystalline form that is insoluble and relatively resistant to α -amylase digestion. During cooking (heating with moisture), this crystalline starch is disrupted (gelatinized) and this renders it much more digestible. During cooling, the starch starts to recrystallize (retrogradation), which reduces its digestibility. There are essentially three reasons why starch may be resistant to digestion in the gut:

1. It may be inaccessible to digestive enzymes because it is enclosed within unbroken cell walls or in partly milled grains or seeds.

2. Some forms of raw crystalline starch, e.g. in raw potatoes and green bananas, are resistant to α -amylase digestion.
3. Retrograded starch in food that has been cooked and cooled is indigestible.

This resistant starch is usually only a small component of total starch in any food. However, it may make a considerable contribution to the apparent NSP content of some foods. For example, resistant starch would represent around 30% of the apparent NSP in white bread, 33% in boiled potatoes and 80% in boiled white rice. The amount of resistant starch in any given food may vary quite a lot,

depending upon slight differences in preparation. In freshly cooked potatoes, around 3–5% of the starch is not digested in the small intestine, but this rises to 12% when potatoes are cooked and cooled before being eaten; in greenish, under-ripe bananas, as much as 70% of the starch reaches the large intestine. It has been estimated that the average intake of resistant starch in Europe is around 4 g/day. If it were considered desirable, it would be possible to alter the resistant starch content of processed food very considerably by manipulating the choice of raw materials and the processing conditions.

Resistant starch is just one of a number of factors that complicate the measurement of dietary fibre intakes of a population. Englyst and Kingman (1993) give values for the estimates of fibre intake in several foods, which vary by as much 100%, depending upon the analytical methods used to estimate it. This clearly represents a major hindrance to epidemiological investigations of the effects of NSP/fibre on health. For example, several oriental populations with rice-based diets have lower fibre intakes than the British because of the very low NSP content of white rice. These oriental diets are, however, often high in starch, and the inclusion of resistant starch with NSP would significantly change the fibre content of these diets. The Englyst method specifically measures the NSP content of food. Starch is completely removed by enzyme hydrolysis and then the amount of sugars released by acid hydrolysis of the food is determined and used as the measure of NSP content. Resistant starch has been reviewed by Asp *et al.* (1996).

DOES DIETARY FIBRE/NSP PROTECT AGAINST BOWEL CANCER AND HEART DISEASE?

The recent preoccupation with the protective effects of dietary fibre against bowel problems such as constipation, diverticular disease, appendicitis, haemorrhoids and perhaps cancer of the colon and rectum can be traced back to the work of Burkitt, Trowell, Painter and others in the 1960s and 1970s. These workers noted the rarity of bowel diseases in rural African populations consuming high-fibre diets and their much greater frequency in populations consuming the 'Western' low-fibre diet. Burkitt (1971) suggested that dietary fibre might protect against bowel cancer by diluting potential carcinogens in faeces and speeding their elimination so that the intestine is exposed to them for less time.

As a general rule, populations with high rates of coronary heart disease also tend to have a high incidence of bowel cancer. This means that much of the evidence that implicates any dietary factor in the aetiology of coronary heart disease could also be used to implicate that factor in the aetiology of bowel cancer. This link between bowel cancer and coronary heart disease is not inevitable. There are populations, such as British Asians, that have high rates of coronary heart disease and low rates of bowel cancer. The modern Japanese population has a relatively high rate of bowel cancer but still has a relatively low rate of coronary heart disease.

- Some starch resists digestion by α -amylase and enters the large bowel undigested – the resistant starch.
- Resistant starch is fermented in the large bowel and behaves like a fermentable component of NSP.
- Some starch resists digestion because it is inaccessible to gut enzymes, e.g. it is within cell walls.
- Raw, crystalline starch is resistant to α -amylase, e.g. the starch in raw potatoes.
- When crystalline starch is cooked, it gelatinizes and becomes digestible but, during cooling, some of it re-crystallizes (retrogradation) and becomes indigestible again.
- The way in which foods are processed can have a significant influence upon the resistant starch content, e.g. cooling boiled rice and potatoes greatly increases their resistant starch content.
- White rice has very little NSP but does contain significant amounts of resistant starch. This could make a significant difference to the functional NSP content of the diets of several Oriental populations that are based upon white rice.

Bingham (1996) has reviewed the role of NSP and other dietary factors in the aetiology of bowel cancer, and Halliday and Ashwell (1991a) have written a general review of NSP; any uncited references may be found in these reviews.

Descriptive epidemiology

In the UK, the bowel is the second most common site for cancer in both men (after lung cancer) and women (after breast cancer). Age-standardized rates for bowel cancer are even higher in the USA than in the UK. There are very large international variations in the incidence of bowel cancer, with generally high rates in the USA, Australasia and northern Europe, but low rates in, for example, India, China and rural Africa. Bingham (1996) suggests that age-standardized rates of bowel cancer may vary between countries by as much as 15-fold.

The large international variations in bowel cancer could be the result of genetic differences among populations or could be due to dietary and other environmental differences amongst them. Migration studies, together with time trends for populations such as the Japanese, clearly indicate the predominance of environmental factors in the aetiology of bowel cancer, which also means that bowel cancer is a potentially preventable disease. Rates of bowel cancer have increased amongst migrants and their descendants who have moved from low to high bowel cancer areas, e.g. Japanese migrants to Hawaii in the late nineteenth and early twentieth centuries. Over the last 40 years, there have been marked changes in the diets of people living in Japan and these have been accompanied by major increases in the incidence of colon cancer. In 1960, age-standardized rates of bowel cancer in Japan were about a third of those in the UK, whereas they are now similar.

Naturally high-fibre diets tend to be low in fat and low-fibre diets tend to be high in fat. This means that much of the evidence linking low-fibre diets to increased risk of bowel cancer could equally well be used to implicate high-fat diets. This is another example of the recurring problem of confounding variables that plagues the epidemiologist. Cross-cultural comparisons indicate that populations with diets high in meat and fat but low in NSP, starch and vegetables are at high risk of bowel cancer. There is thus a general trend for populations with high fibre

intakes to have low rates of bowel cancer, but the negative relationship between fibre intake and incidence of bowel cancer is very sharply reduced when this relationship is corrected for differences in fat and meat consumption. Vegetarian groups, such as the Seventh Day Adventists in the USA, have lower rates of bowel cancer than other Americans and they also have higher intakes of dietary NSP. Once again, low meat and fat intakes or even high intakes of antioxidant vitamins in fruits and vegetables make it impossible to finally attribute these low cancer rates to a direct protective effect of dietary NSP.

According to Bingham (1996), the cross-cultural data suggest that high meat and fat intakes are prerequisites for a population to have a high incidence of bowel cancer, but that in such populations high intakes of NSP seem to exert a protective effect. High starch and vegetable intakes may also be protective; high vegetable intake would mean a high intake of the antioxidant vitamins (e.g. β -carotene, vitamin E and vitamin C) and high starch intake probably also means high intake of resistant starch.

As already noted, there has been a three-fold rise in the incidence of bowel cancer in Japan since 1960. During this period, there has been no significant change in overall NSP intakes, but major increases in the consumption of meat, dairy produce and wheat and a large decrease in rice consumption. Even though fibre intakes have not changed over this period, there has been a very substantial decline in the proportion of calories derived from starch, and most of the functional dietary fibre in white rice is resistant starch. Measures of the fibre intake that have taken no account of residual starch may thus be misleading. The reduction in the amount of carbohydrate in the Japanese diet over this period has been offset by increased fat consumption – the consumption of animal fat has more than doubled over this period.

Case-control and cohort studies

According to Bingham (1990), a large majority of case-control studies indicate that patients with bowel cancer ate less vegetable fibre than controls. According to Willett *et al.* (1990), most case-control studies also indicate a significant positive association between total fat intake and bowel cancer. Bingham (1996) concluded that these studies have indicated trends similar to those found in cross-cultural studies, i.e.

that individuals with high intakes of fat and meat but low intakes of fibre and vegetables are at increased risk of bowel cancer. The limitations of case-control studies have been discussed in Chapter 3. As Bingham (1996) points out, diseases of the large bowel produce pain and altered bowel function and so one of the effects of bowel disease may be to make patients change their diets in order to avoid the symptoms. The low fibre intakes of the patients in these case-control studies might be a consequence of the disease rather than an indication of its cause. Note that many studies do attempt to address this problem by asking patients about their past diets, but other studies suggest that recall of past diet is strongly influenced by current diet.

Willett *et al.* (1990) have reported results from a 6-year follow-up using a cohort of almost 90 000 American nurses. They found no significant association between total fibre intake and risk of developing colon cancer over the 6-year period. They also found no significant associations when they used subdivisions of the fibre from different food sources, i.e. vegetable, fruit or cereal fibre (although there was a non-significant trend with fruit fibre). They did find significant positive associations between total fat intake and colon cancer risk and an even stronger association with animal fat intake (illustrated in Chapter 3, Figure 3.10); 'red meat' (beef, pork and lamb) consumption was also positively associated with bowel cancer risk. They confirmed the expected negative association between fibre intake and animal fat intake in their population; this may explain the apparent contradiction between these results and the fibre results from case-control studies. Their results provide rather weak support for the proposition that low intakes of fruit fibre might contribute to the risk of colon cancer, but stronger evidence in favour of the proposition that high animal fat intake is linked to an increased risk of colon cancer. It should, nonetheless, be borne in mind that, despite the apparently very large sample size, the analysis is based upon only 150 total cases and only 103 cases when those developing cancer within the first 2 years of follow-up were excluded.

In an overall assessment of all of the cohort studies that had published up to 1996, Bingham (1996) concluded that there is only weak evidence for a slightly increased risk for high meat and fat consumers and for a slightly decreased risk in those with high vegetable and NSP intakes. All of the cohort studies published prior to 1996 had major drawbacks, such as:

- poor dietary assessment methodology
- sample populations that had too narrow a range of dietary habits.

Bingham suggests that good additional data from large and well-designed cohort studies using large populations with very diverse dietary habits will not become available until well into the new millennium.

Assessment of the evidence

The overall conclusion from the current evidence is that there is strong support for the proposition that a diet higher in fibre/NSP-containing foods (cereals, vegetables and fruit) will reduce the incidence of bowel cancer in countries such as the UK and USA. Such dietary changes are consistent with current nutrition education guidelines. The resulting diets would not only have more fibre than current diets, but would also have less fat, less animal fat, less meat protein, more starch and more antioxidant vitamins. The evidence supporting a direct protective effect of dietary fibre, or any individual component of the NSPs, is much weaker than the evidence supporting this more generalized change. The current evidence may be interpreted in several ways, such as those listed below.

- High dietary NSP/fibre intake directly protects against bowel cancer.
- High dietary fat/animal fat promotes bowel cancer and any apparent protective effect of a high intake of complex carbohydrates is an artefact due to the inverse relationship between fat and complex carbohydrate intakes.
- NSP protects against the cancer-promoting effects of high dietary fat (or meat protein) in the bowel.
- High fruit and vegetable intakes protect against bowel cancer because of their high content of antioxidants, such as vitamin E, vitamin C and β -carotene.
- High intake of meat protein increases the risk of bowel cancer.

It is also quite possible that several of these statements are true and that several dietary factors promote and help prevent bowel cancer.

As with the link between NSP and bowel cancer, the evidence supporting the beneficial effects of dietary changes that would increase NSP intake upon the risk of coronary heart disease is very much

- High-fibre diets have long been thought to protect against bowel cancer as well as to alleviate other bowel problems such as constipation, diverticular disease and haemorrhoids.
- The bowel is the second most common site for cancer in the UK for men and women.
- High rates of bowel cancer and high rates of coronary heart disease tend to go together, although there are exceptions, such as British Asians and the present-day Japanese population.
- Internationally, rates of bowel cancer vary by 15-fold, with high rates in the USA, Australasia and northern Europe, but low rates in India, China and rural Africa.
- Migration studies and recent changes in bowel cancer frequency in Japan show that environmental factors are largely responsible for the variations in bowel cancer incidence.
- Vegetarian groups and populations with diets low in meat and fat but high in NSP and vegetables have low rates of bowel cancer.
- At both the individual and population levels, intakes of fat and NSP tend to be inversely correlated.
- It is difficult to differentiate the contributions of high meat, high fat, low NSP and low vegetable intakes to the aetiology of bowel cancer.
- Bingham (1996) suggested that high meat and fat intakes are prerequisites for high population rates of bowel cancer and that NSP is protective in these circumstances.
- Most case-control studies indicate higher fat and/or lower NSP intakes in bowel cancer cases. The diets of disease sufferers may reflect their disease rather than indicate its cause.
- Cohort studies provide weak evidence for a slightly increased bowel cancer risk in high fat and meat consumers and a slightly decreased risk in those with high vegetable and NSP intakes.
- Most cohort studies so far conducted have used subjects with a narrow range of dietary habits and/or have used inadequate means for assessing dietary intake.
- Evidence that naturally high-fibre diets are associated with low bowel cancer risk is much stronger than evidence for a specific protective effect of NSP.
- Naturally high-fibre diets tend to be low in meat and fat but high in vegetables and antioxidants.

stronger than the evidence for a direct link between NSP intake and coronary heart disease. These conclusions mean that the consumption of more foods naturally rich in complex carbohydrates has a much higher probability of reducing the risk of these diseases than adding extracted fibre to foods or the use of foods in which the fibre content has been greatly 'amplified' by food manufacturers.

POSSIBLE MECHANISMS BY WHICH DIET MAY INFLUENCE THE RISK OF BOWEL CANCER AND HEART DISEASE

Human faeces contain substances that induce mutation in bacteria (mutagens) and such mutagens are known to be likely carcinogens in animals and people. Some of these mutagens are :

- bile acid breakdown products

- nitrosamines from cured meats and beer
- heterocyclic amines present in cooked meat and fish – more of these are produced if the food is fried, grilled or barbecued than if it is stewed or microwaved.

High-fat and/or high-meat diets may increase the mutagenicity of faeces. A high-fat diet increases the production of bile acids and these may be converted to mutagenic derivatives within the gut. High consumption of meat protein increases the intake of carcinogenic heterocyclic amines and increases the production of nitrosamines and other nitrogen-containing carcinogens within the large bowel.

The following mechanisms have been used to explain how high intakes of NSP and resistant starch could help protect against bowel cancer.

- NSP and resistant starch increase the weight and water content of stools and thus would tend to dilute any potential carcinogens in stools.

- They decrease transit time and thus reduce the time that the gut cells are exposed to these faecal carcinogens.
- The short-chain fatty acids produced during the fermentation of NSP and resistant starch in the colon act as substrates for the epithelial cells of the gut. Butyrate, in particular, may inhibit the proliferation of mutated cells into cancerous tumours. The acid environment they produce in the bowel may reduce the conversion of bile acids to their mutagenic derivatives.

In controlled experimental studies in which animals are exposed to known chemical carcinogens, bran has been consistently reported to reduce the number of tumours. Bran appears to protect against these chemically initiated cancers. In other experimental studies, bran fibre has been reported to reduce the mutagenicity of faeces

produced by human subjects (see Bingham, 1990, 1996).

Some of the possible mechanisms by which NSP might have a specific influence on coronary heart disease risk are listed below.

- Soluble NSP may help to lower the plasma cholesterol concentration, e.g. by interfering with the entero-hepatic recycling of cholesterol and bile acids.
- A high intake of NSP, particularly soluble NSP, improves glucose tolerance and reduces insulin secretion (diabetes greatly increases the risk of coronary heart disease).
- High NSP intake may act indirectly to reduce coronary heart disease by reducing the risk of obesity, e.g. because it reduces the energy density of the diet and increases satiety by slowing gastric emptying and absorption of the products of digestion.

- Human faeces contain carcinogenic substances, including bile acid derivatives, nitrosamines from cured meats and beer, and heterocyclic amines from cooked meat and fish.
- High meat and fat consumption increases the carcinogenicity of faeces.
- Fat increases bile acid production.
- NSP could protect against bowel cancer by causing the dilution and more rapid elimination of faecal carcinogens.
- Increased production of butyrate from the fermentation of NSP and resistant starch may inhibit the proliferation of mutated cells into tumours and reduce the conversion of bile acids to their carcinogenic derivatives.
- Naturally high-fibre diets (low in fat and saturated fat) are also associated with reduced rates of coronary heart disease.
- NSP might reduce coronary heart disease risk by:
 - reducing the re-absorption of bile acids and lowering plasma cholesterol
 - improving glucose tolerance
 - indirectly, by helping to prevent excessive weight gain.
- Choosing a diet that is high in natural fibre/NSP is more likely to reduce the risk of bowel cancer or coronary heart disease than artificially enhancing the fibre content of existing diets.

Protein and amino acids

Traditional scientific aspects of protein nutrition

230

The significance of protein in human nutrition

237

TRADITIONAL SCIENTIFIC ASPECTS OF PROTEIN NUTRITION

Introduction

Experiments stretching back over the last two centuries have demonstrated the principle that all animals require an exogenous supply of protein. In humans, and in other non-ruminating mammals,¹ this protein must be present in the food that they eat. In the early 1800s, Francois Magendie fed dogs with single foods that were regarded as highly nutritious but lacked any nitrogen (e.g. sugar, olive oil or butter). These unfortunate dogs survived much longer than if they had been completely starved, but they lost weight, their muscles wasted dramatically and they all died within 40 days. Several modern writers, including myself, have assumed that Magendie completed the cycle of evidence by showing that the dogs recovered when nitrogen-containing foods were added back into their diets. In his excellent book on the history of protein nutrition, Carpenter (1994) points out that Magendie did not actually do these positive controls. Despite this, he is still usually credited with the discovery that nitrogen-

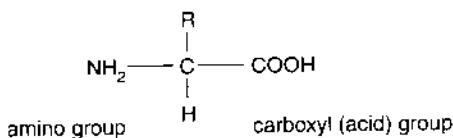
containing foods are essential in the diet. In later years, it became clear that Magendie's initial conclusions were, in fact, correct, i.e. that adequate amounts of nitrogen-containing foods are essential to allow growth and even survival in experimental animals. Such experiments demonstrated the need for nitrogen-containing foods (i.e. protein) even before the chemical nature of dietary protein was known.

Chemistry and digestion

Protein is the only major nitrogen-containing component of the diet. It is composed of long chains of 20 different **amino acids** that are linked together by **peptide bonds**. Figure 10.1 illustrates the chemical nature of amino acids and proteins. The amino acid side chain (the R group in Figure 10.1) is the group that varies between amino acids, and some examples of the R groups in particular amino acids are shown in Figure 10.1. The way in which amino acids are linked together by peptide bonds is also illustrated in Figure 10.1. The nitrogen atom in the **amino group** of one amino acid is linked to the carbon atom of the

- Protein is an essential nutrient for all animals.
- Even before protein was chemically characterized, it was clear that animals need foods that contain nitrogen.

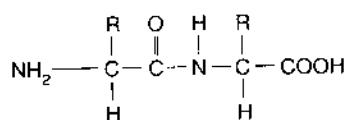
Generalized chemical formula of amino acids



Examples of amino acid side chains

$R = H$	amino acid = glycine
$R = CH_3$	amino acid = alanine
$R = CH_2SH$	amino acid = cysteine
$R = (CH_2)_4NH_2$	amino acid = lysine
$R = CH_3OH$	amino acid = serine

Two amino acids linked by a peptide bond



Diagrammatic representation of a protein

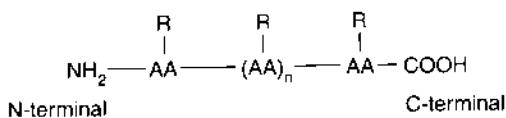


Figure 10.1 The chemical nature of proteins and amino acids.

carboxyl group in the adjacent amino acid. A protein may be comprised of hundreds of amino acids linked together in this way; each protein will have a free amino group at one end (the N-terminal) and a free carboxyl group at the opposite end (the C-terminal).

Dietary proteins are digested to their constituent amino acids prior to absorption in the small intestine. Significant amounts of dipeptides and other small peptide fragments may also be absorbed. (It is also clear that small amounts of intact protein can be

absorbed, e.g. some bacterial toxins and Immunoglobulin G (IgG) antibodies from milk in many newborn mammals.) Peptidase enzymes in gastric and pancreatic juice break (hydrolyse) the peptide bonds between specific amino acids in dietary proteins, yielding small peptides, dipeptides and free amino acids. Other peptidase enzymes located on the mucosal surface of the small intestine continue the digestion of these small protein fragments. Some peptidases are classified as endopeptidases because they hydrolyse the bonds between certain amino acids within the protein or peptide molecule. The major endopeptidases in the gut are:

- pepsin in gastric juice, which breaks the bonds between most amino acids
 - trypsin in pancreatic juice, which breaks the bonds on the carboxyl side of lysine and arginine residues within proteins
 - chymotrypsin in pancreatic juice, which breaks the bonds adjacent to tyrosine, phenylalanine and tryptophan residues as well as some others
 - collagenase (which breaks down collagen) and elastase in pancreatic juice.

The peptidases below break off one of the amino acids at the end of the peptide chain, and these are termed exopeptidases:

- carboxypeptidases in pancreatic juice break off the C-terminal amino acid
 - aminopeptidases located on the surface of the small intestine break off the N-terminal amino acid.

Also located on the surface of the small intestine are dipeptidases and tripeptidases, which break down dipeptides and tripeptides. Dipeptidases and tripeptidases within the cells complete the digestion of any dipeptides and tripeptides that are absorbed intact into the intestinal cells.

- Protein is made up of 20 different amino acids joined together in long chains by peptide bonds.
 - Protein is largely digested to its constituent amino acids in the gut by peptidase enzymes, although some peptide fragments and even some intact protein are absorbed.
 - Endopeptidases, such as pepsin, trypsin and chymotrypsin in gastric and pancreatic juice, break proteins into peptide fragments by breaking peptide bonds within the protein chain.
 - Carboxypeptidases (cleave the C-terminal amino acid), aminopeptidases (cleave the N-terminal amino acid) and a range of dipeptidases and tripeptidases complete the digestion of dietary protein.

Nitrogen balance

ESTIMATION OF PROTEIN CONTENT

Early in the nineteenth century, the German chemist Liebig noted that proteins from different sources all contained approximately 16% by weight of nitrogen. This observation remains the basis of estimating the protein content of foods. The nitrogen content is determined by a relatively simple chemical analysis (the Kjeldahl method²) and then multiplied by 6.25 (i.e. 100/16) and this gives an estimate of the protein content:

$$\text{Weight of protein} = \text{weight of nitrogen} \times 6.25$$

This figure of 6.25 is an approximation, and more precise values may be used for some foods. Some foods may contain non-protein nitrogen, but in most cases this is largely amino acid and so would not produce any dietetic error. The Kjeldahl method can also be used to estimate the nitrogen content of urine and, if this is multiplied by 6.25, it indicates how much protein has been broken down to produce this amount of nitrogen. Sweat and faeces can also be analysed for nitrogen to give a complete picture of protein losses by catabolism or direct protein loss in faeces.

The concept of nitrogen balance

Intakes and losses of nitrogen are used as the measures of protein intake and protein breakdown, respectively, and the difference between intake and total losses is the **nitrogen balance**:

$$\text{nitrogen balance} = \text{nitrogen intake} - \text{nitrogen losses}$$

The nitrogen input is from food protein. Nitrogen losses arise from the breakdown of dietary and body protein. Most nitrogen is lost in the urine,

but small amounts are also lost in faeces and via the skin.

NEGATIVE NITROGEN BALANCE

A negative balance indicates that body protein is being depleted because more nitrogen is being lost than is taken in in the diet. Injury, illness, starvation (including dieting) or inadequate protein intake *per se* may lead to negative nitrogen balance. The minimum amount of protein, in an otherwise complete diet, that enables an adult to achieve balance indicates the minimum requirement for protein. A deficit in energy supplies would be expected to lead to negative nitrogen balance for the two reasons listed below:

1. A diet that has insufficient energy will also tend to have a low protein content.
2. During energy deficit, the protein that is present will tend to be used as a source of energy rather than being used for protein synthesis.

REQUIREMENTS FOR BALANCE

The protein requirement of an adult is the minimum intake that results in nitrogen balance. When adults are put onto a protein-free, but otherwise complete, diet, their rate of nitrogen loss drops during the first few days as they adapt to this situation. Even after adaptation, they excrete a low and relatively constant amount of nitrogen – this is termed the **obligatory nitrogen loss** and represents the minimum replacement requirement. Studies using radioactively labelled amino acids indicate that around 300 g of protein is normally broken down and re-synthesized in the body each day – the **protein turnover**. Most of the amino acids released during the breakdown of body protein can re-enter the body amino acid pool and be re-cycled into new protein synthesis, but a proportion (approximately 10%) cannot be re-cycled and become the obligatory nitrogen loss.

- All proteins contain about 16% nitrogen by weight.
- The protein content of food can be estimated from its nitrogen content, and the rate of protein catabolism can be estimated from the nitrogen loss in urine.
- Nitrogen can be estimated using the Kjeldahl method.

Several amino acids have functions additional to that of protein building and are irreversibly lost from the body amino acid pool, e.g. for the synthesis of non-protein substances such as pigments or transmitters, to maintain acid-base balance, and for conjugation with substances to facilitate their excretion. When amino acids are lost from the pool for such reasons, some other amino acids have to be broken down to retain the balance of the body's pool of amino acids. Some protein will be also lost as hair, skin, sloughed epithelial cells and mucus etc. Such obligatory losses explain why adults have a requirement for dietary protein even though their total body protein content remains constant.

POSITIVE NITROGEN BALANCE

A positive nitrogen balance indicates a net accumulation of body protein. Healthy adults do not go into sustained positive nitrogen balance if dietary protein intake is stepped up, but utilize the excess protein as an energy source and excrete the associated nitrogen as urea. Growing children would be expected to be in positive nitrogen balance, as would pregnant women, those recovering after illness, injury or starvation, and those actively accumulating extra muscle such as body builders. One would expect these groups, particularly rapidly growing children, to have a relatively higher protein requirement than other adults, i.e. they would be expected to require more protein per kilogram of body weight. Lactating women would be expected to have a higher protein requirement because they are secreting protein in their milk. Note that the daily positive nitrogen balance in body builders is very small and would not indicate a significant increase in their protein requirements.

DIETARY ADEQUACY FOR PROTEIN

Table 10.1 shows some of the current UK Reference Nutrient Intakes (RNIs) for protein; in addition to the absolute values, the RNI is also calculated as a percentage of the Estimated Average Requirement (EAR) for energy. To put these values into context, the average UK diet has around 14–15% of the energy as protein (Gregory *et al.*, 1990), compared to the 5–10% of energy as protein indicated as the maximum requirement in Table 10.1. Note that the RNI is a generous estimate of protein requirement in those who need most, whereas the EAR is a less

Table 10.1 Some UK RNIs for protein expressed in absolute terms and as a percentage of average energy requirements

Age (years)	RNI (g/day)	Percentage of energy
1–3	14.5	4.7
7–10	28.3	5.7
<i>Male</i>		
11–14	42.1	7.7
19–50	55.5	8.7
65–74	53.3	9.2
<i>Female</i>		
11–14	41.2	8.9
19–50	45.0	9.3
Pregnancy	+ 6 ^a	9.5 10.5 ^b
Lactation	+ 11 ^a	8.9 9.4 ^b

^aAdditional to that of non pregnant or lactating women

^bVaries as the energy EAR varies during pregnancy and lactation

Data source COMA (1991)

generous estimate of average requirement. This suggests that primary protein deficiency is highly improbable in the UK and, indeed, in other industrialized countries.

To give the values in Table 10.1 an international perspective, rice has around 8% of energy as protein and hard wheat around 17%, whilst most other major cereal staples lie between these two extremes; only finger millet (*Eleusine coracana*) is below this range (7%). Some starchy roots, which are important staples, do fall substantially below this range for cereals:

cassava	3% energy as protein
plantains (cooked bananas)	4% energy as protein
yam	7% energy as protein
sweet potato	4% energy as protein

Table 10.2 shows the protein content of some common foods. Note that only three of the foods on this list have crude protein contents that are less than that in human milk and most have more than 10% of their energy as protein. A comparison of these values with those in Table 10.1 suggests that people meeting their energy needs and eating a varied diet, even a varied vegetarian diet, would seem to run little practical risk of protein deficiency. This is generally true even though fats, oils, alcoholic drinks, soft drinks and sugar-based snacks provide additional energy but little protein so they reduce the nutrient density for protein. Of course, this conclusion might be altered by variations in the quality of different proteins, and this is discussed in the next section.

Table 10.2 The protein content of some foods expressed in absolute terms and as a percentage of energy

Food	Protein (g/100 g food)	Protein (% energy)
Whole milk	3.3	20.3
Skimmed milk	3.4	41.2
Cheddar cheese	26.0	25.6
Human milk	1.3	7.5
Egg	12.3	33.5
Wholemeal bread	8.8	16.3
Sweetcorn	4.1	12.9
Cornflakes	8.6	9.3
Beef (lean)	20.3	66.0
Pork (lean)	20.7	56.3
Chicken (meat)	20.5	67.8
Cod	17.4	91.6
Herring	16.8	28.7
Shrimps	23.8	81.4
Peas	5.8	34.6
Broad beans	4.1	34.2
Baked beans (can)	5.1	31.0
Lentils	23.8	31.3
Potatoes	2.1	9.7
Mushrooms	1.8	55.4
Broccoli	3.3	57.4
Carrots	0.7	12.2
Peanuts	24.3	17.1
Peanut butter	22.4	14.5
Almonds	16.9	12.0
Apple	0.3	2.6
Banana	1.1	5.6
Orange	0.8	9.1
Dried dates	2.0	3.2

Values are generally for raw food

Protein quality

The observation that all proteins contain approximately 16% by weight of nitrogen enabled early workers not only to measure the protein contents of diets, but also to match them for crude protein content by matching them for nitrogen content. Using such nitrogen-matched diets, it soon became clear that all proteins were not of equal nutritional value, i.e. that both the amount and the quality of protein in different foods were variable. For example, pigs fed on diets based upon lentil protein grew faster than those fed barley-based diets even though the diets were matched for crude protein (nitrogen) content; lentil protein is apparently of higher nutritional quality than barley protein.

ESSENTIAL AMINO ACIDS

It is now known that quality differences among individual proteins arise because of variations in their content of certain of the amino acids. In order to synthesize body protein, an animal or person must have all 20 amino acids available. If one or more amino acids is unavailable, or in short supply, the ability to synthesize protein is generally compromised, irrespective of the supply of the others. Proteins cannot be synthesized leaving gaps or using substitutes for the deficient amino acids. Each body protein is synthesized with a genetically determined and unalterable, precise sequence of particular amino acids.

- The nitrogen balance is the difference between nitrogen intake from food and nitrogen losses.
- Positive nitrogen balance indicates a net increase in body protein (e.g. during growth) and a negative nitrogen balance indicates a net loss of body protein (e.g. in starvation).
- Adults require protein to replace that lost from the body pool each day.
- Even under conditions of maximal conservation after adaptation to a protein-free diet, there is still an obligatory nitrogen loss.
- When the RNI for protein is expressed as a percentage of the EAR for energy, values of 5–10% energy as protein is obtained for all age groups.
- Diets in industrialized countries typically contain around 15% of the energy as protein, which makes protein deficiency improbable.
- Only a few staple foods contain low enough levels of protein to make primary protein deficiency a realistic prospect and, in most Third World diets, protein is not the limiting nutrient.

About half of the amino acids have carbon skeletons that can be synthesized by people and these amino acids can therefore be synthesized by transferring amino groups from other surplus amino acids – **transamination** (see Chapter 5). The remaining amino acids have carbon skeletons that cannot be synthesized by people and they cannot, therefore, be made by transamination. These latter amino acids are termed the **essential amino acids** and have to be obtained preformed from the diet. The protein gelatine is lacking in the essential amino acid tryptophan and so dogs fed on gelatine, as their sole nitrogen source, do not fare much better than those fed on a protein-free diet. Despite a surplus of the other amino acids, the dog cannot utilize them for protein synthesis because one amino acid is unavailable. Under these circumstances, the amino acids in the gelatine will be used as an energy source and their nitrogen excreted as urea.

The essential amino acids for humans are:

- histidine (in children)
- isoleucine
- leucine
- lysine
- methionine
- phenylalanine
- threonine
- tryptophan
- valine.

The amino acids cysteine and tyrosine are classified as non-essential even though their carbon skeletons cannot be synthesized. This is because they can be made from methionine and phenylalanine, respectively.

ESTABLISHING THE ESSENTIAL AMINO ACIDS AND QUANTIFYING REQUIREMENTS

Young rats have been shown to grow normally if synthetic mixtures of the 20 amino acids are given as their sole source of dietary protein. Growth is not affected if one of the non-essential amino acids is left out of the mixture because the rats are able to make this missing amino acid by transamination. Growth ceases, however, if one of the essential amino acids is left out. In this way, the essential amino acids for the rat have been identified and the amount of each one

that is required to allow normal growth has also been determined.

Such experiments with children were ruled out on ethical grounds. However, adults have been shown to maintain nitrogen balance if fed mixtures of the 20 amino acids and, like the rats, can cope with the removal of certain amino acids and still maintain this balance; these are the non-essential amino acids. When one of the essential amino acids is left out, there is a net loss of body nitrogen, i.e. a depletion of body protein. The requirement for each of the essential amino acids has been estimated from the minimum amount that is necessary to maintain nitrogen balance when the other amino acids are present in excess. The range and relative needs of the essential amino acids in rats and people are found to be similar and the rat has been widely used as a model for humans in the biological assessments of protein quality.

LIMITING AMINO ACID

In any given dietary protein, one of the essential amino acids will be present in the lowest amount relative to human requirements, i.e. a given amount of the protein will contain a lower proportion of the requirement for this essential amino acid than for any of the others. The availability of this amino will 'limit' the extent to which the others can be utilized if that particular protein is fed alone and in amounts that do not fully supply the needs for the amino acid. This amino acid is called the **limiting amino acid**. The amino acid is limiting because, if supplies of this essential amino acid are insufficient, the use of the others for protein synthesis will be limited by its availability – protein cannot be made leaving gaps where this missing amino acid should be. For example, if only half of the requirement for the limiting amino acid is supplied, then, in effect, only about half of the protein requirement is being supplied; surpluses of the others will be used as energy sources and their nitrogen excreted. In many cereals, the limiting amino acid is lysine, in maize it is tryptophan and in beef and milk it is the sulphur-containing amino acid methionine; one of these three is the limiting amino acid in most dietary proteins. In the past, lysine supplementation of cereal-based diets or developing 'high-lysine' strains of cereals has been widely advocated, and sometimes used, as a means of improving the protein quality of diets based upon cereals.

FIRST-CLASS AND SECOND-CLASS PROTEINS

Whatever measure of protein quality is used, most proteins of animal origin have high values. Meat, fish, egg and milk proteins contain all of the essential amino acids in good quantities; they have been termed **first-class proteins** or complete proteins. Many proteins of vegetable origin, however, have low measured quality – this is true of many staple cereals and starchy roots – and they have been termed **second-class** or incomplete proteins because they have a relatively low amount of one or more of the essential amino acids. Pulses (peas and beans) are a major exception to this general rule and have good amounts of high-quality protein. Some examples of the relative quality values of common dietary proteins are shown in Table 10.3.

MUTUAL SUPPLEMENTATION OF PROTEIN

For the two reasons discussed below, the concept of a limiting amino acid producing large variations in protein quality may have limited practical relevance in human nutrition, especially in affluent countries.

1. The human requirement for protein is low. This is true both in comparison to other species and in comparison to the protein content of many dietary staples.
2. Affluent people seldom eat just single proteins; rather, they eat diets containing mixtures of several different proteins. The probability is that the different proteins consumed, over a period of time, will have differing limiting amino acids and thus that any deficiency of an essential amino acid in one protein will be compensated for by a relative surplus of that amino acid in another. Thus, the quality of the total protein in different human diets will tend to equalize, even though the nature and qual-

Table 10.3 *The relative qualities (NPU, net protein utilization) of some dietary proteins*

Protein quality (NPU %)	
Maize	36
Millet	43
Wheat	49
Rice	63
Soya	67
Cow milk	81
Egg	87
Human milk	94

ity of the individual proteins in those diets may vary very considerably. This is called **mutual supplementation of proteins** and it means that any measure of protein quality of mixed human diets tends to yield a fairly consistent value.

MEASUREMENT OF PROTEIN QUALITY

The simplest method of assessing the quality of a protein is to chemically compare its limiting amino acid content with that of a high-quality reference protein; this is called the **chemical score**. Egg protein has traditionally been used as the reference protein in human nutrition, but breast-milk protein is another possible reference protein. In the protein under test, the limiting amino acid is identified and the amount of this amino acid in a gram of test protein is expressed as a percentage of that found in a gram of egg protein:

$$\text{chemical score} = \frac{\text{mg of limiting amino acid}}{\begin{array}{l} \text{in 1 g of test protein} \\ \text{mg of this amino acid} \end{array}} \times 100$$

in 1 g of egg protein

This chemical score does not necessarily give an accurate guide to the biological availability of the amino acids in the protein. Thus, for example, if a protein is poorly digested, the chemical score could greatly overestimate the real biological quality of the protein.

Net protein utilization (NPU) is a widely used biological measure of protein quality; it is a particularly important measure in agricultural nutrition, but is nowadays considered to be of much less significance in human nutrition. To measure NPU, the protein to be tested is fed as the sole nitrogen source at a level below the minimum requirement for growth of the animal (usually weaning rats). The amount of nitrogen retained under these circumstances is corrected for the nitrogen losses that occur even on a protein-free diet and then expressed as a percentage of the intake. NPU can be directly assessed in human adults using nitrogen balance measurements.

$$\text{NPU} = \frac{\text{amount of retained nitrogen}}{\text{nitrogen intake}} \times 100$$

$$\text{retained nitrogen} = \text{intake} - (\text{loss} - \text{loss on protein-free diet})$$

The NPU values for some proteins are shown in Table 10.3. The NPU of diets, as well as those of individual proteins, may be determined and, because of

the mutual supplementation effect, the NPUs of most human diets come out within a relatively narrow range, even though the qualities of the individual constituent proteins may vary widely. An NPU of around 70 would be typical of many human diets and would not be greatly affected by, for instance, the proportion of animal protein in the diet. The NPU is only likely to be of significance in human nutrition if a single staple with low amounts of poor-quality protein makes up the bulk of the diet and is the source of most of the dietary protein.

THE SIGNIFICANCE OF PROTEIN IN HUMAN NUTRITION

Introduction

The change in attitude to protein deficiency as a suspected key cause of human malnutrition represents

one of the most striking changes in human nutrition over the last 50 years. The author has previously reviewed this topic (Webb, 1989), and the book by Carpenter (1994) contains a detailed and thorough review of both the discovery of protein and changing attitudes to protein in nutrition. The following discussion is a development of my earlier review. This topic can serve as a case study to illustrate the three important general points that have been discussed earlier in the book:

1. The potential costs of premature and ineffective 'health promotion' intervention: these costs may be very real and substantial, even if an unhelpful intervention does not do direct physiological harm (Chapter 1). The financial and time costs of all the measures to solve the illusory 'world protein crisis' were enormous and this meant that other, more real, problems were ignored or deprived of funding.

- Feeding trials using nitrogen-matched diets indicated that proteins vary in their nutritional quality.
- Variation in protein quality is due to variation in the complement of essential amino acids.
- Essential amino acids cannot be made in the body, but non-essential amino acids can be made by transamination.
- Animals can grow and adults can maintain nitrogen balance if provided with complete amino acid mixtures but not if one of the essential ones is missing from the mixture.
- The limiting amino acid is the one present in a food in the lowest amount relative to how much is needed.
- When one essential amino acid is taken in insufficient quantities, it limits the extent to which the others can be used for protein synthesis.
- Lysine is the limiting amino acid in most cereals, and lysine, tryptophan or methionine is the limiting amino acid in most foods.
- Proteins containing good amounts of all essential amino acids are termed first-class proteins, e.g. meat, fish, egg and pulse proteins.
- Proteins that are low in one or more essential amino acid are termed second-class proteins, e.g. gelatine, cereal proteins and most proteins from starchy root staples.
- When mixtures of proteins are eaten, deficits of essential amino acids in individual proteins will usually be compensated for by relative surpluses in others, i.e. mutual supplementation.
- The protein quality of human diets tends to equalize towards the mean because of mutual supplementation.
- The chemical score is a measure of protein quality and is the amount of the limiting amino acid in a gram of the protein expressed as a percentage of that found in a gram of a high-quality reference protein such as egg.
- Net protein utilization (NPU) is a biological measure of protein quality and is the percentage of the protein that growing animals can retain under conditions in which protein intake is limited.

2. The potential dangers of overestimating dietary standards by 'erring on the safe side' (Chapter 3): the 'world protein crisis' seems to have been precipitated by a massive overestimation of the protein needs of children.
3. The dangers of hasty and ill-considered extrapolation of results obtained in animal experiments to people (Chapter 3): observations of very high protein requirements in rapidly growing young animals were probably a factor in creating the illusory belief that the protein needs of children were also very high.

It is largely because it illustrates these points so well that I have given a topic dealing with an historical and primarily Third World issue such prominence in a book whose main focus is on nutrition for promoting health in industrialized countries. These issues are very current in relation to nutrition for health promotion.

Historical overview

For more than two decades (the 1950s and the 1960s), the belief that primary protein deficiency was 'the most serious and widespread dietary deficiency in the world' (Waterlow *et al.*, 1960) dominated nutrition research and education. The consensus of nutrition opinion considered it likely that 'in many parts of the world the majority of young children suffered some protein malnutrition' (Trowell, 1954). The relative protein requirements of young children compared to adults were generally considered to be very high, perhaps five times higher on a weight-for-weight basis. Even where diets contained apparently adequate amounts of total protein, it was believed that protein availability might still be inadequate unless some high-quality protein (usually animal or pulse protein) was taken with each meal. Calculations of the amount of protein required by the world population when compared with estimates

of protein availability gave the impression of a large and increasing shortfall in world protein supplies – **the protein gap**. Jones (1974) suggested that an extra 20 million tons of protein were required each year, and this figure implied that many hundreds of millions of people were protein deficient. This protein gap was so large that a whole new field of research was initiated to try to close it – namely, the mass production of protein-rich foods from novel sources such as fishmeal and microbes. An agency of the United Nations (UN), the Protein Advisory Group, was set up specifically to oversee these developments. As late as 1972, the Chairman's opening address to an international scientific conference contained the following statements:

Every doctor, nutritionist or political leader concerned with the problem of world hunger, has now concluded that the major problem is one of protein malnutrition.

The calorie supply (of developing countries) tends to be more or less satisfactory, but what is lacking is protein, and especially protein containing the essential amino acids.

(Gounelle de Pontanel, 1972).

As estimates of human protein requirements were revised downwards, this protein gap disappeared 'at the stroke of a pen', with no significant change in protein supplies. The above quotations from Gounelle de Pontanel contrast very sharply with the conclusions of Miller and Payne (1969) that almost all dietary staples contain sufficient protein to meet human needs and that even diets based upon very low protein staples are unlikely to be specifically protein deficient. In the 30 years since 1969, this latter view has become the nutritional consensus. Despite this revolutionary change of mind, the greatly reduced emphasis on protein in human nutrition was slow to permeate beyond the ranks of the specialists. Even amongst nutrition educators, some ambivalence towards protein persisted well into the 1990s. The following quotations from a 1991 edition

- Changes in the importance attached to protein illustrate the potential costs of unwarranted health promotion initiatives, the dangers of overestimating dietary standards, and the potential problems of carelessly extrapolating results obtained from laboratory animal studies to people.

of a well-respected US nutrition text seem to sum up that ambivalence.

Most people in the United States and Canada would find it almost impossible not to meet their protein requirements.

Protein is at the heart of a good diet. Menu planners build their meals around the RDA for protein.

(Hamilton *et al.*, 1991).

This exaggerated importance attached to protein in the past led to a huge concentration of research effort, physical resources, educational effort and political priority into protein nutrition and into closing the protein gap. It now seems almost certain that much of this effort was wasted and directed towards solving an illusory problem. McLaren (1974), Webb (1989) and Carpenter (1994) have discussed some of these costs of exaggerating human protein needs, and McLaren was moved to entitle his historically important article 'The great protein fiasco'. Some of the measures taken to try to solve the apparently false protein crisis are summarized below (further extensive details may be found in Carpenter, 1994).

- In 1955, the UN set up the Protein Advisory Group to 'advise on the safety and suitability of new protein-rich food preparations'. This group held regular meetings and produced various reports, until it was finally terminated in 1977.
- In 1957, UN experts suggested that there was an urgent need to develop cheap, locally produced, high-protein foods that would be safe, palatable and easy to incorporate into existing diets from

things like fishmeal, soybeans, peanuts, coconut, sesame and cotton seed. To this was added in the late 1960s the development of protein from yeasts and bacteria (single-cell protein, SCP), the production of genetically improved plants and the use of synthetic amino acids to boost the quality of dietary proteins.

- In 1970, the UN approved a programme to implement these recommendations, despite initial costs estimated at \$75 million (at 1970 prices!).
- Numerous problems delayed and increased the development costs of high-protein foods from existing raw materials such as:

the need to remove lipids from fishmeal to make it palatable and improve its storage
the presence of 'unhygienic' intestinal contents in meal made from small, whole fish
the presence of fungal toxins in peanut products
rising market prices of some of the raw materials.

Ultimately no cheap and widely used product emerged from these expensive undertakings. The poor did not eat the few products that got beyond the trial stage to full marketing.

- Many expensive projects to produce single-cell proteins were begun, but most failed to produce products that were safe and acceptable for human consumption. One fungal product (Quorn) did emerge from these efforts, but it is now marketed as a meat-substitute for affluent vegetarians rather than as a high-protein food for needy Third World children.

- In 1960, protein deficiency was widely regarded as the most serious and widespread form of malnutrition and the majority of children in many developing countries were believed to be protein deficient.
- In 1960, estimates of protein needs greatly exceeded protein supplies and indicated a large and increasing shortfall in world protein supplies – the protein gap.
- As estimates of children's protein needs decreased, so the concept of a protein gap became discredited.
- The current consensus is that few diets are specifically protein deficient and that, if children eat enough food for their energy needs, they will probably get enough protein.
- During the 'protein gap era', huge amounts of time and money were wasted on projects to solve the protein crisis, e.g. the production of high-protein foods from fishmeal, peanuts and microbes, the development of genetically enhanced plants, and programmes to supplement cereals with synthetic amino acids.

- Decades of expensive research did produce high protein quality strains of corn, but these have never been widely grown.
- Chemical production of synthetic lysine for food-supplementation was tried, but proved ineffective in improving nutritional status in human field trials.
- The total costs of these largely unsuccessful and seemingly unnecessary programmes and initiatives must amount to many hundreds of millions of dollars.

Origins of the 'protein gap'

The three assumptions listed below seem to have been critical in creating the illusion of a massive world protein shortage:

1. The belief that children required a high protein concentration in their diets, i.e. that the proportion of total dietary energy derived from protein should be substantially higher in children than in adults.
2. The assumption that the nutritional disease known as **kwashiorkor** was due to primary protein deficiency, i.e. due to a diet in which the proportion of energy derived from protein was too low.

3. The belief that kwashiorkor was the most prevalent form of malnutrition in the world and that cases of frank clinical kwashiorkor represented only the 'tip of the iceberg', with many more children impaired to some extent by lack of protein.

There are now good grounds for suggesting that each of these three assumptions is either dubious or improbable.

PAST EXAGGERATION OF PROTEIN REQUIREMENTS?

Figure 10.2 illustrates the changes in the US RDAs for protein over the past 50 years. In 1943, the protein RDA for a 2-year-old child was substantially more than double its current value (also more than double the current UK RNI). When expressed as a proportion of the energy allowance, the child's protein RNI was well above the adult value in 1943, but is now well below the adult value (see numerical values in Table 10.4).

The current values given in Table 10.4 suggest that children can manage on a lower minimum protein concentration in their diets than adults can. This conclusion seems at first sight to be inconsistent with the notion that growing children have higher relative

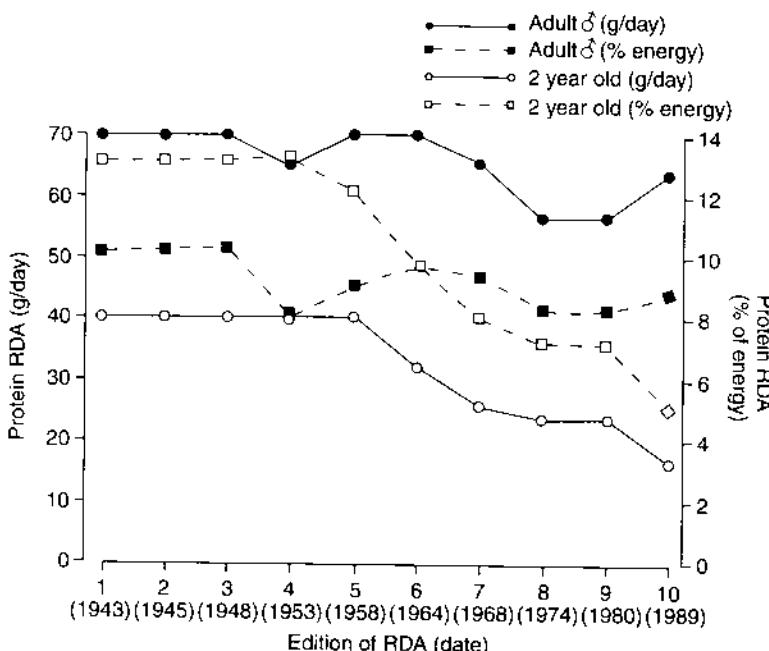


Figure 10.2 Changes in the American RDA for protein for an adult man and a 2-year-old child over ten editions (1943–1989), expressed both in absolute terms and as a percentage of the energy RDA. Source: Webb (1994).

Table 10.4 Changing US estimates of protein needs

	1943	1989
Adult man (g/day)	70	65
2 year old (g/day)	40	18
Adult man (% of energy)	10	8
2 year old (% of energy)	13	5

protein needs than adults. It has been, and still is, assumed that children need more protein on a weight-for-weight basis than adults; only the assumed scale of this difference has declined. In the past, figures of up to five times higher weight-for-weight requirements in children have been suggested (e.g. Waterlow *et al.*, 1960); nowadays, a figure of around double is considered more reasonable. Table 10.5 shows why this higher relative requirement is not inconsistent with the need for a lower dietary protein concentration, as suggested by Figure 10.2. Children require not only relatively more protein than adults, but also more energy. Table 10.5 shows the current UK RNIs for protein and the EARs for energy as a multiple of the standard adult value when expressed on a weight-for-weight basis (i.e. expressed per kilogram of body weight). A 2-year-old child requires almost three times as much energy as an adult per kilogram of body weight, but only about 1.5 times as much protein. The need for increased energy intake (i.e. total food intake) is far greater than the increased relative need for protein. For all ages of children, the increased energy needs cancels

out or, in most cases, greatly exceeds the increased need for protein. The 2-year-old child may need 1.5 times as much protein as an adult, but needs to eat three times as much food; any diet with enough protein for adults should therefore provide enough for children, if they can eat enough of it to satisfy their energy requirements.

Inappropriate extrapolation from animal experiments to humans may have encouraged inflated estimates of protein requirements in children. Primates have slower growth rates than most animals and much lower rates than most common laboratory animals (Table 10.6). The relative protein requirements of these rapidly growing species are thus likely to be higher than those of human infants and children, and this seems to be borne out by the comparison of milk composition of the species also shown in Table 10.5. Up to 80% of the nitrogen requirements of a growing rat are for growth (Altschul, 1965). Rats were widely used to model humans in protein experiments and so this probably encouraged the belief that children's relative protein requirements were up to five times those of adults. As four-fifths of the requirement of young rats is for growth and only one-fifth for maintenance, then, by direct analogy, it might well suggest a five-fold difference in the requirements of children (for growth and maintenance) compared to those of adults (for maintenance only).

Animal experiments also probably played their part in exaggerating the importance of low protein quality as a practical problem in human nutrition. When a diet containing 10% egg protein as the sole nitrogen source is fed to growing rats under test con-

Table 10.5 Weight-for-weight energy EARs and protein RNIs as a multiple of the standard adult value at various ages

Age	Energy EAR	Protein RNI
0-3 months	2.7	2.9
4-6 months	2.7	2.2
7-9 months	2.7	2.1
10-12 months	2.8	2.0
1-3 years	2.9	1.5
4-6 years	2.8	1.5
7-10 years	2.1	1.3
11-14 years	1.5	1.3
15-18 years	1.3	1.2
19-50 years	1.0	1.0
75+ years	0.9	1.0

Calculated from data in COMA (1991).

Table 10.6 Growth rates and milk composition of seven species

Species	Days to double birth weight	Protein in milk (g/dL)	Energy as protein (%)
Human	120-180	1.1	6
Calf	47	3.2	19
Sheep	10	6.5	24
Cat	7	10.6	27
Rat	6	8.1	25
Mouse	5	9.0	21
Chimp	100	1.2	7

Simplified after Webb (1989).

ditions, 1 g of egg protein will bring about an increase in weight of 3.8 g in the animals. Egg protein is said to have a protein efficiency ratio of 3.8. Around 80% of the consumed protein is retained as body protein. Wheat protein has a protein efficiency ratio of only 1 and just 20% of the wheat protein would be retained as body protein (Altschul, 1965). Under the same test conditions, 1 g of wheat protein results in only 1 g of weight gain in the rats. Such experimental demonstrations of the low quality of many cereal and vegetable proteins would have suggested that, unless some high-quality protein was present, protein deficiency would have seemed almost inevitable – 4 g of wheat protein is only equivalent to 1 g of egg protein in growing rats. Of course, most human diets are mixtures of proteins, and mutual supplementation can occur. It was for some time thought, however, that this mutual supplementation could only occur within a meal, but it is now thought that such supplementation can occur over an extended time scale. If a meal lacking in one amino acid is consumed, the liver will donate the deficient amino acid to allow protein synthesis to continue and will replenish its complement of this amino acid at a later meal. This makes it rather unlikely that protein deficiency will occur due to unbalanced amino acid composition, except under very particular conditions.

LOW CONCENTRATION OF DIETARY PROTEIN CAUSES KWASHIORKOR?

The nutritional disease kwashiorkor was first described in the medical literature in the 1930s and a very tentative suggestion was made that it might be due to dietary protein deficiency. Over the next 35 years, it became generally accepted that two diseases known as kwashiorkor and marasmus represented

different clinical manifestations of malnutrition in children, with differing aetiologies. Marasmus was attributed to simple energy deficit (starvation), whereas kwashiorkor was attributed to primary protein deficiency. Children with marasmus are severely wasted and stunted and the clinical picture is generally consistent with what might be expected in simple starvation. Several of the clinical features of kwashiorkor, however, are not superficially consistent with simple starvation, but are consistent with protein deficiency (Table 10.7). Many children suffer from a mixture of these two diseases, and the two conditions were thus said to represent opposite extremes of a spectrum of diseases, with the precise clinical picture depending on the precise levels of protein and energy deficit. The general term **protein-energy malnutrition (PEM)** was used to describe this spectrum of diseases.

A number of studies over the last 25 years have seriously challenged this assumption that kwashiorkor is necessarily due to primary protein deficiency. When the protein needs of children were thought to be high, the protein deficit in the diets of many Third World children seemed to be much greater than any energy deficit. When the estimated protein needs of children were reduced, protein no longer appeared to be the limiting factor in these diets, and energy deficits exceeded any protein deficits. Comparisons of the diets of children developing kwashiorkor or marasmus have reported no difference in their protein concentrations. Detailed analyses of the diets of children in the Third World have indicated that primary deficiency of protein is uncommon and that protein deficiency is usually a secondary consequence of low total food intake (Waterlow and Payne, 1975).

The symptoms of kwashiorkor listed in Table 10.7 are nonetheless persuasively consistent with protein

Table 10.7 Some clinical features of kwashiorkor that are consistent with primary protein deficiency

Oedema – excess tissue fluid	Lack of plasma proteins?
Hepatomegaly – liver swollen with fat	Lack of protein to transport fat?
'Flaky paint syndrome' – depigmented and peeling skin	Lack of amino acids for skin protein and pigment production?
Mental changes – anorexia and apathy	Lack of amino acids for transmitter synthesis?
Subcutaneous fat still present	Inconsistent with simple energy deficit?
Changes in colour and texture of hair	Similar causes to skin changes?

deficiency rather than energy deficit. Acute shortage of energy would, however, lead to the use of protein as an energy source. Thus, dietary energy deficit could well precipitate a physiological deficit of amino acids for protein synthesis and to fulfil other amino acid functions. Symptoms of protein deficiency might be triggered by energy deficit. Why starvation should sometimes result in marasmus and sometimes in kwashiorkor is, as yet, still unclear. Some more recent theories have suggested that the symptoms of kwashiorkor may be triggered in an undernourished child by:

- mineral or vitamin deficiencies
- essential fatty acid deficiencies
- infection
- inadequate antioxidant response (e.g. due to lack of antioxidant vitamins and minerals) to oxid-

ative stress (e.g. exacerbated by infection, toxins, sunlight).

KWASHIORKOR – THE MOST COMMON NUTRITIONAL DISEASE IN THE WORLD?

McClaren (1974) suggested that the widespread assumption that kwashiorkor was the most prevalent manifestation of PEM was caused by faulty extrapolation from a World Health Organization (WHO) survey of rural Africa in 1952. He argued that marasmus was a much more widespread problem than kwashiorkor and that marasmus was becoming increasingly common as breastfeeding declined (marasmus is associated with early weaning). In 1974, even one of the authors of the 1952 WHO report (J.F. Brock) conceded that there had been a relative neglect of marasmus, which was very prevalent even in rural Africa (see Carpenter, 1994).

- The 'protein gap' originated because protein needs were exaggerated and because kwashiorkor was assumed to be both the most prevalent form of protein-energy malnutrition and the result of a primary deficiency of protein.
- The estimated protein needs of children are now less than half what they were in the 1940s.
- Children were thought to need a much higher proportion of their energy as protein than adults do, but this is no longer believed to be true.
- Children were thought to need five times more protein per kilogram of body weight than adults do, but today's figure is less than twice as much.
- Young children require three times more energy per kilogram than adults and so, if they get enough energy, any diet with enough protein for an adult should provide enough for a child.
- Most non-primate animals grow much faster than human babies and so require much more protein, and this is reflected in much higher protein content in their milk.
- Around four-fifths of a young rat's protein requirement is for growth and this may indicate the origin of the belief that human babies require five times more protein per kilogram than adults.
- Laboratory experiments with rats suggested that many cereal and vegetable proteins were inferior to animal proteins in supporting growth, and this probably exaggerated the importance attached to protein quality in human nutrition.
- Many of the symptoms of kwashiorkor are consistent with primary protein deficiency (see Table 10.7), yet there was never any hard evidence to support the belief that kwashiorkor is due to primary protein deficiency.
- Current estimates of protein and energy needs suggest that protein is rarely the limiting nutrient in Third World diets.
- Protein deficiency is usually a secondary consequence of inadequate food intake.
- It is now generally accepted that marasmus, not kwashiorkor, is the most prevalent form of protein-energy malnutrition.

Concluding remarks

Over the last 50 years, opinions as to the significance of protein in human nutrition have undergone some dramatic changes. For a time, protein adequacy was top of the list of international nutritional priorities. Nowadays, primary protein deficiency is generally considered to be a relatively uncommon occurrence, not because of an improved supply, but rather because of a change of nutritional opinion. In the current editions of both the UK and US dietary standards (COMA, 1991; NRC, 1989a), there is even concern voiced about the possible health risks of high protein intakes. It has been suggested, for example, that prolonged excess protein intakes may predispose to renal failure in later life. Both of the panels suggest that it would be prudent to regard a value of twice the RNI/RDA as the safe upper limit for protein intake.

ENDNOTES

1. Note that some ruminants can survive without dietary protein if their diet contains alternative sources of nitrogen. This is because bacteria in the rumen can make protein from non-protein nitrogen sources such as urea and the animals can digest and absorb this bacterial protein.
2. In the Kjeldhal method, the food is boiled with concentrated sulphuric acid and a catalyst, which converts the nitrogen into ammonium sulphate. Alkali is then added to the digest, liberated ammonia is distilled into acid and the amount of ammonia liberated is determined by titration.

Fat

Nature of dietary fat	245	Transport of endogenously produced lipids	258
Types of fatty acids	246	The diet-heart hypothesis	260
Sources of fat in the diet	250	Current 'health images' of different dietary fats	261
Roles of fat in the diet	251	The key tenets of the diet-heart hypothesis	265
Blood lipoproteins	255	Review of the evidence for the diet-heart hypothesis	267
Digestion, absorption and transport of dietary lipids	257	Fish oils	272

NATURE OF DIETARY FAT

Around 95% of dietary fats and oils are in the form of **triacylglycerol** (TAG). (Note that the alternative term **triglyceride** is still widely used.) Most of the storage fat in the body is also in the form of triacylglycerol. TAGs are composed of three **fatty acids** linked to the simple three-carbon molecule glycerol. Glycerol has one hydroxyl (alcohol) group on each of its three carbons, and each of these alcohol groups can form an ester linkage with the carboxyl (acid) group of a fatty acid. The three fatty acids in a TAG can be the same (simple TAG) or different (mixed TAG). A TAG molecule is represented diagrammatically in Figure 11.1a.

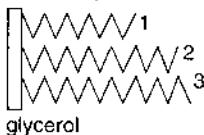
If the third (bottom) fatty acid in Figure 11.1a is replaced by phosphoric acid, this compound is known

as phosphatidic acid, the parent compound for a group of compounds known as **phospholipids**. Phospholipids make up most of the dietary lipid that is not TAG and they have important physiological functions as components of membranes and in the synthesis of a range of important regulatory molecules called the **eicosanoids** (e.g. **prostaglandins**). The phosphate group of phosphatidic acid can link with various other compounds to produce a series of these phospholipids, e.g. phosphatidyl choline (lecithin), phosphatidyl inositol and phosphatidyl serine.

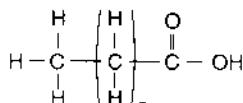
Lipids are usually defined on the basis of their solubility, i.e. those components of the body or food that are insoluble in water but soluble in organic solvents, such as chloroform or ether. Using this definition, the term lipids would encompass not only TAGs and phospholipids, but also the sterols (e.g. cholesterol) and waxes.

- Lipids are organic materials that are insoluble in water but soluble in organic solvents and include triacylglycerols (TAGs), phospholipids, sterols and waxes.
- 95% of dietary fat is TAG and most of the rest is phospholipid.
- TAG is comprised of three fatty acids attached by ester linkages to glycerol.
- If one fatty acid of TAG is replaced by a phosphate-containing moiety, this becomes a phospholipid.

a) Schematic representation of a fatty acid



b) General chemical formula of a saturated fatty acid



c) Diagrammatic representation of a saturated fatty acid (palmitic acid – 16:0)

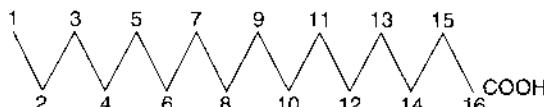
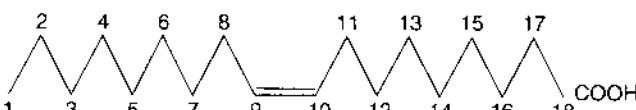
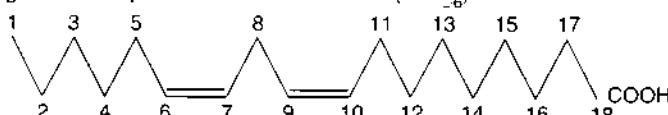
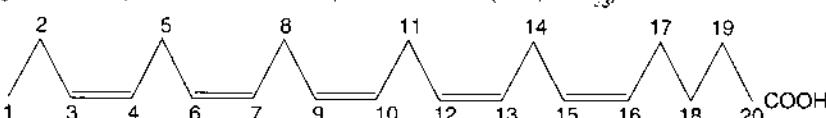
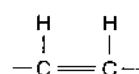
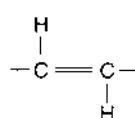
d) Diagrammatic representation of a monounsaturated fatty acid (oleic acid – 18:1₉)e) Diagrammatic representation of linoleic acid (18:2_{6,9})f) Diagrammatic representation of eicosapentaenoic acid (EPA, 20:5₃)g) Diagram to illustrate the *cis* and *trans* configurations of unsaturated fatty acids*cis* configuration*trans* configuration

Figure 11.1 The chemical nature of fats.

TYPES OF FATTY ACIDS

Fatty acids are of three major types: **saturated**, **monounsaturated** and **polyunsaturated fatty acids**. All fatty acids are comprised of a hydrocarbon chain of varying length, with a carboxyl or acid group at one end. The only difference between these three types of fatty acids is in the number of carbon–carbon double bonds present in the hydrocarbon chain.

Saturated fatty acids

Each carbon atom has the potential to form four bonds (i.e. carbon atoms have a valency of four). In saturated fatty acids, all of the carbon atoms in the hydrocarbon chain are joined together by single bonds and all of the remaining valencies (bonds) of the carbon atoms are occupied by bonding to hydrogen atoms, as shown in Figure 11.1b. These fatty acids are termed saturated because all of the available

bonds of the carbon atoms are occupied, or 'saturated', with hydrogen atoms. No more hydrogen atoms can be added to this hydrocarbon chain.

The chemical structure of fatty acids is often represented schematically as shown in Figure 11.1c. Each angle represents a carbon atom, and the carbon atoms are numbered starting with the carbon furthest away from the carboxyl or acid group. The fatty acid depicted in this diagram is called palmitic acid – it has 16 carbon atoms and no double bonds and therefore it can be written in shorthand notation as 16:0.

MONOUNSATURATED FATTY ACIDS

These are fatty acids with a single point of unsaturation, i.e. two of the carbons in the chain are joined together by a double bond and thus there are two fewer hydrogen atoms than in a saturated fatty acid of the same length. Oleic acid is depicted in Figure 11.1d; it is an 18-carbon monounsaturated fatty acid that is widely found in nature, both in animals and plants (it is the principal acid in olive oil). The double bond in oleic acid is between carbons 9 and 10 in the sequence – the shorthand notation for this acid is 18:1_{ω9} (i.e. 18 carbons, 1 double bond and omega-9; the double bond is between carbons 9 and 10). In many sources, the Greek letter omega is replaced by 'n', so that oleic acid becomes 18:1n-9.

POLYUNSATURATED FATTY ACIDS

These are fatty acids that have two or more double bonds in the carbon chain, i.e. they have more than one point of unsaturation. Figure 11.1e shows linoleic acid, which has 18 carbons, two double bonds and the first of these double bonds is between carbons 6 and 7, and thus its shorthand notation is 18:2_{ω6}. Knowing the position of the first double bond makes it possible to predict the position of any further double bonds, because they are normally separated by a single saturated carbon atom (i.e. are methylene interrupted). In linoleic acid, the first double bond is between carbons 6 and 7, carbon 8 is saturated and the second double bond lies between carbons 9 and 10. Figure 11.1f shows a slightly more complicated example: eicosapentaenoic acid (EPA) has 20 carbons, five double bonds and the first double bond is between carbons 3 and 4, hence its shorthand notation is 20:5_{ω3}. In EPA, carbon five is saturated and the second double bond is between carbons 6 and 7; carbon 8 is saturated and the third

double bond is between 9 and 10, the fourth between 12 and 13 and the fifth between 15 and 16, as shown in Figure 11.1f. Each of the double bonds in EPA is separated by a single saturated carbon atom and so, knowing the position of the first double bond (given by the 'ω' or 'n' notation), one can predict the position of all of the others.

CIS/TRANS ISOMERIZATION

Different positional isomers (as shown in Figure 11.1g) are possible wherever a fatty acid has a carbon–carbon double bond. When represented in two dimensions, either both hydrogen atoms can be on the same side, the *cis* form, or the hydrogen atoms can be on opposite sides, the *trans* form. Most naturally occurring fats contain almost exclusively *cis* isomers, although small amounts of *trans* fatty acids are found in butter (4–8%) and in other natural fats, especially those from ruminating animals. Whether a double bond is in the *cis* or *trans* configuration makes a considerable difference to the three-dimensional shape of the molecule. A *cis* double bond causes the molecule to bend back on itself into a U-shape, whereas a *trans* double bond remains in the more linear form characteristic of saturated fatty acids.

It is possible to chemically convert unsaturated fatty acids to saturated fatty acids if the fat is reacted with hydrogen in the presence of an activating agent, or catalyst. This process is called hydrogenation and it causes oils to solidify. Partial hydrogenation is often used in the production of margarine and solid vegetable shortening, i.e. **hydrogenated vegetable oils** (or hydrogenated fish oil). The process of hydrogenation can result in the production of relatively large amounts of *trans* fatty acids in the final product. Many of the early, cheap, 'hard' margarines were particularly high in *trans* fatty acids, but these are being phased out and account for only a small fraction of current margarine sales. Heating of cooking oils also produces some *trans* isomers.

Distribution of fatty acid types in dietary fat

All fat sources contain mixtures of saturated, monounsaturated and polyunsaturated fatty acids, but the proportions vary considerably from fat to fat. Table 11.1 shows the variation in proportion of these

Table 11.1 Typical proportions (%) of saturated (S) monounsaturated (M) and polyunsaturated (P) fatty acids in several dietary fats

Fat	S (%)	M (%)	P (%)	P:S ratio
Butter	64	33	3	0.05
Lard	47	44	9	0.2
Sunflower oil	14	34	53	4
Soybean oil	15	26	59	4
Rapeseed oil (canola)	7	60	33	5
Olive oil	15	73	13	0.8
Coconut oil	91 ^a	7	2	0.02
Palm oil	48	44	8	0.2
Oil in salmon	28	42	30	1
Oil in mackerel	27	43	30	1

^a Includes large amount of medium chain saturates (fewer than 16 carbons).

three types of fatty acids in several common dietary fats.

The two animal fats in Table 11.1 are high in saturates and low in polyunsaturates, and this is typical of fat from farm animals. Fat from ruminating animals (e.g. milk, beef and lamb) is the most saturated and the least amenable to manipulation by alterations in the diet of the animal. The fatty acid composition of the fat from other animals (e.g. pigs) can be manipulated by altering their diets, although this may affect the palatability and the texture of the meat.

Soya oil and sunflower oil are two vegetable oils that are typical of many vegetable oils in that they are low in saturates and high in polyunsaturated fatty acids. In most vegetable oils, the **omega (ω)-6 polyunsaturated fatty acids** predominate, i.e. those with their first double bond between carbons 6 and 7. Olive oil and rapeseed (canola) oil are the most widely eaten examples of fats that are particularly high in monounsaturates and low in saturates. Rapeseed oil also has a relatively high proportion of ω -3 polyunsaturated acids. The compositions of the two tropical oils shown in Table 11.1 (coconut and palm oil) show that there are exceptions to the general observation that vegetable oils are low in saturates and high in unsaturated fatty acids. On the simple basis of the proportions of the three main types of fatty acids, palm oil is quite similar in its make-up to lard. Coconut oil contains a particularly high proportion (around 78%) of medium-chain saturated

fatty acids (fewer than 16 carbons) that make up only a small proportion of the fatty acids in most fats. Palmitic acid (16:0) is the dominant saturated fatty acid in most types of fat.

The two fish oils that are shown in Table 11.1 are typical of oil in most fish, shellfish and even the fat of marine mammals such as seals and whales. These marine oils are relatively high in polyunsaturates, low in saturates and are particularly rich in ω -3 polyunsaturated fatty acids, i.e. those whose first double bond lies between carbons 3 and 4. Fish oils are the principal dietary source of ω -3 polyunsaturated fatty acids that contain 20 or more carbon atoms (long-chain ω -3 polyunsaturates). When fish are farmed, there may be some differences in the fatty acid composition of their oil compared to wild fish, and this reflects the quite different diets of farmed and wild fish.

The physical nature of fats may give some indication of their constituent fatty acids. The presence of *cis* double bonds increases the fluidity of fats. Highly saturated animal fats tend to be solid at room temperature, whereas more unsaturated vegetable oils are usually liquid at room temperature – hence the use of hydrogenated oil to make margarine.

P:S ratio

The P:S ratio of a fat is the ratio of polyunsaturated to saturated fatty acids. In the examples shown in Table 11.1 this ratio ranges from 0.02 in coconut oil

to around 5 in rapeseed (canola) oil. The P:S ratio has traditionally been used as a shorthand guide to the likely effect of a fat or a diet upon the plasma cholesterol concentration. For many years, it has been generally accepted that saturated fatty acids tended to raise and polyunsaturates to lower plasma cholesterol, with monounsaturates having a neutral effect. This would suggest that fats and diets with

higher P:S ratios would have the most favourable (lowering) effects upon plasma cholesterol. Current views on the effects of various fatty acid fractions upon plasma cholesterol are more complex, and this is discussed later in the chapter. Despite this, the simple P:S ratio is still a useful practical indicator of the likely effect of a fat or diet upon plasma cholesterol level.

- Fatty acids are composed of a hydrocarbon chain with a single carboxyl (COOH) or acid group at one end.
- Fatty acids in biological systems have between 2 and 22 carbon atoms, and the carbons are numbered from the end furthest from the carboxyl group.
- In saturated fatty acids, all of the carbons in the hydrocarbon chain are joined by single bonds.
- Monounsaturated fatty acids have a single carbon–carbon double bond in the hydrocarbon chain.
- Polyunsaturated fatty acids have between two and six double bonds.
- The ω -3 or n-3 series of polyunsaturated fatty acids have their first double bond between carbons 3 and 4, whereas the ω -6 or n-6 series have their first double bond between carbons 6 and 7.
- A single saturated carbon atom (methylene group) separates each double bond in a polyunsaturated fatty acid.
- The shorthand notation for fatty acids indicates the number of carbons, the number of double bonds and the position of the first double bond. Thus, 18:2_ω6 (or 18:2n-6) has 18 carbons, two double bonds and the first double bond is between carbons 6 and 7.
- *Trans* fatty acids are those in which the hydrogen atoms at either end of a carbon–carbon double bond are on opposite sides, whereas in *cis* fatty acids they are on the same side.
- In most natural fats, the *cis* configuration predominates, but butter, hydrogenated oil (e.g. margarine) and heated oils contain significant amounts of *trans* fatty acids.
- The presence of a *cis* double bond alters the three-dimensional shape of a fatty acid and causes the carbon chain to bend back on itself, whereas *trans* fatty acids have a linear configuration that is closer to that of saturated fatty acids.
- All dietary fats contain a mixture of saturated, monounsaturated and polyunsaturated fatty acids, but the proportions vary enormously.
- Meat and milk fats are usually high in saturates and low in polyunsaturates.
- Many vegetable oils are low in saturates but high in polyunsaturates and the ω -6 series of polyunsaturates tend to predominate.
- Olive oil and rapeseed oil are particularly high in monounsaturates.
- Tropical oils such as palm oil and coconut oil are high in saturates and low in polyunsaturates.
- Fish oil is low in saturates and high in polyunsaturates and the ω -3 series tend to predominate.
- Fish oil is the only major dietary source of the long-chain ω -3 polyunsaturates (i.e. those with 20 or 22 carbons).
- The presence of *cis* double bonds increases the fluidity of fats. Highly saturated fats tend to be solids (e.g. butter), whereas highly unsaturated fat tends to be liquid (e.g. sunflower oil).
- The P:S ratio of a fat is the ratio of polyunsaturated to saturated fatty acids it contains.
- The P:S ratio is a useful guide to the likely effect of a fat upon plasma cholesterol concentration; those with a high ratio tend to lower it, whereas those with a low ratio tend to raise it.

SOURCES OF FAT IN THE DIET

The four principal sources of fat in the human diet are listed below.

- Milk fat:* in whole milk and yoghurt, cream, butter and cheese. This fat is almost inevitably highly saturated because it is difficult to manipulate the fatty acid content of cow milk by manipulating the cow's diet. The fatty acid profiles of foods made from milk fat are sometimes altered by blending them with polyunsaturated oil, e.g. to produce a softer butter that is easier to spread.
- Meat fat:* including not only carcass meat, but also (often especially) meat products such as sausages and burgers, and cooking fats of animal origin such as lard and beef tallow. This fat also tends to be highly saturated, especially that from ruminating animals, e.g. beef and mutton. Note that, although animal fats are usually referred to as saturated fats, they do contain large amounts of monounsaturated fatty acids (44% in lard), particularly oleic acid.
- Seeds and nuts:* contain fat which provides a concentrated source of energy for germination. These fats and oils are prominent in many diets because as they are the source of the vegetable oils and most margarine. They usually have a relatively high P:S ratio and most of the polyunsaturated fatty acids are predominantly of the

ω -6 type in many of the most widely used vegetable oils. The tropical oils are clearly an exception to these generalities and have P:S ratios that are typical of those found in meat fat (see Table 11.1).

- Fish oil:* from the flesh of oily fish such as mackerel, herring, trout, mullet and salmon. The flesh of white fish (e.g. cod and haddock) is very low in fat, although considerable amounts of fat may be stored in the livers of such fish (hence cod liver oil). Fish oils tend to have a high P:S ratio, the ω -3 acids are generally abundant, and they are the only significant dietary source of the long-chain ω -3 polyunsaturated fatty acids.

Any food that contains any of the above as an ingredient will also contain fat. Fat is a major ingredient of cakes, pastries, biscuits (cookies) and chocolate and thus these foods contribute significantly to total fat intake. Any foods cooked with the fats listed above will also be rich in fat. Thus, potatoes and most other vegetables contain almost no fat but, when fried or roasted with fat, they absorb substantial amounts of the cooking fat.

Table 11.2 shows the sources of fat and of saturated fat in household food purchases in the UK. Foods in the milk group provide only 13% of total energy but 17% of total fat and more than a quarter of the saturated fat. Most dairy fat can be avoided by choosing products made from low-fat milk. Fats and oils contribute 30% of total fat and almost a quarter

Table 11.2 The percentage contribution of various foods to the energy, fat and saturated fat content of UK household diets

Food	Energy (%)	Total fat (%)	Saturated fat (%)
Milk and milk products (including cheese)	13	17	27
Fats (including butter)	12	30	24
Meat and products	14	22	22
Fresh meat and poultry	6	9	9
Cereal foods	35	17	17
Cakes, pastries and biscuits	9	9	11
All other foods (includes fish, eggs, vegetables, fruit etc.)	26	14	10

Data source: NALIS (1997).

- Most dietary fat originates from meat, milk, fish, nuts and seeds or from any food that contains these as an ingredient.
- Nuts and seeds are the source of most cooking oils and margarine.
- Table 11.2 shows the contribution of the major food groups to total fat and saturated fat intake in the UK.
- Until the late 1990s, the proportion of food energy obtained from fat had remained at around 40% for 20 years, but has now started to fall slowly.
- Highly polyunsaturated oils and margarine have largely replaced butter, lard and beef tallow in the British diet and full-fat milk has been largely replaced by lower-fat milk.
- The P:S ratio of the British diet has approximately trebled since 1960.

of the saturated fat, despite representing only 12% of the total energy intake. Meat and meat products together contribute about 22% of both the total and the saturated fat, but note that fresh muscle meat and poultry only provide 9% of total and saturated fat and much of this can be avoided if the meat is trimmed and poultry skinned. Cereal foods provide a surprisingly high proportion of both fat and saturated fat and much of this comes from cakes and biscuits (cookies).

Up until the latter part of the 1990s, the contribution of fat to the UK household diet had remained practically constant for more than two decades at around 40% of total energy (e.g. Chesher, 1990; MAFF, 1993). Total fat consumption did drop very substantially, but only in proportion to the decline in total energy intake that occurred over this period, leaving the proportion of fat unaltered, although there was some decline in the latter part of the 1990s. As discussed in Chapter 4, there have been great changes in the types of fat consumed in Britain. In 1959, the P:S ratio of the British diet was 0.17, still only 0.23 in 1979, but by 1996 it had risen to 0.47. This change in P:S ratio has been brought about principally by the replacement of butter, hard margarine and animal-based cooking fats by soft margarine and vegetable oils. There has also been a huge decrease in the consumption of whole-fat milk. Note that subtle changes in food preparation, e.g. increased trimming of meat, skinning of poultry and skimming of surface fat from meat dishes, would not be reflected in data obtained from the National Food Survey as it measures food purchases.

ROLES OF FAT IN THE DIET

Fat as an energy source

A gram of fat yields more than twice as much metabolizable energy as a gram of either carbohydrate or protein. Vegetable oils and animal-derived cooking fats are almost 100% fat and so yield around the theoretical maximum metabolizable energy of any food, i.e. 9 kcal/g (36 kJ/g). This compares with less than 4 kcal/g (16 kJ/g) for pure sugar, which is the most concentrated form of carbohydrate and is generally regarded as an energy-rich food.

Altering the fat content of foods profoundly affects their **energy density**. Table 11.3 shows some examples of how adding relatively small amounts of fat to portions of food produces disproportionately large increases in the energy yield of the food.

- Adding 8 g of butter or margarine to a slice of bread increases the energy content by around 80%.
- Adding a 35-g portion of double cream to a slice of fruit pie increases the energy content of the dish by 70%.
- Adding 20 g of mayonnaise to a 210-g mixed salad results in a six-fold rise in the energy yield.

Even in the fat-enriched foods in Table 11.3, fat still accounts for a relatively small proportion of the food's weight and yet in most cases it accounts for a high proportion of the energy yield. Thus, fat accounts for:

Table 11.3 The effect of adding typical portions of fats to typical portions of some foods

Food	Weight	Percentage fat by weight	Percentage fat by energy	Percentage energy as fat
Slice bread	315		2.5	11
+ Butter or margarine	563	80	17	61
Jacket potato	617		Trace	0.5
+ Butter or margarine	928	50	5.5	34
Boiled potatoes	504		Trace	1.5
Chips (French fries)	1592	215	11	39
Pork chop – lean only	756		6.5	43
Lean and fat	1462	95	19	66
Skimmed milk	267		Trace	3
Whole milk	542	100	4	53
Chicken – meat only	508		4	25
Meat and skin	773	50	23	58
Fruit pie	937		8	38
+ Double cream	1596	70	17	62
Mixed salad	118		Trace	3
+ Oil and vinegar	365	210	5	70
Mixed salad	118		Trace	3
+ Mayonnaise	697	500	6.5	82.5

Portion sizes and composition are from Davies and Dickerson (1991).

- 4% of the weight of whole milk but over half the energy
- 6% of the weight of salad and mayonnaise but over 80% of the energy
- 6% of the weight of a *lean* pork chop but well over 40% of the energy
- 11% of the weight of chips (French fries) but 40% of the energy
- 17% of the weight of buttered bread but 60% of the energy.

Consumers should be very wary of claims that a product is low fat simply because the percentage of fat by weight is low. A 90% fat-free product has 10% of fat by weight and this may still account for a high proportion of the food's energy. One could, for example, accurately describe whole milk as 96% fat free, despite the fact that more than half its energy comes from fat.

Altering the fat content of a diet profoundly affects its energy density. A high-fat diet is a concentrated diet and a low-fat diet is inevitably bulkier. It may be desirable for people expending lots of energy in prolonged, heavy, physical work to eat a reasonably concentrated diet so that they do not have to eat

inordinate quantities of food (particularly people such as explorers who also have to carry their food). For relatively sedentary adults, low-fat, low energy-density diets are considered desirable because they may reduce the risk of excessive weight gain (see Chapter 8). Very low-fat and low energy-density weaning diets are, on the other hand, considered undesirable because they may precipitate protein-energy malnutrition because children cannot eat enough of the dilute food to meet their energy needs.

Palatability

Fat increases the sensory appeal of foods; it enhances the flavour, colour, texture and smell of food. Many volatile flavour compounds are fat soluble and so are concentrated in the fat fraction of unprocessed foods. Flavour compounds develop when food is cooked in fats and oils, even though some processed oils may be devoid of any inherent flavour. Fat also changes the texture of solid foods and the viscosity of liquids; butter or margarine spread onto bread

changes not only changes the taste but also the texture; full-fat milk has a higher viscosity than low-fat milk that is readily discernible to the human palate. Fat also affects the appearance of some foods, e.g. skimmed milk has noticeably less colour than whole milk and less whitening effect when used in tea or coffee. Volatile, fat-soluble materials contribute to the smell of food as well as its taste. In the foods in table 11.3, the addition of fat generally increases the sensory appeal as well as the energy yield. Boiled potatoes are essentially bland, whereas chips (French fries) have much more sensory appeal to most people. Adding butter to boiled mashed potatoes enhances their flavour, texture, smell and appearance. Adding oily dressings to salads or cream to fruit pie is usually considered to enhance the sensory appeal of these foods. Although some people find the concentrated surface fat on meat nauseating, a moderate amount of intrinsic fat greatly improves the taste and texture of meat, e.g. connoisseurs of beef look for meat with some fat marbled into the flesh.

These effects of fat on the palatability of food mean that, as populations become richer, they tend to increase the proportion of their dietary energy that is derived from fat. When affluence increases the availability of fatty foods, people tend to consume them in amounts that are probably detrimental to their long-term health prospects. This effect of fat on palatability also explains why it has proved difficult to persuade affluent populations to reduce the proportion of fat in their diets. Despite intensifying health education exhortations to reduce fat intake that began over three decades ago, there has been little change in the contribution of fat to the energy yield of the UK diet. What change there has been has been confined to the latter part of the 1990s.

Fat is important in the production of weaning foods of appropriate viscosity for young children (Church, 1979). Children require food of appropriate viscosity for their stage of development. In the immediate postnatal period, they require a liquid feed, then, after weaning (around 6–12 months), they need to be fed with a semi-solid gruel, followed by a gradual transition to a standard adult diet (note that they may require a lower viscosity when ill). After cooking, low-fat, starchy, staple foods require the addition of large amounts of water in order to produce foods of suitable viscosity for newly weaned children. They may only become edible if 70–80% of

their total weight is water and they may require 95% water to make them liquid. Their viscosity may increase sharply as they cool. Thus, thick family pap made from maize may yield 1000 kcal/kg (4.2 MJ/kg,) but as much as 4 kg of a typical maize gruel suitable for a newly weaned child might be required to supply the same amount of energy. Fat not only adds concentrated energy to weaning foods, but also affects their viscosity, such that they become palatable to young children over a wider range of water contents. For example, milk remains drinkable even when evaporated to only 20% water.

Satiation

It has traditionally been said that fat delays stomach emptying and so extends the period of satiation after consumption of the meal. Many more recent studies suggest that, in fact, quite the opposite seems to be true, i.e. that carbohydrate rather than fat may have the greater satiating effect. Not only does fat increase the palatability and energy density of the diet but, calorie for calorie, fat may also have less satiating effect than carbohydrate, thus favouring weight gain and obesity still further (see Chapter 8 for further discussion).

Fat-soluble vitamins

The fat-soluble vitamins (A, D, E and K) are found mainly in the lipid fraction of fat-containing foods and they are also absorbed along with dietary fat in the gut. Dietary deficiency of vitamin A is a major international nutritional problem, even today (see Chapter 12). Vitamin A (**retinol**), is found only in certain fats of animal origin, although humans can convert the pigment β -carotene, from brightly coloured fruits and vegetables, into retinol. Dietary fat facilitates the absorption of carotene, and so extreme fat deprivation increases the risk of vitamin A deficiency in children, even when the diet contains marginally adequate amounts of carotene.

Essential fatty acids

In the 1920s, a deficiency disease was induced in rats by feeding them on a fat-free diet and this disease was cured by the addition of small amounts of fat to

the rats' diets. Fatty acids of the linoleic acid series (ω -6) cure all of the symptoms of this deficiency disease, whereas those of the linolenic acid series (ω -3) give only partial relief of symptoms. Saturated and monounsaturated fatty acids are not effective in relieving the symptoms of this deficiency disease. We now know that there is a dietary requirement for ω -6 polyunsaturated fatty acids – these are the **essential fatty acids**. These essential fatty acids have certain structural functions (e.g. in membranes) and they are necessary for the synthesis of a key group of regulatory molecules (the **eicosanoids**), which include the **prostaglandins**, thromboxanes and leukotrienes. Although we can synthesize fatty acids from other substrates, we do not have the biochemical capability to insert double bonds between any of the first seven carbons of fatty acid molecules. The minimum human requirement for these essential fatty acids is extremely small (around 1% of energy) and it has not been possible to reproduce the symptoms of this deficiency disease of rats in deprivation studies with human volunteers. The practical risk of overt fatty acid deficiency in humans is normally remote.

The extremely low requirement for these essential fatty acids, the presence of considerable reserves of polyunsaturated fatty acids in body fat, and the difficulty of eliminating all traces of fat from the diet compound to make this disease difficult to induce experimentally in people. Even fruits, vegetable and cereals regarded as essentially fat free do contain small amounts of polyunsaturated fat. There have, however, been some reports of essential fatty acid deficiency in humans. In the 1950s, some low-birth-weight babies fed on skimmed milk developed eczema, which was reportedly cured by small amounts of essential fatty acids. In the 1970s, dermatitis caused by essential fatty acid deficiency was reported in several patients maintained on intravenous fat-free nutrition (Rivers and Frankel, 1981). The inability to safely infuse a source of fat hindered the development of long-term **total parenteral nutrition (TPN)** (see Chapter 15). It took more than four decades after the first demonstration in the rat to demonstrate essential fatty acid deficiency in an adult human being, and then only under very unusual circumstances.

Fatty acids of both the ω -6 and ω -3 series are metabolized to longer chain and more highly polyunsaturated acids, as shown in Figure 11.2. The linoleic acid ($18:2_{\omega-6}$) or linolenic acid ($18:3_{\omega-3}$) first has a

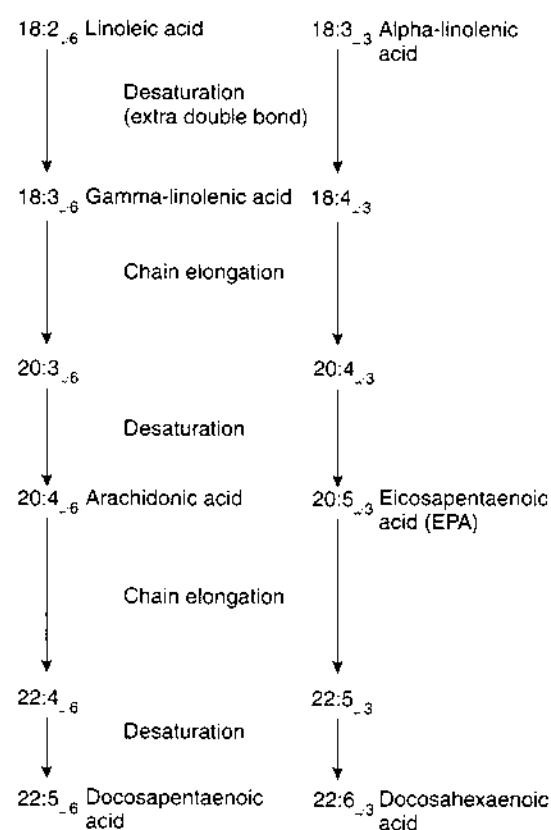


Figure 11.2 The metabolism of essential fatty acids.

double bond inserted at the carboxyl end of the existing sequence of double bonds by an enzyme known as a desaturase. The product of this desaturation reaction then has an extra two-carbon acetate unit added to the carboxyl end by an enzyme system known as an elongase. The rest of the pathway involves two further desaturation reactions (i.e. insertion of two further double bonds) and one further elongation, i.e. addition of a further two-carbon unit.

The fatty acids at the end of these sequences in Figure 11.2 are essential components of cell membranes. The EPA and docosahexaenoic acid (DHA) of the ω -3 series are normal components of membranes in the brain and retina and it is thought that ω -3 acids are probably essential in very small quantities in their own right, not just as partial substitutes for ω -6 acids. Adequate availability of ω -3 fatty acids is probably critical for brain development in infants. The presence of ω -3 fatty acids in the diet is certainly considered to be nutritionally desirable. The long-chain products of the ω -3 series, EPA and DHA, can

be made from the shorter acids in the series. However, the only significant dietary source of these longer acids is from fish oil or, in some populations, from the fat of marine mammals.

Very high intakes of ω -6 acids will tend to competitively inhibit the conversion of shorter-chain ω -3 acids to their longer-chain derivatives, and vice versa. Very high intakes of ω -6 fatty acids from nuts, vegetable oils and soft margarine might thus lead to inadequate rates of production of EPA and DHA, despite the presence of significant amounts of linolenic acid in the diet (Sanders *et al.*, 1984). This might be of practical significance in vegans consuming very large amounts of ω -6 polyunsaturates (e.g. from soft margarine and sunflower oil) or perhaps those on some cholesterol-lowering diets. As was noted earlier, long-chain ω -3 acids are not found in plant foods. Current US and UK diets may have ten times more ω -6 fatty acids than ω -3, and this figure compares to values of two to four times as much in hunter-gatherer societies. The increased use of rapeseed oil in the UK may have partly redressed the balance, because it has a relatively high concentration of ω -3 polyunsaturated fatty acids. Note that human beings cannot interconvert ω -3 and ω -6 fatty acids.

Arachidonic acid (20:4 $_{\omega_6}$) (see Figure 11.2) is the precursor for a range of local regulatory molecules, the eicosanoids (prostaglandins, thromboxanes and leukotrienes). These eicosanoids are involved in the regulation of numerous physiological processes such as the initiation of labour, blood platelet aggregation, blood vessel constriction/dilatation and inflammation. Most of the symptoms of essential fatty acid deficiency in experimental animals are attributed to impaired eicosanoid production and are relieved by linoleic acid or arachidonic acid. A series of ω -3 eicosanoids can also be produced from the corresponding fatty acid on the ω -3 pathway, eicosapentaenoic acid, but some of these have little biological activity.

The second acid on the ω -6 pathway is called gamma (γ)-linolenic acid (18:3 $_{\omega_6}$) and it is the 'active' ingredient in **oil of evening primrose**. Advocates of this product suggest that direct supply of this acid overcomes some block in this step, which is the rate-limiting step in the pathway. Another interpretation could be that the consumption of high amounts of γ -linolenic acid overrides the natural regulatory mechanism!

BLOOD LIPOPROTEINS

Fats are, by definition, insoluble in water and so, in order to transport them around the body in aqueous plasma, they have to be associated with protein carriers and transported as lipid/protein aggregates, the **lipoproteins**. There are several subdivisions of lipoproteins in blood plasma, which can be conveniently separated according to their density – the higher the proportion of fat in a lipoprotein, the lower its density, because fat has low density. The protein components of lipoproteins are known as **apoproteins**.

The subdivisions of the plasma lipoproteins and their approximate compositions are:

- chylomicrons (82% TAG, 9% C, 7% PL, 2% AP)
- very low-density lipoproteins (VLDL) (52% TAG, 22% C, 18% PL, 8% AP)
- low-density lipoproteins (LDL) (9% TAG, 47% C, 23% PL, 21% AP)
- intermediate-density lipoproteins (IDL) (20% TAG, 35% C, 20% PL, 15% AP)
- high-density lipoproteins (HDL) (3% TAG, 19% C, 28% PL, 50% AP)
- TAG = triacylglycerol, C = cholesterol, PL = phospholipid, AP = apoprotein.

Chylomicrons are the form in which ingested fat is transported to the liver and tissues after its absorption from the intestine. They are normally absent from fasting blood. Levels of VLDL should also be low in fasting blood because VLDL is rapidly converted to LDL (see below). This means that blood lipid determinations should be measured in fasting blood samples to avoid any distortion due to variable amounts of postprandial chylomicrons and large amounts of VLDL. IDL is a transitory intermediate that normally represents only a very small fraction of total lipoproteins. The principal lipoproteins in normal fasting blood are LDL, HDL and relatively small amounts of VLDL.

VLDL is rich in TAG and thus measures of blood TAGs are taken to indicate the amount of VLDL. LDL is the major cholesterol-containing fraction (about 60%), and thus measures of total plasma cholesterol are often taken to indicate the LDL concentration. The normal function of VLDL is to transport TAG produced endogenously in the liver (largely from dietary fat) to the adipose tissue or the muscles, where it acts as an energy source. LDL acts as an external

- Fat is the most concentrated source of dietary energy.
- Adding relatively small amounts of fat to a food can produce a disproportionately large rise in its energy yield.
- Claims that a food has a low percentage of fat by weight may disguise the fact that fat accounts for a very large proportion of its energy yield.
- A relatively high-fat, high energy-density diet may be necessary for those expending very large amounts of energy and for active and rapidly growing babies and children.
- For sedentary adults, a high-fat diet probably encourages excessive weight gain and obesity.
- Fat improves the palatability of food and this encourages affluent populations to eat more fat than is good for their long-term health and makes it difficult to persuade them to eat less fat.
- Very low-fat weaning foods may be of such low energy density that babies cannot eat enough of them to meet their energy needs.
- Fat is a concentrated source of energy for weaning foods and produces a suitable viscosity at relatively low water content.
- Calorie for calorie, fat is probably less satiating than carbohydrate.
- Impaired fat absorption or extreme low-fat diets may precipitate fat-soluble deficiencies (especially vitamin A).
- ω -6 polyunsaturated fatty acids are essential nutrients; they have structural functions and are precursors of prostaglandins and other eicosanoids.
- The requirement for essential fatty acids is small and it is difficult to induce overt deficiency in human volunteers.
- The parent compound of the ω -6 series of polyunsaturated fatty acids is linoleic acid 18:2 _{ω -6} and that of the ω -3 series is linolenic acid 18:3 _{ω -3}.
- Linoleic and linolenic acid undergo a series of desaturation and elongation reactions to produce two series of ω -6 and ω -3 polyunsaturated fatty acids, with the end products of these sequences being 22:5 _{ω -6} and 22:6 _{ω -3} respectively.
- Small amounts of ω -3 polyunsaturates are probably essential; the long-chain ones are normal components of brain and retinal membranes and seem to be crucial for normal brain development.
- Long-chain ω -3 polyunsaturates are prevalent in fish oil and they can be made from shorter-chain varieties that are prevalent in some vegetable oils, such as rapeseed oil.
- Humans cannot interconvert ω -6 and ω -3 fatty acids.
- Arachidonic acid (20:4 _{ω -6}) is the precursor of many potent eicosanoids. The corresponding eicosanoids made from eicosapentaenoic acid (20:5 _{ω -3}) often have low biological activity.
- Oil of evening primrose contains large amounts of 18:3 _{ω -6} and this should increase the availability of long-chain ω -6 polyunsaturates because it bypasses the usual rate-limiting step in ω -6 fatty acid metabolism.

source of cholesterol for cells unable to synthesize sufficient for use in membranes or as a precursor for steroid hormone synthesis. The main function of HDL is to transport excess cholesterol from tissues and other blood lipoproteins back to the liver.

Epidemiological evidence and evidence from people with familial hyperlipidaemias suggest that high plasma LDL cholesterol concentrations are

causally associated with increased atherosclerosis and increased risk of coronary heart disease. HDL, on the other hand, inhibits atherosclerosis, and high HDL cholesterol concentrations are protective against coronary heart disease. The LDL-to-HDL ratio is a better predictor of coronary heart disease risk than LDL concentration alone. Simple plasma cholesterol measurements do not indicate the distribution of the

- Fats are insoluble in water and are transported in blood as lipoprotein complexes.
- Chylomicrons are the lipoproteins that carry absorbed fat from the intestine to the tissues and they are absent from fasting blood.
- Very low-density lipoprotein (VLDL) is a TAG-rich fraction that transports fat produced in the liver to the muscles and adipose tissue.
- Low-density lipoprotein (LDL) is rich in cholesterol and acts as an external source of cholesterol for cells.
- LDL is atherogenic.
- High-density lipoprotein (HDL) is anti-atherogenic; it removes excess cholesterol to the liver.
- High plasma LDL is associated with an increased risk of coronary heart disease, but high HDL is associated with a reduced risk.

cholesterol between LDL and HDL, but total cholesterol concentration is used as a crude indicator of LDL cholesterol concentration. Raised VLDL is associated with obesity, glucose intolerance and high alcohol consumption. High VLDL is certainly an indicator of high risk of coronary heart disease, but whether it is a significant independent risk factor for coronary heart disease is still a matter of debate (Grundy, 1998).

DIGESTION, ABSORPTION AND TRANSPORT OF DIETARY LIPIDS

A summary scheme for the digestion, absorption and transport of dietary fats is shown in Figure 11.3. Dietary fat is emulsified by the action of bile salts and phospholipids in the intestine. This emulsified fat is then partially digested by the enzyme lipase, which is secreted into the gut in the pancreatic juice. Note that lipase is a water-soluble enzyme and so can normally only digest water-insoluble fat if it has first been emulsified. Lipase splits off (hydrolyses) fatty acids from the glycerol component of TAG. *In vivo*, this process does not result in total breakdown of TAG; the major end products in the gut are monoacylglycerols and free fatty acids. Lipase does not act on the fatty acid attached to the second (middle) carbon atom of the glycerol component.

Glands at the base of the tongue produce a lingual lipase, an enzyme that works at low pH and can readily digest the fat in milk globules before emulsification by bile salts. Lingual lipase is normally a

minor contributor to fat digestion, although it is important in suckling babies, whose pancreatic function is not fully developed at birth. Churning of fat in the stomach with phospholipids and other digestion products has some emulsifying effect on dietary fat.

The products of lipase action together with bile salts, phospholipids and fat-soluble vitamins form minute molecular aggregates known as micelles, which effectively solubilizes them. Fat-digestion products and fat-soluble vitamins diffuse from the micelles into the mucosal cells that line the small intestine and, once inside, they are re-assembled into TAG and coated with protein to form the chylomicrons. Chylomicrons can be seen as droplets or packets of TAG surrounded by a thin protein coat. These chylomicrons enter the lymphatic vessels that drain the small intestine and pass into venous blood via the thoracic duct, which drains into the right subclavian vein. Most fat is absorbed in the duodenum and jejunum.

In the liver and adipose tissue, the TAG is removed from the chylomicrons by an enzyme called **lipoprotein lipase** that is located on the inner surface of the capillary. The products of this second digestion are absorbed into the cells and again re-assembled into TAG for storage in adipose tissue or for transport in the liver. Liver cells clear the fat-depleted chylomicron remnants from the plasma via a specific chylomicron remnant receptor. Some shorter-chain fatty acids (medium-chain fatty acids) bypass this process and are absorbed directly from the gut into the hepatic portal vein and transported to the liver bound to plasma albumin.

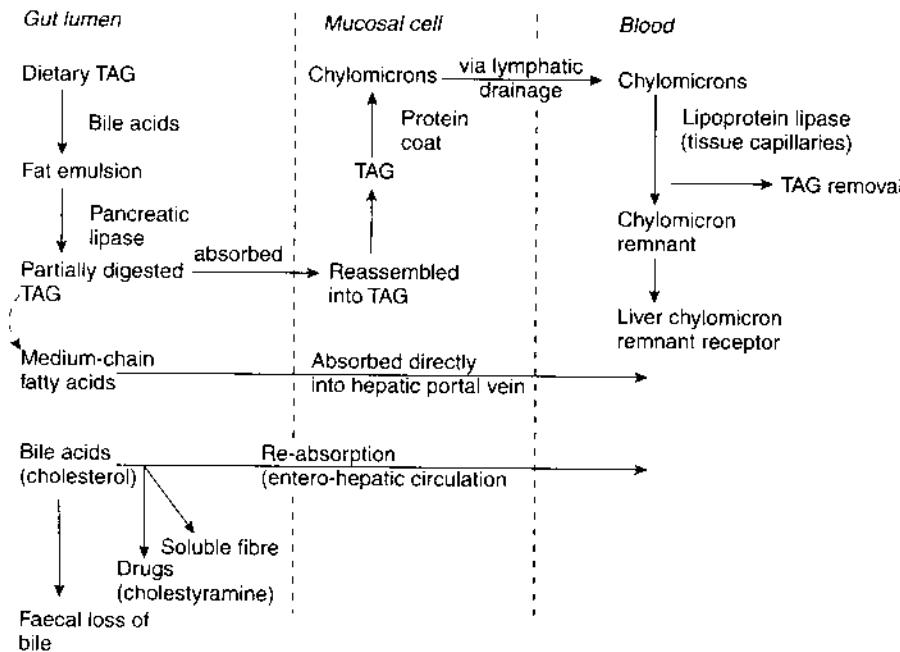


Figure 11.3 A scheme to illustrate the digestion, absorption and transport of dietary triacylglycerol (TAG).

Large amounts of cholesterol-derived bile salts are secreted into the intestine each day in the bile. Most of this bile salt (cholesterol) is reabsorbed in the intestine and recycled to the liver – **entero-hepatic circulation**. Some substances are thought to exert a lowering effect on plasma cholesterol by interfering with this entero-hepatic recycling, leading to increased faecal losses of cholesterol derivatives. Soluble fibre and some plant sterols exert such an effect, and the cholesterol-lowering drug cholestyramine is a resin that binds or chelates bile acids in the intestine, preventing their reabsorption.

TRANSPORT OF ENDOGENOUSLY PRODUCED LIPIDS

Figure 11.4 gives a summary of the origins and inter-relationship of the non-chylomicron lipoproteins. TAG-rich VLDL is produced in the liver by reassembling dietary fats and by endogenous synthesis of fat. VLDL is exported from the liver to tissues and adipose tissue for use in metabolism or storage. The enzyme lipoprotein lipase, in the tissue capillaries, hydrolyses the TAG in VLDL, enabling it to diffuse

- Dietary fat is partly digested by pancreatic lipase in the small intestine; the products of this digestion are absorbed into the mucosal cells and most are re-esterified into TAG.
- Bile salts emulsify dietary fat so that it can be digested and solubilize fat digestion products so that they can be absorbed.
- Bile salts are reabsorbed in the ileum and returned to the liver (the entero-hepatic circulation).
- Chylomicrons carry the absorbed fat into the circulation via the lymphatic system.
- Lipoprotein lipase in tissue capillaries digests the TAG in chylomicrons so that it can be absorbed into the fat or liver cells and again reassembled into TAG.
- The chylomicron remnant is removed from the circulation by the liver.
- Medium-chain fatty acids can be absorbed directly into the hepatic portal vein and transported bound to plasma albumen.

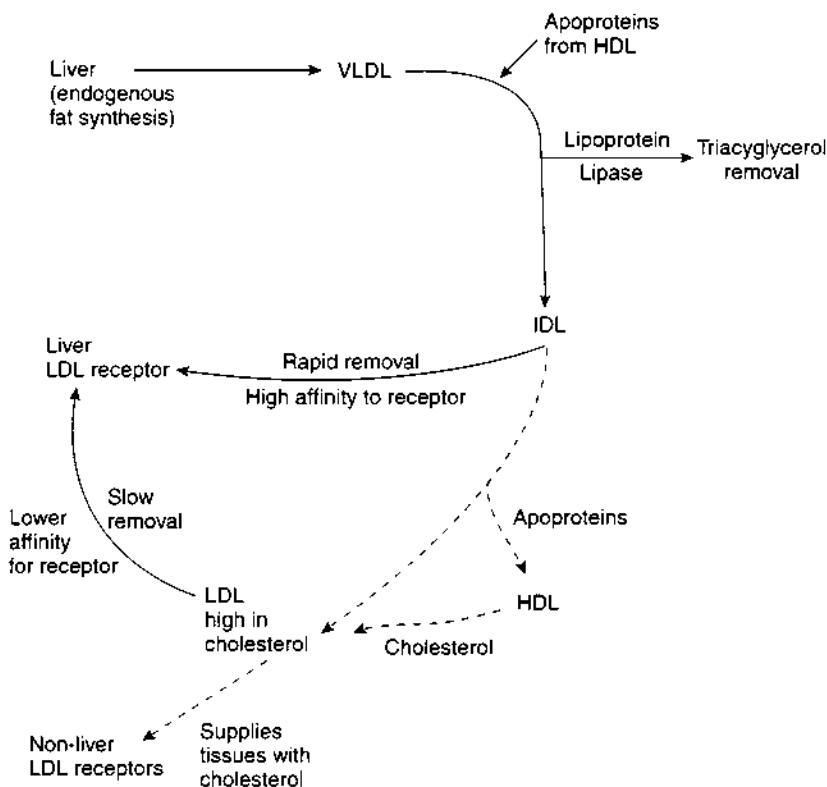


Figure 11.4 The relationship between the various lipoproteins involved in endogenous lipid transport. After Brown and Goldstein (1984). VLDL, very low-density lipoprotein; LDL, low-density lipoprotein; HDL, high-density lipoprotein.

into the tissue cells. VLDL that has been depleted of much of its TAG is called IDL. This IDL is then either rapidly cleared from the plasma by binding with a liver receptor, or some of it is modified within the blood to become LDL. LDL and IDL both bind to a common receptor, called the **LDL receptor**, but this receptor has a higher affinity for IDL than for LDL and so IDL is cleared more quickly (2–6 hours) than LDL (2–3 days). More than 75% of LDL receptors are

found in the liver; other tissues are supplied with cholesterol via binding of cholesterol-rich LDL to their LDL receptors (Brown and Goldstein, 1984). Around 80% of LDL leaving the circulation is normally taken up by the liver and just 20% by all other tissues. The perceived importance of the LDL receptor may be judged by the fact that its discoverers, Brown and Goldstein, were awarded the Nobel Prize for Physiology and Medicine in 1985. Once LDL is

- TAG produced in the liver is exported as VLDL.
- VLDL is depleted of TAG by lipoprotein lipase in tissue capillaries; this depleted VLDL is called intermediate-density lipoprotein (IDL).
- Some IDL is cleared rapidly by the liver and some is converted within the circulation to LDL.
- Both IDL and LDL are cleared by LDL receptors, although LDL is cleared much more slowly.
- Three-quarters of LDL receptors are in the liver; the other quarter are spread amongst the other tissues.
- When cells have surplus cholesterol, they produce less, convert some to cholesterol esters, and suppress their synthesis of LDL receptors.

absorbed into cells and releases its cholesterol, this causes three regulatory responses in the cell:

1. it suppresses the synthesis of LDL receptors in the cell
2. it inhibits the rate-limiting enzyme in cholesterol synthesis – hydroxymethyl glutaryl coenzyme A reductase (**HMG CoA reductase**)
3. it activates an enzyme that converts cholesterol to cholesterol esters that can be stored as droplets in the cell.

THE DIET-HEART HYPOTHESIS

According to Brown and Goldstein (1984), most people living in the Western industrialized countries have elevated plasma cholesterol concentrations, making them prone to atherosclerosis and coronary heart disease. They concluded that this was due to reduced LDL receptor synthesis. This reduced receptor synthesis is one of the mechanisms that protect cells from cholesterol overload, but it also leads to elevated levels of circulating cholesterol. They suggested that dietary and other environmental factors that tend to raise the plasma LDL (e.g. high saturated-fat intake) induce their effect by suppressing LDL-receptor synthesis. Reduced LDL-receptor synthesis would be expected not only to increase the life span of LDL, but also to lead to an increased rate of production because less IDL would be cleared and more would be converted within the blood to LDL. The elevated level of LDL cholesterol in the blood increases the tendency for it to be deposited in artery walls, where it may become oxidized and trigger fibrotic changes, i.e. atherosclerosis, or 'hardening of the arteries'. Reduced LDL-receptor synthesis is seen as initiating a chain of events that can lead to increased atherosclerosis and ultimately increased risk of coronary heart disease and other atherosclerosis-linked diseases – the **diet-heart hypothesis**. The diet-heart hypothesis is outlined in Figure 11.5.

There is overwhelming evidence that raised plasma cholesterol concentrations are linked to an increased risk of coronary heart disease. Not least of this evidence are studies on people with **familial hypercholesterolaemia**, an inherited condition in which there is a defect in the LDL-receptor gene. Even people who are heterozygous for this condition

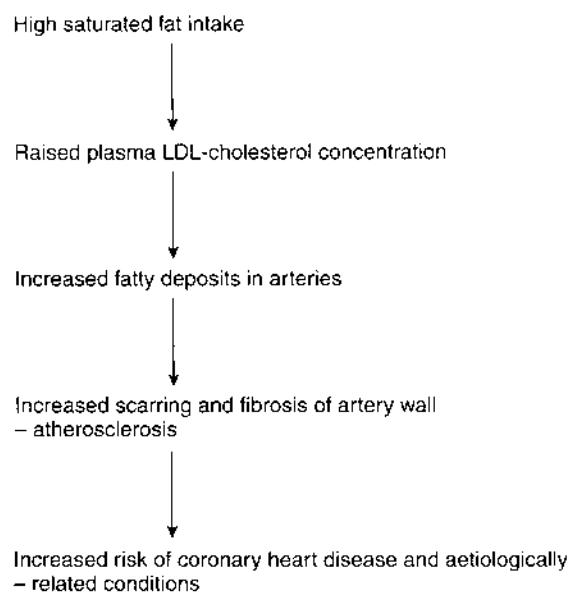


Figure 11.5 The diet-heart hypothesis.

(i.e. one normal and one abnormal gene) become hypercholesterolaemic, and they have a markedly increased probability of dying prematurely from coronary heart disease, particularly if they are male. Those few people (one in a million) who are homozygous for this condition fail to produce any functional LDL receptors. They have to rely entirely upon other, inefficient, methods to clear LDL cholesterol from blood, and their plasma cholesterol concentrations may be up to six times normal. These individuals usually die from coronary heart disease in childhood. Injections of radioactively labelled LDL into such individuals confirm that it remains in the circulation about 2.5 times as long as in unaffected people and that the rate of endogenous LDL production is about doubled because of reduced LDL clearance (Brown and Goldstein, 1984).

This diet-heart hypothesis could reasonably be described as a dominant theme of many nutrition education campaigns in the UK and elsewhere. Dietary measures that would lead to lowered average plasma cholesterol concentrations are widely seen as central to nutrition education for health promotion. A figure of 5.2 mmol/L has been widely suggested as a desirable maximum for individual plasma cholesterol levels, and yet the majority (around two-thirds) of the UK population exceed this level. Some

Table 11.4 Some cut-off points for plasma cholesterol concentration

Plasma cholesterol concentration (mmol/L)	Classification
Less than 4.1	Low
4.1–5.2	Desirable range
5.2–6.5	Mild hypercholesterolaemia
6.5–7.8	Moderate hypercholesterolaemia
Over 7.8	Severe hypercholesterolaemia

frequently used cut-off points for plasma cholesterol concentration are given in Table 11.4. In Chapter 1, it was seen that, in men, the individual risk of coronary heart disease increases exponentially as the plasma cholesterol concentration rises above this target value. However, most of the apparent excess population risk from high blood cholesterol seems to be concentrated in those men whose blood cholesterol levels are only slightly to moderately elevated. There are vast numbers of men (the majority of the population) who are apparently at some relatively small increased individual risk due to slight to moderate cholesterol elevation, but only a small proportion at greatly increased risk due to very high cholesterol concentrations (Table 11.5). The thrust of nutrition education has been to direct cholesterol-lowering dietary advice at the whole population, with the aim of shifting the whole distribution downwards. Additionally, intensive dietary counselling and perhaps cholesterol-lowering drugs may be used for those individuals identified as having substantially

elevated cholesterol levels or existing coronary heart disease.

The distribution of plasma cholesterol concentrations in a representative sample of English adults is shown in Table 11.5. About 70% of men and 71% of women have values that are above the optimal range (i.e. above 5.2 mmol/L). Only around 4% of men and 9% of women are above the 7.8 mmol/L threshold (most of the women in this category are elderly). There is a very pronounced trend in both sexes for plasma cholesterol to increase with age. In 18–24-year-old men and women, the mean value was 4.8 mmol/L, but in 65–74 year olds, this had risen to 6.2 mmol/L for men and 7.0 mmol/L for women.

CURRENT 'HEALTH IMAGES' OF DIFFERENT DIETARY FATS

More than 40 years ago, Keys *et al.* (1959) showed that, in short-term experiments with young men, total plasma cholesterol levels were significantly affected by the degree of saturation of fat in the diet, even when the diets were matched for total fat content. As a result of such experiments, saturated fat has acquired a very negative health image. Saturated fat was found to raise plasma cholesterol concentration in such experiments and this is associated with increased atherosclerosis and an increased risk of coronary heart disease. Polyunsaturated fats, on the other hand, were found to lower plasma cholesterol

Table 11.5 Distribution (%) of plasma cholesterol concentrations in a representative sample of English adults

Group	Percentage with serum cholesterol concentrations (mmol/L)				
	Below 5.2	5.2–6.5	6.5–7.8	7.8+	Mean
<i>Men</i>					
All	31	43	23	4	5.8
18–24 years	66	30	2	3	4.8
65–74 years	16	43	36	5	6.2
<i>Women</i>					
All	30	39	23	9	5.9
18–24 years	70	24	6	—	4.8
65–74 years	7	38	41	14	6.7

Data source White *et al.* (1993).

- A high saturated fat intake leads to reduced LDL receptor synthesis and to elevated levels of circulating LDL cholesterol.
- LDL can enter artery walls, where it may trigger atherosclerosis, e.g. if it becomes oxidized.
- Atherosclerosis predisposes to coronary heart disease.
- High saturated fat intake (and other environmental influences) triggers a chain of events that can ultimately lead to increased risk of coronary heart disease – the diet–heart hypothesis.
- The risk of coronary heart disease rises exponentially with plasma total cholesterol (a crude indicator of LDL cholesterol concentration).
- People who are genetically prone to have very elevated plasma LDL cholesterol concentrations are at greatly increased risk of coronary heart disease.
- 5.2 mmol/L is the ideal upper limit for plasma total cholesterol concentration, but two-thirds of the UK population exceed this level.
- People with severely elevated cholesterol levels are at greatest individual risk from coronary heart disease, but much of the population risk is concentrated in the large numbers of people with mild to moderate cholesterol elevation.
- The health promotion priority has thus generally been to reduce average plasma cholesterol concentrations, rather than solely seeking to identify and treat those with severely elevated cholesterol concentrations.

concentration when they replaced saturated fat. Results from these early experiments suggested that monounsaturated fatty acids were neutral in their effects on blood cholesterol level. Saturated fatty acids were about twice as potent in raising serum cholesterol as polyunsaturated fatty acids were in lowering it. Keys *et al.* (1965) produced an equation (the **Key's equation**) which enabled the change in plasma cholesterol of a group to be fairly well predicted in response to specified changes in its dietary fats:

$$\Delta C = (2.7 \times \Delta S) - (1.3 \times \Delta P) + 1.5Z^{1/2}$$

where ΔC = change in plasma cholesterol concentration; ΔS = change in percentage of energy from saturated fat; ΔP = change in percentage of energy from polyunsaturated fat; Z = dietary cholesterol (mg) per 1000 kcal (4.2 MJ).

As a result of such observations, some vegetable oils and soft margarine, because of their high P:S ratios, acquired a very positive health image and were, for a time, regarded as 'health foods'. In the experiments of Keys *et al.* (1959), switching from a diet very high in saturates (P:S 0.16) to one very high in polyunsaturated fat (P:S 2.8) resulted in a rapid, but variable, reduction in an individual's plasma cholesterol concentration, even though the total

fat intake was kept constant (see Figure 3.11 in Chapter 3).

Since the 1960s, there have been several attempts to produce other such predictive equations that are better or more detailed than those of Keys and his colleagues (e.g. Yu *et al.*, 1995).

Current dietary guidelines relating to dietary fats were detailed and discussed in Chapter 4.

Effects on LDL and HDL cholesterol

It is now known that the total plasma cholesterol is comprised of both LDL and HDL fractions. The LDL fraction is atherogenic, and regarded as 'the bad cholesterol', whereas the HDL fraction is anti-atherogenic, and seen as 'the good cholesterol'. It would thus seem desirable to lower the LDL cholesterol and raise the HDL cholesterol. The LDL cholesterol is generally considered to be readily responsive to dietary manipulation, but the HDL rather less so. Current views on the ways in which the different dietary fat fractions affect these plasma cholesterol fractions and their current 'health image' are summarized below.

TOTAL FAT

In an assessment of current evidence, Mangiapane and Salter (1999) concluded that moderate reductions in total fat consumption without any change in fatty acid composition would not produce any favourable change in blood lipoprotein profiles. There are, of course, several other reasons for reducing our reliance on dietary fats, such as improved weight control, better glucose tolerance and perhaps reduced susceptibility to some cancers (see Chapter 4). Better weight control and improved glucose tolerance would be expected to reduce the risk of heart disease independently of any effects upon plasma lipoprotein profiles.

THE SATURATES

These generally raise the LDL cholesterol concentration and so they have been, and continue to be, viewed very negatively by health educators. Saturated fats also cause a small compensatory rise in HDL cholesterol, but this is considerably outweighed by the effect upon the LDL fraction. It does now seem that saturated fatty acids vary in their effects upon plasma cholesterol. Palmitic acid (16:0) is by far the most common saturated fatty acid in the diet and, together with myristic acid (14:0), the most potent at raising LDL cholesterol. Shorter-chain saturated fatty acids of 12 carbons or fewer (e.g. found in high amounts in coconut oil) have a smaller cholesterol-raising effect. Stearic acid (18:0) is not atherogenic and may be closer to the monounsaturated fatty acid oleic acid (18:1_{ω-9}) in its effects upon plasma cholesterol levels (Yu *et al.*, 1995). Stearic acid is found in large amounts in cocoa butter, which has been found to cause less increase in plasma cholesterol than other saturated fats.

THE MONOUNSATURATES

As a result of the early experiments of Keys *et al.* (1959) and other early workers, monounsaturates were regarded as neutral in their effects upon plasma cholesterol concentration. More recent work suggests that, when they replace saturated fatty acids, they have as great a lowering effect upon the LDL fraction as polyunsaturated fats and are neutral in their effects on the HDL fraction (e.g. Mattson and Grundy, 1985; Yu *et al.*, 1995). At very high intakes, they may actually increase HDL concentrations. Some Mediterranean

men consume high-fat diets (40% of energy from fat) and yet have a very low incidence of coronary heart disease, and these Mediterranean diets are rich in olive oil and monounsaturated oleic acid. These monounsaturated fats are very much 'in fashion' at present, accounting for the current positive image of olive oil and, in the USA, of canola (rapeseed) oil.

THE POLYUNSATURATES

Polyunsaturated fatty acids of the ω-6 or n-6 series (the bulk of those in most vegetable oils) are still thought to lower the plasma LDL cholesterol concentration – a beneficial effect. However, some studies have suggested that they may also lower HDL cholesterol – a potentially harmful effect (Grundy, 1987). Such studies have been widely reported and have helped to increase concern about the overall benefits of diets very high in polyunsaturated fat. Results from some of the early intervention trials using diets very high in polyunsaturated fats also suggested that very high intakes of these acids might have other detrimental effects, including an increased cancer risk. Often, such studies reported a non-significant decrease in cardiovascular mortality and a non-significant increase in mortality from other causes including cancer (see Oliver, 1981). Their high susceptibility to oxidation (see Chapter 12) is one possible mechanism by which high intakes might produce detrimental effects (COMA, 1991). Populations who both consume high-fat diets and have a high proportion of their fat as ω-6 polyunsaturated acids have been historically rare, because the widespread use of large amounts of most extracted vegetable oils is a relatively recent phenomenon. Diets that are high in olive oil, palm oil or coconut oil have long been used, but none of these contains high levels of ω-6 polyunsaturated fatty acids. Safety concerns, together with this lack of historical precedent, have led a number of expert committees (e.g. COMA, 1991) to recommend that it would be prudent not to further raise current average intakes of polyunsaturated fatty acids in the UK and USA. It has also been recommended that individual consumption of polyunsaturated fatty acids should be limited to no more than 10% of total energy. There is little hard evidence of deleterious effects of high polyunsaturated fat intake, but nonetheless it would be fair to say that, during the early 1990s, some health educators viewed them less positively than

they did in the 1960s and 1970s. Recent evidence suggests that polyunsaturated fatty acids are probably neutral in their effect upon HDL, and that any lowering effect upon HDL only occurs at high intakes that are unlikely to be consumed in normal diets (Mangiapane and Salter, 1999).

The ω -3 series of acids are found in large amounts in marine oils and in variable amounts in different vegetable oils, for example:

- 11.4% of total fatty acids in walnut oil
- 7% in soya oil
- 10% in rapeseed oil
- 0.29% in sunflower seed oil.

The ω -3 fatty acids are generally thought to lower blood TAG concentrations and to raise the HDL cholesterol. They are also thought to have other apparently beneficial effects such as reducing inflammation and the tendency of blood to clot and therefore reducing the risk of thrombosis. They are very much 'in fashion' at present. Fish oils, the sole dietary source of the long-chain members of this series, are discussed more fully at the end of this chapter. According to Harris (1989), the overall effect of fish oil on LDL cholesterol is minimal at practical levels of consumption.

TRANS FATTY ACIDS

Diets containing only naturally occurring fatty acids would normally be low in *trans* fatty acids, but certain types of margarine and vegetable shortening may contain quite large amounts of *trans* fatty acids, particularly cheaper, hard margarines. *Trans* fatty acids are also generated when vegetable oils are heated, so frying adds some *trans* fatty acids to foods. There has naturally been concern about whether there are any health implications in consuming relatively large amounts of these *trans* fatty acids because of their low levels in unprocessed foods. In particular, it has been suggested that they may have unfavourable effects on blood lipoproteins similar to those of saturated fatty acids. The precise health implications of consuming large amounts of *trans* fatty acids are still unclear, but they are generally classified as equivalent to saturated fatty acids in their effects upon blood lipids. In short-term experiments, involving both men and women, Mensink and Katan (1990) compared the effects of diets that were matched except for their *trans* fatty acid con-

tent. They concluded that *trans* fatty acids are at least as unfavourable as saturated fatty acids in their effects upon blood lipids, tending not only to raise the LDL levels, but also to lower HDL levels. Willett *et al.* (1993), in a large cohort study of American nurses, reported a highly significant positive association between the risk of coronary heart disease and the dietary intake of *trans* fatty acids. These results achieved widespread publicity and led some margarine manufacturers to change their production methods and produce products that were free of *trans* fatty acids. Other workers have disputed the notion that high consumption of *trans* fatty acids is a major determinant of the current risk of coronary heart disease. In 1989, average intakes of *trans* fatty acids in the UK were estimated at around 5 g/day (2% of calories), but with very wide variation, some individuals consuming as much as five times this average value (COMA, 1991). COMA (1991) recommended that levels should not rise and they have, in fact, been declining in recent years as consumption of hard margarine has declined and been replaced by soft margarine containing much less, and in some cases practically no, *trans* fatty acids. Current average US intakes of *trans* fatty acids have been estimated at around 3% of the total energy intake.

Dietary cholesterol

In short-term experiments, dietary cholesterol does tend to raise the LDL cholesterol level, but, at usual intakes, this effect is small compared to the effects caused by altering the proportions of the various fatty acid fractions in the diet. Some animals, such as rabbits, are particularly susceptible to the hypercholesterolaemic effects of dietary cholesterol. In these animals, dietary cholesterol suppresses LDL receptor production in liver cells and this reduces LDL clearance. Usual intakes of cholesterol are relatively small when compared to the normal rate of endogenous synthesis (around 10–20%). Keys *et al.* (1965) calculated that a change in cholesterol intake from the top to the bottom of the range in 1960s' US diets (i.e. from 1050 mg/day to 450 mg/day) would lower plasma cholesterol by only around 0.23 mmol/L. They concluded that dietary cholesterol should not be ignored in cholesterol lowering, but that changing it alone would achieve very little. Dietary cholesterol is usually given low priority in most sets of nutrition

education guidelines (see Chapter 4). Note, however, that lowering saturated fat intake will probably result in some consequential fall in cholesterol intake, because cholesterol is found only in animal products. Some individuals are genetically susceptible to the hypercholesterolaemic effects of dietary cholesterol and control of dietary cholesterol is more significant for such individuals (Brown and Goldstein, 1984). Note that eggs are the single most important source of cholesterol in most Western diets. Average daily cholesterol intake will be greatly affected by changes in egg consumption, and those who are particularly sensitive to dietary cholesterol are advised to restrict their egg consumption.

Thirty years ago, dietary cholesterol was widely believed to be of predominant importance in the genesis of high blood cholesterol and atherosclerosis (e.g. Frantz and Moore, 1969). The contribution of experiments with herbivorous rabbits to this belief is critically discussed in Chapter 3. Since the 1960s, the evidence linking high plasma cholesterol levels with an increased risk of atherosclerosis has strengthened, but the view of dietary cholesterol as the major influence upon plasma cholesterol concentration has faded.

Even much of the apparent atherogenic effect of cholesterol seen in some rabbit experiments may have been due to the presence of highly atherogenic cholesterol oxides in the cholesterol used in these studies. Some old bottles of chemical cholesterol contain very large amounts of oxidized cholesterol. Oxidized cholesterol is many times more toxic to vascular endothelial cells than purified cholesterol. In fresh foods, cholesterol oxide concentrations are low, but such oxides would be generated if these foods were subjected to prolonged heating and/or exposure to air, e.g. powdered egg, powdered whole milk and some cheeses. Jacobsen (1987) reported that ghee (clarified butter used in Indian cooking) has high levels of cholesterol oxides, whereas fresh butter does not. Damage induced by oxidation of lipids in LDL is now considered to be an important factor in the aetiology of heart disease, and oxidative damage has been implicated in the aetiology of cancer and other chronic diseases (see Chapter 12).

Plant sterols

Plants do not produce cholesterol, but they do produce other sterols such as stigmasterol, ergosterol

and sitosterol. It has been known for decades that these plant sterols, although not themselves absorbed in the gut, do tend to inhibit the absorption of cholesterol. Sitosterol and other plant sterols were identified as possible cholesterol lowering agents more than 30 years ago, but their practical effect was small. More recently, **sitostanol**, a substance produced chemically from sitosterol, has been reported to be a more effective cholesterol-lowering agent than other plant sterols. In a 1-year, double blind trial, intakes of 1.8–2.6 g/day of sitostanol ester incorporated into margarine were reported to reduce average cholesterol concentrations by 10% in mildly hypercholesterolemic men (Miettinen *et al.*, 1995). The effectiveness of sitostanol is probably greater in those with higher dietary cholesterol intakes, even though it may also have some effect upon the reabsorption of biliary cholesterol. A margarine (Benecol) containing sitostanol ester is now available in the UK and has been marketed as a functional food that may help lower blood cholesterol.

THE KEY TENETS OF THE DIET-HEART HYPOTHESIS

The diet heart hypothesis was summarized in Figure 11.5 and can be broken up into the two key tenets listed below.

1. *High total plasma cholesterol concentration causes an increased risk of coronary heart disease.* When the individual lipoprotein fractions in blood are considered, high LDL cholesterol increases atherosclerosis and the risk of coronary heart disease, whereas HDL is thought to be protective against atherosclerosis. High VLDL is associated with increased coronary heart disease risk, but it is difficult to say whether this is a direct harmful effect of high VLDL or a result of the link between high VLDL and obesity and diabetes, which are both strong risk factors for coronary heart disease.
2. *The amount and type of dietary fat are important determinants of the plasma cholesterol concentration.* Current views on the effects of the various dietary fatty acids on blood lipoproteins are summarized in the preceding section.

- It has been known since 1960 that replacing dietary saturated fat with unsaturated fat lowers the total plasma cholesterol concentration.
- It is desirable to lower the plasma LDL cholesterol and to raise the HDL cholesterol.
- A moderate reduction in total fat intake does not produce significant improvement in blood lipoprotein profiles, although it may be desirable upon other grounds.
- Saturated fats and *trans* fatty acids raise the LDL cholesterol concentration.
- Of the saturates, myristic acid (14:0) and palmitic acid (16:0) have the biggest cholesterol-raising effect; shorter-chain acids (e.g. from coconut oil) have a smaller effect, and stearic acid (18:0), found in cocoa butter, has no LDL raising effect.
- Monounsaturates are now thought to lower LDL cholesterol when they replace saturated fat and to have a neutral effect upon HDL. At very high intakes, they may even raise the HDL.
- ω -6 polyunsaturates lower plasma LDL and are probably neutral in their effects upon HDL, although at very high intakes they may lower HDL.
- Some speculative safety concerns coupled with a lack of historical precedent of populations with very high intakes of ω -6 polyunsaturates have persuaded several expert committees to recommend that intakes of these fatty acids should not rise any further.
- The ω -3 series of polyunsaturated fatty acids lower blood TAG concentrations and raise HDL levels. At practical levels of consumption, they have little effect upon LDL levels.
- Dietary cholesterol is now thought to be a relatively minor influence upon plasma cholesterol and lipoprotein profiles.
- Rabbits are very sensitive to the hypercholesterolaemic effects of dietary cholesterol, and some people may also be more sensitive to it.
- A reduction in saturated fat consumption will probably produce a consequent fall in cholesterol intakes, although eggs are the biggest single source of cholesterol in most diets.
- Oxidized cholesterol is much more atherogenic than cholesterol *per se*, and would be present in increased amounts in foods subject to prolonged heating in air, e.g. powdered egg and whole milk, some cheeses and ghee (the clarified butter used in Indian cooking).
- Plant sterols inhibit the absorption of cholesterol from the gut, and sitostanol ester, a derivative of sitosterol, may have a useful cholesterol-lowering effect when incorporated into margarine and other foods.

The evidence in favour of both of these propositions is very substantial, and many would say overwhelming. However, there are several apparent weak links in the chain of evidence supporting the diet-heart hypothesis and the dietary guidelines that have evolved from it. Some examples are listed below.

- Some populations appear to deviate from the usual trend of increasing saturated fat consumption associated with increasing average plasma cholesterol concentrations and increasing mortality from coronary heart disease.

- There is a lack of correlation between plasma cholesterol concentration and the level of saturated fat in the diet when they are measured in individuals from within a population.
- Many large-scale, cholesterol-lowering, intervention trials have failed to produce any measurable reduction in overall mortality in the test group.

In the next section, the evidence that underpins the diet-heart hypothesis is briefly summarized and explanations are offered for these apparent anomalies.

- An elevated plasma cholesterol concentration causes increased risk of atherosclerosis and coronary heart disease. The LDL fraction of plasma cholesterol is atherogenic, whereas the HDL fraction is anti-atherogenic.
- The composition of dietary fats is a key determinant of total plasma cholesterol concentration and its distribution between the different lipoprotein fractions in plasma.

REVIEW OF THE EVIDENCE FOR THE DIET–HEART HYPOTHESIS

Descriptive epidemiology

Cross-cultural comparisons reveal a general tendency for plasma cholesterol concentration and coronary heart disease mortality to rise with increasing saturated fat and cholesterol consumption (e.g. Artaud-Wild *et al.*, 1993). In general, cross-cultural studies tend to support the diet–heart hypothesis, but there are a number of exceptions to these general trends, such as those listed below.

- The Masai of East Africa eat a diet based upon milk, meat and blood, which is high in saturated fat and cholesterol, and yet they are free of advanced atherosclerosis and their rates of coronary heart disease are very low. The traditional Masai are a fit and lean population and this could explain their protection from the effects of their apparently atherogenic diet. It has also been suggested that the presence of large amounts of fermented milk (yoghurt) may be significant (McGill, 1979, contains a referenced summary of early studies on the Masai). Some fermented milk products are now marketed as **functional foods** (see Chapter 17).
- Some studies have reported that farmers have lower levels of coronary heart disease than non-farmers living in the same locality, despite having similar fat intakes and higher plasma cholesterol concentrations. These studies also find farmers to be more active and fitter and to smoke and drink less than non-farmers (e.g. Pomrehn *et al.*, 1982).
- Recorded rates of coronary heart disease in France are low in comparison to other European countries with similar plasma cholesterol levels, and despite saturated fat intake being relatively

high in France. This is known as the ‘French paradox’. One popular explanation is that high intakes of the antioxidants in red wine might prevent the oxidation of lipid infiltrates into artery walls and thus reduce the development of atherosclerotic lesions (Renaud and de Lorgeril, 1992). Low intakes of vegetables, which are more mundane sources of antioxidants, are also found in some of the countries that have unexpectedly high rates of coronary heart disease (Artaud-Wild *et al.*, 1993). It is also possible that this paradox is a measurement artefact caused by under-recording of the coronary heart disease mortality in France.

- McKeigue *et al.* (1985) noted high rates of coronary heart disease amongst British Asians, despite their having lower saturated fat and cholesterol intakes and lower blood cholesterol concentrations than the white population. British Asians also have a much higher prevalence of abdominal obesity, insulin resistance and type 2 diabetes than white Britons, and this probably explains their high rates of coronary heart disease.

Dietary saturated fat intake is clearly only one of several factors that affect plasma cholesterol concentration, and plasma cholesterol is only one factor that governs the progression of atherosclerosis and the risk of coronary heart disease. The study of Morris *et al.* (1980) illustrates this point. In an 8-year cohort study of male British civil servants, they found very low levels of fatal first coronaries in men who did not smoke and who participated in vigorous leisure-time activity. The rates were more than six times higher in cigarette smokers who did not participate in vigorous leisure-time activity. This study took no account of diet or plasma cholesterol and the data were collected before the current nutrition education campaigns aimed at lowering dietary fat intakes had gained much momentum in the UK.

A number of factors that are not part of the diet-heart hypothesis in figure 11.5 can influence plasma cholesterol concentration, the development of atherosclerotic lesions and the risk of coronary heart disease. These factors could obscure the relationship between dietary fats, plasma cholesterol and coronary heart disease risk. Some of these factors have been mentioned above, and selected examples are listed below.

- Plasma total cholesterol is a crude indicator of coronary heart disease risk. It is the relative amounts of LDL (causative) and HDL (protective) that are important.
- Being overweight or obese increases the risk of coronary heart disease, particularly if much of the excess fat is stored in the abdominal cavity.
- Insulin resistance and type 2 diabetes (both linked to abdominal obesity) also increase the risk of coronary heart disease.
- High levels of physical activity and fitness protect against coronary heart disease, perhaps partly mediated by a rise in HDL concentration.
- High blood pressure accelerates atherosclerosis and increases the risk of coronary heart disease. The effect of high blood pressure on atherosclerosis is independent of plasma cholesterol level.
- Some substances increase the oxidative damage to LDL in artery walls and so accelerate the process of atherosclerosis, e.g. some constituents of cigarette smoke. According to Steinberg *et al.* (1989), high levels of LDL only produce atherosclerosis if they are oxidized, which leads to the infiltration and activation of macrophages in the artery wall. Some substances act as antioxidants and may inhibit the oxidation of LDL in arteries and so inhibit the process of atherosclerosis, e.g. **carotenoids** found in coloured fruits and vegetables and flavonoids in red wine (see Chapter 12).
- Soluble fibre (NSP) may have a plasma cholesterol-lowering effect.
- Moderate intakes of alcohol may afford some protection against coronary heart disease, perhaps partly mediated by a rise in plasma HDL.
- Long-chain ω -3 polyunsaturated fatty acids may afford some protection against coronary heart disease.

From a health promotion viewpoint, undue emphasis placed upon the P:S ratio of dietary fats and plasma cholesterol may deflect attention away from

other, perhaps even more beneficial, changes in lifestyle. How significant is a moderately high saturated fat diet or perhaps even moderately elevated plasma cholesterol concentration in lean, active, non-smoking, moderate drinking men who eat good amounts of fruit and vegetables? This question may be even more pertinent for women. How significant are reductions in saturated fat intakes for men who remain overweight, inactive, eat little fruit or vegetables and who smoke and drink heavily? Replacing butter with margarine and frying in oil rather than lard are relatively painless changes that may offer reassurance or salve the conscience, but are unlikely to compensate for failure to alter any of these other risk factors.

Some other descriptive epidemiological evidence that has been used to support the diet-heart hypothesis is summarized below.

- Migrants and their descendants who move to areas where saturated fat intake is higher than in their country of origin have increased rates of coronary heart disease, e.g. Japanese migrants to the USA and British Asians.
- White vegetarians and Seventh Day Adventists in the USA and UK have lower rates of coronary heart disease than the whole population. They also tend to have lower intakes of saturated fat.
- Marked changes in a population's consumption of saturated fat are often associated with corresponding changes in coronary heart disease incidence. In Britain, wartime rationing led to decreased saturated fat consumption and decreased rates of coronary heart disease. In Japan, marked increases in milk and meat consumption since the 1950s have been associated with marked increases in the prevalence of coronary heart disease (note also that Japanese life expectancy has increased markedly since the 1950s).

In addition, the observation that atherosclerotic lesions of both animals and humans contain high concentrations of cholesterol adds weight to the belief that high plasma cholesterol helps to precipitate atherosclerosis.

Experimental studies

Numerous short-term experimental studies have shown that plasma cholesterol concentration, primarily

the LDL fraction, is alterable by manipulating the intake of dietary lipids (e.g. Keys *et al.*, 1959; Mattson and Grundy, 1985; Mensink and Katan, 1990; Yu *et al.*, 1995). Mangiapane and Salter (1999) give a concise, referenced review of studies between 1959 and 1999 that have investigated the relationship between dietary fats and plasma lipoproteins. There is general agreement that an increased intake of saturated fat tends to increase the plasma cholesterol, whereas replacement of saturated with polyunsaturated fat tends to lower the cholesterol concentration. The probable effects of dietary cholesterol, monounsaturated fatty acids and *trans* fatty acids were discussed earlier in the chapter. Susceptibility to dietary manipulation of plasma cholesterol concentration varies amongst individuals and there is a general trend for those with the higher baseline levels to be more responsive to dietary change (Keys *et al.*, 1959). Despite the absence of any demonstrable effect upon total mortality, most of the early cholesterol-lowering dietary intervention studies did, nonetheless, produce significant reductions in plasma cholesterol in the test group (Oliver, 1981). However, according to Oliver (1981), plasma cholesterol was usually reduced by less than 10% in free-living subjects, despite some very major dietary modification and even drug use in many of these early studies.

'Experiments of nature'

People with several inherited hyperlipidaemias have a high propensity to premature coronary heart disease (see Mangiapane and Salter, 1999, for a useful summary). In **familial hypercholesterolaemia** (Fredrickson type IIa), the primary genetic defect is in the LDL-receptor gene. This genetic defect leads to reduced levels of functional LDL receptors and a marked rise in plasma LDL cholesterol concentration. This condition is associated with marked atherosclerosis and greatly increased risk of premature coronary heart disease. This provides persuasive evidence of a direct causal link between high plasma LDL, atherosclerosis and coronary heart disease.

Cohort studies

Cohort studies, whether within or between populations, generally indicate that, at least for young and

middle-aged men, high serum cholesterol concentration is predictive of subsequent risk of death from coronary heart disease over the succeeding 5–20 years. In the Framingham study, conducted in Framingham, Massachusetts, there was a clear, progressive increase in coronary heart disease mortality in all age groups of men with increasing plasma cholesterol level (see Figure 1.1 in Chapter 1; and Mangiapane and Salter, 1999, for a useful summary of the Framingham study). In the famous Seven Countries Cross-cultural Cohort Study, there was a high correlation between dietary saturated fat intake and both plasma cholesterol concentration and incidence of coronary heart disease (summarized in Oliver, 1981).

In general, the relationship between total plasma cholesterol and coronary heart disease is most convincing in middle-aged men and is much weaker in pre-menopausal women (e.g. Crouse, 1989) and elderly men (Shipley *et al.*, 1991). This has caused some writers to question whether universal cholesterol lowering was justified, particularly for women (e.g. in influential editorials in the *Lancet* and the journal *Circulation*, see Dunnigan, 1993; Hulley *et al.*, 1992). The lack of demonstrated holistic benefit in most cholesterol-lowering intervention trials prior to this time was an important reason why these authors recommended such caution. As will be seen below, some more recent intervention trials have provided much stronger evidence that lowering LDL cholesterol can bring substantial net benefits. There are now good theoretical reasons to expect that total plasma cholesterol will be a relatively crude indicator of coronary heart disease risk. Measurement of specific lipoprotein fractions in blood should give a more sensitive indication of coronary heart disease risk. In the Framingham study, total plasma cholesterol was only a weak indicator of coronary heart disease risk in pre-menopausal women and older men, but there was a very strong negative association (i.e. protective effect) between HDL concentration and coronary heart disease risk in these groups.

One of the gaps in the chain of evidence supporting the diet-heart hypothesis has been the failure to find any significant association between measures of dietary fat or saturated fat intake and plasma cholesterol concentration in most within-population studies. Oliver (1981) uses data from the Tecumseh study to illustrate this general finding: dietary intakes of fat, saturated fat, cholesterol and dietary P:S ratio

were not significantly different when those in the highest, middle and lowest **tertiles** (thirds) for serum cholesterol concentrations were compared. In *The Dietary and Nutritional Survey of British Adults*, Gregory *et al.* (1990) found no significant correlation between any measures of dietary fat and serum LDL cholesterol concentration in men, except for a very weak ($r = 0.08$) one with dietary cholesterol; all correlation coefficients were less than 0.1. In women, they found a weak but significant correlation with the percentage of energy from saturated fat and serum LDL cholesterol concentration, but only 2% (r^2) of the difference in women's serum cholesterol could be explained by variation in their percentage of dietary energy from saturated fat.

Dietary differences in individual fat consumption do not appear to be an important factor in determining the position of any individual within the population distribution of plasma cholesterol concentrations. As already stated in Chapter 3, this may be because a host of other factors affect an individual's serum cholesterol concentration, such as:

- genetic variability in plasma cholesterol concentration, including variation in genetic susceptibility to dietary fats
- other dietary influences upon plasma cholesterol
- other lifestyle influences upon plasma cholesterol, e.g. exercise, age, smoking, alcohol, body mass index.

This lack of a significant association is thus not incompatible with the proposal that dietary fat composition is a major determinant of a population's average plasma cholesterol concentration. It is still expected that widespread use of a diet with a higher P:S ratio and less cholesterol will lower the plasma cholesterol concentrations of adherents and lead to a downward shift in the distribution of plasma cholesterol of the population and thus to a reduction in average plasma cholesterol concentration. This general concept is discussed more fully in Chapter 3, using the data in Figure 3.11 to illustrate it.

Intervention trials

Numerous trials of a variety of cholesterol-lowering interventions were carried out between 1960 and

1990. None of these produced a significant reduction in total mortality in the test group. Many did not even produce significant reductions in cardiovascular mortality, although most succeeded in reducing plasma cholesterol to some extent. Some reported a non-significant fall in cardiovascular mortality that was offset by a non-significant rise in mortality from 'other causes'. In several of these early trials, the overall incidence of coronary heart disease was reduced, largely as a result of a fall in the number of non-fatal myocardial infarctions (McCormick and Skrabanek, 1988; Oliver, 1981, 1991). Some early intervention trials that involved the use of cholesterol-lowering drugs actually reported increased mortality in the test group. One of the most spectacular of these 'failed' intervention trials was the Multiple Risk Factor Intervention Trial (1982), which is discussed in Chapter 3.

Smith *et al.* (1993) conducted a combined (meta) analysis of 35 randomized, controlled trials of cholesterol-lowering interventions published prior to 1993. They concluded that net benefit (i.e. reduced total mortality) from cholesterol-lowering intervention was confined to those at very high risk of coronary heart disease. Only those at greater risk than even asymptomatic patients under 65 years with hypercholesterolaemia were reported to derive demonstrable net benefit from cholesterol-lowering interventions. Cholesterol-lowering drugs were used in more than two-thirds of the trials surveyed by these authors. They concluded that these drugs should be reserved for the small proportion of people at very high risk of death from coronary heart disease, perhaps only those with both existing coronary heart disease and hypercholesterolaemia.

The apparent failure of these early intervention trials was explained by suggesting that the sample sizes in these trials were too small to obtain statistically significant reductions in overall mortality. Dunnigan (1993) estimated that a trial involving 80 000 people over a 5-year period would be necessary to demonstrate whether there was any beneficial effect of cholesterol lowering on all-cause mortality, even in patients with asymptomatic hypercholesterolaemia under the age of 65 years. This is, of course, another way of saying that, at least within the time scale of an intervention trial, the difference in risk of dying is only minutely affected by being in the intervention or control group. As a result of such findings, many scientists questioned the benefits of interventions and health promotion campaigns that were aimed at lowering average plasma

cholesterol concentrations, e.g. McCormick and Skrabanek (1988) and Oliver (1981, 1991).

Some more recent intervention trials using a new type of cholesterol-lowering drugs called statins have provided strong evidence that lowering plasma cholesterol can have net beneficial effects both in patients

with existing heart disease and in those with asymptomatic hypercholesterolaemia. These drugs inhibit the key enzyme in cholesterol biosynthesis (HMG CoA reductase) and produce a substantial and sustained reduction in LDL cholesterol, with few side-effects. In a 5-year study involving more than 4000

- The evidence supporting the diet-heart hypothesis is both impressive and varied.
- Populations with high saturated fat and cholesterol intakes tend to have higher plasma cholesterol concentrations and higher death rates from coronary heart disease than those with low saturated fat intakes.
- Some vegetarian groups in the UK and USA with low intakes of saturated fat and cholesterol have lower rates of coronary heart disease than the rest of the population.
- Marked changes in a population's consumption of saturated fat have been accompanied by changes in coronary heart disease rates that are consistent with the diet-heart hypothesis.
- People who migrate from regions of low saturated fat consumption to regions of high consumption tend also to show increased rates of coronary heart disease.
- Experimental or therapeutic manipulations of dietary fat composition produce changes in plasma lipoprotein profiles that are consistent with the diet-heart hypothesis.
- People with genetic disorders that raise their plasma LDL cholesterol are at greatly increased risk of coronary heart disease.
- In cohort studies using single or several populations, a high plasma cholesterol concentration is associated with an increased risk of coronary heart disease.
- Many cholesterol-lowering trials using drugs or diet have resulted in significant reductions in the combined number of fatal and non-fatal coronary events.
- Recent intervention trials using the statin drugs, which have a very potent cholesterol-lowering effect, have produced significant reductions in both deaths from coronary heart disease and total mortality.
- These statin trials suggest that population-wide reductions in plasma cholesterol concentration over the whole life span would increase healthy life expectancy.
- A few anomalous observations are superficially inconsistent with the diet-heart hypothesis:

within populations, there is little correlation between plasma cholesterol and saturated fat intakes

some populations with high saturated fat intakes or high plasma cholesterol concentrations seem to have lower rates of coronary heart disease than might be anticipated until 1994, cholesterol-lowering intervention trials with diet and/or drugs failed to produce significant reductions in total mortality.

- There are plausible explanations for each of these anomalies:

many factors influence plasma cholesterol, and any individual's saturated fat intake is only a minor determinant of where he or she lies in the population distribution of plasma cholesterol concentrations

high levels of fitness and activity, high intakes of antioxidants or consumption of fermented milk products are some of the factors that have been suggested might protect some populations from coronary heart disease.

recent trials using statins have lowered both plasma cholesterol and total mortality.

Scandinavians with existing CHD, the drug simvastatin reduced plasma LDL cholesterol by 35% and reduced total mortality by 30% as a result of reduced cardiovascular mortality (SSSS Group, 1994). In a study of 6500 Scottish men with asymptomatic hypercholesterolaemia, the drug pravastatin reduced LDL cholesterol by 26%, heart attacks by a third and total mortality by 22% (Shepherd *et al.*, 1995). These latter authors estimated that treating 1000 healthy but hypercholesterolaemic middle-aged men with pravastatin for 5 years would result in 20 fewer non-fatal myocardial infarctions and nine fewer deaths, as well as less heart surgery and fewer investigative procedures. These two studies involve the use of drugs, but, nevertheless, they do support the belief that dietary measures that lower average population plasma cholesterol concentration over the whole life span will shift the age distribution of deaths and disability from coronary heart disease upwards and lead to increased healthy life expectancy.

FISH OILS

Overview

A series of papers published in the 1970s by Bang, Dyerberg and their colleagues reported that Greenland Eskimos had low rates of coronary heart disease and low rates of inflammatory joint disease when compared to Danes. The Eskimos ate a traditional diet high in seal meat, whale meat and fish (e.g. Dyerberg and Bang, 1979). This low rate of coronary heart disease was despite a diet with similar total fat content to that of the Danes and despite the fact that most of the fat came from the animal kingdom. The lipid of the Eskimo diet differs from that of the Danish diet in two major respects: firstly, it is much lower in saturated fatty acids and, secondly, it has higher levels of ω -3 polyunsaturated fatty acids (reviewed by Kroinhouwt, 1990).

Fish oils and other marine oils (e.g. from seals and whales) contain large amounts of long-chain ω -3 polyunsaturated fatty acids. The three longest chain members of this series – eicosapentaenoic acid, EPA ($20:5_{\omega-3}$), docosapentaenoic acid, DPA ($22:5_{\omega-3}$) and docosahexaenoic acid, DHA ($22:6_{\omega-3}$) – account for around 20% of the fatty acids in many oils from marine fish. The ω -3 fatty acids, because of their

extra double bond, have greater fluidity at low environmental temperatures than the corresponding ω -6 fatty acid – this may explain their prominence in the fat of cold-water mammals and fish. These acids enter the marine food chain via their synthesis by marine plants and plankton, which then serve as food for higher marine organisms.

These long-chain ω -3 polyunsaturates are widely believed to be the active agents that reduce the risk of coronary heart disease and inflammatory joint diseases. It should be noted that some fish liver oil products contain high levels of vitamins A and D. High intakes of some of them might lead to the consumption of undesirably large amounts of these vitamins. Both vitamin A and D have well-documented toxic effects when taken in large excess.

The reports on the health and mortality of Eskimos served to focus scientific attention upon the protective and therapeutic potential of fish oils and the ω -3 polyunsaturates. The increased scientific interest in fish oils may be gauged by the output of research papers on them: less than ten per year in the 1970s, but rising sharply during the 1980s to a level of several hundred per year in the early 1990s (see Simopolous, 1991). Fish oils can be made available on prescription in the UK for the treatment of certain hyperlipidaemias and for patients with a history of coronary heart disease. The public perception of fish oils as health promoting is illustrated by the massive amount of shelf space occupied by fish-oil products in health food shops, pharmacies and even supermarkets. In 1992, fish oils accounted for over a quarter of the UK market for natural food supplements.

Buttriss (1999) has recently reviewed the health aspects of ω -3 fatty acids. It is assumed in this section that any health benefits of fish oils are due to their ω -3 fatty acid content and, equally, that ω -3 fatty acids from non-fish sources are likely to have similar effects.

Possible protective mechanisms

There are a number of mechanisms by which long-chain ω -3 fatty acids have been suggested to reduce the risk of coronary heart disease and inflammatory joint diseases, and some of these are discussed below.

EFFECTS ON THE PRODUCTION OF EICOSANOIDS

Long chain ω -3 fatty acids inhibit the conversion of linoleic acid to arachidonic acid. They competitively inhibit the rate-limiting enzyme that converts linoleic acid ($18:2 \omega_6$) to gamma linolenic acid ($18:3 \omega_6$) – see scheme for essential fatty acid metabolism in Figure 11.2. The ω -3 equivalent of arachidonic acid (EPA) then partially replaces arachidonic acid as a substrate for the enzyme systems that convert arachidonic acid into active eicosanoids, thus producing eicosanoids with activities different from those produced from arachidonic acid (summarized in Figure 11.6).

The thromboxane TXA-2 is produced from arachidonic acid and this increases blood platelet aggregation and the tendency for blood to clot; the ω -3 equivalent is TXA-3, which has much less activity than TXA-2. The overall effect of partially replacing TXA-2 with TXA-3 is a reduction in the tendency of platelets to aggregate and blood to clot. The prostacyclin PGI-2 is produced in vascular endothelial cells from arachidonic acid and it has an anti-aggregating effect on platelets. The ω -3 equivalent in this case, PGI-3, has only marginally lower activity and so partial replacement of PGI-2 with PGI-3 has little effect

on the anti-aggregating system. The overall effect of these changes shown in Figure 11.6 would be predicted to reduce the risk of thromboses forming in blood vessels. This would reduce the risk of occlusive strokes (due to blockage of cerebral vessels with a clot) and myocardial infarction (heart attacks due to clots lodging in the coronary vessels). Eskimos have extended bleeding times and are prone to frequent and long-lasting nosebleeds and have high stroke mortality (probably due to increased levels of cerebral haemorrhage). In short-term human and animal experiments, high consumption of fish oils has been found to reduce the tendency of blood to clot and to extend the bleeding time (Anon, 1984; Sanders, 1985).

Eicosanoids are also known to be key mediators of inflammation: the leukotriene LTB-4 is produced from arachidonic acid and is a promoter of inflammation. The ω -3 equivalent is LTB-5, which is much less active. The overall effect of fish oil consumption will be the partial replacement of arachidonic acid with EPA as substrate for the leukotriene-producing enzyme system, leading to partial replacement of LTB-4 with LTB-5 and thus a reduced inflammatory response. These changes may reduce the damage caused by the inflammatory response after a myocardial infarction and may also explain the suggested beneficial effects on inflammatory diseases such as arthritis and eczema. In a controlled, secondary intervention trial of fish oils in patients who had previously suffered a myocardial infarction, fish oils significantly reduced both cardiovascular and total mortality, whereas the other treatments (low-fat diets or increased cereal fibre) did not (Burr *et al.*, 1991). In this trial, the fish oil patients did not have significantly fewer re-infarcts, but these new infarcts were less likely to be fatal (consistent with reduced post-infarct inflammatory damage). A more recent intervention trial in Italy found reductions in both total mortality and non-fatal infarction in men receiving large supplements of ω -3 fatty acids (GISSI, 1999).

Kremer *et al.* (1985) reported that high intakes of EPA appeared to improve the clinical features of rheumatoid arthritis and the levels of LTB-4 were reduced by the treatment.

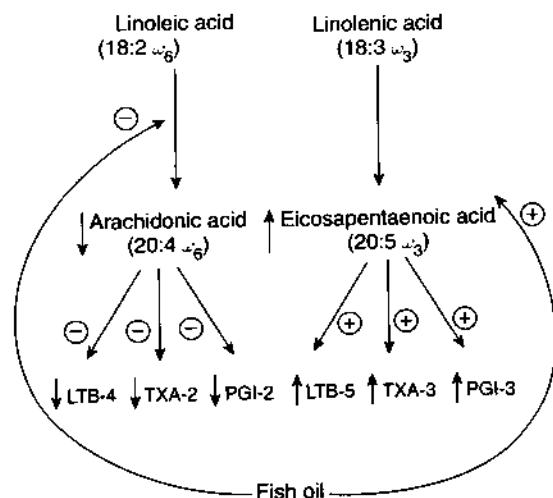


Figure 11.6 Effects of fish-oil consumption on the production of eicosanoids that affect platelet aggregation and inflammation. LTB-4 promotes inflammation; LTB-5 has low activity. TXA-2 causes platelet aggregation; TXA-3 has low activity. PGI-2 and PGI-3 both have anti-aggregating effects on platelets. Net effect of fish oil is reduced inflammation and reduced platelet aggregation.

EFFECTS ON BLOOD LIPOPROTEIN PROFILES

Fish oils reduce both fasting and post-prandial blood TAG levels. They also raise the HDL cholesterol

concentration, and Greenland Eskimos have been reported to have higher HDL levels than those living in Denmark and consuming less marine oil. At normal levels of consumption, their effect on LDL cholesterol is negligible.

EFFECTS ON PLASMINOGEN ACTIVATION

Clots are broken down by plasmin, which is formed when the inactive blood protein plasminogen is activated. Experiments using fish oil supplements in patients with coronary artery disease indicate that they lead to the reduced formation of a substance that inhibits the conversion of plasminogen to active plasmin. Clot-dissolving drugs such as streptokinase are often administered to patients in the period immediately after a myocardial infarction in order to reduce the damage done to the heart. It has been suggested that fish oils have a similar effect, although there is little evidence to support this.

EFFECTS UPON HEART RHYTHM

Irregular electrical activity (arrhythmia) in diseased or damaged heart muscle can have fatal consequences. There is evidence that changes in the balance of ω -3 and ω -6 fatty acids in the heart resulting from increased intakes of ω -3 fatty acids may reduce heart rate variability and arrhythmia and so reduce the risk of death following a heart attack (see Buttriss, 1999).

Review of the evidence of the possible therapeutic and protective effects of fish oils

There is compelling evidence that in short-term, controlled experiments, fish oils affect the blood lipoprotein profile in ways that are currently considered to be beneficial. They also reduce the tendency of blood to clot in such trials and thus would presumably reduce the risk of thromboses forming within the circulation. There is also convincing evidence that, in people with established coronary heart disease, regular consumption of fatty fish or fish oil supplements produces overall benefits, as measured by total mortality in the few years immediately after first infarction (Burr *et al.*, 1991; GISSI, 1999). Eskimos and other populations con-

suming very large amounts of fish have a reduced tendency for blood to clot, prolonged bleeding times and a low incidence of coronary heart disease.

The epidemiological evidence for fatty fish consumption reducing the risk of coronary heart disease or, still further, improving overall life expectancy is rather less convincing. The work involving Eskimos and other populations consuming very large amounts of fish has been some of the most positive of the epidemiological evidence in favour of fish oils having protective effects against the development of coronary heart disease (see Kromhout, 1990; Buttriss, 1999). There is no convincing evidence of an independent relationship between fatty fish consumption and a reduced risk of coronary heart disease in cross-population studies. Those comparisons that have been made have used total fish consumption as the independent variable (a crude predictor of total ω -3 fatty acid consumption), and the weak negative association with coronary heart disease mortality was very dependent upon the inclusion of the Japanese. In the famous Seven Countries Study, there was no significant association between fish consumption and coronary heart disease mortality; some of the populations with negligible fish consumption in inland Yugoslavia had amongst the lowest rates of coronary heart disease, whereas some of those with the highest fish consumption in Finland had amongst the highest rates of coronary heart disease. Fish consumption was related to subsequent mortality in a 20-year cohort study of 850 men in Zutphen, Holland. In this cohort, death from coronary heart disease was inversely related to fish consumption; it was 50% lower in men who ate 30 g of fatty fish per day (equivalent to about 2 servings per week) compared with those who ate no fish at all. Buttriss (1999) has reviewed several other prospective cohort studies and concluded that the majority of these studies report a decrease in the risk of heart disease as fish consumption increases. She further adds that the effects of fish eating are most pronounced in populations that have low habitual fish consumption.

In general, the epidemiological evidence relating fatty fish consumption to reduced coronary heart disease risk is sporadic, but this might not be unexpected, even if the regular consumption of long-chain ω -3 fatty acids is protective, because:

- total fish consumption is a crude indicator of ω -3 fatty acid intake

- the relationship between saturated fat intake and coronary heart disease might well obscure any lesser relationship between fish consumption and coronary heart disease.

At present, the overall evidence seems consistent with the recommendation that fatty fish be included in healthy and diverse diets. A quantitative recommendation of two portions of fatty fish per week has been suggested. A few years ago, fish oil would have been regarded as just another source of fat in our already excessively fatty diets; it now has a much more positive image. There does also seem to be a case for specifically recommending increases in fatty fish consumption or even the use of fish oil supplements by those with established coronary heart disease or some types of hyperlipidaemia. This evidence does not yet warrant recommending the universal consumption of fish oil supplements. Some of the favourable epidemiological evidence for fish oils could be interpreted as their being beneficial in preventing coronary heart disease when they are an

alternative to saturated fat rather than being beneficial *per se*. Some of the most convincing evidence in the fish oil story comes from very reductionist studies of the effects on particular physiological systems or symptoms or from more holistic studies with high-risk groups. The case has been made in Chapter 1 that using such studies as the basis for dietary recommendations for the whole population is fraught with danger. It was also argued in Chapter 1 that an apparently beneficial effect on some disease risk marker only becomes a real benefit if it is ultimately translated into improved overall life expectancy or reduced morbidity.

Some potential dangers of consuming large amounts of fish oil or ω -3 fatty acids are listed below, although it must be emphasized that these are theoretical or speculative risks.

- High fish oil intakes reduce blood clotting and increase bleeding time. This should reduce the risk of thrombosis, but may not be universally beneficial, e.g. for those at risk of cerebral haemorrhage.

- Fish oils are the major dietary source of long-chain ω -3 fatty acids.
- British consumers are advised to eat two portions of oily fish per week and to double their intake of long-chain ω -3 polyunsaturates.
- Low rates of coronary heart disease and inflammatory joint disease amongst Greenland Eskimos focused scientific attention upon the potential health benefits of eating oily fish and the ω -3 polyunsaturates.
- Long-chain ω -3 fatty acids inhibit the production of arachidonic acid and so alter the balance of eicosanoid production in ways that should reduce platelet aggregation, blood clotting and inflammation.
- Fish oils reduce blood triacylglycerol concentrations and increase HDL levels.
- Increased consumption of ω -3 fatty acids may reduce the risk of potentially fatal cardiac arrhythmia.
- Eskimos and other populations consuming large amounts of fish have reduced blood clotting, extended bleeding times and low risk of coronary heart disease.
- Several within-population cohort studies indicate that high fish consumption is associated with reduced risk of coronary heart disease.
- Several secondary intervention trials have reported that fish oil consumption reduces the total mortality of patients with existing coronary disease.
- The widespread consumption of large doses of fish oils as supplements may have potentially adverse effects for some people, e.g. those with impaired blood clotting, immunosuppression and increased susceptibility to oxidative damage by free radicals.

- Fish oils seem to have some immunosuppressive effect because of their effects upon eicosanoid production.
- Omega-3 fatty acids are highly susceptible to oxidation. Foods high in these acids are prone to become rancid, and oxidative damage to them after absorption may have adverse consequences.
- Some sources of fish oil may have too much vitamin A and D.

Vitamins and antioxidants

Some general concepts and principles
The dietetically important vitamins

277
281

Free radicals and antioxidants

297

SOME GENERAL CONCEPTS AND PRINCIPLES

What is a vitamin?

Vitamins are a group of organic compounds that are indispensable to body function. Some vitamins cannot be synthesized in the body and others cannot be synthesized in sufficient quantity to meet the metabolic needs or can only be synthesized from specific dietary precursors. Vitamins or their precursors must therefore be provided in the human diet. Vitamins are only required in small amounts and they do not serve as sources of dietary energy. Most vitamins or their derivatives now have clearly defined roles in particular biochemical pathways or physiological processes, as illustrated by the following examples.

- The active principle of **rhodopsin** (the light-sensing pigment in the rods of the retina) is a derivative of vitamin A. One of the early consequences of vitamin A deficiency is night blindness.

- **Thiamin pyrophosphate** (TPP) is an essential cofactor for several key enzymes in metabolism, including pyruvic oxidase, the enzyme responsible for the conversion of pyruvic acid to acetyl coenzyme A in carbohydrate metabolism (see Chapter 5). In thiamin deficiency, the metabolism of carbohydrates is progressively blocked at this step and the function of other TPP-dependent enzymes is similarly impaired.
- Niacin (nicotinic acid) is a component of the important coenzymes **nicotinamide adenine dinucleotide** (NAD) and the phosphorylated derivative NADP. NAD and NADP serve as hydrogen acceptors and donators in numerous oxidation and reduction reactions. They are essential in many human biochemical pathways.
- Vitamin D (**cholecalciferol**) is the precursor of a hormone produced in the kidney, **1,25-dihydroxy cholecalciferol** (1,25-DHCC) or **calcitriol**. This hormone stimulates the synthesis of proteins essential for calcium absorption in the gut. In vitamin D deficiency (**rickets**), children absorb insufficient calcium for normal skeletal development.

- Vitamins are organic compounds that are only required in small amounts in the diet and do not serve as sources of energy.
- The precise biochemical functions of many vitamins are now known.
- Most vitamins cannot be synthesized at all by the body or can only be synthesized from specific dietary precursors.

Classification

Vitamins are divided into two broad categories according to their solubility characteristics. Vitamins A, D, E and K are insoluble in water but soluble in lipid or lipid solvents and they are thus classified as the **fat-soluble vitamins**. The B group of vitamins and vitamin C are the **water-soluble vitamins**. The term B vitamins actually describes a group of eight different vitamins which were originally grouped together as the B complex because they were found together in, for example, yeast extract. The B vitamins are:

- thiamin (B_1)
- riboflavin (B_2)
- niacin (B_3)
- vitamin B_6 (pyridoxine and related compounds)
- vitamin B_{12} (cobalamin)
- folic acid (folate, folacin)
- pantothenic acid
- biotin.

In general, the fat-soluble vitamins are stored in the liver and thus daily consumption is not required, provided that average consumption over a period of time is adequate. Excesses of these vitamins are not readily excreted and so, if intakes are very high, toxic overload is possible. Concentrated supplements are the most likely cause of toxic overload, but some very rich food sources can also produce symptoms of toxicity. For example, the livers of marine mammals and fish may contain toxic amounts of vitamins A or D and even liver from farm animals may contain enough vitamin A to cause birth defects if eaten by

pregnant women. As fat-soluble vitamins are normally absorbed from the gut and transported along with the fat, fat malabsorption or disorders of fat transport can increase the risk of fat-soluble vitamin deficiency. For example, in cystic fibrosis, the lack of pancreatic fat-digesting enzymes may precipitate fat-soluble vitamin deficiencies (see Chapter 16).

The water-soluble vitamins leach into cooking water and this can result in very substantial losses during food preparation. Stores of these vitamins tend to be smaller than those of the fat-soluble vitamins and excesses tend to be excreted in the urine. A more regular supply is required than for the fat-soluble vitamins, and symptoms of deficiency tend to occur much sooner (vitamin B_{12} is a clear exception to this generalization). Toxicity is much less likely because of their water solubility, and urinary excretion and is all but impossible from natural foodstuffs.

Vitamin deficiency diseases

Dietary deficiencies of vitamins give rise to clinically recognizable deficiency diseases that are cured by administering the missing vitamin. Vitamin discoveries in the first half of the twentieth century greatly increased the perception of good nutrition as a vital prerequisite for good health. Common and frequently fatal diseases could be cured by the administration of small amounts of vitamins or vitamin-containing foods. These cures, probably unreasonably, raised the expectation of the ability of nutrition to cure or prevent other non-deficiency diseases. In many cases, the symptoms of these deficiency diseases, or at least

- Vitamins are subdivided into those that are fat soluble (A, D, E and K) and those that are water soluble (B group and C).
- Fat-soluble vitamins are stored in the liver, so clinical deficiency states usually manifest only after chronic deprivation.
- Fat-soluble vitamins are not readily excreted and so toxic overload is possible, even from rich food sources.
- Fat malabsorption can precipitate fat-soluble deficiency diseases.
- Substantial losses of water-soluble vitamins can occur by leaching into cooking water.
- Stores of most water-soluble vitamins are relatively low and so deficiency states manifest after relatively short periods of deprivation.
- Water-soluble vitamins are excreted in urine and so toxic overload is unlikely.

some of them, are readily explicable from knowledge of the cellular functions of the vitamin, as illustrated by the examples below.

- The impaired night vision associated with vitamin A deficiency is explained by its role as a component of visual pigments.
- The inadequate bone mineralization seen in rickets occurs because vitamin D is required to produce the hormone calcitriol, which is required for calcium absorption and normal bone mineralization.
- The breakdown of connective tissue seen in scurvy can be explained by the role of vitamin C in the production of the structural protein collagen.
- The megaloblastic anaemia seen in vitamin B₁₂ and folic acid deficiency occurs because these two vitamins are required for DNA synthesis and thus for cell division and red cell production in bone marrow.
- The haemorrhagic disease seen in vitamin K deficiency in newborn babies occurs because this vitamin plays an essential role in the synthesis of prothrombin and several other blood-clotting factors.

Several of the deficiency diseases have, in the past, represented major causes of human suffering, such as those listed below:

- beriberi – in Japan, China and several other rice-eating parts of the Far East
- pellagra – in Southern Europe, North Africa, South Africa and even some southern states of the USA where maize was the dominant staple
- rickets – was rife amongst poor children living in industrial cities of Britain and northern Europe
- scurvy – was a major hazard for sailors and others undertaking long sea voyages by sail.

These deficiency diseases are now rare in industrialized countries and are generally confined to those with predisposing medical or social problems, e.g. alcoholics and the frail elderly. In *The Dietary and Nutritional Survey of British Adults*, it was found that average intakes of all of the vitamins exceeded the Recommended Dietary Allowance (RDA) in use at the time (Gregory *et al.*, 1990). However those 2.5% of the sample with the lowest intakes were, in several cases, below the current Estimated Average

Requirement (EAR) (e.g. thiamin for women, riboflavin and folate for men, and vitamin C for both sexes). In a few cases, the recorded intakes of this bottom 2.5% were at or below even the current Lower Reference Nutrient Intake (LRNI) (e.g. vitamin A in both sexes and riboflavin and folate in women). Apparently satisfactory overall averages may obscure the fact that a substantial minority may be receiving seriously inadequate intakes of some vitamins. (Note that there was probably some under-recording in this survey, particularly amongst women.)

Even in developing countries, most of the historically important deficiency diseases are now relatively uncommon, but a few vitamin deficiencies still represent important public health problems in parts of the world, such as the examples listed below.

- Vitamin A deficiency is still widespread amongst children in many parts of the Third World, where it causes blindness and is associated with high child mortality.
- Vitamin D deficiency is an important precipitating factor for osteoporosis and bone fractures amongst elderly housebound people in Britain and some other industrialized countries.
- Lack of folic acid is an important cause of anaemia in several Third World countries.
- Pellagra is still endemic in parts of southern Africa.
- Subclinical thiamin deficiency may still be prevalent in some rice-eating populations.

Precursors and endogenous synthesis of vitamins

Vitamins may be present in food in several different forms and some substances may act as precursors from which vitamins can be manufactured. In a few cases, external supply of the preformed vitamin may make a relatively minor contribution to the body supply of the vitamin. Some examples of these 'alternative' sources of vitamins are given below.

- The plant pigment beta (β)-carotene and several other carotenoids can be converted to vitamin A (retinol) by enzymes in the intestinal mucosa, liver and other human tissues.
- The amino acid tryptophan can be converted to the vitamin niacin. Niacin can be obtained either

- Vitamin deficiencies cause clinically recognizable deficiency diseases that can be cured by vitamin therapy.
- Often, the symptoms of deficiency diseases are explained by their known functions.
- Some deficiency diseases still represent important public health problems in some areas, although in the industrialized countries they are usually confined to those with predisposing social and medical problems.
- Average vitamin intakes of UK adults are above the current RNI, although some individuals may have less than satisfactory intakes.

directly from the vitamin present in food or indirectly by the conversion of tryptophan in dietary protein to niacin.

- The principal source of vitamin D for most people is endogenous synthesis in the skin. Ultraviolet radiation, from sunlight, acts on the substance 7-dehydrocholesterol, which is produced in the skin, and initiates its conversion to cholecalciferol or vitamin D₃.
- Most mammals can convert glucose to ascorbic acid (vitamin C), but primates, guinea-pigs and a few exotic species lack a key enzyme on this pathway and thus they require a dietary supply of this vitamin.

Circumstances that precipitate deficiency

People meeting their energy needs and eating a variety of foods from all of the food groups seldom show clinical signs of vitamin deficiency. Dietary deficiencies arise when either the total food supply is poor or the variety and nutritional quality of the food eaten are low, or both. Vitamin deficiencies may occur as an added problem of general food shortage and starvation. They may also be associated with diets that are very heavily dependent upon a single staple food that is deficient in the vitamin, or be associated with some

other very particular circumstance or dietary restriction. Some examples of circumstances associated with deficiency states are listed below.

- Beriberi (thiamin deficiency) has historically been associated with diets based upon white rice. Rice is inherently low in thiamin, most of which is located in the outer layers of the rice grains, which are removed by milling.
- Pellagra (niacin deficiency) epidemics have usually been associated with diets based upon maize (or sometimes sorghum). Maize protein (zein) is low in the amino acid tryptophan, which can serve as a niacin precursor, and most of the niacin present in maize is in a bound form (**niacytin**) that is unavailable to humans. Sorghum and maize contain large amounts of the amino acid leucine, which inhibits the conversion of tryptophan to niacin.
- Scurvy (vitamin C deficiency) is not a disease of famines, it occurred principally among people living for extended periods upon preserved foods, e.g. those undertaking long sea voyages or taking part in expeditions to places where the fruits and vegetables that normally provide our vitamin C were unavailable.
- Vitamin B₁₂ is naturally confined to foods of animal origin and micro-organisms and so strict vegetarians (**vegans**) are theoretically at risk of

- Some vitamins are present in several forms in the diet and some can be manufactured from other dietary precursors.

- Vitamin deficiency diseases can be a secondary consequence of general food shortage.
- Vitamin deficiencies are usually associated with very specific dietary or other lifestyle circumstances such as the consumption of a restricted range of foods, high alcohol consumption, lack of exposure to sunlight or low absorption of fat from the diet.

dietary deficiency of the vitamin unless they obtain supplementary B₁₂. The usual cause of B₁₂ deficiency is an autoimmune disease, **pernicious anaemia**, that results in damage to the parietal cells of the stomach that produce an **intrinsic factor** that is necessary for the absorption of B₁₂ from the gut.

- Inadequate exposure of the skin to sunlight is regarded as the primary cause of vitamin D deficiency, e.g. in elderly people who are immobile and unable to go outside.
- Primary dietary deficiency of vitamin K is almost never seen, but some anticoagulant drugs (e.g. warfarin) exert their effect by blocking the effects of vitamin K and lead to reduced levels of clotting factors in blood and the risk of bleeding and haemorrhage. Therapeutic overdose of these drugs or the accidental consumption of warfarin-based rodent poisons leads to excessive bleeding and vitamin K acts as an antidote.
- In affluent countries, alcoholics are at particular risk of vitamin deficiencies – most obviously because high alcohol consumption is likely to be associated with low food and nutrient intake. Poor alcoholics spend their limited money upon alcohol rather than food. High alcohol consumption may also increase the requirement for some nutrients and reduce the efficiency of absorption of others. Several of the deficiency diseases listed earlier do occur in alcoholics in affluent countries, and in many cases there are multiple deficiencies. Some of the harmful effects of very high alcohol intakes may be attributable to dietary deficiency rather than to alcohol *per se*. Alcoholics are particularly prone to thiamin deficiency and some neurological manifestations of thiamin deficiency – **Wernicke-Korsakoff syndrome**.
- Very low-fat diets may precipitate fat-soluble vitamin deficiencies. These diets contain low

amounts of fat-soluble vitamins and the absorption of the vitamins may also be impaired by the lack of fat. Poor fat absorption (e.g. in cystic fibrosis and coeliac disease) may also be associated with fat-soluble vitamin deficiencies.

THE DIETETICALLY IMPORTANT VITAMINS

Tables 12.1 and 12.2 summarize some of the important factual information about the dietetically important fat-soluble and water-soluble vitamins respectively, i.e.:

- dietary forms and food sources
- the Dietary Reference Values for adults as an indication of human requirements
- cellular functions
- the effects of deficiency
- the likely risk factors for deficiency.

Readers with limited biochemical background are urged to refer back to Chapter 5 to facilitate their understanding of the significance of particular biochemical lesions that result from vitamin deficiency. In Chapter 3, there is some discussion of the methods used to determine vitamin requirements and assess vitamin status and an indication of normal ranges for these measurements. Table 3.6 contains a summary of some the biochemical tests of vitamin status. The relevant chapters in Groff *et al.* (1995), Garrow and James (1993) and especially Shils *et al.* (1999) are recommended as reference sources for factual information about the vitamins and the pathological consequences of deficiency. Chapters in these texts dealing with individual vitamins are not specifically referenced here.

The sections that follow contain a brief review of the individual vitamins.

Table 12.1 The fat-soluble vitamins: a summary**Vitamin names and dietary forms**

- A **Retinol**; β -carotene and other carotenoids; vitamin A activity is expressed as retinol equivalents
- D **Cholecalciferol** (D_3); **calciferol** (D_2) in supplements and supplemented foods only
- E **α -tocopherol**; other tocopherols have vitamin activity; vitamin E activity expressed as α -tocopherol equivalents
- K **Phylloquinone** from plants; menaquinone produced by intestinal bacteria

Dietary Reference Values

- UK RNI adult male/female; (American RDAs male/female)
- A 700/600 $\mu\text{g}/\text{day}$; (1000/800)
- D None, if exposed to sunlight; (5 $\mu\text{g}/\text{day}$)
- E 'Safe intake' above 4/3 mg/day; (10/8 mg/day)
- K 'Safe intake' 1 $\mu\text{g}/\text{kg}/\text{day}$; (80/65 $\mu\text{g}/\text{day}$)

Dietary sources

- A Retinol in dairy fat, eggs, liver, fatty fish and supplemented foods such as margarine; carotene in many dark green, red or yellow fruits and vegetables
1 cup whole milk – 110 μg retinol equivalents
40 g cheddar cheese – 150 μg
1 portion (90 g) fried lamb liver – 18 500 μg
1 egg – 110 μg
1 tomato – 75 μg
1 serving (95 g) broccoli – 400 μg
1 serving (65 g) carrots – 1300 μg
1 banana – 27 μg
Half cantaloupe melon – 700 μg
- D Dairy fat, eggs, liver, fatty fish and supplemented foods
1 cup whole milk (UK unsupplemented) – 0.03–0.06 μg
1 egg – 1.2 μg
40 g cheddar cheese – 0.1 μg
5 g cod liver oil – 10.5 μg
1 portion (90 g) fried lamb liver – 0.5 μg
100 g fatty fish: herrings (most), salmon, tuna, pilchards, sardines (least) – 6–22 μg
5-g pat of butter/margarine – 0.04/0.4 μg
- E Vegetable oils, wheat germ and wholegrain cereals, dark green leaves of vegetables, seeds and nuts
1 egg – 1 mg

5-g pat of butter/polyunsaturated margarine –

- 0.1/1.25 mg
1 slice wholemeal bread – 0.05 mg
5 g sunflower/olive/rapeseed oil – 2.5/0.26/1.2 mg
- K Liver, green leafy vegetables, milk

Biochemical functions

- A Precursor of 11-cis retinal, a component of visual pigments; maintains integrity of epithelial tissues; synthesis of glycoproteins containing the sugar mannose
- D Precursor of 1,25-dihydroxycholecalciferol (calcitriol), hormone produced in the kidney; increases capacity of gut and kidney tubule to transport calcium; regulates deposition of bone mineral
- E Antioxidant in the lipid phase; scavenges free radicals and prevents lipid peroxidation
- K Cofactor in synthesis of clotting factors

Effects of deficiency

- A Night blindness; xerophthalmia leading to permanent blindness; poor growth; thickening and hardening of epithelial tissue; reduced immunocompetence
- D Rickets/osteomalacia with skeletal abnormalities; low plasma calcium; muscle weakness, possibly tetany; growth failure; increased risk of infection
- E Progressive degenerative neurological syndrome
- K Bleeding, especially brain haemorrhage in the newborn

Risk factors for deficiency

- A Poor diet, based upon starchy staple and unsupplemented with dairy fat, fatty fish or fruits and vegetables; very low-fat diet; common in some Third World countries and major factor in child mortality and blindness
- D Inadequate exposure to sunlight, for variety of causes, e.g. dress customs, housebound, working indoors for long hours; lack of back-up from limited range of animal foods or supplemented foods; pigmented skin?; high-fibre diet?
- E Clinical deficiency almost never seen; confined to people with inherited inability to absorb fat-soluble vitamins
- K Normally confined to newborn, especially premature infants; failure of fat-soluble vitamin absorption

Vitamin A – retinol**NATURE AND SOURCES**

The chemical structure of vitamin A (retinol) is shown in Figure 12.1. If the alcohol (CH_2OH) group at carbon 15 is oxidized to an aldehyde group (CHO), this is retinal, and if oxidized to a

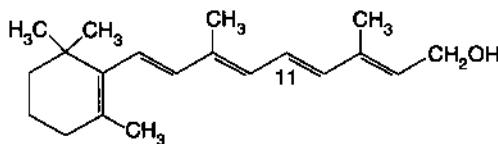
carboxyl group (COOH), this is retinoic acid. All of the double bonds in the side chain of retinol are in the *trans* configuration. Retinol is fat soluble and only found in the fat fraction of some foods of animal origin, e.g. butter and milk fat, liver, egg yolk, oily fish or fish liver oil. There is very little retinol in muscle meat.

Table 12.2 A summary of the dietetically important water-soluble vitamins

<i>Vitamin names and dietary forms</i>	
B ₁	Thiamin
B ₂	Riboflavin
B ₃	Niacin/nicotinic acid; may also be synthesized from the amino acid tryptophan
B ₆	Pyridoxine and related compounds
B ₁₂	Cyanocobalamin and related compounds
-	Folic acid/folate (folacin)
C	Ascorbic acid
<i>Dietary Reference Values</i>	
	UK RNI adult male/female; (American RDAs male/female)
Thiamin	1.0/0.8 mg/day; (1.5/1.1)
Riboflavin	1.3/1.1 mg/day; (1.7/1.3)
Niacin	17/13 mg/day niacin equivalents; (19/15)
B ₆	1.4/1.2 mg/day; (2.0/1.6)
Folate	200 µg/day; (200/180)
B ₁₂	1.5 µg/day; (2.0)
C	40 mg/day; (60)
<i>Dietary sources</i>	
Thiamin	Found in all plant and animal tissues, but only whole cereals, nuts, seeds and pulses are rich sources 1 slice wholemeal bread (UK) – 0.12 mg 1 egg – 0.04 mg 1 pork chop – 0.7 mg 1 cup whole milk – 0.06 mg 30 g peanuts – 0.27 mg 1 serving (165 g) brown/white rice – 0.23/0.02 mg 1 serving (75 g) frozen peas – 0.18 mg 1 serving (200 g) baked beans – 0.14 mg 1 teaspoon (5 g) yeast extract – 0.16 mg
Riboflavin	Small amounts in many foods, rich sources are liver, milk, cheese and egg 1 slice wholemeal bread (UK) – 0.03 mg 1 egg – 0.21 mg 1 cup whole milk – 0.33 mg 40 g cheddar cheese – 0.16 mg 1 portion (90 g) fried lamb liver – 4 mg 1 teaspoon (5 g) yeast extract – 0.55 mg 1 serving (95 g) broccoli – 0.19 mg
Niacin	Widely distributed in small amounts in many foods, but good amounts in meat, offal, fish, wholemeal cereals and pulses; in some cereals, especially maize, much of the niacin may be as unavailable niacytin; the tryptophan in many food proteins is also a potential source of niacin 1 slice wholemeal bread (UK) – 1.4 mg + 0.6 mg from tryptophan
	1 portion (90 g) lamb liver – 13.7 mg + 4.4 mg 1 pork chop – 6.1 mg + 4.9 mg 1 serving (95 g) canned tuna – 12.3 mg + 4.1 mg 1 egg – 0.04 mg + 2.2 mg 1 cup whole milk – 0.16 mg + 1.5 mg 1 portion (75 g) frozen peas – 1.1 mg + 0.7 mg Liver is a rich source; cereals, meats, fruits and vegetables all contain moderate amounts Flesh foods, milk, eggs and fermented foods Liver, nuts, green vegetables, wholegrain cereals are good sources 1 portion (90 g) lamb liver – 220 µg 1 slice wholemeal bread (UK) – 14 µg 30 g peanuts – 33 µg 1 portion (95 g) broccoli – 100 µg 1 banana – 30 µg 1 orange – 90 µg 1 teaspoon (5 g) yeast extract – 50 µg Fruit, fruit juices, salad and leafy vegetables are good sources 1 serving (95 g) broccoli – 30 mg 1 serving (75 g) frozen peas – 10 mg 1 portion (150 g) boiled potatoes – 14 mg 1 portion (45 g) sweet pepper – 45 mg 1 tomato – 15 mg 1 orange – 95 mg 1 banana – 8 mg 1 serving (100 g) strawberries – 60 mg 1 apple – 2 mg 1 glass (200 g) orange juice – 70 mg
<i>Biochemical/physiological roles</i>	
	Thiamin Gives rise to thiamin pyrophosphate, a coenzyme for pyruvic oxidase (carbohydrate metabolism), transketolase (pentose phosphate pathway) and α-oxyglutarate oxidase (Krebs' cycle)
	Riboflavin Component of flavin nucleotides, which are cofactors for several enzymes involved in oxidation-reduction reactions
	Niacin Component of nicotinamide adenine dinucleotide (NAD) and NADP, which are coenzymes involved in many oxidation-reduction reactions
B ₆	Precursor of pyridoxal phosphate, a coenzyme involved in amino acid metabolism
B ₁₂	Interacts with folate in reactions necessary for DNA synthesis; required for nerve myelination
Folate	Involved with B ₁₂ in methylation reactions necessary for DNA synthesis and thus important in cell division

Table 12.2 (continued)

C	Important in synthesis of collagen; promotes iron absorption; functioning of immune system?	<i>Risk factors associated with deficiency</i>
	<i>Effects of deficiency</i>	
Thiamin	Beriberi and the neurological disorder Wernicke–Korsakoff syndrome	Thiamin Diet very heavily dependent upon polished rice; alcoholism
Riboflavin	Cracking at corners of mouth; raw red lips; enlarged nasal follicles plugged with sebaceous material	Riboflavin Absence of milk from diet
Niacin	Pellagra	Niacin Diet heavily dependent upon maize or sorghum and not supplemented by high-protein foods
B ₆	Overt deficiency rare; convulsions reported in infants fed on B ₆ -depleted formula	B ₆ Deficiency rare
B ₁₂	Megaloblastic anaemia; degeneration of spinal cord leading to progressive paralysis	B ₁₂ Vegan diet; failure of absorption in pernicious anaemia
Folate	Megaloblastic anaemia; see also neural tube defects – (Chapter 14)	Folate Poor diet – deficiency common in tropical countries but not developed countries; prolonged heating (e.g. of pulses) destroys folate; poor absorption, e.g. in coeliac disease; some drugs (e.g. anticonvulsants) may interfere with folate functioning
C	Scurvy	C Living upon preserved foods; absence of fruit and vegetables from the diet

**Figure 12.1** The structure of vitamin A (all trans retinol).

Humans can also make retinol from β -carotene and several other members of the carotenoid group of pigments, although 90% of the 600 carotenoids do not have vitamin A activity. β -carotene can be cleaved by humans to yield two units of vitamin A (retinol). The absorption of carotene is less efficient than that of retinol and so 6 µg of carotene (12 µg of other carotenoids with vitamin A activity) is by convention taken to be equivalent to 1 µg of retinol. Note that, in practice, the absorption of carotene varies from food to food (e.g. less than 10% from raw carrots but over 50% when dissolved in palm oil). The vitamin A content of the diet is expressed in **retinol equivalents (RE)**:

$$\begin{aligned} 1 \text{ } \mu\text{g RE} &= 1 \text{ } \mu\text{g of retinol} \\ &6 \text{ } \mu\text{g of } \beta\text{-carotene} \\ &12 \text{ } \mu\text{g of other provitamin A} \\ &\quad \text{carotenoids} \end{aligned}$$

β -carotene, and any other similar plant pigments with vitamin A activity, must all be converted to retinol equivalents to give the total vitamin A activity of foods or diets. Dark green vegetables, orange and yellow fruits and vegetables, and palm oil are rich plant sources of vitamin A.

FUNCTIONS

Rhodopsin is the light-sensitive pigment in the rod cells of the retina. It is comprised of 11-cis retinal (the chromophore) and a protein called opsin. Within the eye, all *trans* retinol (vitamin A) is converted by enzymes to 11-cis retinal and this binds spontaneously with opsin to form rhodopsin. Light induces isomerization of the 11-cis retinal in rhodopsin to all *trans* retinal and this causes the opsin and retinal to dissociate and the pigment to become bleached. It is this light-induced *cis* to *trans* isomerization that generates the nervous impulses that we perceive as vision. Enzymes in the eye then regenerate 11-cis retinal and thus rhodopsin (summarized in Figure 12.2). Rhodopsin is the most light sensitive of the human visual pigments and is responsible for our night vision. In bright light, rhodopsin is completely bleached and other less sensitive pigments in the cones of the eye are responsible for day vision.

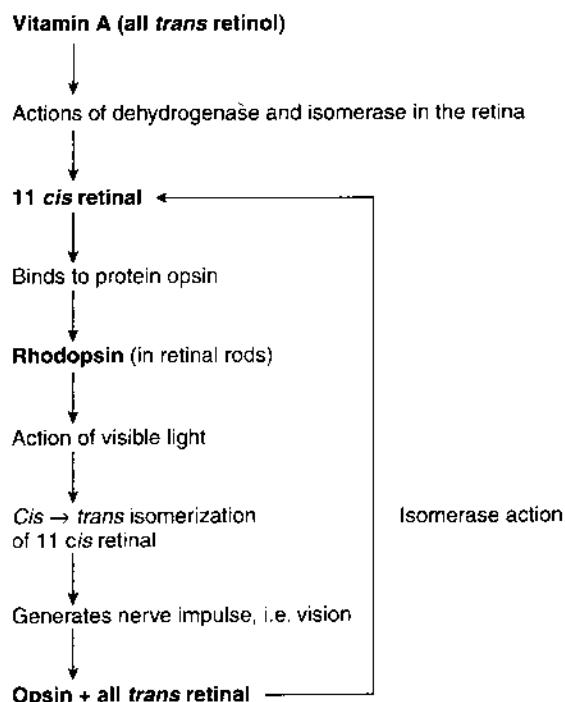


Figure 12.2 Scheme for the visual function of vitamin A in the rods of the retina.

During dark adaptation, rhodopsin is being regenerated. Note that 11-cis retinal is the chromophore in all known visual pigments, including those in the cones of the eye. The light-absorbing characteristics of the chromophore are modified by binding to different proteins, but the essentials of the light response are as described for rhodopsin.

Vitamin A maintains the integrity of epithelial tissues in the eye, respiratory tract, gut and elsewhere. In vitamin A deficiency, there are pathological changes in the integrity of epithelial tissues. Normal epithelial tissue has columnar or cuboidal cells, and goblet cells produce mucus, which moistens the epithelial surface. In vitamin A deficiency, there is overgrowth of undifferentiated flat cells, which are dry and keratinized (hard). Retinoic acid acts as a hormone, which binds to specific receptors within epithelial cells and controls the expression of numerous genes. This control of gene expression and protein synthesis regulates the proliferation and differentiation of epithelial cells. Retinol also plays a role in the synthesis of cell surface glycoproteins, which contain the sugar mannose.

Vitamin A is essential for proper functioning of the immune system. It is well established that vitamin A deficiency is associated with reduced resistance to infection and increased child mortality. In experimental vitamin A deficiency in rats, both antibody and cell-mediated immune responses are impaired and they improve when vitamin A is administered. Loss of integrity of epithelial surfaces also increases vulnerability to infection.

Vitamin A is necessary for normal fetal development. Fetal abnormalities occur in both vitamin A deficiency and vitamin A overload and this has been confirmed by experimental studies with animals. These effects upon development are mediated by the effects of retinoic acid upon gene expression in the fetus.

This brief account of the effects of vitamin A is largely descriptive and readers who are interested in a more mechanistic account are referred to Shils *et al.* (1999), which will also provide an entry into the specialist literature.

REQUIREMENTS AND ASSESSMENT OF STATUS

The estimated adult requirements for vitamin A are given in Table 12.1 and the origins of these values is discussed in Chapter 3. In practice, the biochemical assessment of individual vitamin A status (plasma retinol concentration – Table 3.6) is insensitive because plasma levels only change after prolonged deprivation or oversupply. Healthy adults have large stores of vitamin A in their livers, which take many months to deplete. Poor dark adaptation and lack of goblet cells and other histological changes in smears of corneal epithelium are often used as indicators of inadequate vitamin A status.

DEFICIENCY STATES

The most obvious effects of vitamin A deficiency are upon the eye. Night blindness due to reduced production of rhodopsin is an early indication of vitamin A deficiency. The corneal epithelium is also affected by vitamin A deficiency. There is drying and hardening of the cornea and the appearance of **Bitot's spots** (white clumps of keratinized epithelial cells on the surface of the cornea). In the early stages, these changes in the eye can be reversed by vitamin A, but as they progress they become irreversible and permanent blindness results. Infection and necrosis of the cornea may eventually result in the extrusion

of the lens. The clinical terminology used to describe the various levels of ocular pathology in vitamin A deficiency is complex, but the term **xerophthalmia** is used to cover all of the ocular manifestations of vitamin A deficiency.

Other consequences of vitamin A deficiency are:

- impaired growth
- depressed immune function and increased susceptibility to infection
- overgrowth and keratinization of other epithelial tissues, e.g. in the trachea.

Vitamin A deficiency is very prevalent in South East Asia, the Indian subcontinent and several developing countries in Africa and South America. It is the major cause of child blindness in these regions and also causes high child mortality from infections. The World Health Organization (WHO) has estimated that there may be 6–7 million new cases of xerophthalmia each year, with about 10% of these being severe enough to cause corneal damage. More than half of this 10% with ocular damage die within a year and three-quarters of the survivors are totally or partially blind. As many as 20–40 million children suffer from mild vitamin A deficiency that leads to reduced resistance to infection and increased child mortality. Eradication of vitamin A deficiency is a major aim of the WHO.

RISK FACTORS FOR DEFICIENCY

A diet that is almost entirely composed of starchy roots or cereals (e.g. rice or cassava) has no significant source of either retinol or the provitamin A carotenoids. Such a diet is also very low in fat and this reduces the absorption of any carotene that is present in the diet. Vitamin A deficiency can also be a secondary consequence of cystic fibrosis, coeliac disease and obstructive jaundice because of poor absorption or cirrhosis of the liver due to an inability to store the vitamin. Deficiency does not usually manifest while children are being breastfed, even if it is only partial breastfeeding.

BENEFITS AND RISKS OF HIGH INTAKES

The carotenoids may have protective effects against cancer and other chronic diseases because of their antioxidant properties and independent of their

provitamin A functions. This is discussed later in the chapter.

Retinol is relatively toxic when taken in excess, but carotene is not. Chronic consumption of ten times the Reference Nutrient Intake (RNI); (Recommended Dietary Allowance, RDA) of retinol causes:

- loss of hair
- dry, itchy skin
- Swollen liver and spleen with permanent liver damage possible.

Very large single doses can also be toxic. Excess retinol in pregnancy is teratogenic (causes birth defects).

The epithelial changes seen in vitamin A deficiency resemble some of those seen in the early stages of epithelial carcinogenesis. Some prospective case-control studies published in the 1970s and early 1980s (e.g. Wald *et al.*, 1980) reported that low plasma retinol concentrations were associated with an increased risk of cancer. There are many reasons why the results from such studies might not necessarily indicate a protective effect of dietary vitamin A on cancer risk (see Chapter 3 for a fuller discussion of this example). Two very obvious criticisms of such studies are:

- changes in plasma retinol concentration may be an early consequence of cancer rather than a cause
- plasma retinol concentration is not a sensitive indicator of individual vitamin A intake and is affected by hormonal and other non-dietary factors.

In general, studies in which intake of retinol have been related to subsequent cancer risk have not found any association, although a weak association between total vitamin A intake and cancer risk has been reported (Rutishauser, 1992; Shils *et al.*, 1999). There are overwhelming grounds, unrelated to cancer risk, for measures to ensure adequacy of vitamin A intake in areas of the world where deficiency exists. There is, however, no convincing evidence that supplemental intakes of retinol, well in excess of the RNI/RDA, are likely to be beneficial; given the known toxicity and teratogenicity of vitamin A, they are undesirable.

Natural and synthetic retinoids have been widely and successfully used in the treatment of skin diseases including acne, psoriasis and skin cancer. They

have been used both for topical application and oral administration. When given orally, these retinoids have toxic effects like those described for vitamin A and this limits the doses that may be used. Pregnancy is a contraindication for the use of these retinoids because of their teratogenic effects.

Vitamin D – cholecalciferol

NATURE, SOURCES AND REQUIREMENTS

Vitamin D₃ (cholecalciferol) is formed in the skin of animals by the action of ultraviolet light (sunlight) upon 7-dehydrocholesterol. This latter compound is made by inserting an extra double bond into cholesterol (in the B ring of the steroid nucleus). The chemical structures of 7-dehydrocholesterol and cholecalciferol are shown in Figure 12.3. Almost all of the vitamin naturally in food is vitamin D₃ and it is, like retinol, obtained only from foods of animal origin such as eggs, butter, liver, oily fish and fish liver oil. Irradiation of the plant sterol ergosterol (e.g. from yeast) yields vitamin D₂, which has similar activity to cholecalciferol in mammals. This form of vitamin D is often used in the production of vitamin tablets and to supple-

ment foods such as margarine, milk in the USA, baby foods and breakfast cereals.

For most people, endogenous synthesis of cholecalciferol in the skin is the primary source of the vitamin and a dietary supply is not necessary. Dietary supply of the vitamin becomes much more important in people who do not get adequate exposure to sunlight, e.g. the housebound elderly. In the UK, winter sunlight does not contain the correct ultraviolet wavelengths for vitamin D production, but production and storage during the summer months should be adequate to last through the winter. For those with white skins, short periods of exposure of the hands and face to summer sunlight are sufficient to maintain adequate vitamin D status, but much longer exposure is required if the skin is black or brown. One theory suggests that the evolution of white races in northern latitudes and those with darker skins near the equator might have been due to differences in ability to produce vitamin D.

FUNCTIONS

Vitamin D is required for the synthesis of the hormone 1,25-dihydroxycholecalciferol (calcitriol). Cholecalciferol is hydroxylated to 25-hydroxycholecalciferol in the liver and this is the main circulating form of the vitamin in plasma. This compound is further hydroxylated to calcitriol in the kidney by an hydroxylase whose activity is regulated by the parathyroid hormone.

Calcitriol acts like a steroid hormone. It binds to receptors within the nucleus of its target tissues and affects gene expression, i.e. the synthesis of particular proteins. Overall, the best documented effects of vitamin D are in the maintenance of calcium homeostasis. In the intestine, it increases the synthesis of several proteins that are involved in the absorption of calcium. In vitamin D deficiency, the gut absorbs calcium very inefficiently. In the bone, calcitriol has several effects. It induces undifferentiated bone cells to become osteoclasts and these cells release calcium and phosphorus from bone when plasma calcium is low and they are activated by parathyroid hormone. Calcitriol also induces the synthesis of some of the proteins that are prominent in the bone matrix, e.g. osteocalcin. There is little evidence that vitamin D is directly involved in bone mineralization, although, because of its effects upon calcium and phosphorus supply, it is essential for the development and

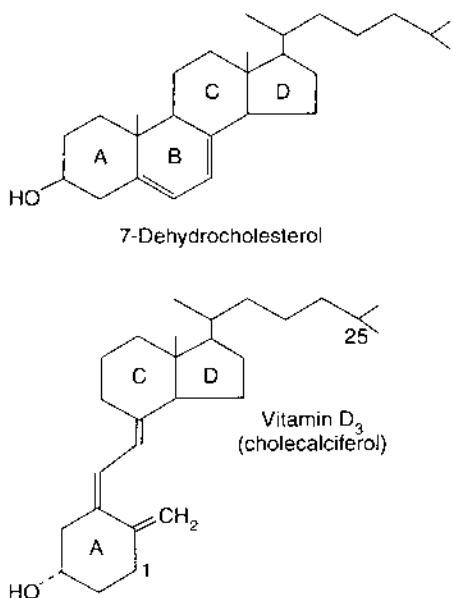


Figure 12.3 The structures of 7-dehydrocholesterol and cholecalciferol (vitamin D₃).

maintenance of a healthy skeleton. If rats are deprived of vitamin D but their blood levels of calcium and phosphorus are maintained within the normal range, bone mineralization appears to be normal (see the relevant chapter in Shils *et al.*, 1999).

Several tissues that are not associated with the maintenance of calcium homeostasis have cellular receptors for calcitriol, including lymphocytes and skin cells. Calcitriol plays a role in the regulation of cell proliferation and differentiation in several tissues. It has been used in the treatment of psoriasis, a condition in which there is excessive proliferation and keratinization of epidermal cells in the skin. In experimental animals, calcitriol has immunosuppressive effects and in pharmacological doses reduces several autoimmune conditions. This suggests that it may play a role in the normal regulation of immune responses.

DEFICIENCY STATES

Rickets in children and **osteomalacia** in adults are due to lack of vitamin D. During the late nineteenth and early twentieth centuries, rickets was extremely prevalent amongst poor children living in the industrial cities of Britain. The symptoms arise from a failure to absorb calcium from the gut, which leads to reduced plasma calcium concentration and reduced calcification of bones. Muscle weakness and gastrointestinal and respiratory infections are general symptoms of rickets. There are a series of skeletal abnormalities that are characteristic of the disease, such as:

- bowing of the legs, or knock knees
- swelling of the wrist
- beading of the ribs at the normal junction of cartilage and bone (rickety rosary)
- pelvic deformities, which may cause problems in childbirth in later years.

In severe rickets, plasma calcium concentration may fall low enough to induce uncontrolled convulsive muscle contractions (tetany). Chronic vitamin D deficiency is now regarded as a major cause of osteoporosis in very elderly, housebound people (see Chapter 13 for further discussion of osteoporosis).

Inadequate exposure of the skin to sunlight is regarded as the primary cause of vitamin D deficiency. Poor dietary supply of the vitamin is usually regarded as an exacerbating factor. Average intakes of vitamin D in UK adults are 2–3 µg/day, which is considerably less than the RNI of 10 µg/day that is given

for those adults who do not get out in the sun. The risk of vitamin D deficiency is increased by any factor that reduces exposure of the skin to the correct ultraviolet wavelengths, such as:

- living in northern latitudes
- atmospheric pollution
- skin pigmentation
- some traditional costumes
- being housebound because of illness or immobility
- working indoors for very long hours and other cultural or social factors that prevent people getting out of doors during the daytime.

Relatively high levels of rickets and osteomalacia have been a problem amongst Asian women and children who have migrated to Britain from the Indian subcontinent.

TOXICITY

Vitamin D is the most toxic of the vitamins and as little as five times the dietary standard (50 µg/day) may be enough to cause toxicity. Vitamin D poisoning does not occur if the skin is overexposed to the sun because of natural regulation. Poisoning is due to overfortification of foods or excessive use of supplements. Excess vitamin D causes hypercalcaemia and this leads to anorexia, nausea, vomiting, hypertension, kidney stones, renal failure and possible calcification of soft tissues. It may have fatal consequences if severe and prolonged. In the 1950s, some British babies developed a vitamin-D-induced hypercalcaemia because of excessive fortification of milk and other foods. Milk in the UK is no longer fortified with vitamin D, although it is in the USA.

Vitamin E – α -tocopherol

There are several compounds synthesized by plants that have vitamin E activity and the most active and important is α -tocopherol (the structure of which is shown in Figure 12.4). Vitamin E is widely distributed in plant foods and is particularly concentrated in plant oils that typically contain 10–50 mg of vitamin E per 100 g. Animal fats contain small amounts of vitamin E. The estimated requirements for vitamin E are shown in Table 12.1.

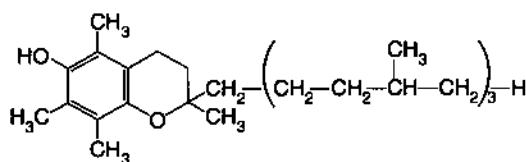


Figure 12.4 The structure of vitamin E (α -tocopherol).

Vitamin E is readily oxidized and so acts as an antioxidant. Its primary function is to prevent oxidative damage by free radicals to polyunsaturated fatty acids in membrane phospholipids. Free radicals are discussed more fully at the end of this chapter.

Overt deficiency of vitamin E is rare and is usually associated with some disorder of fat absorption or fat transport. As discussed later in the chapter, there is considerable debate about whether low intakes of vitamin E may allow increased oxidative damage to membranes and thus increase the risk of chronic disease. Many people take very large supplements of vitamin E and doses up to 800 mg/day seem to produce no acute toxic effects. Nerve degeneration, muscle atrophy and retinopathy are seen in both animals and people with prolonged vitamin E deficiency. Early studies of experimental vitamin E deficiency in rats found that it produced sterility, and this is the origin of its ineffective use in preparations that are claimed to increase sexual potency or to act as aphrodisiacs.

Vitamin K – phylloquinone

Phylloquinone (vitamin K₁) is found largely in plant foods, particularly green, leafy vegetables, although it is found in smaller amounts in many foods. An alternative form of the vitamin, menaquinone, is produced by bacteria. Human gut bacteria produce significant amounts of vitamin K, which probably contribute to vitamin K intakes. The structure of phylloquinone is shown in Figure 12.5.

Several clotting factors, including prothrombin, only become functional when several residues of the amino acid glutamic acid within their structures are carboxylated (addition of the COOH group); this carboxylation occurs after translation. The enzyme responsible for this carboxylation requires vitamin K as a coenzyme.

Vitamin K deficiency leads to a reduced efficiency of the blood-clotting system and thus to a tendency to bleed. Primary dietary deficiency of vitamin K is almost never seen, but malabsorption, for example in coeliac disease, can precipitate deficiency and lead to bleeding. The coumarin group of anticoagulant drugs (e.g. warfarin) exert their effect by blocking the effects of vitamin K and lead to reduced levels of clotting factors in blood. Therapeutic overdosage of these drugs or accidental overdosage from warfarin-based rodent poisons leads to bleeding and vitamin K acts as an antidote.

Some newborn babies, especially premature babies, may develop intracranial haemorrhages shortly after birth; this tendency to haemorrhage can be prevented by the prophylactic administration of vitamin K, and the routine administration of vitamin K is common in maternity units. In 1992, Golding *et al.* reported that vitamin K injections in the newborn were associated with an increased risk of childhood cancer. This report naturally raised concerns about the safety of this prophylactic use of vitamin K, but subsequent studies (e.g. Eklund *et al.*, 1993) have suggested that these concerns were probably unwarranted.

Thiamin – vitamin B₁

NATURE AND SOURCES

The chemical structure of thiamin is shown in Figure 12.6. Thiamin is water soluble and so can leach into cooking water. It is oxidized when heated in alkaline solution. It is found in wholegrain cereals, where it is

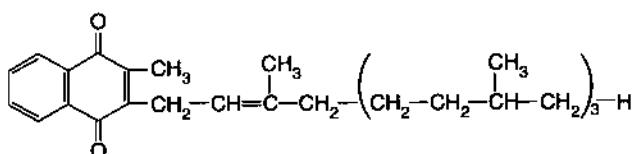


Figure 12.5 The structure of vitamin K (phylloquinone).

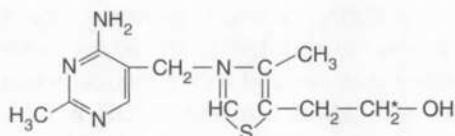


Figure 12.6 The structure of thiamin.

concentrated in the bran and germ layers. Much of the thiamin is removed from cereals when they are refined, e.g. when milled to produce white flour or white rice. White rice is particularly low in thiamin, but white flour is fortified with thiamin in the UK and USA. Thiamin is present or added to many breakfast cereals and is found in pulses, vegetables, milk, organ meat and pork.

FUNCTIONS

Thiamin is the precursor of the important coenzyme thiamin pyrophosphate (TPP). TPP is formed by addition of pyrophosphate (two phosphate groups) at the position marked by an asterisk in Figure 12.6. It is an essential coenzyme in several key reactions in metabolism, including (see also Chapter 5):

- the conversion of pyruvate to acetyl coenzyme A (CoA) in carbohydrate metabolism
- the conversion of α -ketoglutarate to succinyl CoA in the Krebs' cycle
- the reactions catalysed by transketolase in the pentose phosphate pathway.

In thiamin deficiency there is a progressive block at these steps in metabolism. The pyruvate dehydrogenase enzyme that converts pyruvate to acetyl CoA seems to be particularly sensitive to thiamin deficiency. Thiamin deficiency leads to an inability to metabolize carbohydrate aerobically, and pyruvate and lactate accumulate in the blood and tissues. The block in the pentose phosphate pathway restricts the production of reduced nicotinamide adenine dinucleotide phosphate (NADP), which is essential for lipid biosynthesis, including the synthesis of **myelin** (the fatty sheath that surrounds many nerves).

REQUIREMENTS AND ASSESSMENT OF STATUS

The estimated adult requirements for thiamin are given in Table 12.2. The requirement for thiamin

increases as energy expenditure increases and the UK RNI is set at 0.4 mg/1000 kcal (4.2 MJ). The **erythrocyte transketolase activation coefficient** is used to assess thiamin status (see Chapter 3).

DEFICIENCY STATES

Beriberi is caused by thiamin deficiency. The major symptoms of thiamin deficiency are summarized briefly below:

- lactic acidosis
- peripheral neuropathy – degeneration of peripheral nerves with tingling and loss of sensation at the extremities followed by gradual loss of motor function, muscle wasting and paralysis (dry beriberi)
- heart failure and oedema (wet beriberi)
- Wernicke–Korsakoff syndrome.

Wernicke–Korsakoff syndrome is the neurological and psychiatric manifestation of thiamin deficiency that is usually associated with alcoholism. This syndrome is characterized by paralysis of eye movements or rapid, jerky eye movements (nystagmus), unsteadiness when walking or standing and mental derangements, including loss of memory and the ability to learn. *Post mortems* upon such patients show lesions in various parts of the brain, ranging from loss of myelin to complete necrosis.

The clinical manifestations of beriberi can be largely explained by the known biochemical functions of TPP. Except during starvation, nervous tissue relies upon carbohydrate metabolism for its energy generation, and the metabolism of heart muscle also relies heavily upon the aerobic metabolism of glucose. As thiamin deficiency impairs carbohydrate metabolism, it is thus not surprising that many of the symptoms of beriberi are associated with changes in nerve and cardiac function. Production of the fatty sheath (myelin) covering nerve fibres is dependent upon the enzyme transketolase in the pentose phosphate pathway; this enzyme requires TPP as a coenzyme, and demyelination of nerves is a feature of beriberi and Wernicke–Korsakoff syndrome. The accumulation of lactic acid in blood and tissues that results from impaired pyruvic oxidase functioning may exacerbate the oedema of wet beriberi.

Beriberi has been known in China and the Far East for several thousand years, but large-scale epidemics

occurred in the late nineteenth and early twentieth centuries and were probably precipitated by the introduction of machine milling for rice. Epidemics of beriberi are associated with diets that are based upon white (polished) rice. Rice is low in thiamin and, if rice is efficiently milled, most of this thiamin is removed with the bran. The high carbohydrate content of rice may also increase requirements. The poorest members of these Eastern societies often escaped beriberi because they ate the cheaper brown rice with its thiamin still present. Only those who could afford the more expensive and higher prestige white rice were vulnerable to this condition, including the armed forces of Japan and China.

In the late nineteenth century, a Japanese naval surgeon noted that many supplementary foods were effective in curing and preventing beriberi and he proposed that it might be caused by protein deficiency. In 1890, Eijkmann noted that beriberi amongst patients at a military hospital in Java seemed to be associated with eating polished rice, but not with brown rice. He induced similar symptoms to human beriberi in chickens when he fed them with boiled polished rice. These symptoms could be prevented by feeding them rice bran. In 1926, Jansen and Donath, working in Eijkmann's laboratory in Java, purified thiamin from rice bran. Most parts of India escaped epidemics of beriberi because of different milling practices. The rice was soaked and then steamed or boiled prior to milling (parboiled), and this greatly increases the retention of B vitamins within the rice grain after milling. The introduction of less harsh milling practices, the fortification of white rice with thiamin and diversification of the diet have all contributed to the decline in beriberi. Intakes of thiamin are, however, still low in many rice-eating countries and offer little margin of safety.

It is very common for alcoholics to be thiamin deficient and Wernicke-Korsakoff syndrome is largely confined to alcoholics. There are several reasons for this:

- their diet is poor and thiamin intake is low
- high alcohol intake impairs the absorption of thiamin
- some alcoholics may be particularly sensitive to thiamin deficiency and more susceptible to Wernicke-Korsakoff syndrome because their transketolase enzyme has a low affinity for TPP.

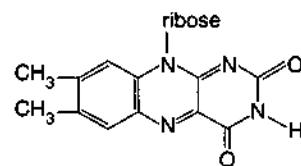


Figure 12.7 The structure of riboflavin.

Riboflavin – vitamin B₂

NATURE AND SOURCES

The structure of riboflavin is shown in Figure 12.7. Dairy products, meat, fish, eggs, liver and some green vegetables are rich sources, although it is widely distributed in many foods. Cereals are generally a poor source unless they are fortified.

FUNCTIONS

Riboflavin gives rise to **flavin mononucleotide (FMN)** and **flavin adenine dinucleotide (FAD)**. These flavin nucleotides are essential components (prosthetic groups) of several key **flavoprotein** enzymes involved in oxidation-reduction reactions. They are involved in both the Krebs' cycle and oxidative phosphorylation, and so riboflavin deficiency leads to a general depression of oxidative metabolism. Glutathione reductase, an enzyme involved in the disposal of free radicals, also has a flavin prosthetic group.

REQUIREMENTS AND ASSESSMENTS OF STATUS

The adult requirements for riboflavin are given in Table 12.2. Riboflavin status may be assessed by the **erythrocyte glutathione reductase activation coefficient** that is discussed in Chapter 3.

RIBOFLAVIN DEFICIENCY

Mild riboflavin deficiency is relatively common, even in industrialized countries, but no severe deficiency disease has been described for this vitamin. This may be because the wide distribution of small amounts of the vitamin in food prevents mild deficiency progressing to major life-threatening illness. Also, riboflavin deficiency is usually associated with concurrent beriberi and the severe symptoms of beriberi mask the symptoms of riboflavin deficiency. In volunteers

deliberately made riboflavin deficient, the symptoms were: chapping at the corners of the mouth (angular stomatitis), raw red lips (cheilosis) and enlarged follicles around the sides of the nose, which become plugged with yellow sebaceous material (nasolabial seborrhoea).

Niacin – vitamin B₃

NATURE AND SOURCES

Niacin is also called nicotinic acid and is designated vitamin B₃. Its chemical structure is shown in Figure 12.8. Preformed niacin is found in red meat, liver, pulses, milk, wholegrain cereals and fish and it is added to white flour and many breakfast cereals. Much of the niacin present in cereals may be in a bound form that is not readily available to humans.

The amino acid tryptophan can also act as a source of niacin. Niacin can therefore be obtained from the diet either directly as vitamin present in the food or indirectly by the conversion of tryptophan in dietary protein to niacin. Thus, early suggestions that pellagra was due to protein deficiency were partly correct because high-protein foods, by providing tryptophan, could permit endogenous niacin synthesis and cure or prevent deficiency. The niacin content of foods and diets is expressed in milligrams of **niacin equivalents**. In calculating the niacin equivalents in a food, 60 mg of tryptophan is usually taken to be equivalent to 1 mg of niacin. In the typical British or American diet, tryptophan would make up just over 1% of the total dietary protein.

$$\begin{aligned} 1 \text{ mg niacin equivalent} &= 1 \text{ mg niacin} \\ &\quad 60 \text{ mg tryptophan} \end{aligned}$$

FUNCTIONS

Niacin is a component of the important coenzymes NAD and the phosphorylated derivative NADP. NAD and NADP serve as hydrogen acceptors and

donators in numerous oxidation and reduction reactions in all human biochemical pathways. It is during the re-oxidation of the reduced NAD, produced during the oxidation of foodstuffs, that most of the adenosine triphosphate (ATP) yielded by aerobic metabolism is produced in the mitochondria of human cells (see Chapter 5).

DIETARY REQUIREMENTS AND ASSESSMENT OF STATUS

The adult dietary standards for niacin are given in Table 12.2. The requirement for niacin varies with energy expenditure and the UK RNI is set at 6.6 mg of niacin equivalents per 1000 kcal (4.2 MJ). There is no ideal method for assessing niacin status, but the urinary excretion of N'-methyl nicotinamide is one method that can be used (see Chapter 3, Table 3.6).

NIACIN DEFICIENCY

Niacin deficiency causes the disease **pellagra**. The symptoms of the disease are often referred to as the 3Ds – dermatitis, diarrhoea and dementia. An early sign of pellagra is itching and burning of the skin, which becomes roughened and thickened; skin exposed to sunlight is most liable to show dermatitis and a ring of affected skin around the neck (Casal's collar) is a common manifestation. Gastrointestinal symptoms, including diarrhoea, are a characteristic manifestation of pellagra. Early neurological symptoms are weakness, tremor, anxiety, depression and irritability and, if there is severe prolonged deficiency, eventually dementia. The disease has high mortality and at its peak in the late 1920s and early 1930s caused thousands of deaths in the southern USA. It is still endemic in parts of southern Africa.

Pellagra has usually been associated with diets based upon maize. Maize protein is low in tryptophan and the preformed niacin present in maize is present in a bound form (**niacytin**) that is unavailable to humans. The introduction of maize from the Americas has been associated with major epidemics in those parts of the world where it became established as the dominant staple, e.g. in Southern Europe, North Africa and South Africa. In the early decades of the twentieth century, it was so prevalent in some southern states of the USA that it was thought to be an infectious disease.

Davies (1964) gives an analysis of the factors that led to the rise and fall of pellagra in the southern

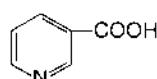


Figure 12.8 The structure of niacin (nicotinic acid).

states of the USA. At the turn of the twentieth century, the diets of poor Southerners were based upon corn (maize), pork fat and molasses. The first recorded outbreak of pellagra in the southern USA was in 1907 and Davies attributes this to changes in the way that the corn was milled. At the turn of the twentieth century, large-scale milling replaced the old, inefficient water-driven mills and this harsher milling produced a finer and more palatable corn meal, but it also removed the germ of the corn grain and reduced its vitamin content. Paradoxically, the disease started to decline at the height of the Great Depression. Davies suggests that this was because demand for cotton, the principal agricultural product of the region, declined very sharply, leading many farmers to cultivate gardens and smallholdings to produce food crops. He thus attributes the decline in pellagra, at a time of great economic deprivation, to departure from the cash-crop monoculture rather than to increased knowledge of the causes of the disease and methods of prevention.

Note that the introduction of machine milling may have been an important factor in precipitating epidemics of both beriberi and pellagra. Harsh machine milling is very efficient at removing the nutrient-rich outer layers of the cereal grain. Traditional hand-milling methods were less efficient in removing these outer layers and left more of the vitamins in the final product.

Amongst the people of Central America, where maize originated, its consumption was not associated with pellagra. The traditional method of tortilla preparation resulted in the release of niacin from its bound state; the maize grains were mixed with slaked lime (calcium hydroxide) and subject to prolonged heating prior to grinding and cooking.

Sorghum consumption is associated with pellagra in some parts of India and China, despite the fact that sorghum contains reasonable amounts of tryptophan (almost as much as rice, which is not pella-

gragic). Bender (1983) concluded that this is because of the high content of another amino acid, leucine, in sorghum, which inhibits key enzymes in the conversion of tryptophan to niacin. The high leucine content of maize may also increase its potential to cause pellagra.

Vitamin B₆ – pyridoxine

NATURE AND SOURCES

Vitamin B₆ is present in food as pyridoxine, pyridoxal or pyridoxamine, which are all biologically active and interconvertible. The structures of these are given in Figure 12.9. The different forms of vitamin B₆ are widely distributed in both plant and animal foods. Liver, eggs, meat, fish, green leafy vegetables, pulses, fruits and wholegrain cereals all contain significant amounts of B₆.

FUNCTIONS

Vitamin B₆ is a precursor of the important coenzyme pyridoxal phosphate. This coenzyme is essential in many biochemical reactions, particularly those involving the metabolism of amino acids. Some of the reactions/pathways in which pyridoxal phosphate acts as a coenzyme are:

- transamination reactions – transfer of amino groups to produce non-essential amino acids (see Chapter 5)
- decarboxylation of amino acids, e.g. to produce the nerve transmitters gamma amino butyric acid (GABA), histamine, dopamine and 5HT (serotonin)
- the synthesis of niacin from tryptophan
- the breakdown of glycogen by the enzyme glycogen phosphorylase

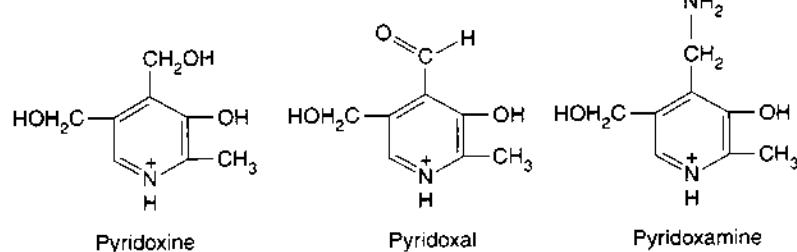


Figure 12.9 The structures of the three forms of vitamin B₆.

- the synthesis of the haem component of haemoglobin.

REQUIREMENTS AND ASSESSMENT OF STATUS

The adult requirements for vitamin B₆ are given in Table 12.2. A wide range of tests of vitamin B₆ status has been developed, including the measurement of pyridoxal phosphate concentrations in blood or erythrocytes. Erythrocyte enzyme reactivation tests (see Chapter 3) using the enzymes aspartate amino transferase or alanine amino transferase are used as measures of long-term B₆ status.

DEFICIENCY AND TOXICITY

Primary dietary deficiency of vitamin B₆ is rare. In the 1950s, vitamin B₆ deficiency was seen in some American babies who had been fed formula depleted of its vitamin B₆ by a fault in the processing. These babies suffered from convulsions and had abnormal electroencephalograms (EEGs). Symptoms of deficiency in adults include anaemia, raw, red lips, smooth, inflamed tongue, lesions at the corners of the mouth, dermatitis and fatigue.

Certain drugs used in the treatment of tuberculosis, rheumatoid arthritis and epilepsy increase the requirement for B₆ and supplements of the vitamin are normally given with these drugs.

There has been much speculation about whether vitamin B₆ alleviates the symptoms of premenstrual syndrome (because of its involvement in the synthesis of nerve transmitters). It has been widely used as a self-medication for this purpose. Some cases of B₆ toxicity have been reported in women who chronically consume very large doses of this vitamin. The symptoms are sensory neuropathy of the hands and feet. The UK government has controversially tried to address this problem by seeking to limit the amount of B₆ that 'over the counter' preparations may contain.

Vitamin B₁₂ – cobalamins

Vitamin B₁₂ is comprised of a group of complex molecules that contain the element cobalt. All of the vitamin B₁₂ ordinarily present in the diet is in foods of animal origin. The ultimate source of the cobalamins is their synthesis by micro-organisms and any foods contaminated by micro-organisms will contain B₁₂.

FUNCTIONS

Together with folic acid, vitamin B₁₂ is essential for the synthesis of the nucleotide thymidylate. This nucleotide is found exclusively in DNA and so its production is essential for DNA synthesis and cell division. This means that cell division, including the production of red blood cells, will be impaired by lack of B₁₂ or folic acid. Vitamin B₁₂ is also required as a coenzyme for a reaction that is essential for the metabolism of fatty acids with odd numbers of carbon atoms (methylmalonyl CoA conversion to succinyl CoA).

REQUIREMENTS AND ASSESSMENT OF STATUS

Human requirements for vitamin B₁₂ are only 1–2 µg/day and healthy adults have several years' supply stored in their livers. Vitamin B₁₂ status is assessed by the measurement of serum B₁₂ concentrations (see Table 3.6).

DEFICIENCY OF VITAMIN B₁₂

Lack of vitamin B₁₂ or folic acid causes a macrocytic (large cell) or **megaloblastic anaemia**. In this type of anaemia, red blood cells are large, irregular and fragile. The blood haemoglobin concentration may fall to less than a third of the normal level. Dietetic lack of B₁₂ is rare and usually associated with a strict vegetarian (vegan) diet. Folic acid deficiency is, on the other hand, common in some Third World countries and high intakes of folic acid can compensate for lack of B₁₂ and mask the haematological consequences of B₁₂ deficiency. Prolonged B₁₂ deficiency also leads to irreversible damage to the spinal cord – **combined subacute degeneration**. This nerve damage manifests initially as loss of sensation or tingling sensation in the fingers and toes, which leads to progressive paralysis. These consequences of B₁₂ deficiency upon the nervous system are not corrected by folic acid and so misdiagnosis and treatment of B₁₂ deficiency with folic acid may mask the haematological consequences but allow the insidious progression of the irreversible spinal cord damage.

Vitamin B₁₂ is naturally confined to foods of animal origin and to micro-organisms, although meat substitutes must contain added B₁₂ in the UK. This means that vegans may avoid all foods that naturally contain B₁₂ and thus be at risk of dietary deficiency of the vitamin unless they eat supple-

mented foods or take B_{12} supplements. In practice, even amongst those theoretically at risk, dietary deficiency is uncommon because requirements are very small and an omnivore has stores that represent several years, turnover of the vitamin. Contamination of food with micro-organisms, insects, faecal matter or perhaps synthesis by gut bacteria may contribute significant amounts of B_{12} to the diet of even the strictest vegetarian (e.g. mould on nuts, insects or insect droppings on fruit or vegetables). Some lactovegetarian Asian women in the UK develop signs of deficiency despite the presence of B_{12} in milk – this is explained by suggesting that the traditional practice of boiling milk with tea may destroy the B_{12} in the milk. It may take many years of dietary deficiency of vitamin B_{12} before clinical symptoms of deficiency manifest in adults.

The usual cause of B_{12} deficiency is not dietary lack but an autoimmune disease, **pernicious anaemia**, which results in damage to the parietal cells of the stomach that produce an **intrinsic factor** that is necessary for the absorption of B_{12} from the gut. Until the 1920s, pernicious anaemia was an inevitably fatal condition; nowadays, the effects of the resulting B_{12} deficiency can be alleviated by monthly injections of the vitamin.

Folic acid – folate, folacin

NATURE AND SOURCES

Folic acid (or pteroylmonoglutamic acid) is the parent compound for a group of compounds known collectively as folate or folacin in the USA. It is made up of three components – pteridine, para-amino benzoic acid (PABA) and glutamic acid (Figure 12.10). Within the body, metabolically active folates

have multiple glutamic acid residues conjugated to them. Folates are found in good amounts in green vegetables, liver, yeast extract, mushrooms, nuts and whole grains.

FUNCTIONS

Folate interacts with vitamin B_{12} in the transfer of methyl groups. The synthesis of thymidylate, the nucleotide necessary for DNA synthesis, involves methyl transfer and requires both folate and B_{12} . As with B_{12} , deficiency of folic acid limits DNA synthesis and thus cell division.

REQUIREMENTS AND ASSESSMENT OF STATUS

The estimated adult requirements for folate are given in Table 12.2. Folate status can be assessed by measurements of serum or red cell folate concentrations (see Table 3.6 for normal limits).

FOLATE DEFICIENCY

Megaloblastic or macrocytic anaemia is a symptom of both folic acid and B_{12} deficiency. The bone marrow's ability to produce blood cells is affected by the reduced ability to synthesize thymidylate and thus DNA. Large, immature and unstable red cells are found circulating in the blood in megaloblastic anaemia, as well as abnormal white cells and platelets. Other rapidly dividing tissues are also affected, such as the gut lining. High doses of folic acid will mask the haematological consequences of B_{12} deficiency but will not prevent the neurological consequences, so it is important that the cause of megaloblastic anaemia is determined before treatment is undertaken.

Some degree of folate deficiency is relatively common in those living on a poor diet in developing countries, i.e. a diet lacking in green vegetables and

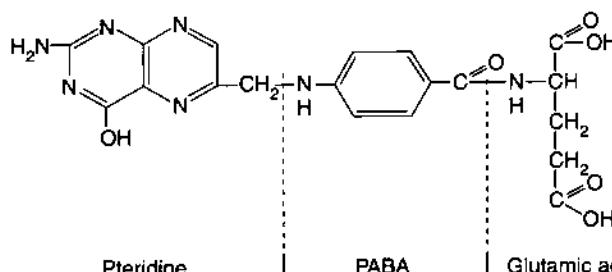


Figure 12.10 The structure of folic acid.

pulses and made up largely of a low-folate, starchy staple. Several drugs, including some anti-epileptics, increase folate requirements and alcoholics often have poor folate status. Folate deficiency may also be precipitated by diseases that affect the absorption of folate, e.g. coeliac disease and tropical sprue.

There is now overwhelming evidence that folate supplements given preconceptually and in early pregnancy reduce the risk of the baby developing a **neural tube defect**. All women are now advised to take folic acid supplements in the first trimester of pregnancy and ideally before they become pregnant. Some foods are now fortified with folic acid in an attempt to increase the folic acid status of women who have unplanned pregnancies (see Chapter 14 for further discussion of this issue).

Biotin

Biotin was originally known as the 'anti-egg white injury factor' because it cured a dermatitis induced in rats by feeding them large amounts of raw egg white. A protein present in raw egg white inhibits the absorption of biotin, but this is denatured when eggs are cooked. Biotin deficiency in humans is rare and usually associated with eating large amounts of raw egg or other bizarre diets. It has been reported in the past in patients being fed by total parenteral nutrition for long periods. Human milk is relatively low in biotin and some cases of biotin-responsive dermatitis have been reported in babies breastfed by poorly nourished mothers.

Biotin functions as a coenzyme for several important carboxylase enzymes, i.e. enzymes that add a carboxyl (COOH) group via the fixation of carbon dioxide. The enzyme pyruvate carboxylase is important in gluconeogenesis (the production of glucose from pyruvate and amino acids – see Chapter 5). Other carboxylase enzymes are important in fatty acid synthesis and the metabolism of several amino acids.

Biotin requirements have not been clearly established, partly because gut bacteria may make some indeterminate contribution to biotin supply.

'Safe intakes of biotin': USA 30–100 µg/day
for adults UK 10–200 µg/day

Biotin is present in good amounts in liver, cooked eggs, cereals and yeast.

Pantothenic acid

Pantothenic acid is a precursor of coenzyme A, which is essential for many metabolic processes – note the acetyl CoA in carbohydrate and fatty acid metabolism and the succinyl CoA of the Krebs' cycle. The vitamin is widely distributed in food and clinical deficiency is rare. A 'burning feet' syndrome has been reported in severely malnourished prisoners of war and in experimental deficiency. Requirements for pantothenic acid are difficult to estimate accurately.

Safe intakes of pantothenic : USA 4–7 mg/day
acid for adults UK 3–7 mg/day

Vitamin C – ascorbic acid

NATURE AND SOURCES

The structure of vitamin C (ascorbic acid) is shown in Figure 12.11. Most species are able to make ascorbic acid from glucose, but primates, guinea-pigs, bats and some birds lack a key enzyme (gulanolactone oxidase) on this synthetic pathway. Details of this synthetic pathway may be found in Groff *et al.* (1995). Ascorbic acid is water soluble and in solution is readily oxidized to dehydroascorbic acid, and thus it acts as an antioxidant. It is sometimes used in food processing as an antioxidant.

Fruits and vegetables (including potatoes) are the classical sources of vitamin C. Significant amounts of vitamin C are also present in offal and fresh cow milk. There may even be traces in fresh raw meat and fish. Because it is used as a food additive, it may also be present in foods like ham, sausages and pate.

FUNCTIONS

Vitamin C is necessary for the functioning of the enzyme **proline hydroxylase**. This enzyme hydroxy-

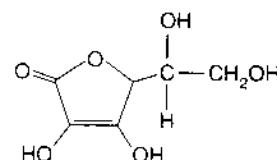


Figure 12.11 The structure of vitamin C – ascorbic acid.

lates residues of the amino acid proline to hydroxyproline once it has been incorporated into the important structural protein **collagen**. Collagen is a major structural protein. It has the capacity to form strong, insoluble fibres; it is the most abundant protein in mammals and serves to hold cells together. It is a structural component of bones, skin, cartilage and blood vessels and, indeed, fulfils a structural role in most organs. The hydroxylation of proline residues occurs after incorporation of the amino acid into procollagen (the precursor of collagen) and the hydroxylation is necessary for the structural properties of collagen; hydroxyproline is almost exclusively confined to collagen. A similar vitamin-C-dependent hydroxylation of lysine residues in procollagen also occurs.

Some other functions of vitamin C are listed below.

- It acts as an antioxidant in the aqueous phase (discussed later in this chapter).
- It is a cofactor in the synthesis of the important nerve transmitter noradrenaline (norepinephrine).
- It is a cofactor in the synthesis of carnitine, which is necessary for the transfer of long-chain fatty acids into the mitochondrion, where β -oxidation occurs.
- Large doses of vitamin C increase the absorption of non-haem iron from the gut.
- It is involved in the synthesis of several peptide hormones and transmitters, e.g. oxytocin, gastrin, calcitonin and antidiuretic hormone. These hormones all have a glycine residue at their C-terminals which has to be converted to an amide to make the active hormone.

REQUIREMENTS AND ASSESSMENT OF STATUS

The dietary standards for vitamin C are shown in Table 12.2. Vitamin C status can be assessed by the measurement of plasma vitamin C concentration; standards of normality are shown in Table 3.6.

DEFICIENCY STATES

Scurvy is the disease caused by lack of vitamin C. Most of the symptoms of scurvy can be attributed to impaired synthesis of the structural protein collagen, i.e. bleeding gums, loose teeth, subdermal haemor-

rhages due to leaking blood vessels, impaired wound healing and breakdown of old scars. Large areas of spontaneous bruising occur in severe scurvy, and haemorrhages in heart muscle or the brain may cause sudden death. Sudden death from haemorrhage or heart failure may occur even when the outward manifestations of scurvy do not appear particularly severe. Anaemia is usually associated with scurvy, almost certainly because vitamin C is an important promoter of iron absorption from the gut. The organic bone matrix is largely collagen and so osteoporosis may also occur in scurvy.

Scurvy occurs in people deprived of fresh foods for extended periods, especially with diets lacking in both fruits and vegetables. Outbreaks of scurvy commonly occurred among people living for extended periods upon preserved foods, e.g. those undertaking long sea voyages in sailing ships or taking part in expeditions to places where fresh foods were unavailable. The curative and preventative effects of fresh vegetables, fruits and fruit juices were documented almost two centuries before the discovery of vitamin C. In 1795, the regular provision of lime juice (or some suitable alternative) to British naval personnel resulted in the 'limey' nickname for British sailors. This nickname is still sometimes used to describe the British.

Around 10 mg/day of vitamin C is enough to prevent scurvy, but some sources recommend intakes of several grams per day for 'optimal health'. This issue is discussed in Chapter 3 and again later on in this chapter when the role of vitamin C as an antioxidant is discussed. Many people take very large doses of vitamin C (1–5 g/day) without any obvious toxic effects. When the body pool of the vitamin exceeds 1500 mg, it starts to be excreted in the urine, along with its metabolites, because the reabsorption system for vitamin C in the kidney becomes saturated.

FREE RADICALS AND ANTIOXIDANTS

Overview of free radicals and their disposal

WHAT ARE FREE RADICALS?

Free radicals are highly reactive species that have an unpaired electron, e.g. the **hydroxyl** and **superoxide** radicals. The electrons in an atom or molecule orbit

the nucleus in shells or layers and the most stable configuration occurs when these electrons are in pairs that orbit in opposite directions. If an atom or molecule within the body loses or gains an electron, the resulting product is highly reactive and can react with and damage DNA, proteins, lipids or carbohydrates. Cellular damage caused by oxygen-derived free radical species has been implicated in the aetiology of a range of diseases, including cancer, atherosclerosis, cataract and retinopathy. It has also been suggested that many of the degenerative changes associated with ageing may be due to the cumulative effects of free radical damage.

Oxygen free radicals can react with DNA to cause breaks in the DNA chain and alteration of bases (mutation); this could initiate carcinogenesis. Free radicals can peroxidize polyunsaturated fatty acid residues in **low-density lipoprotein (LDL)**. This altered LDL is taken up by macrophages and generates foam cells and this ultimately leads to the scarring and fibrosis of artery walls seen in atherosclerosis. Unoxidized LDL is considered relatively benign in its effects upon artery walls. The reactions of free radicals involve the loss or gain of an electron and this creates another free radical, which can initiate a damaging chain reaction unless the free radical is quenched by antioxidant systems and the chain reaction halted. For example, peroxidation of a polyunsaturated fatty acid will generate another unstable compound (the **lipid peroxy radical**) and this reacts with another fatty acid to produce a stable lipid peroxide and another peroxy radical, and so on (Figure 12.12). Their supposed susceptibility to free radical damage is one of the concerns about diets with very high levels of polyunsaturated fat (COMA, 1991).

Oxygen free radicals are a normal by-product of the oxidative processes of the cell. Some of the processes that generate free radicals are listed below.

- Free radicals are a by-product of the electron transport chain.
- Dissociation of oxygen from haemoglobin generates superoxide radicals.
- Certain environmental factors increase the generation of free radicals, e.g. cigarette smoke, exposure to high oxygen tension, exposure to ionizing radiation, including sunlight, some chemicals, including excesses of certain antioxidant nutrients such as iron and perhaps β -carotene.

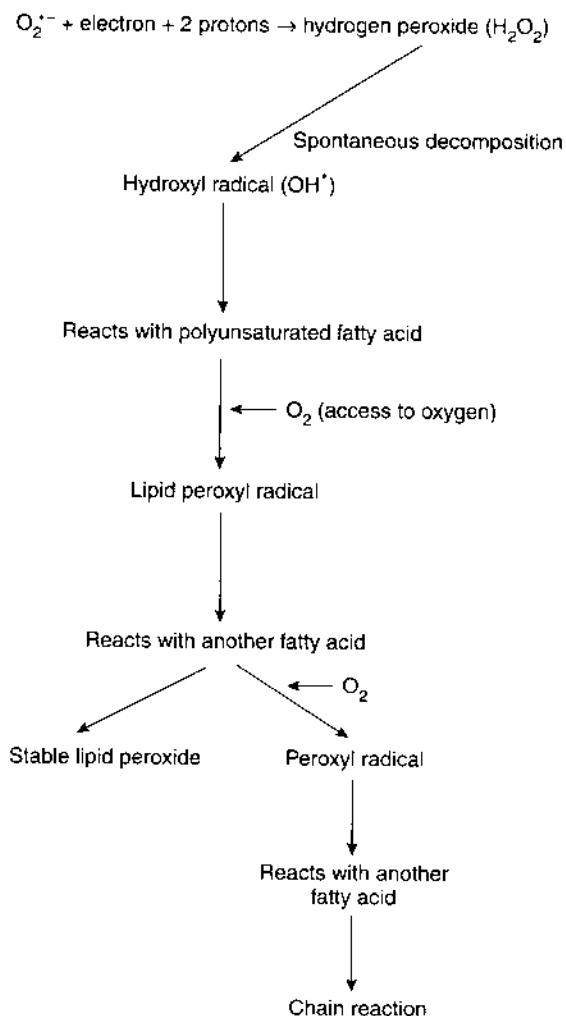


Figure 12.12 Reactions of the superoxide radical and its role in lipid peroxidation.

- Phagocytic white cells generate oxygen free radicals to kill ingested bacteria and destroy other 'foreign bodies'. They can also secrete these reactive species into surrounding tissues (e.g. to kill large parasites) and this can cause significant damage to surrounding tissues. Injured and diseased tissue thus has high levels of free radicals.

PHYSIOLOGICAL MECHANISMS TO LIMIT FREE RADICAL DAMAGE

Free radicals are normal by-products of the oxidative processes in cells and there are thus necessarily several physiological mechanisms whose specific role is to neutralize these free radicals and limit their tissue-

damaging effects. There are other mechanisms for the repair of the cellular damage induced by free radicals such as the mechanisms to repair damaged segments of DNA.

There are a number of metal-containing enzymes whose function is to scavenge and dispose of free radicals. Several essential nutrients are components of, or cofactors for, enzymes that are involved in free radical disposal. Some examples are listed below.

- Zinc and copper are components of the enzyme **superoxide dismutase**, which disposes of the superoxide radical by converting two superoxide radicals to hydrogen peroxide and oxygen.
- Selenium is a component of the enzyme **glutathione peroxidase**, which neutralizes hydrogen peroxide and converts it to water and oxygen. It also converts peroxidized lipids into stable and harmless products, thus breaking the chain reaction of free radical production.
- Iron is a component of the enzyme **catalase**, which converts hydrogen peroxide to water and oxygen.
- The enzyme **glutathione reductase** regenerates glutathione, which is oxidized by the glutathione peroxidase reaction mentioned above.

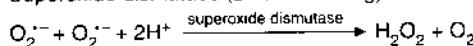
This enzyme is a flavoprotein and utilizes a riboflavin derivative as a prosthetic group.

In addition to these enzyme systems, some vitamins and other plant pigments have innate antioxidant properties and so have the capacity to scavenge free radicals, e.g. vitamin E in the lipid phase and vitamin C in the aqueous phase. Vitamins E and C are used as antioxidant food additives by food manufacturers. Some of the substances in food that are known to have, or probably have, an antioxidant effect are:

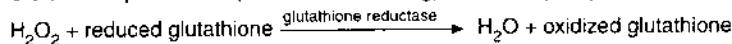
- the essential minerals selenium, zinc, copper and iron
- vitamins C and E
- β -carotene and several other carotenoids, including lycopene (abundant in tomatoes), lutein (found in green vegetables), α -carotene, zeaxanthin and cryptoxanthin
- other plant pigments, such as the polyphenols found in some fruits, tea, olive oil and red wine and the flavonoids found in grapes, nuts, oranges and strawberries.

Figure 12.13 summarizes some of the mechanisms involved in the disposal of free radicals.

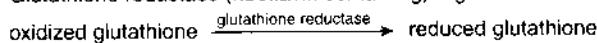
Superoxide dismutase (zinc containing) converts superoxide radicals to hydrogen peroxide:



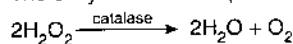
Glutathione peroxidase (selenium containing) converts hydrogen peroxide to water:



Glutathione reductase (riboflavin containing) regenerates reduced glutathione:



The enzyme catalase (iron containing) converts hydrogen peroxide to water and oxygen:



Vitamin E can quench free radicals when it is oxidized:



Vitamin E can be regenerated by a mechanism that involves vitamin C:

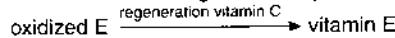


Figure 12.13 Some of the mechanisms for the disposal of free radicals.

SITUATIONS THAT INCREASE DAMAGE BY FREE RADICALS

The following circumstances may increase the risk of disease due to free radical damage of cellular components:

- increased generation of free radicals beyond the capacity of the mechanisms for their safe disposal and repair of the damage that they induce
- impaired capacity of the disposal mechanisms to handle any free radicals that are generated, for example due to dietary deficiency of a key antioxidant nutrient.

If premature babies are exposed to elevated oxygen concentration in their incubators, this can cause damage to the retina and lead to blindness. **retinopathy of prematurity.** This is one of the relatively few examples of a disease process being fairly strongly and causally linked to excess free radical production. High oxygen concentration is thought to lead to excessive generation of oxygen free radicals, which are responsible for the retinal damage. There is evidence that vitamin E, a free radical scavenger, protects these babies from the damaging effects of oxygen. High oxygen concentration of inspired air also produces lung damage in adults. When water is exposed to ionizing radiations, this generates hydroxyl radicals, which are responsible for much of the damage to living cells caused by ionizing radiations. There is also evidence that the toxic effects of excess iron may be due to free radical effects: *in vitro*, excess iron in serum stimulates lipid peroxidation and the formation of damaging hydroxyl radicals from hydrogen peroxide. Cigarette smoke and other air pollutants may exert some of their harmful effects by generating free radicals.

In most diseases, there is increased formation of oxygen free radicals as a consequence of the disease process. Infiltration of white cells into damaged or diseased tissue will result in excess oxidant load because superoxide generation is used by these cells to kill pathogenic bacteria. Mechanical injury to tissue also results in increased free radical reactions. This means that injured or diseased tissue generates free radicals much faster than healthy tissue and so reports that biochemical indices of free radical damage are raised in disease states should be treated with caution. These raised free radical indices may be a reflection of the disease process rather than an indi-

cation that free radical 'attack' is a key factor in initiating the disease, and they do not necessarily indicate that free radical scavengers such as vitamin E will prevent or ameliorate the condition. For example, in experimental deficiency of vitamin E in rats, there is a muscular atrophy that resembles that seen in the fatal, inherited human disease muscular dystrophy. Indices of free radical damage are, indeed, also raised in the wasting muscles of boys afflicted with muscular dystrophy, but vitamin E does not alter the progress of the disease. In this instance, the free radical damage is almost certainly a result of the disease rather than its primary cause.

The antioxidant minerals, together with riboflavin, are essential components of enzyme systems involved in the safe disposal of free radicals. Deficiencies of these micronutrients might be expected to reduce the activity of these disposal mechanisms and thus increase tissue damage by free radicals. However, given this mode of action, it seems unlikely that high intakes would exert any extra protective effects. Selenium or zinc deficiency might be expected to increase free radical damage, but it seems improbable that supplements of these minerals, in well-nourished people, would supply any extra protection against free radicals by increasing the synthesis of their dependent enzymes. **Keshan disease** may be an example of a deficiency disease due to lack of one of these antioxidant minerals. In this condition, there is a potentially fatal degeneration of the heart muscle (cardiomyopathy). This disease has been reported in certain areas of China where the soil is very low in selenium. Activity of the selenium-dependent enzyme glutathione peroxidase is low and the condition responds to selenium supplements, and has therefore been attributed to lack of dietary selenium with consequent failure of the free radical disposal mechanism.

The other antioxidant vitamins (E and C) together with carotene and some of the non nutrient antioxidants have the capacity to react with and quench free radicals without themselves becoming highly reactive. They thus have a free-standing antioxidant effect, independent of the enzyme mechanisms. There has, therefore, been considerable speculation as to whether high intakes of these micronutrients and plant pigments might protect against free radical damage and, in particular, reduce the long-term risk of cardiovascular disease and cancer. Epidemiological studies consistently report a negative

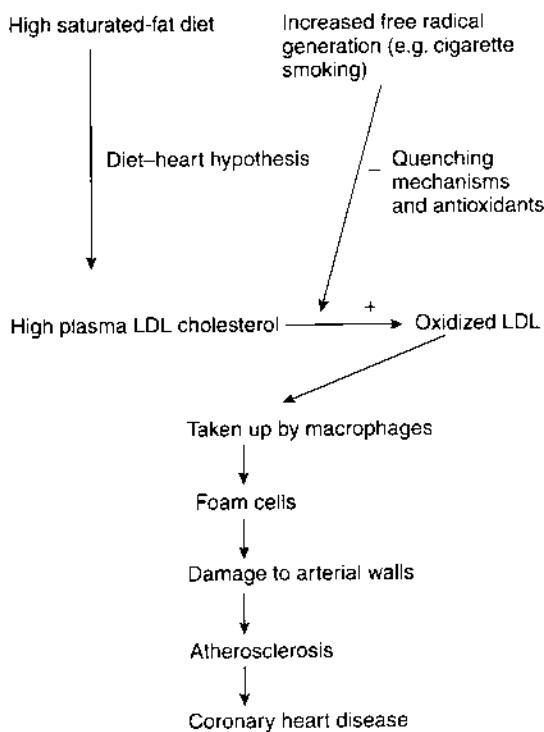


Figure 12.14 A hypothetical scheme to illustrate how free radicals might contribute to the risk of cardiovascular disease.

association between high intakes of fruits and vegetables and cancer risk, and these foods are the principal dietary sources of carotene, vitamin C and other plant pigments with antioxidant activity. There has also been much speculation about whether high intakes of vitamin E might exert a protective effect against the subsequent risk of cardiovascular disease.

Figure 12.14 shows a theoretical scheme that seeks to explain how a free radical mechanism might be involved in the development of atherosclerosis. Such a scheme could explain the apparent interaction between several risk factors in the development of cardiovascular disease and it might also help to explain why plasma LDL cholesterol concentration is such an imperfect predictor of cardiovascular risk. A high-saturated-fat diet tends to raise the LDL cholesterol concentration; cigarette smoking increases free radical production; and lack of dietary antioxidants limits the capacity to scavenge free radicals and thus to minimize the damage that they can induce. Any unquenched free radicals can react with LDL to convert it to much more atherogenic oxidized forms. Oxidized LDL is taken up very readily by macrophages and these LDL-loaded macrophages become foam cells, which are

- Free radicals are highly reactive species that are normal by-products of metabolic processes and immune defence mechanisms.
- Free radicals can react with DNA, membrane phospholipids, proteins and other cellular components.
- Oxidative damage caused by the reaction of free radicals with cellular components has been implicated in the aetiology of several chronic diseases and perhaps even in ageing.
- Free radical reactions generate further free radicals, and this can set up a damaging chain reaction leading to extensive cellular damage.
- Several cellular enzymes quench free radicals and these enzymes have riboflavin or a dietary mineral as an essential cofactor.
- Vitamins E and C and some plant pigments such as the carotenoids have inherent antioxidant and free radical-quenching activity.
- Increased free radical damage occurs when there is excess free radical production or impaired free radical disposal.
- Free radical production is increased in injured or diseased tissue and by certain environmental exposures such as radiation, sunlight, smoke or high oxygen pressure.
- Deficiencies of some antioxidant nutrients will impair the functioning of free radical-quenching enzymes.
- There is much speculation that high intakes of those vitamins and plant pigments with inherent antioxidant activity may afford extra protection from free radicals and thus reduce the risk of chronic diseases such as cancer and atherosclerosis.

characteristically found in atheromatous arterial lesions.

Conning (1991) and Diplock (1991) have written concise, referenced reviews of the role of antioxidant nutrients in disease. The VERIS website provides a wealth of information and analysis about dietary antioxidants (www.veris-online.org) and within this site three research summaries are recommended:

- the efficacy of carotenoids (August 1997)
- an overview of vitamin E efficacy (October 1998)
- the role of antioxidants in cancer prevention and treatment (December 1999).

Do high carotenoid intakes prevent cancer and other chronic diseases?

In an influential paper published in 1981, Peto *et al.* suggested that high intakes of β -carotene and other carotenoids might reduce the risk of human cancer. This effect was thought to be independent of their provitamin A activity. In the succeeding 20 years, hundreds of research papers have addressed the role of carotenoids in the prevention of cancer and other chronic, degenerative diseases.

The carotenoids are a group of more than 600 naturally occurring plant pigments and about 10% of these have provitamin A activity. β -carotene, α -carotene and cryptoxanthin act as provitamin A, but others found in significant amounts in human blood and tissues do not, e.g. lutein and lycopene. All of the carotenoids have antioxidant activity and are able to quench oxygen free radicals. Lycopene has the most potent antioxidant effect. Coloured fruits and vegetables are the richest sources of carotenoids, e.g. carrots, green vegetables, tomatoes, cantaloupe melon:

- carrots are rich in β -carotene
- tomatoes and water melon are rich in lycopene
- green, leafy vegetables and red peppers are rich in lutein
- mangoes and sweetcorn are rich in cryptoxanthin.

There is an overwhelming body of evidence that suggests that diets high in fruits and vegetables are associated with a low risk of cancers and other chronic diseases. This means that diets high in carotenoids have been associated with reduced cancer risk in many case-control and in more reliable cohort studies. Low blood levels of β -carotene

and other carotenoids (especially lycopene) are also associated with an increased risk of some cancers (see Ziegler, 1991, and www.veris-online.org for a review of this epidemiological evidence). In some studies, carotenoids have also been reported to protect against experimentally induced cancers in laboratory animals, despite the fact that β -carotene is very poorly absorbed by experimental animals. The huge body of epidemiological evidence is generally consistent with the proposition that dietary carotenoids protect against cancer and, indeed, other chronic degenerative diseases in which oxidative damage by free radicals is implicated. However, there are several possible explanations for these epidemiological findings such as those listed below.

- The carotenoids are directly protective against cancer and degenerative disease because of their antioxidant and perhaps other effects.
- Other components of fruits and vegetables (e.g. other plant pigments or other nutrients) might be exerting a protective effect, and the raised carotenoid intakes are thus a secondary consequence of this.
- High fruit and vegetable consumption is not directly protective, but is merely a marker for a generally healthy diet, e.g. lower in fat, saturated fat or meat protein.
- High fruit and vegetable consumption is a marker for a generally healthy lifestyle; people who eat the most fruits and vegetables are the most affluent, physically active and health conscious.

In order to test the cancer–carotenoid hypothesis more directly, a number of large-scale intervention trials of supplements have been carried out. Unfortunately, several of these studies have used multiple supplements, which hinders their interpretation. In general, those intervention studies carried out using well-nourished populations have failed to show any significant benefit for β -carotene supplements. Some studies have actually reported a significant increase in the risk of lung cancer in high-risk groups such as smokers and asbestos workers (i.e. those initially expected to be most likely to benefit). These findings have led COMA (1998) to specifically recommend the avoidance of β -carotene supplements and ‘to exercise caution in the use of purified supplements of other micronutrients’.

The one major trial of a β -carotene-containing supplement that did indicate a beneficial effect upon cancer risk was conducted in China (Blot *et al.*, 1993). This trial used 30 000 middle-aged people living in the Linxian province of China; some were given supplements containing β -carotene, selenium and vitamin E. A 13% reduction in cancer was reported in the group receiving this supplement and a 20% reduction in gastric cancer mortality. The participants in this study had low baseline intakes and so the supplements could be seen as correcting an existing deficiency and the benefits as eliminating the adverse consequences of inadequacy. The effects could have been caused by any of the three components of the supplements.

A Finnish study published in 1994 (Group, 1994) was the first to indicate possible adverse consequences of β -carotene supplements. This was a randomized, double-blind, placebo-controlled trial of the effects of vitamin E and/or β -carotene supplements in a sample

of 30 000, 50–69-year-old, male smokers. Neither supplement appeared to afford any protection against the development of lung cancer during the follow-up period (5–8 years). The rates of lung cancer in the β -carotene-supplemented men were, in fact, significantly higher than the rates in those not receiving β -carotene. There was no significant difference in total mortality between those receiving and not receiving vitamin E, but total mortality was significantly higher in those receiving β -carotene supplements (8%) due to increased deaths from lung cancer, heart disease and stroke. The authors concluded that supplements may well have harmful as well as beneficial effects.

A 12-year controlled study of β -carotene supplements involving 22 000 middle-aged, male American doctors showed no benefits of β -carotene supplements upon the overall incidence of cancer or cardiovascular disease (Hennekens *et al.*, 1996).

In another study in the USA, the CARET study, half of 18 000 men identified as being at high risk of

- The carotenoids are a group of more than 600 plant pigments that are found in green and other brightly coloured fruits and vegetables.
- A few carotenoids, such as β -carotene, have provitamin A activity and all of them are antioxidants that can quench free radicals.
- Diets high in fruits and vegetables (thus also high in carotenoids) are associated with a reduced risk of cancer and chronic disease in both case-control and cohort studies.
- Low blood levels of carotenoids are also associated with an increased cancer risk.
- It has been proposed that high intakes of β -carotene and other antioxidant carotenoids protect against cancer and chronic disease, and that this is independent of their provitamin A activity.
- There are several other possible explanations for the epidemiological association between high carotenoid intake and low cancer risk, e.g. other components of fruits and vegetables may be protective or a high intake of fruit and vegetables may simply be a marker for a generally healthy diet or lifestyle.
- Large-scale intervention trials of β -carotene-containing supplements in industrialized countries have failed to show any beneficial effects and have suggested that these supplements increase lung cancer risk in high-risk groups such as smokers and asbestos workers.
- Large supplements of β -carotene may do more harm than good, and this suggests more generally that one should exercise caution in the use of purified micronutrient supplements unless habitual dietary intakes are inadequate.
- Increased consumption of fruits and vegetables is recommended and would increase carotenoid intakes.
- Experiments with rats suggest that large doses of β -carotene may induce certain enzymes that activate environmental carcinogens and thus act as co-carcinogens.
- A co-carcinogenic action of β -carotene might explain the increased cancer risk in those exposed to high levels of environmental carcinogens and given β -carotene supplements.

lung cancer were given supplements of β -carotene and retinol. This trial was discontinued after only 4 years because the rates of lung cancer in the supplemented group were significantly higher than those in the placebo group (Omenn *et al.*, 1996).

As with cancer, the epidemiological evidence generally supports the view that a high intake or blood level of β -carotene is associated with a reduced risk of coronary heart disease. However, at least one intervention trial has reported findings that would suggest caution in the use of β -carotene supplements (Rapala *et al.*, 1997). In this study, 18 000 male, Finnish smokers with a history of myocardial infarction were given supplements of β -carotene and/or vitamin E or a placebo. There was an increased death rate from coronary heart disease in those receiving supplemental β -carotene.

The epidemiological evidence suggests that carotenoids may protect against cancer and cardiovascular disease, but several intervention trials suggest that, in those with adequate baseline intakes, supplements may do no good and may do significant harm to smokers and others at high risk of lung cancer and perhaps coronary heart disease.

As a result of some short-term experiments with rats, Paolini *et al.* (1999) have suggested that β -carotene might exert a co-carcinogenic effect by inducing certain enzymes that activate environmental carcinogens. This theory could explain the apparent contradictions: carotenoids might protect DNA from oxidative damage in the early stages of cancer development, but, nonetheless, very large doses might help activate environmental carcinogens and have detrimental effects upon those with high exposure to these, such as smokers.

Are vitamin E supplements beneficial?

Vitamin E is a potent antioxidant and it has been shown to protect against some conditions whose symptoms are confidently thought to be due to free radical-induced tissue damage, e.g. the retinopathy induced by exposure to high oxygen concentrations in premature babies. Riemersma *et al.* (1991) identified around 100 men suffering from undiagnosed angina pectoris (chest pain due to atherosclerotic narrowing of the coronary blood vessels) by using a self-administered chest pain questionnaire, and matched them with around 400 controls who

appeared to be free of angina. They found that plasma concentrations of the three antioxidant vitamins (vitamin C, vitamin E and carotene) were inversely related to the risk of angina. Only the relationship for vitamin E remained statistically significant when allowance was made for confounding variables such as smoking. Similarly, Gey *et al.* (1991) reported an inverse correlation between blood levels of vitamin E and mortality from coronary heart disease in a cross-sectional study of men from 16 European countries. In both of these studies, it is assumed that variations in vitamin E intakes are largely the result of dietary variations rather than the effects of supplements.

Two large cohort studies have reported that, in both women (Stampfer *et al.*, 1993) and men (Rimm *et al.*, 1993), the consumption of high levels of vitamin E was associated with a reduced risk of coronary heart disease. The associations were not removed by correction for expected confounding variables. In neither of these studies was any significant association found between vitamin E intake from food and the risk of coronary heart disease, and thus the apparent benefit was only observed at doses of vitamin E beyond those consumed in typical US diets. In the women's study, the follow-up period was 8 years and detailed information was given on other causes of mortality during this period. Neither vitamin E supplements nor multivitamin supplements were associated with any significant reduction in overall mortality, despite a sample size of almost 90 000 women – there was a non-significant reduction in those taking vitamin E supplements, but, equally, there was a non-significant increase in those taking multivitamins. The results from intervention trials using vitamin E are mixed (see www.veris-online.org). One randomized trial of vitamin E supplements involved 2000 English patients assessed as being at high risk of myocardial infarction on the basis of angiograms. This study reported a highly significant reduction in cardiac events, but no significant change in cardiovascular mortality (a small, non-significant rise) (Stephens *et al.*, 1996). In an Italian study of vitamin E and n-3 (ω -3) polyunsaturated fatty acid supplements involving more than 11 000 patients who had experienced a previous myocardial infarction, there was no indication of any benefit from the vitamin E supplements after 3.5 years (GISSI, 1999).

The results from the β -carotene supplementation studies discussed earlier suggest that one should be very wary of advising long-term, mass consumption

of pharmacological doses of vitamins unless the long-term and holistic benefits have been unequivocally demonstrated and the long-term safety confirmed. COMA (1991) felt that there was still insufficient information about human vitamin E requirements to offer a full set of UK Dietary Reference Values, and only a safe intake was given.

Vitamin E is found in leafy vegetables, wholegrain cereals, liver and egg yolks, but the richest sources are vegetable oils and the fat fractions of nuts and seeds. Recommending diets specifically aimed at raising vitamin E concentrations could thus conflict with the general advice to reduce total fat intakes, although increased intakes of vegetables and wholegrain cereals are consistent with other recommendations. The requirement for vitamin E is increased by high intakes of polyunsaturated fatty acids because of their susceptibility to peroxidation by oxygen free radicals.

Are high intakes of vitamin C beneficial?

The UK panel on Dietary Reference Values (COMA, 1991) listed at least eight benefits that have been suggested for intakes of vitamin C that are well in excess of those needed to prevent or to cure scurvy. They were, nevertheless, not persuaded to allow any of these suggested benefits to influence their decisions about the reference values for vitamin C. The size of the body pool of vitamin C is maximized at intakes of 100–150 mg/day. At higher intakes, the absorption of vitamin C from the intestine becomes saturated

and there is increased renal excretion of vitamin C. It has been suggested at one time or another that high intakes might have the following beneficial effects:

- a serum cholesterol-lowering effect
- improved physical work capacity
- improved immune function
- reduced formation of carcinogenic substances called nitrosamines in the stomach and thus a reduced risk of stomach cancer
- improved male fertility
- extended survival time in patients with terminal cancer when gram quantities are administered. There is no consistent or persuasive evidence to support this.

Perhaps the most persistent of these claims is that intakes of more than ten times the RNI/RDA for vitamin C might help prevent the development of the common cold. This view was popularized by a book published in the early 1970s and written by the double Nobel Prize winner Linus Pauling (Pauling, 1972). Since 1970, there have been dozens of controlled trials on the efficacy of vitamin C in preventing colds and/or reducing their duration. Despite this plethora of research effort, there remains no convincing evidence that high vitamin C intakes affect either the frequency of colds or their duration. The original basis for advocating very high intakes of vitamin C was largely cross-species extrapolations of either the amounts synthesized by the rat or the amounts consumed by other vitamin-C-requiring primates such as gorillas. More recent recognition that the antioxidant properties of vitamin C may be important in protecting against free radical damage

- Vitamin E is a potent antioxidant.
- High blood levels and dietary intakes of vitamin have been associated with a reduced risk of coronary heart disease.
- Two large cohort studies have reported that supplemental vitamin E is associated with a reduced risk of coronary disease in men and women.
- Intervention trials of vitamin E supplements have shown mixed results.
- Current evidence does not seem to justify the widespread use of vitamin E supplements.
- Leafy vegetables and wholegrain cereals contain vitamin E, but the richest sources are vegetable oils, and increased consumption of these is not consistent with general nutrition education guidelines.
- The increased consumption of polyunsaturated fatty acids increases vitamin E requirements.

is more persuasive and, once again, the epidemiological evidence supports an association between high dietary intakes of vitamin C and a reduced risk of chronic disease.

Claims for the beneficial effects of vitamin intakes that are orders of magnitude higher than the reference values or usual intake are really the province of the pharmacologist rather than the nutritionist. The vitamin is being used as a potential drug rather than as a nutrient. Vitamin C is one of the antioxidant vitamins, and there is a fairly consistent body of epidemiological evidence that diets rich in vitamin C (i.e. high in fruits and vegetables and relatively low in fat) are associated with a reduced risk of cancer and cardiovascular disease. There is a good case, both theoretical and observational, for recommending diets rich in vitamin C, but there is little convincing evidence that vitamin C supplements *per se* will confer any significant health benefits.

Many thousands of people (maybe a third of Americans) are unconvinced by arguments against the efficacy of vitamin C supplements and, in such individuals, the placebo effect may mean that they do, in practice, derive benefit from these preparations. Such widespread consumption of doses that are far in excess of those that can be obtained from

normal human diets means that the potential toxicity of these preparations needs to be considered. Despite the widespread use of supplements, reports of adverse effects are few and the vitamin appears to be non-toxic, even at doses many times the RNI/RDA, say 1 g/day, and much of the excess is excreted in the urine. Some concerns about intakes of gram quantities of the vitamin are listed below (see Shils *et al.*, 1999).

- Very large excesses may lead to diarrhoea and gastrointestinal symptoms because the absorption system is saturated and vitamin C remains in the gut.
- Products of vitamin C metabolism may increase the risk of gout and kidney stones in susceptible people. The risk for healthy people is probably small.
- Large doses of vitamin C may interfere with anti-coagulant therapy.
- Large amounts of vitamin C excreted in the urine may make urinary tests for diabetes and faecal tests for occult blood unreliable.

Doses of up to 1 g/day can probably be tolerated for very long periods with no obvious ill-effects in most people.

- Numerous claims have been made for beneficial effects of very large intakes of vitamin C.
- The claim that high vitamin C intakes improve immune function has been particularly persistent.
- Vitamin C is a potent antioxidant.
- Vitamin C intakes of 100–150 mg/day maximize the body pool and excesses beyond this are readily excreted in the urine.
- A diet rich in vitamin C (high in fruits and vegetables) is associated with a reduced risk of chronic disease, but the evidence that megadoses of vitamin C confer extra benefits is generally inconclusive or weak.
- The use of gram quantities of vitamin C or any other vitamin is the province of the pharmacologist rather than the nutritionist.
- The use of large vitamin C supplements is widespread and there is little evidence that these are harmful for most people.
- Large vitamin C supplements may be contraindicated under some circumstances and in some individuals.

The minerals

Introduction	307	Calcium, diet and osteoporosis	317
Iodine and iodine deficiency diseases	311	Salt and hypertension	327
Iron and iron-deficiency anaemia	313		

INTRODUCTION

In the UK, Dietary Reference Values (DRVs) are given for 11 minerals and safe intakes are listed for a further four (COMA, 1991). These same 15 minerals are also deemed to warrant a recommendation of some type in the American Recommended Dietary Allowances (RDAs) (NRC, 1989a). There is, therefore, international acceptance that at least 15 dietary minerals are essential. Several of these essential minerals are required only in trace amounts and in some cases requirements are difficult to establish with confidence. COMA (1991) considered the possible essentiality of a further 19 minerals, most of these they classified as non-essential or of unproven essentiality, a few they regarded as probably essential in trace amounts.

Clinically recognizable, acute dietary deficiency states are extremely rare for most essential minerals. Only two such deficiency diseases either have historically been or are currently seen as anything other than extremely rare or localized, i.e. goitre, a swelling of the thyroid gland caused by lack of dietary iodine, and iron-deficiency anaemia. These are both discussed below. This has not prevented very active debate about whether suboptimal intakes and sub-clinical deficiencies of other minerals are widespread, particularly for calcium, selenium and zinc.

Table 13.1 gives a selective summary of factual information about the dietetically important minerals, i.e. their functions, sources, adult reference values and deficiency states.

Table 13.2 shows the measured intakes of five dietetically important minerals in a representative sample of British men and women, together with the corresponding Reference Nutrient Intakes (RNIs) and Lower Reference Nutrient Intakes (LRNIs). Several of the mean values are below the RNI, i.e. iron and magnesium in women and potassium in both sexes. The median for iron intake of women is substantially below the mean because of skewing of the distribution by supplemental iron, and this median intake represents only around two-thirds of the RNI. As with the vitamins, the intakes of the lowest 2.5% of the population are in several cases below the LRNI. Note that there was probably some under-recording of habitual intakes in this survey. Note also the very high intakes of iron of the top 2.5%; in the case of men, these intakes were over three times the RNI.

The role of salt in the aetiology of hypertension and the significance of dietary calcium in the aetiology of osteoporosis are important mineral-related topics, which have been singled out for discussion at the end of this chapter. The role of dietary fluoride in dental health is discussed in Chapter 9. Fluoride has not been proven to be an essential nutrient, and in both the UK and the USA, only safe intakes are given. In the UK, even this safe intake is restricted to infants. The roles of zinc, selenium, iron and copper as prosthetic groups for antioxidant enzyme systems are discussed in Chapter 12, in the section dealing with free radicals and antioxidants.

Table 13.1 Selective summary of the dietetically important minerals

<i>Dietary Reference Values</i>	
Calcium	UK adult RNI male/female; (American RDAs male/female) 700 mg/day; (800)
Phosphorus	550 mg/day, equivalent in molar terms to the calcium RNI; (800 same as calcium)
Magnesium	300/270 mg/day; (350/280)
Sodium	1.6 g/day; (estimated minimum requirement 500 mg/day)
Potassium	3.5 g/day; (estimated minimum requirement 2 g/day)
Chloride	2.5 g/day, equivalent in molar terms to the sodium RNI; (estimated minimum requirement 750 mg/day)
Iron	8.7/14.8 mg/day; (10/15)
Zinc	9.5/7 mg/day; (15/12)
Copper	1.2 mg/day; (safe intake 1.5–3 mg/day)
Selenium	75/60 µg/day; (70/55)
Iodine	140 µg/day; (150)
<i>Dietary sources</i>	
Calcium	Milk and milk products are rich sources with high biological availability; fish, especially if fish bones can be eaten as in tinned fish; green vegetables; pulses; nuts; wholegrain cereals (white flour is supplemented in the UK); the calcium in vegetables and in cereals may have relatively low biological availability 1 cup (195 g) whole milk – 225 mg 40-g slice of cheddar cheese – 290 mg 1 serving (20 g) almonds – 50 mg 1 slice wholemeal bread – 19 mg 1 serving (95 g) broccoli – 72 mg 1 serving (75 g) cabbage – 40 mg 1 portion (85 g) canned sardines – 390 mg 1 serving (80 g) peeled prawns – 120 mg 1 portion (130 g) grilled cod – 13 mg
Magnesium	Vegetables; pulses; fruits
Sodium	Major sources of dietary salt (sodium chloride) are processed foods; small amounts in natural foods; discretionary salt, added during cooking and at the table 1 slice wholemeal bread – 0.5 g salt 40-g slice cheddar cheese – 0.6 g salt 40-g slice Danish blue-type cheese – 1.4 g salt 1 serving (45 g) All-Bran – 1.9 g salt 1 serving (25 g) cornflakes – 0.7 g salt 4 fish fingers – 0.9 g salt 1 portion (85 g) canned sardines – 1.5 g salt 1 (80 g) portion prawns – 3.2 g salt 1 (120 g) gammon steak – 6.5 g salt 2 slices (60 g) corned beef – 1.5 g salt 2 pork sausages (UK) -1.5 g salt Small packet salted peanuts – 0.3 g salt 1 bowl (145 g) canned tomato soup – 1.7 g salt Small packet crisps (chips) – 0.4 g salt 1 portion (200 g) baked beans – 2.4 g salt 1 serving (60 g) liver pâté – 1.1 g salt
Potassium	Fruits and vegetables are the best sources; some in milk and flesh foods 1 orange – 490 mg 1 banana – 470 mg 1 apple – 145 mg 1 portion (95 g) broccoli – 210 mg 1 portion (75 g) frozen peas – 100 mg

Table 13.1 (Continued)

	1 tomato – 220 g
	1 portion (265 g) chips (French fries) – 270 mg
	1 glass (200 g) canned orange juice – 260 mg
Iodine	The richest natural sources are seafood; amounts in fruits, vegetables and meat depend upon soil content of iodine; modern farming and processing methods greatly increase amounts in milk and some bakery products; iodized salt
Zinc	Meats; wholegrain cereals; pulses; shellfish 1 slice wholemeal bread – 0.7 mg 1 pork chop – 2.8 mg 1 portion (75 g) frozen peas – 0.5 mg 1 portion fried lamb liver – 4 mg 1 serving (70 mg) canned crabmeat – 3.5 mg
Copper	Green vegetables; many types of fish; liver
Selenium	Meat; cereals; fish
Iron	Meat (particularly offal); fish; cereals; green vegetables; the biological availability of iron is much higher from meat and fish than from vegetable sources 1 pork chop – 1 mg 1 portion (90 g) fried lamb liver – 9 mg 1 egg – 1.1 mg 1 serving (80 g) peeled prawns – 0.9 mg 1 serving (85 g) canned sardines – 3.9 mg 1 portion (75 g) frozen peas – 1.1 mg 1 portion (95 g) broccoli – 1 mg 1 50-g bar of dark (plain) chocolate – 1.2 mg 1 serving (130 mg) spinach – 5.2 mg 1 slice wholemeal bread – 1 mg
<i>Biochemical/physiological functions</i>	
Calcium	99% of body calcium is in bone mineral; important in nerve and muscle function; release of hormones and nerve transmitters; blood clotting; intracellular regulator of metabolism
Phosphorus	80% of body phosphorus is in bone mineral; phosphorylation–dephosphorylation reactions are an important feature of all biochemical pathways (see ATP functioning in Chapter 5); component of phospholipids, which are important in membrane function; buffering systems
Magnesium	Most of body magnesium is in bone; important role in calcium metabolism; many magnesium-requiring enzymes
Sodium	Major cation in extracellular fluid; maintains fluid and electrolyte balance; nerve impulse transmission; acid–base balance
Potassium	Major cation in intracellular fluid; nerve impulse transmission; contraction of muscle and heart; acid–base balance
Chloride	Major anion in the body; maintains fluid and electrolyte balance
Iron	Component of haemoglobin, myoglobin and cytochromes; there are several iron-containing enzymes such as catalase
Zinc	Numerous zinc-containing enzymes, including superoxide dismutase; insulin is stored as complex with zinc in the pancreas
Copper	There are numerous copper-containing enzymes
Selenium	Component of enzyme system glutathione peroxidase, involved in disposal of free radicals; other selenium-containing enzymes
Iodine	A component of the thyroid hormones
<i>Effects of deficiency</i>	
For most minerals, overt deficiency states are rarely, if ever, seen	
Calcium	Overt, primary calcium deficiency is rare; vitamin D deficiency leads to poor absorption of calcium and the disease rickets; osteoporosis suggested as a long-term consequence of calcium insufficiency
Iron	Iron deficiency anaemia

Table 13.1 (Continued)

Copper	Anaemia due to copper deficiency has been reported in premature babies and patients fed by total parenteral nutrition
Selenium	Keshan disease, a progressive and potentially fatal cardiomyopathy seen in parts of China where soil is low in selenium; experimental deficiency in animals is similar to vitamin E deficiency, presumably because of their common role in disposal of free radicals
Iodine	Goitre/iodine deficiency diseases
<i>Risk factors for deficiency</i>	
Iron	Chronic blood loss due to, for example, heavy menstrual bleeding, malignancy, intestinal parasites or repeated pregnancies; a vegetarian diet that is low in iron bioavailability and lacks promoters of absorption such as vitamin C
Iodine	Living in areas where soil iodine content is low
Calcium	Vitamin D deficiency

Table 13.2 Comparison of measured intakes and Dietary Reference Values for selected minerals in British adults

Mineral	Mean intake	Lower 2.5 percentile	RNI (19–50 years)	LRNI
Iron (mg)				
Men	14.0	6.5 (27.1 ^b)	8.7	4.7
Women	12.3 (10.0 ^a)	4.7 (30.7 ^b)	14.8	8.0
Calcium (mg)				
Men	940	410	700	400
Women	730	266	700	400
Magnesium (mg)				
Men	323	156	300	190
Women	237	105	270	150
Zinc (mg)				
Men	11.4	5.7	9.5	5.5
Women	8.4	3.6	7.0	4.0
Potassium (g)				
Men	3.2	1.7	3.5	2.0
Women	2.4	1.2	3.5	2.0

EAR is midway between RNI and LRNI.

^a Median (middle value if arranged in rank order) may be more useful than mean with a skewed distribution.^b Upper 2.5 percentile.Data sources: COMA (1991); and Gregory *et al.* (1990).

- There are at least 15 essential minerals.
- Iodine deficiency and iron deficiency are two of the most prevalent micronutrient deficiencies in the world.
- Acute, clinical deficiency of most other minerals is rare or localized.
- The average UK intakes of several minerals are below the RNI, and there is further evidence of low mineral intakes in a large number of individuals.
- Iron supplements skew the average intakes of British women upwards.
- The top 2.5% of British men have iron intakes that are three times the RNI.

IODINE AND IODINE DEFICIENCY DISEASES

Distribution and physiological function of body iodine

Iodine is a component of the thyroid hormones **thyroxine** and **triiodothyronine**. The thyroid gland traps and concentrates absorbed iodine very efficiently; if radioactively labelled iodine is taken in, it can be shown to be rapidly concentrated within the thyroid gland. The ability of the thyroid to concentrate 'labelled' iodine is used diagnostically as a measure of thyroid function. High doses of radioactive iodine may be used as an alternative to surgery to selectively destroy thyroid tissue in some cases of hyperthyroidism or thyroid carcinoma, especially in the elderly. Nuclear emissions released after accidents or explosions contain large amounts of relatively short-lived radioactive iodine, and this accumulates in the thyroid, where it can ultimately cause thyroid carcinoma. The administration of iodine-containing tablets can be used as a public health measure after nuclear accidents to compete with and competitively reduce the uptake of the radioactive iodine into the thyroid.

Iodine deficiency

In goitre, there is a swelling of the thyroid gland. This swelling ranges from that which is only detectable to the trained eye to massive and sometimes nodular overgrowth. The thyroid swells because of excessive stimulation by the pituitary hormone **thyrotropin**. Thyrotropin output is normally inhibited by circulating thyroid hormones and so any reduced output of thyroid hormones caused by iodine deficiency leads to compensatory increases in thyrotropin output and consequent thyroid growth (Figure 13.1). In some cases, the swelling of the thyroid is the only apparent symptom of iodine deficiency.

In severe iodine deficiency, there are symptoms of thyroid hormone deficiency. In children, iodine deficiency can lead to **cretinism**, a condition characterized by impaired mental and physical development. In adults, iodine deficiency causes not only goitre, but can also lead to impaired mental functioning,

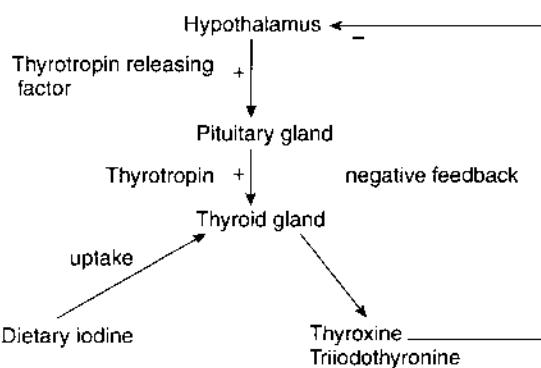


Figure 13.1 The regulation of thyroid hormone output and the origins of goitre.

low metabolic rate, hypotension, weight gain and other symptoms of thyroid hormone deficiency. In places where goitre is endemic, it causes increased rates of spontaneous abortion and stillbirth and the birth of large numbers of children with congenital physical and neurological abnormalities, e.g. deaf-mutism, spasticity and mental deficiency. There is a general impairment of intellectual performance and psychomotor skills in children from iodine-deficient areas. If a goitre is very large, it may lead to problems with breathing or swallowing that may require surgery, and occasionally large nodular goitres may become malignant. Controlled studies by David Marine in Ohio in the 1920s showed that iodine supplements were effective in the treatment and prevention of goitre in schoolchildren. Women and girls are more susceptible to goitre than males.

Causes and scale of the problem

The richest sources of dietary iodine are seafoods. The smaller amount of iodine in other foods varies according to the iodine level in the soil. The iodine content of drinking water gives a good indication of the general iodine level in the soil in an area. In many regions of the world, the soil iodine content is low, especially in mountainous regions and flooded river valleys, where glaciation and run-off from melting snow and heavy rain have leached much of the iodine from the soil. The area around the Great Lakes in the USA and Derbyshire and the Cotswold areas of England were once goitre areas. The Himalayas, Andes, Alps and the mountains of China, together

with the Ganges valley and other river valleys in Asia, are all areas where soil iodine content is inadequate. In areas where the soil iodine content is low and where seafood is not consumed, the iodine deficiency diseases, manifesting most apparently as goitre, have been, and often still are, endemic. Iodine deficiency is one of the most common nutrient deficiencies in the world and is the most common cause of preventable mental deficiency in children. The World Health Organization (WHO) estimated that, in 1990, there were over a billion people in the Third World at risk of iodine deficiency, with around 200 million suffering from goitre and around 20 million of these suffering from some degree of resultant mental defect – more than 5 million from gross cretinism and mental retardation. In 1993, the WHO revised the number of people at risk of iodine deficiency diseases up to 1.6 billion. Iodine deficiency is endemic in large parts of Central and South America, large parts of India, China and Southeast Asia, and some regions of Africa and the Oceanic islands; around 85% of those suffering from overt cretinism are in Asia (Hetzl, 1993; Hetzel and Clugston, 1999).

Nowadays, few affluent communities rely solely upon food produced in their locality and this reduces

the risk of goitre. In both the UK and USA, the iodine content of cow milk is high because of the iodine supplementation of feed and the use of iodine disinfectants. In many goitre areas, including Switzerland and the USA, iodized salt has been effectively used as a prophylactic measure to prevent goitre. In some other areas, injections of iodized oil have been used and a single injection gives protection for 2–3 years. Iodized oil can be administered orally, but the duration of effectiveness is reduced to about 1 year.

Requirements and intake of iodine

The adult RNI (RDA) for iodine is 140 (150) µg/day and intakes are typically double this in the UK and more than double in the USA. The very high intakes in the USA, which reached a peak of 800 µg/day in 1974, have been a cause for concern. Very high intakes can cause toxic nodular goitre and hyperthyroidism. COMA (1991) suggested a safe upper limit of 1000 µg/day. Some Japanese populations that habitually consume seaweed have intakes many times this suggested maximum with no apparent ill-effects, and Hetzel and Clugston (1999) suggest that

- Iodine is essential as a component of the hormones thyroxine and triiodothyronine.
- In iodine deficiency, reduced thyroid hormone output causes increased thyrotropin output from the pituitary, which stimulates the thyroid and causes it to swell (goitre).
- Severe iodine deficiency in children causes impaired mental and physical development (cretinism), and it leads to symptoms of thyroid insufficiency in adults.
- In goitre areas, there is general impairment of mental performance and increased numbers of abortions, stillbirths and births of children with mental and physical defects.
- Iodine deficiency occurs in mountainous areas and frequently flooded river valleys where melting snow and floodwater have leached the iodine from the soil.
- One to two billion people in the world are at risk of iodine deficiency and 20 million suffer some degree of mental defect as a result of iodine deficiency.
- Iodine deficiency is the most common cause of preventable mental deficiency in children.
- Access to seafood and food from various geographical regions reduce the risk of iodine deficiency.
- Iodized salt and injections of iodized oil have been used to prevent and treat iodine deficiency.
- Very high intakes of iodine may cause toxic goitre and hyperthyroidism, but those people with chronically low iodine intakes may be particularly susceptible.
- Certain foods such as cabbage and cassava contain goitrogens that increase the requirement for iodine.

populations that have been iodine deficient may be much more sensitive to the toxic effects of excess iodine.

Certain foods of the cabbage (*Brassica*) family and cassava contain goitrogenic substances that either block the uptake of iodine by the thyroid gland or block the synthesis of thyroid hormones. The presence of such goitrogens may precipitate deficiency where intakes of iodine are at the borders of adequacy. Cassava, in particular, may be a significant factor in precipitating iodine deficiency in some tropical areas, where it is the principal energy source for around 200 million people. Hetzel (1993) suggested that intakes of iodine should be doubled to 200–300 µg/day where there is a high intake of these goitrogens.

IRON AND IRON-DEFICIENCY ANAEMIA

Iron nutrition

DISTRIBUTION OF BODY IRON

An average, well-nourished male body will contain about 4 g of iron. Around 70% of this total will be present in the respiratory pigments **haemoglobin** in blood (67%) and **myoglobin** in muscles (3–4%). Most of the body's non-haem iron will be stored as the iron–protein complex **ferritin**. Small amounts of iron will also be found in a variety of important iron-containing enzymes and proteins and bound to the protein **transferrin** in plasma that is responsible for the transport of iron around the body. The first effect of iron deficiency is depletion of iron stores and this is followed by reduced levels of the iron-containing respiratory pigments haemoglobin and myoglobin, i.e. iron-deficiency anaemia.

REQUIREMENT FOR DIETARY IRON

Iron is very efficiently conserved by the human body, with daily losses of body iron in a typical, healthy male amounting to only around 1 mg/day (i.e. only 0.025% of normal body iron content). This tiny net loss of iron compares with around 20 mg of iron in the ageing red blood cells that are removed from the circulation and destroyed each day. The iron released from broken down red cells is recycled and the small

amount of iron lost from the body is that present in lost skin, nails, hair, body fluids and sloughed cells from the alimentary and genitourinary tracts. In pre-menopausal women, losses of menstrual blood and the losses of iron at parturition and during lactation also contribute to body iron loss. Average menstrual blood loss represents an additional increment of around 80% to the general iron losses of women, but in some women (about 10%) these extra menstrual losses may represent an addition of 150% or more to the general iron loss. Any condition which results in chronic blood loss will deplete body iron stores and increase the risk of anaemia, e.g. ulcers, intestinal parasites, intestinal cancer.

Average intakes of iron in the UK are around 14 mg/day in men and 12.3 mg/day in women (see Table 13.2), i.e. around 14 times average daily losses in men and, even in those women with particularly high menstrual losses, still around five times estimated losses. The iron RNI (RDA) is 8.7 (10) mg/day for men and 14.8 (15) mg/day for pre-menopausal women. According to COMA (1991), even this RNI may be insufficient for those women with particularly high menstrual losses. The very large differences in the estimated requirements for and losses of iron occur because only a small and variable proportion of dietary iron is absorbed. The proportion of dietary iron that is absorbed varies with several factors listed below.

- The iron status of the individual: inorganic iron absorption is two to three times higher in subjects who are iron deficient compared to those with good iron stores.
- The form in which the dietary iron is present: haem iron from meat and fish is much better absorbed than inorganic iron from vegetable foods.
- The presence of promoters of inorganic iron absorption: these include vitamin C, alcohol and meat protein.
- The presence of inhibitors of iron absorption in the diet: these include phytate (e.g. from unleavened bread), tannin (e.g. from tea) and perhaps even large amounts of dietary fibre.

Haem iron from meat and fish is far better absorbed (10–30%) than the iron in vegetable foods (2–10%). The absorption of haem iron is also unaffected by the presence of inhibitors and promoters.

Vitamin C, stomach acid and meat and fish all promote iron absorption. Vitamin C is a major factor in promoting the absorption of iron from vegetable foods. Fibre, phytate (from unleavened bread) and tannin in tea all tend to hinder iron absorption. Alcohol increases gastric acid secretion and facilitates iron absorption.

COMA (1991) assumed an average efficiency of absorption of iron of about 15% when making their calculations of DRV.

REGULATION OF IRON BALANCE

The principal physiological mechanism for regulating body iron balance is by controlling the efficiency of absorption. When iron stores are low, iron is absorbed more efficiently than when body iron stores are full. The efficiency of iron absorption may more than double when iron stores are depleted. This regulatory system is usually sufficient to prevent iron toxicity from dietary sources and even when moderate doses of therapeutic iron are consumed chronically by people who do not need them. Chronic iron overload can lead to cirrhosis of the liver. Single very high doses of iron can cause diarrhoea, vomiting, gastrointestinal bleeding, circulatory collapse and liver necrosis. Iron poisoning resulting from children consuming pharmaceutical preparations of iron is a common cause of accidental poisoning in the UK. The consumption of alcoholic drinks containing large amounts of iron can precipitate iron overload and contribute to liver cirrhosis. Some wines and Normandy cider contain large amounts of iron and cirrhosis due to iron overload has been common amongst Bantu people of South Africa who consumed large amounts of beer brewed in iron pots. White *et al.* (1993) found a positive association between alcohol consumption and level of iron stores in both men and women. As there is no physiological mechanism for excreting excess iron, iron overload and toxicity are major problems for those with certain conditions that require repeated blood transfusions, e.g. the inherited blood disorder thalassaemia. Drugs that bind or chelate iron and facilitate its excretion can be used to treat or prevent iron overload. Hereditary haemochromatosis is an inherited condition in which there is chronic over-absorption of iron, which can lead to liver and other tissue damage unless it is treated by regular therapeutic bleeding.

DETERMINATION OF IRON STATUS

Iron deficiency will eventually impair the ability to synthesize haemoglobin. Iron status has traditionally been diagnosed by the measurement of blood haemoglobin concentration. Iron-deficiency anaemia is characterized by reduced blood haemoglobin concentration and reduced red blood cell size – a **microcytic anaemia** (small cell) as compared to the **macrocytic anaemia** (large cell) of vitamin B₁₂ or (folic acid) deficiency. A blood haemoglobin concentration of less than 12 g/100 mL of blood has traditionally been taken as 'the level below which anaemia is likely to be present' and the prevalence of anaemia has usually been established by the use of a cut-off value like this.

Other measures of iron status have been increasingly used in recent years. The plasma ferritin level indicates the state of body iron stores and these stores may become depleted without any fall in blood haemoglobin concentration. Plasma ferritin is a more sensitive measure of iron status than blood haemoglobin concentration. Plasma ferritin levels of less than 25 µg/L are said to be indicative of low body iron stores (White *et al.*, 1993) and values of 12 µg/L are taken by COMA (1991) to indicate frank depletion. In iron deficiency, the concentration of the iron transport protein transferrin in the plasma rises, but the amount of iron in plasma falls, and so a low ratio of iron to transferrin in plasma, the **transferrin saturation**, indicates low iron status.

The symptoms of anaemia result from the low haemoglobin content of blood leading to impaired ability to transport oxygen and they include general fatigue, breathlessness on exertion, pallor, headache and insomnia.

Iron deficiency

PREVALENCE OF IRON DEFICIENCY AND ANAEMIA

A major problem in determining the prevalence of iron deficiency is to decide upon the criteria that are to be used to define it. In a series of studies, Elwood and his colleagues found that women with haemoglobin concentrations that would indicate mild to moderate anaemia showed little or no evidence of clinical impairment, even cardiopulmonary

- An adult male body contains about 4 g of iron, with 70% in haemoglobin and myoglobin, small amounts in iron-containing enzymes and the rest stored as the iron-protein complex ferritin.
- A healthy male loses less than 1 mg of iron per day, but menstrual losses, pregnancy and lactation increase the average losses of younger women.
- Iron requirements greatly exceed iron losses because only a proportion of dietary iron (about 15%) is absorbed.
- The efficiency of iron absorption increases in iron deficiency and decreases when iron stores are full; the gut is the main site of regulation of body iron balance.
- Haem iron is absorbed more efficiently than inorganic iron from vegetable sources.
- Substances such as phytate and tannin inhibit inorganic iron absorption, whereas alcohol, gastric acid and vitamin C promote its absorption.
- There is no route for excreting body iron.
- Chronic high consumption of iron can lead to cirrhosis of the liver and other tissue damage due to iron overload.
- Single high doses of iron can cause acute poisoning, typically in children accidentally consuming iron tablets.
- Iron deficiency eventually impairs haemoglobin synthesis and the capacity of blood to transport oxygen – iron-deficiency anaemia.
- Iron-deficiency anaemia is a microcytic (small red cells) anaemia and leads to pallor, fatigue, breathlessness on exertion and headache.
- Iron status may be assessed by measuring blood haemoglobin concentration, plasma ferritin concentration or plasma transferrin saturation.
- Plasma ferritin is a more sensitive indicator of iron status than blood haemoglobin concentration because iron stores become depleted before anaemia occurs.

function under exercise appeared to be little affected by this degree of apparent anaemia (reviewed by Lock, 1977). More recently, it has been suggested by several groups that functional impairment due to iron deficiency (as measured by plasma ferritin) may occur in the absence of anaemia (COMA, 1991). Reduced work capacity and changes in certain brain functions that adversely affect attention, memory and learning in children are some of the suggested adverse consequences of iron deficiency.

Iron deficiency is considered to be the most common nutrient deficiency in the world. WHO estimates put the worldwide incidence of iron deficiency at around 40%, with perhaps a third of these suffering from iron-deficiency anaemia. There may be something like 700 million people with iron-deficiency anaemia in the world. The incidence is higher in women than in men because of the extra

demands of menstruation, pregnancy and lactation upon body iron stores. In the developed countries, the condition is considered to be uncommon in young men. In the UK, it was reported that, in 1991, only 4% of women had haemoglobin levels below 11 g/100 mL, but 34% had serum ferritin levels below 25 µg/L and 16% less than 13 µg/L (the level taken by COMA, 1991, to indicate depletion of body stores). The serum ferritin levels of pre-menopausal women were lower than those of older women, with almost half of those under 45 years having serum ferritin levels below 25 µg/L (White *et al.*, 1993). Several studies have found high levels of anaemia in young children in the UK, especially in children of Asian origin, in whom rates of around 25% have been reported (COMA, 1991). Adolescents, pregnant women and the elderly are other groups regarded as being at increased risk of anaemia.

PREVENTING IRON DEFICIENCY

In people consuming the same type of diet, iron intake is likely to be directly related to total energy intake. On a good mixed European or American diet, iron intake will be around 5–7 mg/1000 kcal (4.2 MJ). This means that if the diet provides enough iron for women to meet their dietary standard, men consuming the same type of diet are likely to take in around double theirs. Young vegetarian women who restrict their food intake to control their weight are clearly at increased risk of iron deficiency. Any dietary supplementation programme that seeks to eliminate iron-deficiency anaemia in those women with particularly high needs will increase still further the surplus intake of men.

Some risk factors for iron deficiency are listed below.

- A diet that is low in absorbable iron, i.e. a diet that is low in haem iron: has low amounts of vitamin C or relatively high amounts of tannins or phytate that inhibit iron absorption.
- High menstrual blood losses, repeated pregnancies or prolonged lactation.
- Infestation of the gut with parasites that cause blood loss: hookworm infestation is a major precipitating factor in many tropical countries.
- Lack of stomach acid (achlorhydria) or gastrectomy.
- Any other condition that leads to chronic blood loss.

There are essentially three strategies that can be used to try to prevent and treat iron deficiency:

1. Provision of iron supplements to those with or at high risk of deficiency.
2. Fortification of a core food(s) with iron.
3. Education aimed at general improvement in the dietary iron content or its availability.

Of course, where chronic blood loss is identified as a precipitating factor, this problem should also be appropriately dealt with.

In Sweden, the prevalence of iron deficiency was reduced by three-quarters over a 10 year period from the mid 1960s (Anon, 1980). Increased fortification of flour with iron, the widespread use of prophylactic iron preparations and increased intakes of vitamin C were considered to be some of the factors that contributed to this decline in anaemia. In 1974, sales of pharmaceutical iron preparations in Sweden

amounted to almost 6 mg per head per day. The indiscriminate use of pharmaceutical preparations leads to the risk of minor side-effects, accidental poisoning and chronic iron overload in susceptible people. In the UK, all flour other than wholemeal must be fortified with iron and there is also iron fortification in the USA. The iron added to British and American flour is in a form that is very poorly absorbed and so probably makes little contribution to preventing iron deficiency. Any fortification programme designed to eliminate anaemia in women with high iron needs inevitably results in men and even many women receiving iron intakes greatly in excess of their needs. This excess iron intake might cause harm, particularly to subjects inherently sensitive to iron overload.

Another factor thought to have contributed to the decline in iron deficiency anaemia in Sweden was increased vitamin C intake; high intakes of vitamin C may account for the relatively low levels of anaemia found in vegetarians.

If pre-menopausal women and other groups at increased risk of anaemia are advised to eat a diet containing more available iron, then such a diet would probably contain more meat (and/or fish). This could lead to an increased intake of saturated fat and might not be entirely consistent with other nutrition education guidelines. It might also be unacceptable or impractical for many women. Nevertheless, lean meat and fish are important sources of iron and other nutrients and moderate amounts should be promoted as important components of a healthy diet, especially for those at high risk of anaemia. Many young women mistakenly view red meat as particularly fattening and nutrition education should aim to correct this misconception. Current advice to increase consumption of fruits and vegetables would increase vitamin C intake and thus increase iron absorption. Several fruits and vegetables are also good sources of available iron, e.g. tomatoes, cabbage and broccoli. Large increases in consumption of cereal fibre might tend to reduce iron availability. In developed countries, the indiscriminate use of iron supplements is probably not justified, although they are clearly appropriate for those with poor iron status. The case for higher levels of intervention will clearly be much stronger in developing countries, where the scale and severity of iron deficiency are much greater than in developed countries.

- Some past studies have reported little impairment of physiological functioning in women with mild anaemia, but more recent studies suggest that symptoms of iron deficiency may occur in the absence of anaemia.
- Iron deficiency is the most common micronutrient deficiency in the world and 700 million people worldwide may suffer from frank anaemia.
- Even in the UK, a third of pre-menopausal women may have relatively low iron stores and half of these have frank depletion of their iron stores.
- Children, adolescents, pregnant women and the elderly may all be regarded as at increased risk of anaemia, although it is generally uncommon in young men.
- The risk of anaemia is increased by chronic blood loss, lack of available iron in the diet, repeated pregnancies, prolonged lactation and lack of stomach acid.
- Strategies for tackling iron deficiency include the increased use of iron supplements, food fortification and nutrition education aimed at improving iron availability in the diet.
- On a normal, mixed diet, iron intake will be roughly proportional to energy intake.
- Any supplementation programme aimed at increasing the iron intakes of high-risk women will lead to some men consuming large excesses of iron.
- Iron from fortified bread (UK and USA) is often poorly absorbed.
- The increased use of prophylactic iron preparations may result in some women taking supplements that they do not need.
- Excess iron from fortified food or supplements may cause harm, particularly to people who are inherently sensitive to iron overload.
- Increased consumption of meat is the most obvious way of improving iron availability.
- Modest amounts of lean meat are consistent with a healthy diet, and nutrition education should counteract the view of many girls and young women that red meat is very fattening.
- Increased consumption of fruits and vegetables would increase vitamin C intake and thus increase inorganic iron uptake; several fruits and vegetables are also good sources of available iron.

CALCIUM, DIET AND OSTEOPOROSIS

Halliday and Ashwell (1991b) have written a general review of calcium nutrition, and Prentice (1997) has reviewed osteoporosis and the role of diet in its aetiology. The National Osteoporosis Society web-site (www.nos.org.uk) is also recommended a good source of information about osteoporosis; this site also lists a range of inexpensive publications produced by the society that can be ordered online. Most of the uncited primary sources used for this section are listed in these reviews.

Distribution and functions of body calcium

A typical adult human body contains more than 1 kg of calcium and 99% of this calcium is in the skeleton.

The skeletal calcium is present as the substance **hydroxyapatite**, which gives bone its mechanical strength. The small amount of body calcium that is outside the skeleton has a multitude of other vital functions, such as:

- it has a major role in the release of hormones and neurotransmitters
- it is an important intracellular regulator
- it is an essential cofactor for some enzymes and is necessary for blood clotting
- it is involved in nerve function and in muscle and heart contraction.

Any variation in the calcium concentration of body fluids can thus have diverse effects on many systems of the body, and the calcium concentration in extracellular fluids is under very fine hormonal control. The skeletal calcium not only strengthens bone, but also serves as a reservoir of calcium. The skeleton

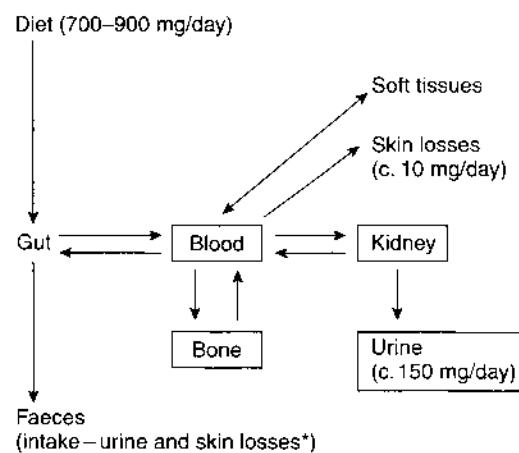
can donate or soak up excess calcium and thus enable the concentration in extracellular fluid to be regulated within narrow limits. A fall in blood calcium concentration, as seen after removal of the parathyroid glands or in extreme vitamin D deficiency, leads to muscle weakness and eventually to tetany and death due to asphyxiation when the muscles of the larynx go into spasm. A rise in plasma calcium (hypercalcaemia) is seen in vitamin D poisoning and in hyperparathyroidism. This hypercalcaemia can lead to gastrointestinal symptoms, neurological symptoms, the formation of kidney stones and, in some cases, to death from renal failure.

Hormonal regulation of calcium homeostasis

The hormone **1,25-dihydroxycholecalciferol** (1,25-DHCC or calcitriol) is produced in the kidney from vitamin D and is essential for the active absorption of dietary calcium in the intestine. Calcitriol promotes the synthesis of key proteins involved in the active transport of calcium in both the gut and kidney. Some passive absorption of calcium occurs independently of vitamin D and this passive absorption increases with increasing dietary intake. Parathyroid hormone and calcitonin (from the thyroid gland) regulate the deposition and release of calcium from bone and also the rate of excretion of calcium in the urine. Parathyroid hormone also indirectly regulates the intestinal absorption of calcium by activating the renal enzyme that produces calcitriol. Typical daily fluxes of calcium are illustrated in Figure 13.2 and the hormonal influences upon these calcium fluxes are summarized in Figure 13.3.

Requirement and availability of calcium

The adult RNI for calcium in the UK is 700 mg/day for both males and females, slightly less than the American RDA of 800 mg/day. The extra increments allowed for growth and pregnancy are also generally higher in the USA than in the UK. Almost two-thirds of non-pregnant women in the USA fail to meet the current RDA. Adult intakes of UK men and women are shown in Table 13.2 and these suggest a very substantial minority of British women also take in much



* This assumes calcium balance. During skeletal growth, pregnancy and lactation, loss via the faeces will be less than this value (values greater than this indicate net skeletal loss).

Figure 13.2 Typical daily calcium fluxes in UK adults.

less than the RNI. In the UK, COMA arrived at the RNI by calculating the amount needed to replace typical urinary and skin losses (see Figure 13.2), assuming that around 30% of the dietary intake is absorbed. In Western countries, 50–70% of dietary calcium is from dairy produce and the calcium in dairy produce tends to be well absorbed. This predominance of dairy produce as the major source of dietary calcium makes calcium intake vulnerable to changes in the image of dairy produce induced by health education messages. In the UK, daily calcium

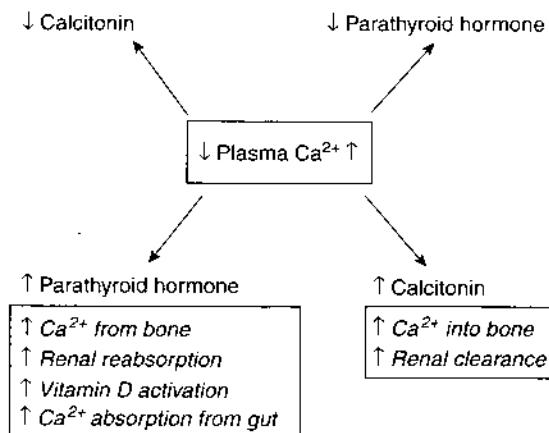


Figure 13.3 Some hormonal influences upon calcium homeostasis.

intake fell by 20% over the period 1975–1995 as a result of reduced total food intake and milk consumption; this trend was most pronounced in young women and teenaged girls. Note that lower-fat milks contain just as much calcium as whole milk, and so reduced intake of dairy fat does not inevitably lead to reduced calcium intake. Table 13.1 shows the calcium content of some common foods.

The efficiency of absorption of calcium varies according to the phase of the life cycle, the food source of calcium and the total intake (see examples below).

- The efficiency of calcium absorption increases markedly in pregnancy, but is reduced in the elderly.
- Calcium from vegetable sources is generally less well absorbed than that from milk and cheese. The efficiency of absorption from breast-milk is higher than from infant formula.

- At low calcium intakes the active, vitamin-D-dependent absorption process predominates, whereas at high intakes the less efficient passive process predominates, so that the proportion of the oral load that is absorbed tends to decline with increasing intake.

Many populations whose adults do not consume large amounts of dairy produce have average calcium intakes that are less than half the current UK RNI and thus below even the current LRNI. Despite this, overt symptoms of acute dietary deficiency are not seen. For example, Prentice *et al.* (1993) reported calcium intakes in pregnant and lactating women in rural Gambia in 1979 of only around 400 mg/day. A follow-up report by Jarjou *et al.* (1994) suggested that this had fallen still further, to less than a quarter of the current UK RNI for lactating women. Despite this, there is no overt evidence of acute calcium deficiency and

- An adult body contains more than 1 kg of calcium and 99% of this is in the bone mineral.
- Calcium has numerous functions in addition to being a component of bone mineral.
- Large fluctuations in the calcium concentration of extracellular fluid have adverse effects on several systems and are subject to fine hormonal control.
- Bone mineral acts as a large reservoir of calcium.
- Calcitriol is the hormone produced in the kidney from vitamin D.
- Calcitriol promotes the synthesis of key proteins necessary for the active absorption of calcium.
- Parathyroid hormone and calcitonin control the flow of calcium into and out of bone and the rate of calcium excretion in the urine.
- Parathyroid hormone indirectly controls the intestinal absorption of calcium by regulating the renal enzyme that produces calcitriol.
- The UK RNI for calcium is 700 mg/day for men and women and substantial numbers of UK women do not take in this much.
- This RNI is calculated as the amount needed to replace daily losses assuming that 30% of the dietary intake is absorbed.
- More than half of the calcium in Western diets is from milk and dairy products.
- UK intakes of calcium have fallen in recent decades as total food intake and milk consumption have dropped.
- Low-fat milk has just as much calcium as full-fat milk.
- Calcium from dairy products is much better absorbed than that from vegetable sources.
- The efficiency of calcium absorption increases in pregnancy but declines in the elderly.
- At low calcium intakes the active vitamin-D-dependent process predominates, but at high intakes this becomes saturated, and passive absorption becomes increasingly prominent.
- Calcium intakes of many populations are much lower than in the UK and well below the current LRNI; however, these populations show no overt signs of calcium deficiency and often have very low rates of bone fractures.

rates of bone fractures in the Gambia are low. There is evidence that individuals who consume chronically low amounts of calcium can adapt to these low intakes by increasing the efficiency of absorption and by reducing their urinary losses of calcium.

Calcium and bone health

THE NATURE OF BONE

The skeletal calcium is present as **hydroxyapatite**, a crystalline complex of calcium, phosphate and hydroxide ions. These crystals are laid down upon an organic matrix that is largely composed of the protein collagen. Hydroxyapatite makes up around half the weight of the bone and is responsible for conferring its mechanical strength. There are two types of mature bone, 80% is cortical or compact bone and 20% is trabecular bone. Cortical bone is on the outside and trabecular bone, which has a honeycomb structure with marrow filling the spaces, is found in the centre of bones such as the pelvis and on vertebrae and on the ends of the long bones.

EFFECTS OF AGE AND SEX UPON BONE DENSITY AND FRACTURE RISK

Bones are not fixed and inert, but living, dynamic structures. The processes of bone deposition and bone breakdown continue throughout life. Bones are continually being remodelled by cells that synthesize new bone (osteoblasts) and cells that reabsorb bone and cartilage (osteoclasts). Thus, if a bone's weight remains constant, this is the result of a dynamic equilibrium, a balance between synthesis and breakdown.

Bone mass increases throughout childhood as the skeleton grows and the rate of deposition exceeds the rate of breakdown. A newborn baby has 25–30 g of calcium in its skeleton and by the time the **peak bone mass (PBM)** is reached, the skeletal calcium content has risen to well over 1 kg. People reach their peak bone mass in their thirties and for a few years bone mass remains fairly constant because of a dynamic equilibrium between synthesis and breakdown. During middle and old age, the rate of breakdown exceeds the rate of bone formation and there is a steady decline in bone mass in older adults (Figure 13.4). This decrease in bone mass involves the loss of both organic matrix and mineral matter. As people age, their bones become thinner, less strong and increasingly prone to fracture when subjected to trauma. Smith (1987) suggested that thinning of the bones may take them below a 'fracture threshold' at which they become liable to fracture when subjected to relatively minor trauma such as a fall from a standing position. This thinning of the bones and consequent increase in fracture risk is termed **osteoporosis**.

In healthy men and pre-menopausal women, the age-related decline in bone mass is relatively slow – around 0.3–0.4% of bone lost per year. In women, the decade immediately after the menopause is associated with a marked acceleration in the rate of bone loss. This post-menopausal acceleration in the rate of bone loss is due to the decline in oestrogen production and means that women usually reach the 'fracture threshold' bone density long before men (Figure 13.4). Osteoporosis-related fractures are much more common in women than in men, e.g. Spector *et al.* (1990) recorded a 4:1 female to male ratio in the age

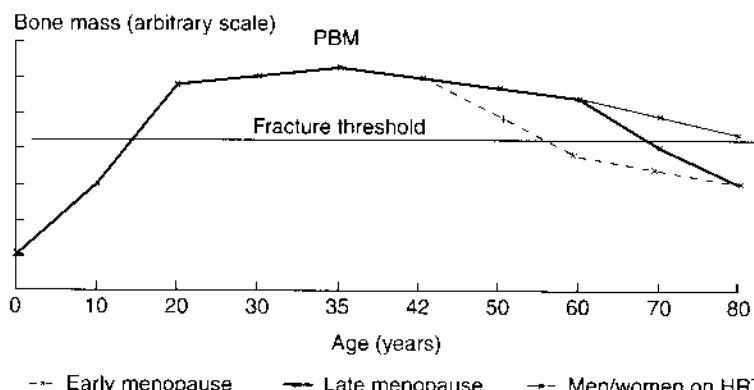


Figure 13.4 Changes in bone mass throughout life. PBM, peak bone mass; HRT hormone replacement therapy. From Webb and Copeman (1996).

specific incidence of hip fracture in the UK. The high female to male ratio in fracture rates is only seen in countries in which rates of osteoporosis are high (Cummings *et al.*, 1985).

Bone density can be measured in living subjects by techniques such as dual-energy X-ray absorptiometry. These bone density measurements are regarded as reliable indicators of bone strength. Although a subthreshold density is a necessary permissive factor for osteoporotic fracture, bone mineral density measurements are imperfect predictors of individual susceptibility to fracture, especially for fractures of the hip (Cummings *et al.*, 1985; Smith, 1987). Leichter *et al.* (1982) measured bone density and the shear stress required to fracture the femoral neck using isolated human bones obtained from cadavers. As expected, they found a high statistical correlation between shear stress at fracture and both bone density and bone mineral content. Breaking stress, bone density and bone mineral content all declined with the age of the subject. However, they found that breaking stress declined with age much faster than either bone density or bone mineral content. They concluded that changes in bone strength

are apparently influenced by factors other than just bone density and mineral content, e.g. changes in the micro-architecture of the bone.

The imperfect association between fracture risk and measured bone density needs to be borne in mind when evaluating reports of weak associations between environmental/lifestyle variables and bone density. Likewise, evidence that interventions can have acute effects on bone density need to be treated with caution. Will any short-term increase be maintained in the longer term? Will small measured increases in bone density significantly reduce fracture risk?

INCIDENCE OF OSTEOPOROSIS

The thinning of the bones in osteoporosis is responsible for increased rates of bone fracture in older people, especially fractures of the vertebrae, wrist and the neck of the femur (hip fracture). Rates of osteoporotic fractures increased very substantially in many industrialized countries during the second half of the twentieth century. Adult bone mass declines with age, so the increased prevalence of osteoporosis

- Bone consists of crystals of the calcium phosphate salt hydroxyapatite, deposited upon an organic matrix consisting largely of the protein collagen.
- Hydroxyapatite makes up half of the bone weight and increases its mechanical strength.
- Bones are living, dynamic structures that are continually being broken down and resynthesized.
- Cells called osteoblasts synthesize new bone and osteoclasts cause bone breakdown.
- During childhood, the rate of bone synthesis exceeds breakdown as the skeleton grows.
- In early adulthood, the peak bone mass is reached and for a few years the rates of synthesis and breakdown are matched.
- In middle and old age there is a net loss of bone and the bones become thinner.
- Thinning of bones in old age weakens them and can take them below a 'fracture threshold', where they break in response to relatively minor trauma – this is called osteoporosis.
- There is a marked acceleration in bone loss in women around the menopause that is due to decreased oestrogen production.
- Women usually reach their 'fracture threshold' before men and so elderly women are more prone to bone fractures than elderly men.
- Non-invasive bone density measurements are indicators of bone strength.
- Bone strength decreases more rapidly than bone density, probably due to age-related changes in bone micro-architecture.
- Bone density is thus an imperfect measure of fracture risk and this should be borne in mind when assessing the results of studies in which bone density has been used as the measured outcome.

is partly accounted for by the increased numbers of people surviving into old age. Wickham *et al.* (1989), in a prospective study of 1400 elderly British people, found that the incidence of hip fracture was four times higher in women over 85 years than in women in the 65–74-year age band. Figure 13.5 shows the age and sex distribution of hip fractures in one region of England. The proportion of fractures in each age group rises exponentially with increasing age and the higher prevalence in females is also illustrated. The number of people aged over 85 years doubled in the UK in last two decades of the twentieth century and this group now represents 2% of the UK population. There has, however, also been a real increase in the number of fractures, as demonstrated by the large increases in the age-specific incidence of fractures seen in the UK (Spector *et al.*, 1990).

Up to 3 million people in the UK may be affected by osteoporosis, which results in around 70 000 hip fractures, 50 000 wrist fractures and more than 40 000 vertebral fractures each year. As many as one in three women and one in 12 men in the UK will develop osteoporosis during their lifetime (see www.nos.org.uk). Fractures of the hip are currently considered to be the major public health priority. A significant proportion of those affected die within a few months of the initial fracture and many more are permanently disabled or never regain full mobility after the fracture. Up to 20 000 deaths each year in the UK may be the consequence of hip fracture or the associated surgery. Prentice (1997) suggested that a third of all orthopaedic beds in UK hospitals were occupied by patients with osteoporosis and this accounted for an annual £750 million cost to the National Health Service. In the USA, there are over 200 000 hip fractures each year that are attributable to osteoporosis. Cummings *et al.* (1985) and Johnell

et al. (1992) give comparative rates of osteoporotic fracture in different countries.

TYPES OF OSTEOPOROSIS

Osteoporosis has been divided up into the two main types listed below.

1. *Type I or menopausal osteoporosis:* this occurs in the 50–75-year age group and results mainly in wrist and vertebral fractures. Loss of oestrogen is considered the major precipitating factor and it is six times more common in women than in men.
2. *Type II or senile osteoporosis:* this manifests mainly as fractures to the hip and vertebrae in those aged over 70 years. The incidence in women is only twice that in men. It is suggested that vitamin D deficiency and consequent increased parathyroid hormone secretion are important causative factors.

GENERAL AND LIFESTYLE RISK FACTORS FOR OSTEOPOROSIS

Table 13.3 lists some of the known or possible risk factors for osteoporosis. It is clear from Figure 13.4 that one should expect the risk of osteoporosis and fractures to increase with age and elderly women to be at increased risk compared to elderly men. Projected increases in the elderly populations of many countries (see Chapter 14) make it likely that the numbers of affected people will rise over the coming decades unless prevention and treatment become more effective. The over-85-year age group has been increasing, and will continue to increase, more rapidly than the total elderly population. Women outnumber men by three to one in the over-85-year age group.

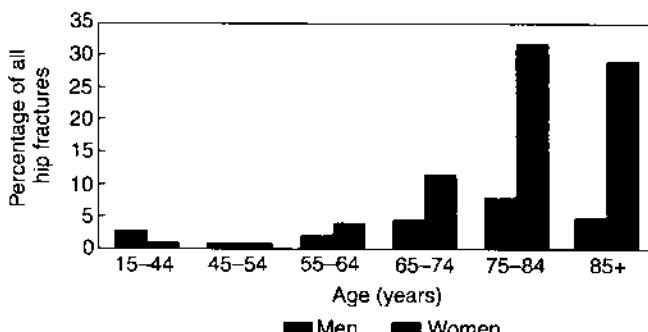


Figure 13.5 Age distribution of hip fractures in the Trent region of England 1989/1990. (Data source Kanis, 1993). From Webb and Copeman (1996).

Table 13.3 Risk factors for osteoporosis

Old age: risk of osteoporosis increases with age and accelerates in women when they reach the menopause
Being female: rates are higher in women than in men in Western Europe and North America
Lack of sex hormones: factors that reduce sex hormone secretion, e.g. early menopause or ovarian removal, increase risk; hormone replacement therapy reduces risk
Being white: there are distinct racial differences in bone density and considerable genetic variations within races
Never having borne a child
A sedentary lifestyle
Smoking
High alcohol consumption
Being small or underweight
Inadequate calcium intake in childhood?
High consumption of carbonated, cola drinks?
Being an omnivore: vegetarians have lower rates of osteoporosis

The accelerated bone loss associated with menopausal oestrogen loss suggests that anything that diminishes sex hormone output might be likely to increase bone loss and fracture risk. Many conditions in which there is reduced sex hormone output are indeed associated with reduced bone density, for example:

- early menopause or surgical removal of the ovaries
- cessation of menstruation due to starvation or anorexia nervosa
- amenorrhoea due to extremely high activity and low body fat content in women athletes
- hypogonadism in men.

Hormone replacement therapy (HRT) in post-menopausal women prevents or reduces the post-menopausal acceleration in the rate of bone loss and reduces the risk of osteoporotic fracture. This treatment is widely accepted as being effective in the prevention of osteoporosis. The controversy about HRT is about whether this and other benefits of the treatment are sufficient to outweigh any possible increased risks such as that of uterine cancer or perhaps even of breast cancer.

There are clear racial differences in bone density; people of European and Asian origin have bones that are less dense than those of people of African origin. In the USA, white women are much more likely to

suffer from osteoporotic fractures than black women. Even within racial groups there is almost certainly a very large inherited variability in bone density and susceptibility to fracture. According to Prentice (1997), around 80% of the variation in bone mineral status can be attributed to genetic factors.

The increasingly sedentary lifestyle of those living in the industrialized countries is widely believed to be a major factor in the increased age-specific incidence of osteoporotic fracture. Load-bearing activity is known to increase bone density, and immobilization or extreme inactivity (e.g. confinement to bed or the weightlessness experienced by astronauts) is known to result in loss of bone mass. When bones are repeatedly mechanically stressed, they thicken and become stronger. Thus, tennis players have much thicker bones in their racket arms than in their other arm and runners have higher than average bone density in their vertebrae. Experimental studies have shown that jumping up and down or the use of a small trampoline increases bone density in pre-menopausal women. Increased activity in the early years increases the peak bone mass and maintaining activity in later years slows down the rate of bone loss.

Children who spent more than 2 hours daily in weight-bearing activity were reported to have greater bone mineral density than those who spent less than 1 hour daily in such activity (see Cooper and Eastell, 1993). Pocock *et al.* (1986) found a significant correlation between measured fitness level and bone mineral density in normal women. In the post-menopausal women in this sample, fitness was the only measured variable that correlated significantly with bone mineral density. In 1400 elderly British people sampled by Wickham *et al.* (1989), the risk of hip fracture was reduced in those people with higher levels of outdoor physical activity.

Being underweight is associated with a higher risk of osteoporotic fracture and in the elderly this is often as a result of low lean body mass caused by inactivity. More active and fitter elderly people may also be less prone to osteoporotic fracture because they are better co-ordinated and stronger and therefore less prone to falling or less liable to injury if they do fall.

Many studies have suggested that cigarette smoking is associated with reduced bone density and with increased fracture risk. In their sample of elderly British people, Wickham *et al.* (1989) reported a

strong association between cigarette smoking and risk of hip fracture. Smoking is associated with:

- lower body weight
- reduced physical activity
- early menopause
- reduced blood oestrogen concentration (even in women receiving HRT).

These are all known associates of high fracture risk.

High alcohol consumption has also been widely reported as a risk factor for osteoporosis. Heavy drinking is associated with extensive bone loss, even in relatively young adults (Smith, 1987). High alcohol consumption might also be expected to increase the risk of falling.

DIETARY RISK FACTORS FOR OSTEOPOROSIS

A vegetarian lifestyle is generally associated with higher bone density and a reduced risk of osteoporosis (Marsh *et al.*, 1988; see Chapter 16). To what extent, if any, this is a result of meat avoidance *per se* is difficult to establish. There are numerous differences between Caucasian vegetarians and the general white population that would be confounding variables in this association between meat avoidance and osteoporosis. See Chapter 16 for a discussion of the health aspects of vegetarianism.

The extent to which calcium nutrition affects the risk of osteoporosis has been widely investigated and debated, but without yet reaching universally accepted conclusions. Cross-cultural comparisons indicate that it is unlikely that low calcium intake is of itself a major causative factor for osteoporosis. Osteoporosis is uncommon in some populations with low calcium intakes, but its prevalence is high in many populations with relatively high calcium intakes, e.g. in the USA, UK and northern Europe. For example, rural Bantu women suffer only one-tenth the fracture incidence of Western women despite taking in only half as much calcium (Rolls, 1992), and the similar findings of Prentice *et al.* (1993) in rural Gambia were discussed earlier. Gatehouse and Webb (unpublished observations) found a significant *positive* association between age-standardized hip fracture rates in women and *per capita* consumption of fresh milk products in eight western European countries; the consumption of fresh milk products is a good predictor of calcium intakes in such countries. Hegsted (1986) also found a *positive* correlation between cal-

cium intake and the incidence of hip fracture in a genetically and culturally more diverse sample of nine countries in Europe, the Far East, Australasia and North America. International comparisons may be of limited usefulness in proving cause-and-effect relationships, especially when used rather anecdotally as above, but they are more persuasive when used to argue against such a relationship. If high calcium intake is associated internationally with high rates of osteoporosis, this makes it improbable that calcium deficiency is an important causal factor. It is still possible, of course, that high calcium intakes or calcium supplements might have some role in offsetting an increase in osteoporosis risk even though it is primarily due to other causes.

PREVENTION AND TREATMENT OF OSTEOPOROSIS

There are several strategic approaches that may be employed to try to reduce the number of fractures attributed to osteoporosis.

- Promote changes in the diet and lifestyle of children and young adults that increase the PBM. This increases the amount of bone that can be lost in later life before the fracture threshold is reached.
- Try to encourage dietary and lifestyle changes that slow down the rate of bone loss in older people, particularly in post-menopausal women. This means that it will take longer to reach the fracture threshold from any given PBM.
- Try to reverse the process of osteoporosis in older people, by promoting dietary, behavioural and pharmacological measures that will increase bone density in those with low bone density or frank osteoporosis.
- Try to reduce the risk of bones being subjected to the traumas that cause them to fracture. Fractures of the hip usually occur after a fall. A high risk of falling contributes towards the increased risk of all fractures in the elderly. In someone with advanced osteoporosis, measures that reduce the tendency to fall or protect the faller from injury may offer the greatest potential for reducing the risk of fractures. Measures designed to improve muscle strength in the elderly may reduce the risk of falling. In contrast to the enormous amount of work done looking at factors that affect bone density, there has been relatively little work done on the causes of falls in the elderly.

It is often argued that a high intake of calcium in childhood together with ensuring good vitamin D status are likely to increase PBM and thus reduce the later risk of osteoporosis. Calcium intakes substantially below the dietary standard during childhood may limit PBM, and poor vitamin D status certainly has adverse effects upon skeletal development. However, in observational studies, there is little or no correlation between calcium intake and bone density in children. According to Prentice (1997), calcium supplementation studies in children suggest that supplements can have a small effect upon bone density. This effect seems to be greatest in the first few months of supplementation and disappears quickly once the supplements are stopped. In contrast, there is clear evidence that active children have denser bones than those who do little load-bearing exercise. Increasing the activity levels of children can be promoted on many grounds and would seem to be the most effective long-term strategy for improving bone health. Good intakes of milk and cheese in growing children will, in most cases, be necessary for them to achieve calcium intakes that reach dietary standard values. Vitamin D is produced by the action of sunlight on the skin and so outdoor activity should have a double effect upon bone health.

There is little evidence that normal variation in dietary calcium intake has any major effect on bone density in adults (Cummings *et al.*, 1985; Smith, 1987). Wickham *et al.* (1989) found no relationship between dietary calcium intake and risk of osteoporotic fracture in a sample of elderly British adults. Supplemental calcium in women during the menopause is generally ineffective at preventing oestrogen-related bone loss, but supplements in the elderly may reduce the rate of bone loss at the hip (Prentice, 1997). There is also compelling evidence that elderly people in Britain often have poor vitamin D status and this is probably a major contributory factor to type II osteoporosis and to the high rates of hip fractures in the elderly. According to COMA (1991), average blood levels of 25-hydroxycholecalciferol in Britons over the age of 75 years were unsatisfactory and indicated widespread insufficiency for this vitamin. They attributed this to inadequate exposure to sunlight during the summer months. Elderly housebound Britons are unlikely to obtain adequate amounts of vitamin D from their diet. Up to 40% of elderly British people living in residential homes have biochemical indications of poor vitamin status.

Chapuy *et al.* (1994) showed that calcium and vitamin D supplements given to institutionalized, elderly women in France substantially reduced the risk of hip and other fractures over a 3-year period. Many of their sample had raised parathyroid levels and low serum 25-hydroxycholecalciferol concentrations, which are indicative of poor vitamin D status. The treatment lowered serum parathyroid hormone levels, whereas levels rose in the placebo group. High parathyroid hormone levels were associated with lowered bone density measurements.

Wickham *et al.* (1989) reported a lower fracture risk in elderly people who engaged in outdoor physical activity; higher exposure to sunlight, and thus better vitamin D status, could be a contributory factor to this risk reduction. An expert report in the UK (CWT, 1995) recommended that architects designing residential accommodation for older people should take particular account of the need of residents for regular exposure to sunlight. They suggested the provision of sheltered alcoves on the south side of buildings and well-paved pathways with hand rails and no steps. There is a clear case for vitamin D supplements in elderly, housebound people. The case for inactivity as a cause of osteoporosis was made earlier and so the case for encouraging increased amounts of load-bearing exercise in all age groups is overwhelming. There are many other reasons for encouraging increased activity and physical fitness and these are summarized in Chapter 16.

There is compelling evidence that the administration of oestrogen (HRT) to menopausal women prevents or reduces the post-menopausal acceleration in the rate of bone loss and reduces fracture risk. Detailed discussion of this issue is beyond the scope of this book.

The treatments for osteoporosis include:

- administration of sex hormones or synthetic analogues – oestrogen replacement therapy in women and testosterone injections in men
- supplements of vitamin D and calcium
- use of calcitonin injections or the bisphosphonate group of drugs, both of which inhibit bone breakdown by osteoclasts.

CONCLUSIONS

An active lifestyle, absence of cigarette smoking, moderate alcohol consumption and a diet that meets the standards for essential nutrients are both

consistent with general health education guidelines and offer the expectation of reducing the risk of osteoporosis and bone fractures. There seems to be no convincing evidence that calcium supplements *per se* will prevent osteoporosis or be effective in its treatment. Good intakes of calcium in childhood and early adulthood may contribute to increased

PBM, but other lifestyle factors (e.g. activity level) are probably more significant and the case for calcium supplements seems unconvincing. Widespread vitamin D deficiency in the elderly is probably an important cause of increased fracture risk and so increased exposure to sunlight and/or supplements of vitamin D and calcium should be encouraged.

- Osteoporotic fracture rates rose during the twentieth century, not only because of the growth of the elderly population but also because the age-specific incidence increased.
- In the UK, 150 000–200 000 fractures each year are attributed to osteoporosis.
- Many of the 70 000 people who annually suffer hip fractures die within a few months of the initial fracture or never regain full mobility.
- A third of all orthopaedic beds in the UK are occupied by people with osteoporosis.
- Menopausal oestrogen loss is thought to be a major precipitating factor for type I osteoporosis and vitamin D deficiency for the type II that occurs principally in the very elderly.
- The projected growth of the very elderly population together with the 3:1 ratio of women to men in this age group indicate that fracture rates are likely to increase in the coming decades.
- Lack of sex hormones in men or women increases bone loss, and oestrogen replacement therapy slows bone loss.
- There are inherent racial and individual differences in bone density and perhaps 80% of the variation in bone mineral status can be attributed to genetic factors.
- Bones thicken when exposed to repeated stress and they become thinner when immobilized.
- Experimental and epidemiological studies suggest that load-bearing exercise:
 - increases bone deposition in children and increases peak bone mass
 - slows the rate of bone loss in older adults
 - increases bone density in older adults.
- Cigarette smokers and alcoholics have reduced bone density, but vegetarians usually have higher than average bone density.
- International comparisons suggest that osteoporosis is often most prevalent in those countries with the highest calcium intakes, which makes it unlikely that low calcium intake is a primary cause of osteoporosis.
- Calcium intakes seem to have a small positive effect upon bone density in children, but only as long as the supplements continue to be taken.
- Many elderly people have little exposure to sunlight and low vitamin D status.
- Combined supplements of vitamin D and calcium have been shown to reduce fracture risk in very elderly, housebound women.
- Osteoporosis would be reduced by:
 - increased activity in children and ensuring they have good vitamin D and calcium status
 - increased participation in load-bearing activity in all adults
 - ensuring the vitamin D status of elderly people by increasing access to sunlight or the provision of supplements
 - measures to reduce the risk of falling in the elderly or to decrease the risk of injury in elderly fallers
 - increased use of hormone replacement therapy by post-menopausal women
 - less tobacco use and alcohol abuse.

Detailed discussion of the pros and cons of HRT is beyond the scope of this book but, on the narrow question of whether it reduces the risk of osteoporosis, the evidence is compelling.

SALT AND HYPERTENSION

Overview

The monograph by MacGregor and de Wardener (1999) is recommended to interested readers as a starting point for entry into the specialist literature concerning the health consequences of high salt intake. It lists many of the landmark studies in this field and also contains much interesting background material about the past and present economic, political and cultural significance of salt.

THE PROBLEMS WITH SALT

High salt intake has long been suspected of being involved in the aetiology of essential hypertension. Note that essential hypertension is the most common type of high blood pressure for which the cause is unknown and the hypertension is not a secondary consequence of some other known pathology. Salt (sodium chloride) is made up of the two mineral elements sodium and chlorine. It is the sodium component that has been generally regarded as significant in causing hypertension, although this has been questioned. This discussion generally refers to salt rather than to sodium alone. Even in patients with hypertension, salt (sodium) balance is very effectively hormonally regulated; if salt intake is raised or lowered, urinary excretion rapidly rises or falls so as to restore the balance between intake and output. This homeostatic mechanism is so effective that salt excretion is generally regarded as the best way of measuring salt intake. Despite this homeostatic mechanism, it has been argued that prolonged excess salt intake can lead to excess salt retention and therefore increased fluid retention and hypertension. Evidence from dialysis patients and those with endocrine disorders confirms that salt retention does, indeed, lead to increased blood pressure.

MacGregor and de Wardener (1999) suggest that the kidneys of patients with essential hypertension have an impaired ability to excrete salt. When large amounts of salt are consumed by susceptible people,

their kidney defect means that there is a tendency to retain excess salt, which triggers several compensatory mechanisms that increase salt excretion and restore homeostasis. A rise in blood pressure and consequent increase in renal filtration of salt is one of these compensatory mechanisms. As evidence for this theory, MacGregor and de Wardener cite cross-transplantation studies in rats. When the kidneys of genetically hypertensive and normal rats were surgically swapped, the blood pressure rose in the normal rats with hypertensive strain kidneys but remained normal in the hypertensive strain rats with normal kidneys. They claim that limited observations on patients receiving therapeutic kidney transplants show a similar trend.

High salt intake has also been implicated in the aetiology of gastric cancer. One theory suggests that salt acts as a gastric irritant and that continued irritation due to high salt intake leads to atrophy of the acid-secreting glands in the stomach. Reduced acid secretion in the stomach then allows the proliferation of bacteria, which produce carcinogenic substances (nitrosamines) from some components of food. Again, there is considerable support for the latter part of the hypothesis: patients with pernicious anaemia who produce less gastric acid have markedly increased rates of gastric cancer.

MacGregor and de Wardener (1999) review evidence that high salt intake may be a factor in the aetiology of osteoporosis. In short-term studies, increases in salt intake and excretion also increase the renal excretion of calcium, although one would expect homeostatic mechanisms to compensate for this by increasing the intestinal absorption of calcium. One study has reported a positive correlation between rates of bone loss and urinary salt excretion in older women.

Affluent populations have generally been advised to substantially reduce their dietary salt intake. The World Health Organization (WHO) have suggested a move towards a population average salt intake of 5 g/day. In both Britain (COMA, 1991, 1994) and the USA (NRC, 1989b), it is recommended that salt intake should be limited to 6 g/day. The principal aim of such recommendations is to reduce the 15–20% incidence of essential hypertension in the adult populations of affluent countries and thus to reduce the morbidity and mortality from diseases precipitated by high blood pressure, namely strokes, myocardial infarction and renal failure. In a survey of

English adults, about 15% were classified as hypertensive, but only 76% were normotensive and not taking drugs that could affect their blood pressure. The average blood pressure of this sample and therefore the prevalence of hypertension, increased progressively with age, e.g. in the 16–24-year group 99% were classified as normotensive and untreated, whereas in those aged over 75 years this had dropped to around a third of the sample (White *et al.*, 1993).

There is a wide variety of antihypertensive drugs now available that are very effective in lowering blood pressure. These drugs also seem to be effective in reducing the renal and cerebrovascular complications of hypertension, although some of them, particularly the older drugs, are less effective in reducing the coronary conditions associated with high blood pressure and may even be associated with increased coronary risk. As with any drug, there may be acute side-effects, and possible long-term problems are of particular concern because antihypertensives may need to be taken for the rest of the patient's life. There have been suggestions that the use of some antihypertensives can be associated with increased overall mortality (e.g. MRFIT, 1982). Despite the availability of these powerful antihypertensive drugs, the idea that moderation of dietary salt intake might prevent hypertension or be a useful alternative to drugs in the treatment of hypertension is still an attractive proposition. Salt restriction may also be used in conjunction with antihypertensive drugs to increase their effectiveness or reduce the doses needed.

High intakes of potassium and calcium have also been suggested as having possible protective effects against the development of hypertension.

REQUIREMENT FOR SALT

Salt is an essential nutrient. Sodium is the principal cation in extracellular fluid; the standard 70 kg male has around 92 g of sodium in his body, equivalent to around 250 g of salt. Around half of total body sodium is in the extracellular fluid, 38% in bone and the remaining 12% in the intracellular fluid; potassium is the dominant cation in intracellular fluid. Some sodium is essential in the diet, but the habitual intake of most affluent populations greatly exceeds minimum requirements. Almost all ingested salt is absorbed in the intestine, and the homeostatic control of salt excretion by the endocrine system enables individuals, even those with hypertension, to remain in salt balance over a huge range of intakes. Some populations, such as the Kalahari Bushmen and New Guinea Highlanders, manage on habitual salt intakes of around 1 g/day, whereas other populations (e.g. in some regions of Japan and Korea) habitually take in more than 25 g/day. The average urinary excretion of salt in a sample of adult Yanomamo Indians in Brazil was found to be only 50 mg/day, which is less than 1% of that excreted by an average Briton (see MacGregor and de Wardener, 1999, for substantial detail on the comparative salt intakes of different populations).

- It is widely believed that high salt intake is an important cause of essential hypertension.
- High blood pressure increases the risk of strokes, myocardial infarctions and renal failure.
- One theory suggests that the kidneys of hypertensives have an impaired ability to excrete salt and that a rise in blood pressure is a compensatory mechanism that increases renal filtration and salt excretion.
- There is very effective renal regulation of daily salt balance, even in hypertensives.
- Salt has also been implicated in the aetiology of gastric cancer and, less convincingly, in the aetiology of osteoporosis.
- In the UK (and USA), a maximum salt intake of 6 g/day has been recommended, compared to the current average of around 9 g/day.
- Reduced salt intake is expected to reduce the prevalence of hypertension and thus ultimately the risk of the hypertension-linked diseases.
- Salt restriction may also be a useful adjunct to, or alternative to, antihypertensive drug therapy.
- 15–25% of all UK adults may be hypertensive and the proportion rises steeply with age.

NRC (1989a) estimated that minimum requirements for sodium chloride were only around 1.25 g/day. This minimum figure did not provide for large, prolonged losses from the skin through sweat. Individuals adapt to salt restriction during heavy exercise or in a hot environment by producing very dilute sweat. The UK RNIs for sodium and chloride amount to a salt intake of around 4 g/day; the panel felt that this value allowed for losses in sweat during exercise and high temperatures once adaptation had occurred, but that, in the short term, extra intake might be required under these circumstances.

AMOUNT AND SOURCES OF DIETARY SALT

It is generally accepted that **discretionary salt**, used in cooking and at table, makes the traditional methods of dietary analysis using food tables an unreliable method of assessing salt intake. The best method of estimating salt intake is by measurement of sodium excreted over a 24-hour period. This method relies upon the assumption that, whatever the intake, individuals are in salt balance on a day-to-day basis and thus that sodium excretion equals sodium intake. The method usually involves collecting a 24-hour urine specimen from the subjects. Normally, well over 90% of total body sodium loss is via the urine and the figure can be corrected for losses via the skin and in faeces, which are small under most circumstances. If there is severe diarrhoea or unaccustomed heavy sweating due to physical activity, a high ambient temperature or fever, losses by these other routes may, of course, be very considerable. Ideally, a marker substance should be used to ensure that the 24-hour urine specimen is complete, i.e. a substance that is consumed by the subject and excreted in the urine; only where there is complete recovery of the marker is the sample considered to be complete.

In a large sample of British adults, Gregory *et al.* (1990) found a very poor correlation ($r = 0.25$) between salt intake as measured by 24-hour urinary sodium excretion and that calculated from the average salt content of foods recorded during a 7-day weighed inventory. There are several factors that may contribute to this poor agreement between the two methods:

- The sodium excretion method involved just 1 day, which was different from the time of the weighed inventory.
- No allowance was made for salt added during cooking or at the table in the weighed inventory.

- No marker was used to ensure completeness of the 24-hour urine specimen.

Nevertheless, the lack of agreement between these methods underlines the difficulty of estimating a subject's habitual salt intake in order to correlate this with, for example, blood pressure in epidemiological studies. Both methods have been used as the measure of habitual intake in epidemiological studies. Nutrient intakes fluctuate from day to day and several days' intake needs to be recorded to obtain a useful estimate of average daily intake. The length of the necessary recording time varies from nutrient to nutrient, depending upon its distribution in food (see Chapter 3). This means that this disagreement between the two methods is in line with what is found with other nutrients. For example, potassium intake in food and 24-hour excretion of potassium were also both measured in this sample of people. In this case there is no discretionary intake from table and cooking use. Despite this, the correlation between the two methods for potassium intake was almost as poor as that for sodium. This strongly suggests that, even though 24-hour sodium excretion may be an accurate reflection of sodium intake on that day, large day-to-day variations may make it a relatively poor predictor of the habitual salt intake of an individual.

Average 24-hour sodium excretion should be a much better reflection of the habitual intake of a population or group. Using this method, James *et al.* (1987) concluded that average salt intake in the UK was around 10.7 g/day for men and 8.0 g/day for women. These estimates of average intakes agree well with other recent UK estimates using this method (e.g. Gregory *et al.*, 1990), but are less than some earlier estimates (e.g. those assumed by NACNE, 1983). The authors suggested this was because those using more traditional methods had not recognized that as much as 75% of cooking salt is not actually consumed and most is discarded with cooking water.

James *et al.* provided their subjects with pots of table and cooking salt that had been 'labelled' with lithium; recovery of lithium in urine enabled them to estimate what proportion of total excretion was from these sources. By comparing the amount removed from the pot with the amount recovered, they were also able to make their estimate of the proportion of this salt that is not consumed, e.g. that is discarded in cooking water or in uneaten food. They concluded

that discretionary salt made up only about 15% of total intake, 9% from table salt and 6% from cooking salt. They estimated that a further 10% of total intake was from salt naturally present in foods, but that the remaining 75% was derived from salt added by food manufacturers during processing (the salt content of some common foods is shown in Table 13.1). Such figures indicate that there is very limited scope for individuals to reduce their salt intake by restricting their discretionary use of salt. If salt intake is to be substantially reduced, either the salt content of manufactured foods must be reduced or the contribution of high-salt manufactured foods to the total diet must be reduced.

Reduced-salt versions of several salty foods are now widely available in Britain, e.g. cured meats and some canned foods. In late 1998, an announcement by a major British supermarket chain that they were going to reduce the salt content of their own-brand prepared foods by 25% achieved much publicity. Many other supermarket chains and food producers claimed that they were already reducing the salt content of their manufactured food products. Encouraging food manufacturers to use less salt may not be entirely without problems. Salt is used by manufacturers not only as a flavouring agent but also as a preservative and processing aid. Reduced salt

content of processed foods may have effects other than on their palatability. In particular, shelf life may be adversely affected, thus increasing costs, and the microbiological safety of the food may also be threatened. If salt is replaced by alternative preservatives, this increased intake of alternative preservatives may itself be undesirable. In a chapter entitled 'The industrial conspiracy', MacGregor and de Wardener (1999) argue that salt producers and food processors have resisted pressure to reduce the salt content of manufactured foods for commercial reasons rather than out of any necessity to maintain food safety or quality.

A review of the evidence for a salt-hypertension link

In some populations that have very low salt intakes of about 1 g/day, hypertension is practically unknown and there is no general increase in blood pressure with age, as is seen in the UK and USA (e.g. Kalahari Bushmen and the Yanomamo Indians of Brazil). In countries such as Japan and Korea, average salt intake may be two to three times that seen in the UK and there is a very high incidence of strokes and gastric cancer, which have both been causally linked to high salt intake. The low-salt populations often have diets

- A normal man's body contains around 250 g of salt and sodium is the major cation in extracellular fluid.
- Sodium and chlorine are essential minerals, but current salt intakes are ten times minimum requirements.
- Individuals adapt to low salt intakes by reducing the sodium concentrations of sweat and urine.
- The UK RNIs for sodium and chlorine are equivalent to about 4 g/day of salt.
- 24-hour sodium excretion is regarded as the best measure of salt intake.
- There is very poor correlation between salt intake as measured by conventional dietary analysis and 24-hour sodium excretion.
- 24-hour sodium excretion is a good measure of any day's salt intake or the average intake of a population; it is a poor measure of an individual's habitual salt intake because of large day-to-day variations.
- More than 75% of UK dietary salt is from manufactured foods and only around 15% is salt added at the table and during home cooking.
- Substantial reductions in total salt intake require either that less is used by food processors or that the consumption of high-salt processed foods is reduced.
- The salt content of many processed foods has been reduced and low-salt versions of a number of processed foods are now available.

and lifestyles that are very different from those of affluent Westerners. Any number of confounding variables could explain the apparent link between salt and hypertension, e.g. high physical activity, low levels of obesity, low alcohol and tobacco use and high potassium intake of those populations with low salt intake. There are, however, some examples of fit, active populations still leading a fairly traditional lifestyle who nonetheless have high salt intakes and they have blood pressure and hypertension profiles that are more typical of those seen in Western populations. For example, Solomon islanders who cooked their food in sea-water had eight times the incidence of hypertension compared to those living on nearby islands whose salt intake was low. Tobian (1979) and MacGregor and de Wardener (1999) provide interesting summaries of the early work in this area.

The level of salt in these low-salt populations, such as the Yanomamo Indians, is so low as to be a totally unrealistic goal for those living in affluent countries. However, more extensive cross-population studies, like those of Giebermann (1973), suggested that there was a near-linear relationship between the average salt intake of a population and the average blood pressure in middle-aged and elderly people (see Figure 3.8 in Chapter 3). This would mean that even modest reductions in population salt intake could be expected to lead to reduced average blood pressure, reduced incidence of hypertension and thus, eventually, to reduced morbidity and mortality from those conditions that are precipitated by hypertension. In Giebermann's paper, the reliability and validity of many of the measures of salt intake were questionable. As in all such cross-cultural studies, there was the recurring problem of confounding variables. Giebermann's data also included some black populations who are genetically salt sensitive and prone to hypertension, and this could account for some of the most divergent points on her graphs.

In a more recent cross-population study, Law *et al.* (1991a) related blood pressure to salt intake from reported data on 24 populations around the world. They overcame some of the criticisms of Giebermann's earlier work by:

- restricting their analysis to data in which salt intake was measured by 24-hour sodium excretion
- excluding African and other black populations from their study

- analysing data from developed and economically underdeveloped countries separately so as to reduce the problem of confounding lifestyle variables
- including in their study people across a wide age range, enabling them to look at the relationship between salt intake and blood pressure in different age groups.

Despite these improvements in methodology, their conclusions were similar to those of Giebermann. They found a highly significant relationship between average blood pressure and population salt intake in both developed and economically underdeveloped populations. The effect of salt intake on blood pressure increased with age and with initial blood pressure level. They derived an equation that they suggested would predict the average change in blood pressure of a group that would result from a specified change in salt intake. The variables in this equation were the group's age, current salt intake and initial blood pressure. They predicted, for example, that in a developed country like the UK, a 60% reduction in salt intake in 60–69 year olds with a systolic pressure of 183 mmHg might be expected to lead to a 14 mmHg reduction in systolic pressure.

Studies of migrant populations confirm that much of the international variation in blood pressure and incidence of hypertension-related diseases is due to environmental rather than genetic factors. People of Japanese origin in the USA, where salt intake is much lower than in Japan, have a reduced incidence of strokes as compared to Japanese people living in Japan. Samburu tribesmen in Kenya have a traditional low-salt diet in their villages, but if they join the Kenyan army, the high-salt army rations are associated with a rise in their average blood pressure after 2–3 years (see Tobian, 1979).

The reduced reliance upon salting as a means of food preservation during the twentieth century has been associated with falling mortality from strokes and gastric cancer in both the USA and UK and, indeed, in most industrialized countries. On an international basis, there is a strong positive correlation between stroke and gastric cancer mortality rates in industrialized countries. This is consistent with there being a common causative factor for the two diseases. Joossens and Geboers (1981) regressed mean death rates for gastric cancer in 12 industrialized countries with those for stroke; they found a very strong linear

correlation between 1955 and 1973 (see Figure 3.9 in Chapter 3). In a more recent study, Joossens *et al.* (1996) reported a strong positive correlation between average salt intake and death rate from stomach cancer in a study of 24 populations from around the world.

The weak point in the evidential link between salt intake and hypertension has traditionally been the lack of association between salt intake and blood pressure in individuals from within the same community. This lack of association has been a fairly consistent finding in many within-population studies (Tobian, 1979; Law *et al.*, 1991b) and may be because of the general tendency for the influence of other factors upon blood pressure to obscure the relationship with salt intake (see Chapter 3). The difficulty of reliably measuring habitual salt intake would also be an important factor – many studies have used 24-hour sodium excretion as the measure of habitual salt intake, but it is almost certainly a very poor measure of this. Others have argued that this lack of association is because only a relatively small proportion of the population are salt sensitive. There will inevitably be genetic variation in salt sensitivity but, if this variation is so great that effectively only a small subgroup of the population are salt sensitive, this would greatly weaken the case for population intervention to reduce salt intake.

Frost *et al.* (1991), in a complex re-analysis of 14 published within-population studies of the relationship between salt intake and blood pressure, concluded that the results were consistent with the results predicted from their equation derived from between-population analysis (Law *et al.*, 1991a). They argued that day-to-day variation in an individual's salt intake tends to obscure the true association between 24-hour sodium intake and blood pressure in within-population studies. This is consistent with the poor correlation between 24-hour sodium excretion and salt content of food in weighed inventories reported by Gregory *et al.* (1990), which was discussed earlier in the chapter.

There have been several intervention trials that have aimed to reduce the salt intake of a population as a means of reducing blood pressure and the incidence of hypertension. The results of trials conducted in Japan and Belgium are consistent with the belief that salt restriction will have beneficial effects upon blood pressure, but are less than conclusive (see MacGregor and de Wardener, 1999). Forte *et al.* (1989) report the

effects of an intervention trial using two small, rural populations in Portugal. Over a 2-year period, the intervention succeeded in reducing average salt intake by at least 25% in the test village, whereas that in the control village did not change. By the end of the study period, the average systolic blood pressure in the test village had fallen by 13 mmHg relative to the control village.

Denton *et al.* (1995) conducted a 2-year controlled study of the effects of salt on the blood pressure of chimpanzees. Salt was added to the normal vegetable and fruit diet of the experimental group and, over the first 18 months of the study, blood pressure rose substantially in 10 of the 13 animals in the experimental group. When the salt supplements were withdrawn, the blood pressure of these chimps returned to normal within 20 weeks.

IS SALT RESTRICTION AN EFFECTIVE ANTIHYPERTENSIVE THERAPY?

Moderation of salt intake is widely promoted as a preventative measure, but it has also long been used as a method of treating hypertension. Before the advent of the antihypertensive drugs, extreme salt restriction was one of the few treatments available for severe hypertension. The Kempner rice and fruit diet aimed to restrict salt intake to 1 g/day. This diet was effective in some individuals, but it was associated with very poor compliance, anorexia and side-effects that greatly limited its practical usefulness. These early therapeutic studies of patients with severe (malignant) hypertension generally suggested that, although extreme salt restriction was effective in some hypertensives, more moderate salt restriction (say 2.5+ g/day) was not effective. The advent of antihypertensive drugs heralded a general loss of interest in this method of treating severe hypertension. There is now, however, renewed interest in the use of moderate salt restriction as a therapy for those people with asymptomatic mild to moderate hypertension where the principal aim is to reduce the long-term risk of hypertension-linked diseases. Drug therapy is effective in reducing hypertension in such subjects, but lifelong consumption of antihypertensive drugs is clearly not an ideal solution and may be associated with a new range of hazards.

A large number of trials of the effectiveness of salt restriction in the treatment of hypertension have been conducted in recent decades, e.g. note the rigor-

ous double-blind, random, cross-over study of MacGregor *et al.* (1982) discussed in Chapter 3. Law *et al.* (1991b) analysed the results of 78 controlled salt reduction trials. They concluded that, in those trials in which salt restriction was for 5 weeks or more, the results obtained closely matched those that they had predicted from the cross-population analysis of Law *et al.* (1991a). The effects of salt restriction were less than predicted when shorter periods of salt restriction were used. They concluded that in 50–59 year olds in industrialized countries, a reduction in salt intake of 3 g/day could be expected to lead to an average reduction in systolic blood pressure of 5 mmHg, with greater reductions in those who were hypertensive. They argue that such a level of salt reduction would, if universally adopted, lead to major reductions in stroke and heart disease mortality.

OTHER FACTORS INVOLVED IN THE AETIOLOGY OF HYPERTENSION

Much of the evidence implicating salt in the aetiology of hypertension could also be used to support the proposition that high potassium intake is protective against hypertension. In acute studies of sodium loading in both salt-sensitive rats and people, high potassium intake seems to ameliorate the effects of high sodium loads. Some correlation between sodium:potassium ratio and blood pressure may be found in within-population studies where no relationship can be shown between blood pressure and sodium intake alone.

There is some epidemiological evidence that low calcium intake may be associated with increased risk of hypertension. People who suffer from osteoporosis also have a higher incidence of hypertension, a finding consistent with a common aetiology. Some have used this latter observation to suggest that high salt may cause osteoporosis, whereas others have argued that low calcium intake might be a factor in hypertension. Several intervention trials have suggested that calcium supplements may reduce blood pressure, at least in a proportion of hypertensives.

Overweight and obesity are generally agreed to lead to increases in blood pressure, and weight loss reduces blood pressure in the overweight hypertensive. White *et al.* (1993) reported that, in their representative sample of English adults, blood pressure rose with increasing body mass index. Increased physical activity reduces the tendency to gain excess-

sive weight and also contributes directly to a reduction in blood pressure. Excessive alcohol consumption also leads to a rise in blood pressure and increases the risk of stroke, even in the absence of hypertension. White *et al.* (1993) found that blood pressure was positively associated with alcohol consumption in male drinkers, but they also found that non-drinkers had higher blood pressures than low and moderate drinkers.

A list of several of the factors that are associated with an increase in blood pressure and an increased risk of hypertension is given in Table 13.4.

CONCLUSIONS

An active lifestyle, moderate alcohol consumption and good control of body weight are undoubtedly factors that will lessen the risk of hypertension and the hypertension-linked diseases. Increased intakes of fruits and vegetables will increase potassium intakes and lower the sodium:potassium ratio in the diet; processing of foods generally involves an increase in this ratio by some loss of potassium and the addition of salt. There are general grounds for assuring that calcium intakes meet dietary standard levels, and this may have some beneficial effect on population blood pressure.

There is compelling evidence of some causal link between dietary salt intake and hypertension. The argument tends to be about how significant and widespread the benefits of salt restriction would be: would salt intake moderation lead to a general fall in average population blood pressure or would effects be limited to a relatively small proportion of salt-sensitive people? Moderate salt restriction

Table 13.4 Factors associated with increased risk of hypertension

Being black: prevalence is higher amongst blacks in the USA and UK than in whites; there are almost certainly also large individual variations in susceptibility within races
Being overweight
High alcohol consumption
A sedentary lifestyle
High population salt intake
High dietary fat intake
Low potassium or calcium intake?

will almost certainly be of some value to some salt-sensitive people, and the general consensus is that it would have wider benefits. Salt restriction may also be useful in the treatment of mild to moderate hypertension, either as an alternative to drug use

or as an adjunct to drug therapy. Processed foods contain most of the salt in Western diets and so any substantial reduction in salt intake requires either less reliance upon these high-salt foods or reduced use of salt by food manufacturers.

- Cross-population studies show that the average blood pressure of adult populations is highly correlated to average salt intakes.
- This correlation remains strong even when developed and developing countries are analysed separately.
- When migration is accompanied by changes in salt intake, measured changes in blood pressure or stroke mortality are consistent with salt intake being a cause of hypertension.
- Falls in salt consumption over the last century have been associated with reductions in stroke mortality.
- Internationally, death rates from stroke and gastric cancer are highly correlated and this suggests a common aetiology.
- An international correlation between average salt intake and gastric cancer mortality has also been reported.
- In general, there is no correlation between salt intake and blood pressure in individuals from the same population. This is not unexpected and is partly because the usual measure of salt intake, 24-hour sodium excretion, is a poor measure of habitual salt intake.
- Limited evidence from population interventions is consistent with the salt-hypertension hypothesis.
- When salt is added to the fruit and vegetable diets of chimpanzees, their blood pressure rises relative to control animals and falls when the supplemental salt is withdrawn.
- Several clinical trials suggest that sustained moderate reductions in salt intake reduce the blood pressures of people with mild to moderate hypertension.
- Very severe salt restriction was the first partly effective therapy for severe hypertension, although it fell into disuse with the advent of effective antihypertensive drugs.
- Other factors associated with an increased risk of hypertension are listed in Table 13.4 and attention to these will contribute to a reduction in hypertension prevalence.
- Reduced salt consumption will almost certainly help to reduce the prevalence of hypertension; the debate is really about how widespread these benefits will be.

Part 4

Variation in nutritional requirements and priorities

Nutrition and the human life-cycle

Introduction	337	Infancy	349
Nutritional aspects of pregnancy	338	Childhood and adolescence	358
Lactation	348	The elderly	361

INTRODUCTION

When I show my students Table 14.1 and ask for their comments, they almost invariably consider it a very bad diet, which they doubt anyone could actually be eating. This is, in fact, the approximate composition of mature human milk, the ideal food for babies. This extreme example illustrates the general point that nutritional needs and priorities are likely to change during the human life cycle. As another example, one would expect the relative nutritional needs of an active, rapidly growing adolescent to be markedly different from those of an elderly, housebound person or even of a sedentary, middle-aged person. Such differences mean that the Dietary Reference Values vary with age, sex and reproductive status, but they also mean that there should be differences in the priority and nature of the general dietary guidelines offered for different life-cycle groups. This chapter focuses on the differing requirements and nutritional priorities of different life-cycle groups.

Table 14.1 An unbelievably bad diet?

	Energy (%)
Fat	53
Saturated fat	23
Carbohydrate	39
Sugars	39
Starch	0
Protein	8
Fibre (g)	0

Overt deficiency diseases are uncommon in industrialized countries, but this is no guarantee that many people are not more subtly affected by sub-optimal nutrition. Those people who are subject to particular nutritional stresses, such as pregnancy, lactation or rapid growth, are likely to be the most sensitive to any marginal deficiencies or suboptimal intakes of nutrients.

- Different life-cycle groups have widely differing nutrient needs, and the nutritional priorities for health promotion also vary with age, sex and reproductive status.
- Rapid growth, pregnancy or lactation may make people more sensitive to marginal nutrient deficiencies.

NUTRITIONAL ASPECTS OF PREGNANCY

Overview

Sadler (1994) has written a concise, referenced review of nutrition in pregnancy and lists many of the uncited references used for this section.

Healthy, well-fed women have the best chance of having healthy babies and of lactating successfully after the birth. One would expect pregnancy to be a time of significantly increased nutritional needs for the mother; the phrase 'eating for two' has, in the past, been widely used to sum up this expectation. The health and well-being of many pregnant women and their babies might be adversely affected by sub-optimal nutrition, despite the low frequency of overt malnutrition within the population. Morning sickness affects over half of women in the early part of pregnancy and this could compound with other stresses to deplete nutrient stores in some women.

Cross-species comparisons show that human babies are much smaller than would be predicted from the size of the mother and that gestation is also longer than would be predicted. Thus, the nutritional burden of pregnancy is likely to be less in women than in other, non-primate species. The small size and relatively slow postnatal growth of human babies mean that similar arguments apply to lactation. This difference in the relative burdens of pregnancy and lactation in different species is illustrated in Table 14.2. An extreme comparison is between a female mouse and a woman. The mouse produces a litter that represents around 40% of her pre-pregnant weight in 21 days of gestation. She then provides enough nutrients in her milk to enable this litter to double its birth weight within 5 days. A

woman produces a baby that represents 6% of her body weight after 9 months of pregnancy, and a breastfed baby might then take 4–6 months to double its birth weight. Extrapolating or predicting the extra nutritional needs of pregnant or lactating women from studies of non-primate animals may be misleading and tend to greatly exaggerate the extra needs of these women.

Behavioural and physiological adaptations to pregnancy and the utilization of maternal nutrient stores also tend to minimize the extra nutritional needs of pregnant women. These extra nutritional needs may thus turn out to be small in comparison to the needs of non-pregnant women. 'Eating for two' implies a much greater increase in need than is indicated by measurements of the actual increase in intakes that occurs in pregnant women and those indicated by current Dietary Reference Values (DRVs) for pregnant women.

Table 14.3 summarizes the current UK and US recommendations of the extra nutritional needs of pregnant women. This table highlights some significant differences between the UK and US dietary standards. The US increments are often substantially higher than those in the UK, e.g. those for iron, calcium and the B vitamins. As noted in Chapter 3, baseline adult values also tend to be higher in the USA. This means that, for all nutrients, the absolute US Recommended Dietary Allowance (RDA) for pregnancy equals or exceeds the corresponding UK value.

Where the percentage increase in the reference value for a nutrient in Table 14.3 significantly exceeds the percentage increase in the energy value, this implies the need for increased nutrient density or supplementation. For example, if a non-pregnant American woman is just meeting her RDA for cal-

Table 14.2 Species variations in the relative burdens of pregnancy and lactation

Species	Maternal weight (kg)	Litter weight (% of maternal)	Gestation length (days)	Days to double birth weight
Mouse	0.025	40	21	5
Rat	0.200	25	21	6
Guinea-pig	0.560	68	67	14
Rabbit	1.175	19	28	6
Cat	2.75	16	64	7
Woman	56	6	280	120–180

Table 14.3 Daily extra increments for pregnancy

Nutrient	UK RNI*	US RDA
Energy (kcal)	200 (10) ^b	300 (14) ^c
(MJ)	0.8	1.26
Protein (g)	6 (13)	10 (20)
Vitamin A (μg)	100 (17)	0
Vitamin D (μg)	10 ^d	5 (100)
Thiamin (mg)	0.1 (13) ^b	0.4 (36)
Riboflavin (mg)	0.3 (27)	0.3 (23)
Niacin (mg)	0	2 (13)
Folic acid	Supplements now recommended in UK and USA	
Vitamin C (mg)	10 (25)	10 (17)
Calcium (mg)	0	400 (50)
Iron (mg)	0	15 (100)
Zinc (mg)	0	3 (25)

^a EAR for energy.^b Last trimester only.^c Last two trimesters.^d No RNI for non-pregnant/lactating women.

Values in parentheses are percentages of the value for other women.

Source: COMA (1991) and NRC (1989).

cium, then, on becoming pregnant, she would need to increase the prominence of calcium-rich foods in her diet otherwise she would need to increase her total food intake by 50%. Even many non-pregnant American women fail to meet the calcium RDA, and therefore few pregnant women are likely to meet the now substantially higher RDA. Some values, such as the US RDA for iron and the UK Reference Nutrient Intake (RNI) for vitamin D, require the use of supplements to be realistically achievable. In other cases, average intakes of relatively affluent UK and US women will comfortably exceed the pregnancy RNI/RDA for many nutrients prior to pregnancy, and so, in practice, avoid the need for dietary change. For example, average intakes of riboflavin in British women are 1.8 mg/day, comfortably in excess of even the raised RNI for riboflavin in pregnancy. Average intakes of all vitamins in British women in 1987 were comfortably in excess of the then RDAs (Gregory *et al.*, 1990).

Many well-nourished women take supplements during pregnancy and the assumption is that, even if they do no good, then neither will they do any harm. Apart from folic acid, which is discussed later in the chapter, there is little evidence that such supplements do any good. On the contrary, there is clear evidence that high levels of retinol are teratogenic and that

concentrated protein supplements reduce birth weight. There is also some evidence that routine iron supplements can increase the risk of low birth weight and the risk of prematurity when given to women who do not need them (see Mathews, 1996).

There is a strong association between infant mortality rate and birth weight. Babies who weigh less than 2.5 kg are classified as **low birth weight (LBW)** babies and they have higher mortality and morbidity rates. Thus, the birth weight of babies is a useful, quantitative measure of pregnancy outcome. Anything that restricts fetal growth is liable to be detrimental to the chances of the baby's survival and subsequent well-being. Many of the risk factors for having a low-birth-weight baby are listed in Table 14.4. When the birth weight of babies is plotted against mortality, there is a U-shaped curve. As babies get heavier, initially, the mortality risk drops, until an optimum birth weight is reached that is slightly higher than the average birth weight of around 3.5 kg. As birth weight rises above this optimum, so mortality risk starts to increase again (see Symonds and Clarke, 1996). Many low-birth-weight babies are born prematurely, and gestational diabetes can be a cause of high birth weight.

The ideal outcome of pregnancy is not only a healthy baby, but also a healthy mother who is nutritionally well prepared to lactate. The preparedness of the mother to begin lactation and the general well-being of the mother after parturition are not so readily quantifiable as birth weight, but can be important determinants of successful reproduction. In poorer countries, the presence or absence of a healthy, lactating mother may be a key determinant of the baby's chances of survival.

Table 14.4 Some risk factors for having a low-birth-weight baby

Low social class of the mother
Mother less than 18 years or over 35 years
Previous birth of a low-birth-weight baby
Maternal drug or alcohol abuse
Cigarette smoking
Mother being underweight at the time of conception
Low maternal weight gain in pregnancy
Dieting or restricted energy intake during pregnancy
Poor quality of maternal diet during pregnancy
Excessive exercise or work during pregnancy

Maternal weight gain during pregnancy may be one quantifiable measure of the mother's preparedness for lactation. In rats, there is a 50% increase in the maternal fat store during pregnancy and this fat store is rapidly depleted during the subsequent lactation. This increase in the maternal fat store is a physiological adaptation to enable the female rat to provide enough milk for her large and rapidly growing litter. The increases in maternal fat typically seen in human pregnancy can also be assumed to be physiological. In women in industrialized countries, who have access to an essentially unlimited food supply, this store of energy for lactation may not be critical. It may be much more important to women who are subsisting on a marginal diet and who may have a limited opportunity to boost their intakes to meet the extra needs of lactation.

For those nutrients that accumulate during the pregnancy, either in the mother's stores or in the products of conception, it is possible to use factorial calculations (see Chapter 3) to predict the likely extra requirements for these nutrients during pregnancy. However, it should be remembered that these are only theoretical predictions and, whilst they may

provide a useful basis for discussion, they may not truly indicate extra dietary needs because of the behavioural and physiological adaptations to pregnancy.

There has, over the years, been a very considerable amount of discussion about how any extra nutritional requirements of pregnant women are distributed throughout the pregnancy. At the end of the first trimester, the fetus weighs around 5 g and around 500 g at the end of the second trimester. This compares with an average birth weight of 3.5 kg. Thus, fetal growth accelerates throughout pregnancy, with most of the fetal weight gain concentrated in the final trimester. Should any extra nutritional allowances be concentrated in the last trimester when fetal growth rate is highest? Several adaptive mechanisms ensure that the burden is distributed more evenly throughout the pregnancy. In the first dietary standards (NRC, 1943), the increases in RDAs in pregnancy were even higher than current US values, but the increases in 1943 were all confined to the second half of pregnancy, whereas most current increases are for the whole of the pregnancy.

- The extra nutritional needs of pregnant women are modest and much less than the phrase 'eating for two' implies.
- Behavioural changes, physiological adaptations and utilization of maternal reserves all tend to minimize extra requirements in pregnancy and the impact of suboptimal intakes.
- The nutritional burden of human pregnancy is much less than that in non-primate species.
- US allowances for pregnancy tend to be higher than those in the UK (see Table 14.3).
- Large increases in nutrient intakes may be difficult to achieve if there is only minimal increase in total food intake in pregnancy. They may only be realistic if supplements are used.
- For many nutrients, the habitual intakes of non-pregnant women exceed the UK RNI for pregnant women.
- Apart from folic acid, there is little evidence that nutrient supplements are beneficial in pregnancy unless women are undernourished.
- Overdoses of some nutrients are harmful during pregnancy.
- Low-birth-weight babies (under 2.5 kg) have higher postnatal mortality and morbidity; some of the risk factors for low birth weight are listed in Table 14.4.
- Too high a birth weight increases obstetric complications and also increases mortality risk.
- A successful pregnancy should produce a healthy, normal-weight baby and a healthy mother who is well prepared for lactation.
- Several mechanisms may serve to distribute the nutritional burden of pregnancy more evenly than the concentration of fetal growth in the last trimester might imply.

Energy needs in pregnancy

A typical British woman gains around 12.5 kg in weight during a typical pregnancy. This weight gain is made up of around 3.5–4 kg of extra maternal fat, around 6 kg as the increased weight of the uterus and contents, and the remainder is due to increased body fluid. The total theoretical energy cost of this weight gain in a typical pregnancy has been estimated at about 70 000–80 000 kcal (around 300 MJ). In under-nourished women in developing countries, the total weight gain during pregnancy may be only half of this value, with no net increase in maternal fat stores.

The theoretical energy cost of the average British pregnancy with 3.5 kg of maternal fat gain works out at just under 300 kcal (1.3 MJ) per day. If this theoretical energy cost is distributed throughout the pregnancy in proportion to the rate of fetal growth, this would amount to an extra energy cost of over 750 kcal/day (3.1 MJ/day) in the last trimester. However, most of the gain in maternal fat occurs in the first trimester and this would tend to even out any extra energy requirements during the pregnancy. The maternal fat laid down in early pregnancy represents not only a reserve for lactation, but also a potential energy reserve for the last trimester if intake at this time is inadequate.

The current UK Estimated Average Requirement (EAR) for pregnant women is only 200 kcal (0.8 MJ) above that for non-pregnant women and is for the last trimester only (see Table 14.3). The UK panel on Dietary Reference Values (COMA, 1991) did suggest that women who are underweight at the start of pregnancy or who remain active during their pregnancy may need more than suggested by this EAR. The US energy RDA for pregnant women is significantly more generous than the UK equivalent. An extra 300 kcal/day (1.25 MJ/day) is recommended for the second and third trimesters. Several studies have suggested that pregnant women in Britain only tend to eat more than non-pregnant women in the last few weeks of pregnancy, and then less than 100 kcal/day (0.4 MJ/day) extra (COMA, 1991). These and similar findings in other industrialized countries suggest that the spontaneous increases in energy intake during pregnancy are very small in previously well-nourished women. As energy balance is physiologically regulated, this strongly implies that the extra energy needs of pregnant women are also very small. Clearly, many women in

industrialized countries markedly reduce their activity levels during pregnancy and so save much of the anticipated energy costs of the pregnancy. Studies on marginally nourished pregnant women, using the doubly labelled water method of measuring long-term energy expenditure, further suggest that physiological energy-sparing mechanisms operate during pregnancy (Prentice, 1989). Lack of maternal fat storage during pregnancy and energy-sparing physiological mechanisms may eliminate any theoretical energy costs of pregnancy in developing countries.

Any reduced energy supply to the fetus during pregnancy is likely to restrict fetal growth and thus increase the risk of low birth weight. There is strong evidence linking maternal weight gain during pregnancy with birth weight of the baby. Low weight gain is associated with an increased risk of low birth weight and perinatal mortality. Excessive weight gain is associated with prolonged labour, birth injury, higher rates of death and complications in the neonatal period and an increased risk of maternal obesity. In a survey of 10 000 births in the USA in 1968 (summarized in Morgan, 1988), the chances of having a low-birth-weight baby were inversely related to maternal weight gain in pregnancy. Low-birth-weight babies were born to:

- 15% of women who gained less than 7 kg
- 8% of women with 7–12 kg weight gain
- 4% of women with 12–16 kg weight gain
- 3% of women who gained 16+ kg.

Restricting maternal weight gain by dieting during pregnancy may increase the risk of having a low-birth-weight baby. In animal studies, even food restriction of obese mothers during pregnancy reduced fetal weight. This suggests the need for extreme caution in the use of energy-restricted diets, even if women are overweight at the start of pregnancy. The risk of having a low-birth-weight baby is also inversely related to maternal weight at the onset of pregnancy. In one study quoted by Morgan (1988), the heaviest 10% of women had only one-third (2.3%) the incidence of low-birth-weight babies of the lightest 10%. Weight reduction should not be attempted during pregnancy, but the American National Academy of Sciences has published guideline values for recommended weight gains for women during pregnancy that vary with their pre-pregnant weight, i.e.:

- underweight at conception – 12.5–18 kg weight gain
- normal weight – 11.5–16 kg gain
- overweight – 7–11.5 kg gain
- obese – at least 6 kg gain.

Pre-eclampsia, or pregnancy-induced hypertension, is not precipitated by high weight gain in pregnancy unless that weight gain is as the result of excessive fluid retention, but obesity prior to becoming pregnant is a significant risk factor for pre-eclampsia. Ideally, women intending to become pregnant should try to ensure that their body mass index (BMI) is within the normal range (20–25) prior to conception. Women seeking pre-conception advice can be counselled to normalize their weight before becoming pregnant. Menstruation ceases and fertility is reduced in women who are underweight and only 10% of women with a BMI of less than 18 are fertile. Underweight women who do manage to conceive have an increased risk of having a low-birth-weight baby.

Glucose that has been transferred from maternal blood in the placenta is the primary energy source of the fetus. The concentration of glucose in fetal blood is always lower than that in maternal blood, and the transfer of glucose is by facilitated diffusion down a concentration gradient. This means that, if the maternal blood glucose concentration is low, this can limit the availability of glucose to the fetus. Active transfer of nutrients from maternal blood makes this less likely to restrict supplies of other nutrients such as amino acids to the fetus. Starvation or heavy exercise during pregnancy may restrict fetal growth by reducing glucose concentrations in placental blood. Fetal nutrient supply depends not only upon nutrient concentration in maternal blood, but also upon placental blood flow; therefore, anything that reduces placental blood flow may also restrict fetal growth and nutrient supply. One effect of smoking may be to reduce placental blood flow and thus retard the supply of nutrients to the fetus and fetal growth. Purely anatomical factors, such as the precise site of implantation and development of the placenta, may affect fetal blood flow and may account for some unexplained incidence of low-birth-weight babies. Uncontrolled gestational diabetes can increase the supply of glucose to the fetus and cause babies to be born too heavy.

Poverty is associated with low birth weight. The average weights of babies born in developing coun-

tries are lower than in industrialized countries and the proportion of low-birth-weight babies is also much higher in many developing countries. These differences are not removed by correcting for differences in maternal stature. Even in industrialized countries, social class and economic status affect birth weight. Leon (1991) found progressive increases in the percentage of low-birth-weight babies and increased postnatal mortality with decrease in the socioeconomic status of the mother. Evidence from a number of studies indicates that nutritional factors are a likely explanation for some of these differences. Women from a poor community in Britain who had a low-birth-weight baby had lower average energy intakes than those who had had babies of ideal weight. During the Dutch famine at the end of World War II, the average birth weight of babies was reduced by over 300 g. Studies in Guatemala have shown that women who were marginally nourished, with typical unsupplemented intakes of 1600 kcal/day (6.7 MJ/day), had significantly heavier babies and fewer low-birth-weight babies when they were given food supplements prior to conception and throughout the pregnancy. A similar study in Taiwan with women whose unsupplemented intake was higher, typically 2000 kcal/day (8.4 MJ/day), showed no significant effect on average birth weight (both studies are reviewed in Naismith, 1980). In The Gambia, there are significant differences in average birth weight between the hungry season and the harvest season. This difference can be eliminated by the use of protein-energy supplements, and the perinatal mortality is also reduced.

Dietary intervention studies generally suggest that, at very low baseline levels of intake, energy supplements do favourably affect birth weight and reduce the incidence of low-birth-weight babies. However, this favourable effect on birth weight is not seen at slightly higher baseline levels, despite the use of very substantial supplements. A lack of a measurable effect of supplements on birth weight in the less undernourished women does not necessarily mean that supplementation had no beneficial effects, because the supplements may have made mothers better prepared to lactate. In a review of such intervention trials, Morgan (1988) concluded that the supplements probably have a beneficial effect, despite their disappointing effect on birth weight, because of increased maternal weight gain and improved lactation.

- British women gain an average 12.5 kg in pregnancy and this represents a theoretical energy cost of 70 000–80 000 kcal (300 MJ) or 300 kcal (1.3 MJ) per day.
- The UK energy EAR for pregnancy rises by only 200 kcal (0.8 MJ) per day in the last trimester.
- The US RDA rises by 300 kcal (1.3 MJ) per day for the last two trimesters.
- Spontaneous increases in food intake by well-nourished women may rise by only 100 kcal (0.4 MJ) per day in the last few weeks of pregnancy.
- Reduced activity and perhaps energy-sparing physiological changes may account for the difference between the theoretical energy costs of pregnancy and the apparent increases in intake; under-reporting of intakes is also possible.
- In undernourished women, lack of maternal fat gain and energy-sparing mechanisms may eliminate even the theoretical energy cost of pregnancy.
- Low weight gain during pregnancy and being underweight at conception both reduce birth weight.
- Attempts at weight reduction should be avoided during pregnancy.
- Being overweight at conception increases the risk of pre-eclampsia.
- Being underweight reduces fertility and increases the risk of a low-birth-weight baby in women who do conceive.
- Glucose transfer across the placenta is passive and fetal blood glucose concentration is lower than maternal.
- Low maternal blood glucose or low fetal blood flow will restrict glucose supply to the fetus and so limit fetal growth.
- Poverty and low social class are both associated with reduced birth size and increased postnatal mortality of babies.
- Malnourished women have smaller babies, more low-birth-weight babies and higher postnatal mortality.
- Energy supplements increase birth weight and reduce low-birth-weight in the babies of women with low baseline intakes (<1600 kcal/6.7 MJ per day).
- In marginally undernourished women, supplements do not usually have a measurable effect upon birth weight, but they may increase maternal weight gain and fitness to lactate.

Protein in pregnancy

The current UK RNI for protein is increased by 6 g/day throughout pregnancy. The US RDA is increased by 10 g/day in women who are over 25 years and by 14 g/day in women who are 19–24 years old. If the UK RNI is expressed as a percentage of the EAR for energy, then, in the first two trimesters (when the energy EAR is unaltered), it represents a rise from 9.3% of energy to 10.7% of energy as protein. Protein represents around 15.2% of total energy intake in UK women (Gregory *et al.*, 1990) and so protein seems unlikely to be a limiting nutrient for pregnant women in the UK or, indeed, in other industrialized countries. In dietary intervention studies, protein-rich supplements appeared to offer

no additional benefits over and above their energy value. Even in marginally nourished women, it is energy rather than protein that seems to be the limiting factor. Amino acids are actively transported across the placenta and so adequate levels can be maintained in fetal blood even when levels in maternal blood are low. Some studies even suggest that concentrated protein supplements may depress birth weight (see Mathews, 1996).

Factorial calculations suggest a total 'protein cost' of about 925 g for a typical pregnancy. Theoretical calculations of protein accretion rates during the four quarters of pregnancy suggest rates of 0.6, 1.8, 4.8 and 6.1 g/day respectively (NRC, 1989a). Naismith (1980) provides direct evidence that, in the rat, extra protein is laid down in maternal muscle in

- The habitual protein intakes of women in industrialized countries exceed even the slightly raised RNI/RDA for protein in pregnancy.
- Even in developing countries, the supply of energy is more likely to be limiting than that of protein.
- Physiological mechanisms may help to distribute the protein costs of pregnancy more evenly through the gestation period.
- Intervention studies with protein-rich supplements provide no evidence that they are useful and some indications that they may depress birth weight.

early pregnancy and that this is released in late pregnancy when the demands from the growing fetuses are highest. He argues that limited observations in women are consistent with these observations in the rat, which may serve to distribute the extra protein need more evenly throughout the pregnancy. Measurements of nitrogen balance in pregnant women show that a small and fairly constant positive balance occurs throughout the pregnancy. Excretion of the amino acid 3-methyl histidine is taken as an indicator of muscle protein breakdown and it increases in the second half of human pregnancy.

The habitual consumption of protein in industrialized countries is almost always in excess of requirements and, thus, in practice, it is unlikely that protein supply will be limiting during pregnancy. Pregnant women do not usually need to take specific steps to increase their protein intake, despite their increased RNI/RDA. Even in developing countries, energy is more likely than protein to be the limiting factor in pregnancy.

Minerals in pregnancy

The current RNIs for minerals in the UK indicate no routine extra increment for any mineral in pregnancy. High priority has historically been given to increased mineral intakes during pregnancy and in the USA, and the RDAs for all minerals are increased during pregnancy. Details for iron, calcium and zinc are given in Table 14.3.

CALCIUM

Around 30 g of calcium accumulates in the fetal skeleton during pregnancy. This represents just over

100 mg/day when spread throughout the pregnancy. If one assumes that only a fraction of ingested calcium is retained (say 25–35%), then this factorial approach gives some idea of the origins of the 400 mg increase in the US RDA for calcium in pregnancy. The 'calcium cost' of pregnancy represents only around 2–3% of the calcium in the maternal skeleton, and there is clear evidence that the efficiency of calcium absorption in the gut increases during pregnancy. The UK panel on DRVs has concluded that, in most women, the extra calcium cost of the pregnancy can be met by the utilization of maternal stores and by improved efficiency of absorption (COMA, 1991). In certain cases (e.g. adolescent pregnancy), the panel suggests that some extra increment of calcium might be advisable.

IRON

During pregnancy, the plasma volume increases by around 1.5 L and the red cell mass by 200–250 mL. This means that, during a normal pregnancy, the large increase in plasma volume causes a fall in haematocrit and blood haemoglobin concentration, despite an increase in the number of circulating red cells and the amount of haemoglobin. This physiological fall in haemoglobin concentration in pregnancy was historically interpreted as iron-deficiency anaemia. Iron supplements will reduce or prevent the fall in haemoglobin concentration and have traditionally been given to pregnant women. The US RDAs clearly suggest that universal iron supplementation in pregnancy should continue. Iron supplements often cause adverse gastrointestinal symptoms and it is likely that many of them are not actually taken.

The net iron cost of pregnancy, allowing for the savings due to the cessation of menstruation, has

been estimated at between 500 mg and 1000 mg, i.e. 2–4 mg/day throughout the pregnancy (Sadler, 1994, suggested a figure of 680 mg). The UK panel on DRVs suggested that, in most women, the iron requirements for pregnancy can be met by the following adaptations:

- utilization of maternal iron stores
- a considerable increase in the efficiency of iron absorption during pregnancy – the intestinal absorption of iron may rise from a normal 10% to over 50% in the last trimester
- the cessation of menstruation

In the UK, there is considered to be no need for routine iron supplementation during pregnancy, although targeted supplementation (e.g. to women whose iron status is poor when they enter pregnancy) may be necessary.

The need for and benefits of iron supplements in pregnancy are still a matter of scientific debate. The US and UK panels have thus come up with very different practical 'recommendations' using essentially the same body of factual knowledge. Mathews (1996) drew the following conclusions from a review of the literature.

- Iron-deficiency anaemia is associated with an increased risk of low birth weight and prematurity.
- High haemoglobin and haematocrit levels are associated with similar increased risks, although this may not be a simple cause-and-effect relationship.
- Randomized, controlled trials have shown no benefit from routine iron supplements in pregnancy.

Screening and targeted supplementation are a feasible option for industrialized countries, but in developing countries where iron-deficiency anaemia is very prevalent and often severe, routine supplementation may be the only practical option.

Folic acid and neural tube defects

The collective term **neural tube defects (NTDs)** is used to describe a group of conditions (most commonly anencephaly and spina bifida) in which the brain, spinal cord, skull and/or vertebral column fail to develop properly early in embryonic life. In anencephaly, there is almost complete absence of the brain and skull and these babies die *in utero* or shortly after birth. In spina bifida, the spinal canal in the vertebral column fails to close and in severe cases the spinal cord bulges out of the back. Babies affected by this condition suffer from a range of physical disabilities that in severe cases include paralysis, bladder and bowel incontinence and neurological damage caused by hydrocephalus.

About 1 in 250 UK pregnancies are affected by NTDs and, despite screening and termination of affected pregnancies, around 1 in 4000 babies born in the UK have a NTD. In the 1970s, before widespread screening and termination, 1 in 700 births were affected. The risk is ten times higher in women who have already had an affected pregnancy, but most babies born with NTDs are nevertheless first occurrences.

The UK DRV published in 1991 suggested a 50% increase in the RNI for folic acid in pregnancy (to 300 µg) and the US RDAs published in 1989 recommended a 122% increase (to 400 µg). After these recommendations were formulated, a report was published that provided very strong evidence that folic acid supplements during pregnancy could reduce the risk of NTDs in babies. This was a randomized, double-blind, prevention trial using around 2000 women identified as being at high risk of having babies with NTDs because of a previous affected pregnancy (MRC, 1991). There were only six affected babies in the folic acid supplemented groups, but 21 in the same number of women not

- In the UK, no routine increases in mineral intakes are considered necessary during pregnancy, although they may be necessary for some individuals or groups.
- In the USA, there are substantial increases in mineral RDAs during pregnancy.
- In the UK, it is assumed that the mineral costs of pregnancy can be met by physiological adaptation, increased efficiency of absorption and utilization of maternal mineral reserves.

receiving folic acid supplements, i.e. a 72% protective effect. This study convincingly demonstrated the benefits of supplements for such high-risk women.

Studies published in the 1960s indicated that women who had taken a folic acid antagonist during pregnancy had an increased risk of having a baby affected by a NTD. Others had shown that women who gave birth to NTD-affected babies showed biochemical evidence of defective folic acid metabolism. In 1965, Smithells and his colleagues put forward the hypothesis that folic acid supplements might reduce the risk of NTDs by compensating for some defect in folic acid absorption or metabolism. In the early 1980s, results of intervention trials using preconceptual supplements that contained 400 µg or 4000 µg of folic acid were effective in preventing the recurrence of NTDs (references to these early studies may be found in Report, 1992).

Since the MRC (1991) study was published, other studies have confirmed the protective effect of folic acid, both in preventing the recurrence of NTDs in previously affected women and also in preventing primary occurrence in women with no history of an NTD-affected pregnancy (Czeizel and Dudas, 1992).

The dose of supplements (5 mg/day) used in the MRC trial represents around 25 times the current adult RNI for folic acid. At this dose, some rare but possibly serious and specific side-effects of folic overdose have been predicted and others are possible. In 1992, an expert advisory group in the UK recommended that all women in the high-risk category should take supplements of 5 mg of folic acid daily if

they plan to become pregnant and during the first 12 weeks of pregnancy. All other women were recommended to take a more modest 0.4 mg extra folic acid when planning to become pregnant and in the first 12 weeks of pregnancy (Report, 1992). However, as many as half of all pregnancies may be unplanned and so, to have maximum impact on the incidence of NTDs, all women of reproductive age would need to substantially increase their intake of folic acid. There is currently an active campaign in the UK to increase the range and level of folic acid fortification of foods such as bread and cereals. Such fortified foods are the only realistic way that many women will reach a total target intake of 0.6 mg/day of folic acid when they first become pregnant (note average intakes of UK women were about 200 µg/day in 1990). In 1996, the US Food and Drug Administration (FDA) announced that it would require fortification of all cereal grain products with folic acid. An expert report in the UK (COMA, 2000) has also recommended that there should be universal fortification of flour with 240 µg/100 g of folic acid. It estimated that this measure would reduce the incidence of NTD-affected pregnancies by around 40%.

Since the previous edition of this book, concerns about the safety of modest increases in folic acid consumption by the general population have eased and there is even some evidence that they may be beneficial in the prevention of heart disease. Progress has also been made in elucidating the mechanism by which folic acid exerts its protective effect against NTDs. The enzyme methionine synthase requires

- In the 1980s, around 1 in 250 UK pregnancies were affected by a NTD.
- Folic acid supplements in early pregnancy and preconceptually reduce both the recurrence and primary occurrence of NTDs.
- Supplements of 400 µg of folic acid are recommended for the first trimester of pregnancy and for those planning to become pregnant.
- Many pregnancies are unplanned and so the fortification of flour and other cereal products has been recommended in both the UK and USA.
- High doses (>1 mg) of folic acid may have well-documented detrimental consequences in some circumstances, but the whole population might benefit from moderate increases in folic acid intake.
- Folic acid supplements exert their protective effect by compensating for an hereditary low activity of the enzyme methionine synthase, for which folic acid is a co-factor.

folic acid as a cofactor and its activity is low in women who have had a NTD-affected pregnancy. Folic acid supplements may overcome a genetic defect in the regulation of this enzyme's activity (see Eskes, 1998).

Other vitamins in pregnancy

The increases in vitamin RNIs and RDAs in Britain and the USA are summarized in Table 14.3.

Table 14.3 shows that the UK RNI for vitamin A is increased by 100 µg in pregnancy. There is no extra increment for pregnancy in the USA, but the baseline value for American women is the same as that for pregnancy in the UK. It is, however, the possible consequences of excessive doses of retinol in pregnancy rather than potential inadequacy that are the current focus of concern about vitamin A in pregnancy in the UK. Very high intakes of retinol during pregnancy are **teratogenic**, i.e. they increase the risk of birth defects. Women in the UK are currently advised not to take vitamin-A-containing supplements during pregnancy and are advised to avoid eating liver or products containing liver (e.g. pâté) because of the high levels of retinol in animal livers (see COMA, 1991). American women are advised not to take supplements that contain more than the RDA of any vitamin (except folic acid).

It is assumed by the UK panel on DRV that most of the adult population obtain their vitamin D principally from endogenous production in the skin via the action of sunlight on 7-dehydrocholesterol, so there is no RNI for these groups. Pregnant women are one of the groups for which the panel felt that endogenous production could not be relied upon

and therefore an RNI was given (see Table 14.3). This RNI for vitamin D is over three times average UK intakes and therefore supplements are necessary to achieve it.

Alcohol and pregnancy

Fetal alcohol syndrome (FAS) is a recognizable clinical entity that is caused by heavy drinking during pregnancy. Babies with this syndrome have the following characteristics. They:

- are small
- have characteristic facial abnormalities
- are often mentally retarded
- are immunodeficient
- show slow postnatal growth.

This syndrome starts to occur with habitual alcohol intakes in excess of 50 g/day (four glasses of wine) and the frequency increases as alcohol intake rises. The alcohol probably exerts a direct toxic effect on the fetus and may also increase the risk of oxygen or glucose deficit to the fetus.

Moderate amounts of alcohol in pregnancy (habitual intakes of 10–50 g/day) may increase the risk of low birth weight. It may be prudent to advise women to limit their alcohol consumption during pregnancy to the occasional small amount, and the Royal College of Obstetricians and Gynaecologists advise that they should avoid alcohol altogether. Alcohol binges should certainly be avoided during pregnancy.

Heavy drinking prior to conception not only reduces fertility but also affects fetal growth, even if drinking stops after conception. The fetus is also most

- There is some increase in the dietary standards for several vitamins during pregnancy (see Table 14.3).
- Retinol is teratogenic, so pregnant women should avoid concentrated retinol supplements and in the UK are advised to avoid all liver, liver products, cod liver oil and any vitamin supplements that contain retinol.
- Dietary standards for vitamin D in pregnancy can only realistically be achieved by taking supplements, although for other adults it is assumed that they can manufacture enough if adequately exposed to sunlight.

- Heavy drinking during pregnancy has serious adverse effects upon the fetus – fetal alcohol syndrome.
- The effects of moderate and low alcohol intakes are less clear.
- Ideally, women should avoid alcohol during pregnancy and when planning to become pregnant.

vulnerable to the harmful effects of alcohol in the early stages of pregnancy. Taken together, such observations suggest that moderation of alcohol consumption should ideally begin prior to conception.

LACTATION

Lactation is nutritionally far more demanding for the mother than pregnancy. A woman will be producing up to 800 mL of milk, containing as much as 9 g of protein and 700 kcal (2.9 MJ) of energy, as well as all the other nutrients necessary to sustain the growing infant. The extra increments for lactation in the US and UK dietary standards are summarized in Table 14.5.

Table 14.5 Extra increments for lactation in the UK and USA dietary standards

Nutrient	UK RNI ^a	US RDA ^b
Energy (kcal) (MJ)	520 (27) ^b 2.18	500 (23) 2.09
Protein (g)	11 (24)	15 (30)
Vitamin A (µg)	350 (58)	500 (63)
Vitamin D (µg)	10 ^c	5 (100)
Thiamin (mg)	0.2 (25)	0.5 (43)
Riboflavin (mg)	0.5 (45)	0.5 (38)
Niacin (mg)	2 (15)	5 (33)
Folate (µg)	60 (30)	100 (56)
Vitamin C (mg)	30 (75)	35 (58)
Calcium (mg)	550 (79)	400 (50)
Iron (mg)	0	0
Zinc (mg)	6 (86)	7 (58)

^a EAR for energy.

^b Average for first 3 months.

^c No RNI for non-pregnant/lactating women.

Values in parentheses are percentages of the value for other women.

Sources: COMA (1991) and NRC (1989a).

The amounts by which the UK Estimated Average Requirements (EARs) for energy exceed those of other women during the first 3 months of lactation are:

month 1: 450 kcal/day (1.9 MJ)

month 2: 530 kcal/day (2.2 MJ)

month 3: 570 kcal/day (2.4 MJ)

After 3 months, the EAR depends upon whether breastfeeding is the primary source of nourishment for the baby or whether the baby is receiving substantial amounts of supplementary feeds. The US RDA for lactating women is increased by a set 500 kcal/day (2.1 MJ)/day.

The UK EARs have been estimated by measuring the energy content of milk production and assuming an efficiency of 80% for the conversion of food energy to milk energy. The contribution from the maternal fat stores laid down during pregnancy has been taken as 120 kcal/day (0.5 MJ/day). These increases in the EARs for energy during lactation are said by COMA (1991) to closely match the extra intakes observed in lactating women in affluent societies.

The RNI for protein is increased by 11 g/day throughout lactation of up to 6 months' duration. When this is expressed as a proportion of the EAR for energy, it does not rise significantly above that for other women, i.e. 9.3% of energy. As in pregnancy, it seems unlikely that protein will be a limiting nutrient in lactation because habitual intakes are usually well in excess of this.

The RNIs for eight vitamins and six minerals are also increased during lactation, reflecting the panel's judgement of the extra needs of women at various stages of lactation. Some of these increases are substantial proportions of the standard RNIs for adult women, and the increases in the US RDAs are generally even bigger (see Table 14.5).

Moderate weight loss of around 0.5 kg per week does not affect milk output, and changes in the proportions of fat, carbohydrate and protein in the

- Lactation is nutritionally demanding and this is reflected in substantial increases in the dietary standards for energy and many nutrients (detailed in Table 14.5).
- Successful lactation is compatible with moderate weight loss.
- The macronutrient and major mineral contents of the maternal diet have little effect upon milk composition.
- The fatty acid profile of milk reflects that of the maternal diet.
- Maternal intake and stores of the vitamins do affect milk composition.

maternal diet do not affect milk composition. The fatty acid profile of milk is affected by the fatty acid profile of the maternal diet. The concentration of the major minerals in milk is not affected by maternal intake; maternal stores are depleted if intake is inadequate. Maternal intake and the size of maternal stores affect the milk content of selenium, iodine and the vitamins. During periods of low maternal intake, milk vitamin content is initially maintained at the expense of maternal stores.

INFANCY

Breastfeeding versus bottle feeding

The assumption of most biologists would be that, through the process of natural selection, evolution will have produced, in breast-milk, a food very close to the optimum for human babies. Only in a few circumstances, such as those listed below, will breast-milk not be the preferred food for a baby:

- Babies with an inborn error of metabolism, which leaves them unable to tolerate lactose or some amino acid in milk protein.
- Babies born to women infected with agents such as the human immunodeficiency virus (HIV) virus that may be transmissible in the milk.
- Babies of mothers taking drugs that may be transmitted in the milk.

Prevalence of breastfeeding

In the 1960s, the majority of babies born in the UK would not have been breastfed after leaving hospital. Perhaps only a quarter of babies born in the USA and parts of Western Europe in the 1960s would have been breastfed for any significant time after delivery. Since then, there has been some increase in breastfeeding, led by women in the upper socioeconomic groups, the same groups that led the earlier trend away from breastfeeding.

Table 14.6 shows the estimated prevalence of breastfeeding in babies of different ages in England and Wales at 5 yearly intervals between 1975 and 1995. In 1995, just over two-thirds of women made some initial attempt to breastfeed their babies, but only 44% of babies were still being breastfed 6 weeks after delivery and, by 4 months, this had dropped to just over a

Table 14.6 Prevalence (%) of breastfeeding in England and Wales

	1975	1980	1985	1990	1995
Birth	51	67	65	64	68
1 week	42	58	57	54	58
2 weeks	35	54	53	51	54
6 weeks	24	42	40	39	44
4 months	13	27	26	25	28
6 months	9	23	22	21	22

Data sources DHSS (1988) and Foster *et al.* (1997).

- With few exceptions, breastfeeding is the best way of feeding newborn babies.

- About two-thirds of UK women start to breastfeed their babies, but a third of these give up within the first 6 weeks.
- The prevalence of breastfeeding in the UK rose substantially between 1975 and 1980, but current rates are still similar to those of 1980.
- By 6 weeks of age, many breastfed UK babies have received additional feeds and only around a fifth are being wholly breastfed.
- There is wide international variation in the prevalence of breastfeeding, and rates in Scandinavia are amongst the highest in the industrialized world – 90% of Norwegian babies are still receiving some breast-milk at 3–4 months of age.

quarter of all babies. These low figures nevertheless represent major increases over the figures for 1975, but there have been no sustained improvements since 1980. The substantial improvement shown between 1975 and 1980 has not been continued. The figures in Table 14.6 actually underestimate the use of bottle feeds, because around 46% of breastfed babies received some additional bottle feeds at 6 weeks of age and, by 6–10 weeks, only 21% of babies were being wholly breastfed (Foster *et al.*, 1997).

Most of the statistical data used in this discussion of infant feeding practices are for England and Wales (or for the UK), but it is intended to illustrate principles, patterns and trends that are likely to be common to several industrialized countries. The prevalence of breastfeeding in Scotland and Northern Ireland is substantially below that in England and Wales. In 1995, only 55% of Scottish babies were initially breastfed and only 36% were still being breastfed at 6 weeks (corresponding figures for Northern Ireland 45% and 25%, respectively). In *Healthy people 2000* (DHHS, 1992), the goal is for 75% of US mothers to be breastfeeding in the immediate post-partum period by the year 2000 and for 50% to continue to breastfeed for 5–6 months. Assumed baseline figures for 1988 were that just 54% of mothers were breastfeeding on discharge from hospital and 21% were breastfeeding at 5–6 months. Only 32% of low-income US mothers breastfed their babies at the time of hospital discharge and less than 10% were still breastfeeding at 5–6 months. These figures are comparable to the 1990 figures for England and Wales. Much higher breastfeeding rates have been reported for Canada, the Netherlands, Scandinavia, New Zealand and parts of Australia. In Sweden in 1973, only 20% of babies were being breastfed at 2 months but, by 1993, this had risen to 85%, which shows what

can be achieved by effective promotion and facilitation measures. According to Laurent (1993), 99% of Norwegian mothers breastfeed their babies at birth and 90% are still at least partially breastfeeding at 3–4 months. International data on breastfeeding and the primary references can be found in Wharton (1998) and King and Ashworth (1994).

Factors influencing choice of infant feeding method

Choice of infant feeding method is very strongly dependent upon social class in the UK. In 1995, 90% of women in social class 1 attempted to breastfeed their babies and 73% were still breastfeeding at 6 weeks. The corresponding figures for women in social class 5 were 50% and 23%, respectively. Women with higher levels of education are also more likely to breastfeed their children in Britain. The mother's choice of feeding method for her first baby is a major influence on feeding of subsequent children. Mothers who had breastfed their first babies usually breastfed subsequent children, whereas women who had bottle fed the first baby usually did not. Women with previous experience of successful breastfeeding were also less likely to stop prematurely (Foster *et al.*, 1997).

Many factors will have compounded to decrease the prevalence of breastfeeding, but the availability of a cheap, simple and 'modern' alternative was clearly an essential prerequisite. There may have even been a temptation to think that science had improved upon nature if bottle-fed babies grew faster than breastfed babies. Some other factors that have probably contributed to the decline in breastfeeding are listed below.

- Breastfeeding may be inconvenient to the lifestyle of many women, especially those who wish to resume their careers soon after birth or those whose income is considered to be a vital part of the family budget. Statutory paid maternity leave should have decreased this influence in more recent years in the UK. According to Foster *et al.* (1997), a third of UK women who stopped breastfeeding between 3 and 4 months gave returning to work as their reason. Only a third of women in this survey were on paid maternity leave when their babies were 6 weeks old.
- Breastfeeding prevents fathers and others sharing the burden of feeding, especially night feeds in the early weeks after birth.
- There are undoubtedly some women (and men) who find the whole idea of breastfeeding distasteful. There is a strong taboo in this country against breastfeeding in public or even amongst close friends or relatives. This, coupled with the poor provision of facilities for breastfeeding in many public places, can only increase the inconvenience of breastfeeding and restrict the range of activities that women can participate in during lactation.
- The figures in Table 14.6 clearly indicate that many women who initially try to breastfeed give up after a very short time. Many women experience discomfort in the early stages of lactation and many believe that they are unable to produce sufficient milk. Delay in starting suckling (e.g. to allow the woman to rest after delivery) or any prolonged break from suckling is likely to reduce the likelihood of initiating and maintaining lactation.
- Breastfeeding is promoted in many Third World countries as the modern, sophisticated, 'Western' way of feeding babies. It is promoted as the high prestige option. Yet, in industrialized countries, breastfeeding is more common amongst the higher socioeconomic groups and so in these countries is, in practice, the higher prestige option.

In a 1995 survey of British women (Foster *et al.*, 1997), the following were given by women as important reasons for choosing to breastfeed their babies:

- 83% because they thought this would be best for the baby
- 37% because breastfeeding is more convenient
- 33% of second or later births because of the mother's previous experience

- 20% because it promotes bonding between mother and baby
- 21% because it is cheaper
- 12% because it is more natural
- 12% because it is good for the mother.

Listed below are the most common reasons given for choosing to bottle feed:

- 36% because someone else can feed the baby
- 34% did not like the idea of breastfeeding or would be too embarrassed to breastfeed
- 44% of second or later babies because of the mother's previous experience.

The most common reasons given for stopping breastfeeding within the first week were:

- lack of milk or failure of the baby to suckle
- the discomfort caused by breastfeeding or dislike of breastfeeding
- too time consuming or too tiring (note that 90% of feeding should occur within the first 5 minutes of suckling)
- the mother was ill.

The most common reasons given by women for stopping breastfeeding between the third and fourth month were:

- lack of milk or failure of the baby to suckle
- the mother returning to work
- too time consuming or too tiring
- the mother was ill.

In a 1990 survey, 44% of breastfeeding women reported that they had experienced problems finding somewhere to feed their babies in public places. Almost a quarter of all women with unweaned children, both breastfed and bottle fed, reported that they only went out between feeds (White *et al.*, 1992).

Early feeding experiences in the hospital also strongly influence the likelihood of early cessation of breastfeeding. Mothers attempting to breastfeed but whose babies were also given some bottle feeds within the first week were more than three times as likely to give up breast feeding within a fortnight, as were women whose babies received no bottle feeds. Delay in initiating breastfeeding is also strongly associated with early cessation. Early cessation of breastfeeding occurred in 30% of cases where there was a delay of 12 hours or more in putting the baby to the breast,

but in only 14% of cases where this occurred immediately after birth; even a delay of 1 hour had a significant effect (Foster *et al.* 1997). White *et al.* (1992) reported that women were almost twice as likely to continue breastfeeding if feeding on demand rather than at set times was practised in the hospital.

Foster *et al.* (1997) noted that there had been some improvements in practices in UK maternity hospital between 1990 and 1995. For example:

- The number of women starting to breastfeed within 1 hour had increased from 63% in 1990 to 68% in 1995.
- The proportion of mothers who had their babies with them continuously in hospital rose from 63% to 74%.
- The proportion of breastfed babies who were given some bottles in hospital dropped from 45% to 36%.
- In 1995, most first-time mothers said that they had been given some useful help or advice about breastfeeding their babies.

The benefits of breastfeeding

Table 14.7 illustrates just how different are human milk, cow milk and infant formula. Cow milk is an

ideal food for rapidly growing calves but, unless substantially modified, is now regarded as an unsuitable food for human babies. Unmodified cow milk is now not recommended for babies until they are at least 6 months old. The table also demonstrates the scale of changes that are necessary to convert cow milk into a formula that is compositionally similar to human milk. Even though modern 'humanized' formula is closer to human milk than formulations such as National Dried Milk that were used up until 1976 in the UK, they are still quite different from the human milk that they are trying to imitate.

Cow milk has around three times the protein content of human milk. This means that babies will need to excrete substantial amounts of urea if fed on cow milk, whereas when breastfed, the bulk of the nitrogen is retained for growth. Around 20% of the nitrogen in cow milk and breast milk is non-protein nitrogen, e.g. urea, creatine and free amino acids; its biological significance is unclear, but this fraction will be absent from infant formula. The amino acid **taurine** is present in relatively large amounts in human milk and is added to some infant formula. In breastfed infants, the principal bile acid is taurocholic acid, but in infants fed on cow milk it is glycocholic acid and this may adversely affect the emulsification and therefore digestion and absorption of milk fat.

- Breastfeeding in Britain decreases in prevalence with decreasing social class and with decreasing educational attainment of the mother.
- Women choose to breastfeed because it is good for the baby and for them, because it promotes bonding between mother and baby and because it is convenient, natural and cheap.
- Women choose to bottle feed their babies because someone else can help with feeding or because they do not like the idea of breastfeeding or are embarrassed by it.
- The feeding method used for second and subsequent births is markedly influenced by the mother's experience with her first baby.
- Women often give up breastfeeding prematurely because they have difficulties successfully and comfortably feeding the baby.
- Returning to work becomes a more important factor in stopping breastfeeding as time after the birth increases.
- Delay in first putting the baby to the breast and the use of supplementary bottle feeds are both associated with early cessation of breastfeeding.
- Hospital practices in the UK are changing so as to minimize these apparent hindrances to successful lactation and most women are now given support and guidance in breastfeeding.

Table 14.7 The composition of human milk, cow milk and modern infant formula

	Human milk	Cow milk	Infant formula	1970s Formula^a
Energy				
(kcal)	70	67	67	55
(kJ)	293	281	281	231
Protein (g)	1.3	3.5	1.4–1.9	2.5
Carbohydrate (g)	7	4.9	7.2–7.7	5.9
Fat (g)	4.2	3.6	3.3–3.6	2.4
Sodium (mg)	15	52	19–25	40
Calcium (mg)	35	120	51–71	83
Iron (μg)	76	50	670–680	480
Vitamin A (μg)	60	40	80	23
Vitamin C (mg)	3.8	1.5	8	2.4
Vitamin D (μg)	0.01	0.03	1.1–1.2	0.8

^a UK National Dried Milk that was withdrawn in 1976.

All values are per 100 mL as consumed.

Data sources Poskitt (1988) and Wharton (1998).

The principal proteins in milk are **casein** and **whey**. The casein:whey ratio is much higher in cow milk than in human milk. Casein forms hard clots in the stomach and is also relatively low in cysteine (the usual precursor of taurine). Infant formulae vary in the extent to which the casein:whey ratio has been modified to make it closer to that of human milk. Up to 6 weeks of age, **casein-dominant formula** was used by 60% of UK mothers and **whey-dominant formula** by 37% (Foster *et al.*, 1997). In casein-dominant formula the casein:whey ratio is 80:20, as in cow milk, but in whey-dominant formula it is closer to the 40:60 ratio in human milk. In older infants, the use of casein-dominant formula became more prevalent, perhaps because of unsubstantiated claims by some manufacturers that a casein-dominant formula is more satisfying than a whey-dominant formula. Follow-on formulae are produced as an alternative to cow milk for infants who are over 6 months; they are not intended to be a complete diet but a component of a weaning diet.

Cow milk fat is poorly digested and absorbed compared to that in breast milk; it is also much lower in essential polyunsaturated fatty acids. Some or all of the butterfat will be replaced by vegetable oils by infant formula manufacturers to give a fatty acid composition closer to that of human milk. Cow milk has only around 1% of energy as linoleic acid, whereas the figure for human milk is up to 16%, depending upon the diet of the mother. The differences in

digestibility of cow and breast milk fat may be due to the three factors listed below.

1. Differences in the nature of the fat itself; palmitic acid is generally not well absorbed by infants, but in human milk most of it is in the 2 position and is absorbed as monoacylglycerol.
2. Breast-milk contains lipase, which is activated by bile salts: lingual lipases make a major contribution to fat digestion of bottle-fed and breastfed babies.
3. The differences in bile acid secretion mentioned earlier may contribute to the better fat digestion by breastfed babies.

Breast-milk is particularly high in lactose and this lactose content may increase calcium absorption and act as a substrate for *Lactobacilli* in the lumen of the gut, thus creating an acid environment that will inhibit the growth of pathogens. Breast-milk contains factors that stimulate the growth of *Lactobacillus bifidi* – small amounts of oligosaccharides and nucleotides may be important in this respect. The stools of healthy breastfed babies have a low pH and contain almost exclusively *Lactobacilli*, whereas formula-fed babies have a more mixed gut flora. Formula manufacturers increase the carbohydrate content so that it corresponds to that of human milk – they may or may not use lactose for this purpose. Oligosaccharides of glucose produced by partial hydrolysis of starch are often used.

Cow milk contains much higher concentrations of the major minerals than human milk. The sodium content is over three times higher; this is a high solute load for the immature kidney to deal with and it can precipitate **hypernatraemic dehydration**. The risk of dehydration would be exacerbated by fever, diarrhoea or by making up the formula too concentrated. High sodium content was the single most important reason why, in the UK in 1976, several formulae, including National Dried Milk, were withdrawn.

The calcium contents of cow milk and breast-milk differ by even more than the sodium content, but calcium uptake is homeostatically regulated – calcium absorption from cow milk may be poor despite the high concentration. Unabsorbed fat from cow milk or formula hinders calcium absorption, whereas the presence of lactose increases it.

The vitamin and trace mineral content of formula is generally set so as to match or exceed that in human milk. The iron content of formula is higher than in breast-milk, even though it is normal for breastfed babies to be in negative iron balance for the first few months of their lives. The iron content of breast-milk is considered insufficient for older babies (4–6 months). All formulae are fortified with vitamin D.

Table 14.7 shows that formula is still very different from human milk. Even if manufacturers had succeeded in producing a formula that exactly corresponded to breast-milk in its nutrient content, there would still be substantial residual benefits to be gained from breastfeeding. In many Third World countries, the choice of feeding method may greatly affect the chances of survival of the baby. Mortality rates of bottle-fed babies are often much higher than those of breastfed babies (Walker, 1990). Even in industrialized countries, there are suggestions that deaths from sudden infant death syndrome are lower in breastfed babies (see Laurent, 1993). Breastfeeding can reduce the risk of infection, both by preventing exposure to contaminated alternatives and because of the inherent anti-infective properties of breast-milk. Although the former may be the dominant benefit in some developing countries, there is evidence that it also reduces gut and respiratory infection rates in industrialized countries and that the latter mechanisms are predominant here (Filteau and Tomkins, 1994).

Some additional benefits of breastfeeding are discussed below.

- **Bonding.** Breastfeeding is thought to increase bonding between mother and baby.
- **Anti-infective properties.** Breast-milk contains immunoglobulin A, which protects the mucosal surfaces of the digestive tract and upper respiratory tract from pathogens. It also contains immune cells, which aid in the immunological protection of the infant gut. The iron-binding protein lactoferrin has a bacteriostatic effect. The enzyme lysozyme breaks down bacterial cell walls and oligosaccharides and nucleotides promote the growth of *Lactobacilli*. **Colostrum**, the secretion produced in the first few days of lactation, is particularly rich in these anti-infective agents. Colostrum is quite different in its composition from mature human milk. Anti-infective agents in cow milk will probably be destroyed during processing and many are in any case likely to be species specific.
- **Hygiene.** It is taken for granted in most industrialized countries that infant formula will be made up with clean, boiled water, in sterile bottles, and the final mixture pasteurized by being made up with hot water. The risk of contamination in many other countries is a very serious one – the purity of the water, facilities for sterilization of water and bottles and storage conditions for pre-mixed feeds may all be poor. This means that bottle-fed babies may have a greatly increased risk of gastrointestinal infections. Gastroenteritis is the single most common cause of death for babies in developing countries. This is a major reason why breastfeeding should be particularly encouraged in these countries. This is an important cause of the higher mortality of bottle-fed babies in developing countries.
- **Restoration of maternal pre-pregnant weight.** It is generally assumed that maternal fat laid down during pregnancy serves as a store to help meet the energy demands of lactation upon the mother. In countries like the UK and USA, where obesity is prevalent, the assumed effect of lactation in reducing this maternal fat after delivery may be an important one.
- **Cost of formula.** In wealthy, industrialized countries, the cost of formula will usually be relatively trivial. Nevertheless, modern infant formula is a high-technology product and in many Third World countries feeding an infant with formula will absorb a very substantial proportion of the total household budget. This may seriously reduce

resources available to feed the rest of the family or it may lead to the baby being fed over-diluted formula as an economy measure, effectively starving it.

- **Anti-fertility effect.** Although breastfeeding cannot be regarded as a reliable method of contraception, it does reduce fertility. In cultures in which suckling occurs for extended periods rather than in discreet 'meals', this effect may be increased. In many Third World countries, prolonged breastfeeding will have an important birth-spacing effect and will make a contribution to limiting overall population fertility.
- **Reduced risk of breast cancer.** A large case-control study in the UK has suggested that breastfeeding reduces the risk of developing breast cancer in young women. Breast cancer risk declined with both the duration of breastfeeding undertaken by the woman and the number of babies that she had breastfed (UK National Case-Control Study Group, 1993).

Weaning

WHEN TO WEAN?

Weaning is the process of transferring an infant from breast-milk or infant formula onto solid food.

The process begins when the child is first offered food other than breast-milk or formula. The process may be very gradual, with other foods forming an increasing proportion of total energy intake over several months until breast or formula is eventually phased out completely. During weaning, there is ideally gradual transition from a very high-fat, high-sugar, liquid diet to a starchy, moderate-fat, low-sugar and fibre-containing, solid diet. The magnitude of the compositional and textural changes involved would seem to indicate the advisability of a gradual transition to allow babies to adapt to them.

In the report 'Present day practice in infant feeding' (COMA, 1988), it was concluded that very few infants require solid foods before 3 months of age, but that the majority require them by 6 months of age. It was recommended in this report that weaning should not begin before 3 months, but that infants should be offered a mixed diet by 6 months of age. After about 6 months, it is thought that breast-milk can no longer supply all of the nutritional needs of the infant and growth is likely to be impaired if the baby receives only breast-milk. Breast-milk or formula may continue to make a contribution to total food supply long after weaning has begun.

The COMA panel considered too early introduction of solid foods undesirable because:

- Breast-milk is very different in composition from either cow milk or even modern infant formula.
- The casein:whey ratio is higher in cow milk and formula can either be casein dominant or whey dominant, like breast-milk.
- In infant formula, some of the highly saturated butterfat of cow milk is replaced by vegetable oil.
- Human milk fat is much better digested and absorbed than that in cow milk.
- Breast-milk has more lactose than cow milk, and infant formula is supplemented with carbohydrate, which may not be lactose.
- Cow milk is high in sodium and some early infant formulae were withdrawn in 1976 because of their high sodium content.
- Mortality and infection rates of breastfed babies are much lower than those of bottle-fed babies in developing countries.
- Even in industrialized countries, breastfed babies have lower rates or severity of infection.
- Breastfeeding reduces the risk of infection because it reduces exposure to contaminated alternatives and because of its inherent anti-infective properties.
- Infant formula is an expensive commodity for mothers in developing countries.
- Breastfeeding may decrease fertility, help the mother to regain her pre-pregnant weight and reduce the mother's risk of breast cancer.

- some babies do not properly develop the ability to bite and chew before 3–4 months
- the infant gut is very vulnerable to infection and allergy
- the early introduction of energy-dense weaning foods may increase the likelihood of obesity.

It has been suggested that full production of pancreatic amylase does not occur in human infants until 6–9 months of age. As milk contains no starch but most weaning foods are starchy, this may be a physiological indicator that a relatively late introduction of starches into the diet is desirable. Introducing starchy solid foods too early may produce symptoms similar to those of infectious gastroenteritis because of poor digestion and absorption due to the lack of pancreatic amylase (COMA, 1988).

In 1975, 18% of British babies had been given some food other than milk in their first month and 97% by the time they were 4 months old; corresponding figures for 1995 were 2% and 91%, respectively (COMA, 1988; Foster *et al.*, 1997). The practice of introducing solid foods to infants very soon after birth is now much less common. Nevertheless, 55% of British babies still receive solid food before the 3 months suggested by the COMA panel, although this is a significant improvement on the 68% recorded in 1990 (Foster *et al.*, 1997).

More recently, Morgan (1998) has suggested that 14 weeks should be the minimum age for weaning or when the infant's weight reaches 7 kg.

WHAT WEANING FOODS?

In the UK, the first foods for most babies are cereals, rusks or commercial weaning foods rather than home-prepared weaning foods (COMA, 1988). At 6 weeks, only 4% of mothers who had offered solid food to their babies had used a home-made food and even at 4 months, when over 90% had offered some solid food, only about a quarter of them had used home made food (Foster *et al.*, 1997).

THE PRIORITIES FOR WEANING FOODS

Morgan and her colleagues asked a large sample of British mothers for their views on good infant feeding practices and some of the findings of this survey are summarized in Table 14.8.

Clearly, most of these women think that a high fibre and low-fat diet should be the priority for

Table 14.8 *The priorities of a sample of UK mothers for infant feeding*

	Percentage of mothers' responses Not important	Percentage of mothers' responses (Very) important
Wide variety of foods	3	95
Plenty to drink	1	98
Plenty of calories	20	76
Low-fat intake	10	88
High-fibre intake	15	83

Adapted from Morgan (1998).

weaning foods and a fifth do not think that an energy-rich diet is important for rapidly growing infants. These maternal priorities are very different from the current consensus of expert nutritional opinion. For example, an expert group in the UK (COMA, 1994b) considered that:

- the provision of adequate dietary energy for growth is the principal determinant of the diets of the under fives
- adult recommendations on reducing fat consumption should not apply to the under twos and should fully apply from the age of 5
- the diets of the under twos should not contain fibre-rich foods to the extent that they limit energy content and the absorption of micronutrients.

In a recent review, Poskitt (1998) suggested that weaning foods should be:

- rich in energy, including energy from fat
- rich in vitamins and minerals
- fed frequently
- initially used as a supplement to milk rather than a replacement
- fed in a form that develops the child's feeding skills whilst still allowing assistance.

One of the main aims of weaning is to raise the energy density of the infant's diet above that for breast-milk. The weaning food should have a suitable texture, but be of high enough energy and nutrient density for the baby to meet its nutritional needs without having to consume an excessive volume of food. If very viscous food is introduced too early in the weaning process, the infant may reject it by spitting it out. A typical Third World weaning food made up to give a suitable viscosity from a starchy cereal or root staple, such as cassava or millet flour, might

contain only 0.3 kcal/g (1.3 kJ/g). This compares with around 0.7 kcal/g (3 kJ/g) for breast-milk and perhaps 1.5 kcal/g (6 kJ/g) for a typical UK weaning diet (Church, 1979). At the lower extreme, the child is incapable of consuming the volume of food required to meet its energy needs. This problem may be exacerbated if the child is fed infrequently, has frequent periods of infection and anorexia, and perhaps by the poor sensory characteristics of the food itself. In industrialized countries, this could be a problem if parents mistakenly apply the recommendations for low-fat, low-sugar and high-starch diets in adults too rigorously to infants. Some strict vegetarian weaning diets in industrialized countries may also be of insufficient energy density because of their low fat and high starch content. The lower limit for the energy density of weaning foods should be 0.7 kcal/g (3 kJ/g). Pureed fruit or vegetables are not suitable as weaning foods unless they have their energy density enhanced, e.g. by the addition of a source of fat. Skimmed milk and semi-skimmed milk are not suitable for very young children.

Weaning foods should be clean and not contaminated with infective agents. Poverty, poor hygiene and contaminated food precipitate much malnutrition in the Third World. Even when dietary intakes are judged sufficient to permit normal growth, infection and diarrhoea may be indirect causes of dietary

deficiency. One survey reported that 41% of traditional weaning foods and 50% of drinking water specimens in rural Bangladesh were contaminated with faecal micro-organisms (see Walker, 1990).

In affluent countries, such as the UK and USA, other aims are also considered important for infant feeds. They should be:

- low in salt
- low in added sugar
- perhaps gluten free.

High-salt foods expose the immature kidney to a high solute load, increase the risk of hypernatraemic dehydration and may increase the later risk of high blood pressure. Sugar is regarded as empty calories and is detrimental to the baby's new teeth. Over-consumption of sugar in infancy may also be creating bad preferences for the future. There is very strong evidence that fluoride is protective against dental caries. The UK panel on DRVs suggested a safe fluoride intake for infants of 0.05 mg/kg/ per day – around 50% of the amount likely to cause fluorosis. To achieve this safe intake, most UK infants would need supplements. Swallowing fluoridated toothpaste is one way that many young children receive supplemental fluoride.

Whereas most babies will suffer no harm from early exposure to the wheat protein called gluten,

- Weaning should start some time between 3 and 6 months after birth.
- Breast-milk is not adequate for babies after 6 months, but can usefully continue to contribute to their diets.
- Very early introduction of solid foods is undesirable and is now much less common in the UK than it was.
- Mothers in Britain usually use some form of commercial weaning food in the early stages of weaning.
- Weaning foods should be free from contamination and should be energy and nutrient dense.
- Weaning foods should not be low in fat or too high in fibre and they should be low in salt and added sugar.
- Babies should be fed frequently during weaning.
- Many British mothers mistakenly believe that weaning foods should be low in fat and high in fibre.
- Pureed fruit and vegetables are unsuitable weaning foods unless their energy density has been increased.
- In many developing countries, the traditional starchy weaning foods have such low energy density that they limit energy intake and precipitate malnutrition.

those sensitive to gluten and thus at risk of coeliac disease (see Chapter 15) cannot be identified in advance. The incidence of coeliac disease in children has been falling in recent years, at the same time as there have been trends towards later introduction of solid foods and towards the use of gluten-free, rice-based weaning cereals.

CHILDHOOD AND ADOLESCENCE

Even when fully weaned, children still have nutritional needs and priorities that differ from those of adults. They need enough energy and nutrients to enable them to grow and develop, especially during the intensely anabolic adolescent period. Too heavy an emphasis on a low-fat, low-sugar and high-fibre diet may limit growth. These differing needs have to be reconciled with the need to start good dietary habits early and to prevent the development of obesity. Note that inactivity is now regarded as the key factor in childhood obesity (see Chapter 8).

Table 14.9 shows selected UK RNIs for children in the 1–3 and 4–6-year-old category. The values have

also been expressed as an approximate percentage of the value for adults. These values show that children require proportionately much more energy than adults. A 1–3-year-old girl weighs about a fifth of an adult woman, but requires three-fifths as much dietary energy as the woman. The values for the other nutrients are generally in line with this proportionately higher energy requirement (much less in the case of protein). A diet that is compositionally suitable for adults should therefore contain adequate amounts of nutrients for young children provided they eat enough of it to satisfy their energy needs. Of course, children may be more sensitive to the effects of nutrient deficiencies as, for example, they may have smaller stores than adults or their growth may be impaired.

In a survey of the diets of a representative sample of UK pre-school children (1.5–4.5 years old), Gregory *et al.* (1995) made the following observations.

- Energy intakes of the children were less than the EARs, which is probably because the EARs are set too high.
- The proportions of food energy contributed by the major nutrients was 13% from protein, 51% from carbohydrate and 36% from fat.

Table 14.9 Selected Dietary Reference Values (per day) for UK children and their percentage of the corresponding adult values

Nutrient	1–3 years	Percentage of adult ^a	4–6 years	Percentage of adult ^a
Boys' energy (kcal)	1230 (5.15 MJ)	48	1715 (7.16 MJ)	67
Girls' energy (kcal)	1165 (4.86 MJ)	60	1545 (6.46 MJ)	86
Protein (g)	14.5	29	19.7	40
Thiamin (mg)	0.5	55	0.7	78
Riboflavin (mg)	0.6	50	0.8	50
Niacin (mg)	8	53	11	73
Folate (µg)	70	35	100	50
Vitamin C (mg)	30	75	30	75
Vitamin A (µg)	400	62	500	77
Vitamin D (µg) ^b	7	—	—	—
Calcium (mg)	350	50	450	64
Iron (mg)	6.9	60	6.1	52
Zinc (mg)	5.0	61	6.5	79

^a In most cases, this is calculated using the average of the male and female value for 19–50 year olds.

^b For adults and older children there are no reference values as it is synthesized in the skin when exposed to sunlight. Note that the body weight of a 1–3 year old is about 20% of that of an adult, and that of a 4–6 year old is about 30% of the adult value.

Data source COMA (1991).

- Within the major macronutrient categories, sugars provided 29% of energy, non-milk extrinsic sugars 19% and saturated fat 16%.
- Average intakes of most vitamins and minerals were generally well above the RNI, except for vitamin A, iron and zinc.
- About half of the children had intakes of vitamin A that were below the RNI and around 8% below the Lower Reference Nutrient Intake (LRNI).
- A large majority of children had iron intakes below the RNI and a fifth of the very young ones had intakes that were below the LRNI.
- Inadequate iron intakes in many of the sample were confirmed by blood analyses, which showed that 10% were anaemic and 20% had less than satisfactory iron stores (serum ferritin below 10 µg/L).
- Intakes of vitamin D were under 2 µg/day and this is clearly an insufficient amount if children do not get adequate exposure to summer sunlight.

Whereas the macronutrient composition of the diets of UK pre-school children may be close to the guideline values for adults, the levels of added sugars and saturated fat are high. More than half of the added sugar in the diet came from soft drinks and sweets. These are practically devoid of nutrients, lower the overall nutrient density of the diet and are damaging to children's teeth. Milk and cheese contributed about half of the vitamin A in children's diets. Liver, when eaten, made a major contribution to average daily vitamin A intake. Fruits and vegetables together only accounted for 16% of total vitamin A intake. Cereals, which are often fortified, contributed about half of the children's iron intake and meat and fish only about 20%. Milk and milk products provided about 40% of the saturated fat and meat and fish about 15%. Biscuits, cakes, chips and savoury snacks provided 20% of the saturated fat.

The message from this survey is what most parents probably would have guessed and many are trying to implement. Children's' diets would be improved if they consumed less sugary drinks, sweets, cakes, biscuits and chips (fries) and ate more fruit, vegetables, lean meat, fish, cereals and boiled potatoes.

The rate of growth of children decelerates steadily from birth until puberty. During adolescence, there is a very pronounced and sustained growth spurt. Between the ages of 12 and 17 years, boys gain an

average 26 cm in height and 26 kg in weight; girls gain 23 cm in height and 21 kg in weight between the ages of 10 and 15 years. Adolescence is thus an intensely anabolic period and a time when there is inevitably a relatively high demand for energy and nutrients to sustain this rapid growth. In a proportion of adolescents, high levels of physical activity because of participation in games and sports will still further increase their needs for energy and perhaps other nutrients. Adolescence is also a time of major psychological changes, brought on by the hormonal changes of puberty, and these may have major effects on children's attitudes to food. Adolescent girls are the most common sufferers from anorexia nervosa and, paradoxically, obesity is also a common nutritional problem of adolescence.

The increasing secretion of sex hormones at puberty produces a major divergence in the body composition of boys and girls. Male sex hormones cause boys to gain more muscle than fat and to adopt a more abdominal distribution of their body fat. Female sex hormones cause girls to gain more fat than muscle and to increase their deposition of fat in the hip and thigh region. This increase in fatness, which is a natural part of female adolescence, may be an important contributor to the unhealthy preoccupation with body image and dieting experienced by many adolescent girls. Many adolescent girls perceive themselves as too fat when they are not and diet when it is inappropriate. They sometimes adopt unhealthy dietary and other strategies for losing weight, e.g. fasting, avoidance of meat and other staple foods and even smoking (e.g. Ryan *et al.*, 1998). Girls need to be better prepared for these changes and to be reassured that they are a normal consequence of maturation.

As with rapidly growing infants, very low energy-density diets (e.g. diets very low in fat and high in fibre) may tend to limit growth during adolescence. Strict vegetarian diets tend to be bulky and have low energy density; this is, indeed, often cited as one of their advantages. Vegan children tend to be smaller and lighter than other children (Sanders, 1988). The phrase 'muesli belt malnutrition' has been widely used in the UK to describe undernutrition in rapidly growing children precipitated by the overzealous restriction of energy-dense foods by health conscious, middle-class parents.

During the 5-year period of adolescence, boys accumulate an average of around 200 mg/day of

calcium in their skeleton, with a peak value of about double this average. They also accumulate an average of 0.5 mg/day of iron and 2 g/day of protein during the 5 years of adolescence. The onset of menstruation represents a further major stress on the iron status of adolescent girls. Table 14.10 shows some DRVs for 11–14 year olds, 15–18 year olds and adults. The energy EAR for 15–18-year-old boys is 8% higher than that for adult men, despite the smaller size of the boys. The calcium RNI is 43% higher for both age groups of boys than that for men and clearly reflects the DRV panel's assumptions of the extra calcium required for skeletal growth. This calcium RNI would be difficult to meet without drinking good amounts milk. Similar relatively high DRVs for girls compared to women are also shown in the table. Similar trends are also seen in the US RDAs: the relative allowance (i.e. allowing for size) is higher in adolescents than in adults. American values for all groups tend to be higher than the corresponding British values. Note particularly that the calcium RDAs for both ages of boys and girls are 1200 mg/day (c.f. adult value 800 mg/day); the calcium RDAs for American girls are thus 50% higher than the corresponding UK values.

COMA (1989b) published a survey of the diets of a nationally representative sample of British girls and boys at 10–11 years and 14–15 years. They concluded that the main sources of energy in the diets

of British children were bread, chips (french fries), milk, biscuits (cookies), meat products (e.g. sausages, burgers, meat pies), cakes and puddings. These foods together accounted for about half of the total energy intake. Energy intakes of the two age groups of girls were not significantly different, whereas intakes of the older boys were 20% higher than those of the younger boys. Mean recorded energy intakes were within 5% of the EAR for all groups when these EARs were interpolated from the actual mean body weights of the groups using figures in COMA (1991). Given the tendency for dietary surveys to underestimate total intakes and the satisfactory heights and weights for ages of the population, energy intakes would seem to be adequate to meet their needs.

Fat made up about 38% of the total energy intake of these British children, very similar to the adult figure recorded by Gregory *et al.* (1990). Three-quarters of all children exceeded the target of 35% of food energy from fat. Surprisingly, meat and meat products contributed less than 15% of the total fat, marginally less than that contributed by chips (french fries) and crisps (potato chips).

Average intakes of all four groups of children exceeded the current RNIs for most of the nutrients surveyed, i.e. for protein, vitamin A, vitamin C, nicotinic acid, riboflavin and thiamin. Pyridoxine intakes were at or marginally below current RNIs. In the younger children, calcium intakes were well above the RNIs for 10 year olds but below those of the 11–14-year age bands. Average calcium intakes of the older children were about 10% below current RNI. Milk and cheese contributed 40% or more of total calcium intakes for all groups. Milk also contributed about 30% of riboflavin intakes. Iron intakes of the older girls were only 62% of the current RNI. Intakes of the younger girls met the RNI for 10 year olds, but represented only 58% of the 11–14-year old figure, which should presumably apply from the start of menstruation.

The overall impression from this survey of British children is that their diets meet most current criteria for adequacy, although there are a few areas for concern, such as the iron intakes of the adolescent girls. In terms of current health education priorities, particularly dietary fats, the diets of children seem to be very similar to those of their parents. These conclusions are as one might expect, and the general conclusions would probably apply to most other industrialized countries.

Table 14.10 A comparison of some UK Dietary Reference Values for adolescent boys and girls with those for adults

	Age group		
	11–14 years	15–18 years	Adult
Energy EAR (kcal)	Male	2200	2755
	Female	1845	2110
Energy EAR (MJ)	M	9.21	11.51
	F	7.92	8.83
Calcium (mg)	M	1000	1000
	F	800	800
Iron (mg)	M	11.3	11.3
	F	14.8	14.8
Zinc (mg)	M	9.0	9.5
	F	9.0	7.0
Niacin (mg)	M	15	17
	F	12	13

Data source: COMA (1991).

- Children need enough dietary energy and nutrients to allow them to grow and develop properly.
- Young children need proportionately much more energy than adults and so good energy density is still an important priority.
- Inactivity rather than overeating is the major factor in child obesity.
- Diets that are adequate for adults should be adequate for young children if they meet their energy requirements.
- The average diets of young British children are high in added sugar but low in vitamin A, iron and zinc.
- 20% of young British children had biochemical evidence of poor iron status.
- Young British children need regular exposure to sunlight to achieve satisfactory vitamin D status.
- British schoolchildren and pre-schoolchildren should consume less sugary drinks, sweets, cakes, biscuits and chips but more fruit, vegetables, lean meat, cereals and potatoes.
- Adolescence is an intensely anabolic period and there are substantial increases in body weight, body protein and body calcium content.
- At puberty, boys increase their lean:fat ratio.
- Female sex hormones cause a substantial increase in the body fat content of girls, particularly in fat around the hips and thighs.
- Girls who are unprepared for the pubertal increases in body fat may be encouraged to diet inappropriately and to adopt unhealthy practices to try to lose weight.
- The average diets of British schoolchildren are similar in their macronutrient composition to those of their parents and contain adequate amounts of most essential nutrients.
- The calcium intakes of older children and, more especially, the iron intakes of adolescent girls are less than satisfactory.

THE ELDERLY

Demographic and social trends

Much of this section is a summary of material in the first chapter of Webb and Copeman (1996).

Table 14.11 gives an approximate breakdown of the proportion of the population of England and Wales in various age bands. The proportion of elderly people in the populations of industrialized countries rose substantially during the twentieth century and is projected to continue rising. By the end of the 1990s, around 16% of the population of England and Wales was over 65 years and almost 2% were over 85 years of age. At the start of the twentieth century, life expectancy at birth was around 47 years in both the UK and USA, but had risen to over 76 years by the end of the century. In 1901, only 4.7% of the population of England and Wales was over 65 years. The increase in life expectancy during the twentieth century has not

been confined to the younger age groups. Between 1901 and 1991, the life expectancy of:

- a 1-year-old child rose from 55 to 76 years (38% increase)
- a 60 year old increased from 14 to 19 more years (36% increase)
- a 75 year old increased from 7.5 to 10.5 more years (40% increase).

The ratio of females to males in the elderly population increases progressively with age and in the over-85-year age group there are three times as many women as men.

By 2026, the over-65s are expected to represent 19% of the UK population and the over-85s almost 2.5% of the population. In the USA in 1990, about 13% of the population was over the age of 65 and this proportion is projected to rise to around 22% by 2040.

Around 36% of the over-65s in England and Wales lived alone in 1989 and the proportion living by themselves increases markedly with advancing age

Table 14.11 The percentage of the population of England and Wales in various age bands

Age band (years)	Percentage of total population*
0–14 (children)	19.2
15–44 (young adults)	43
45–64 (middle-aged)	21.9
65+ (elderly adults)	15.9
65–74	8.8
75–84	5.4
85–89	1.2
90+	0.5

* total population of 51.3 million in 1992

Reproduced from Webb and Copeman (1996).

as a result of the deaths of spouses. Almost 60% of women and 45% of men aged over 85 lived alone. The number of elderly people living in residential homes for the elderly increased from around 150 000 in 1977 to around 235 000 in 1990. This 1990 figure represents around 3% of all people over the age of 65 years and almost 0.5% of the total population of England and Wales. The chances of an elderly person living in care accommodation rise steeply with age, and the increase in the number of people living in care accommodation is entirely due to increases in the size of the elderly and very elderly populations. In 1995, the proportions of elderly people living in all types of long-term residential care (residential homes, nursing homes and long-stay hospitals) were:

- 1% of 65–74 year olds
- 6% of 75–84 year olds
- 27% of those aged over 85 years.

Catering for the dietary and nutritional needs of the elderly is thus a topic that should warrant an increased educational and research priority.

The household income and expenditure of elderly people in England and Wales tend to be concentrated at the lower end of the range. In both single adult and one man, one woman households, retired persons account for 40–80% of the three lowest household income groups used by the General Household Survey, but they become progressively less well represented in the higher income categories. Retired persons in the upper income groups inevitably receive some income in addition to the state pension. In 1992, over 60% of the households with an elderly head had an income in the lowest

band, compared to only about 15% of young adult households. Less than 20% of older households had an income in the highest band, compared to about 60% of young adult households. Amongst the elderly, income declines with increasing age. The average expenditure per person in households in which the head was aged over 75 years was only two-thirds of that in households in which the head was aged 50–64 years. Many elderly people spend their later years in relative poverty. The increasing ratio of retired to working people in the population makes it difficult to foresee any immediate significant improvement in the finances of elderly people in the UK who are largely dependent upon the state retirement pension.

As income declines, so absolute *per capita* expenditure on food also declines, but food accounts for an increasing share of expenditure. The proportion of income spent upon housing and fuel also increases with age in Britain, but the proportion spent on clothing, transport, alcohol and tobacco declines. The food expenditure of retired couples in the lowest income group in the UK is close to the minimum estimated cost of a 'modest but adequate' diet that is also broadly in line with the usual UK diet (see Chapter 2). Webb and Copeman (1996) compared the food expenditure of elderly households with other one-adult and two-adult households in the UK. They made the following observations:

- elderly households spend less upon food than the equivalent younger households
- elderly households get more calories for each penny of their expenditure, a general trend with lower income
- the elderly spend less upon cheese, vegetables (excluding potatoes), fruit, soft drinks and alcohol, but more upon sugars and fats
- elderly households are less likely to buy low-calorie soft drinks and low-fat milk, but more likely to buy butter.

More than half of all the over-65s report being affected by some long-standing illness and the proportion rises with advancing age. Bone and joint diseases, including arthritis and rheumatism, are easily the most common group of long-standing illnesses. Other important causes of long-standing illness in the elderly include heart diseases, hypertension, respiratory diseases such as bronchitis and emphysema, stroke and diabetes. Ageing inevitably leads to increasing rates of mortality, morbidity and disabil-

ity. Despite this, the majority of elderly people consider themselves to be in good or fairly good general health. In 1985, about 80% of men and women aged over 65 years perceived their own health as good or fairly good, and even in the 85+ age group only around 27% of men and 33% of women described their health in the previous year as 'not good'. This suggests that some level of disability is seen by elderly people as being inevitable and does not stop many of them from perceiving their overall health as good. More than 80% of elderly people still report seeing friends or relatives at least once a week and even in the over-85s this figure is still around 75%. About 97% of people aged over 65 years still live either in their own homes or with their families. The widespread image of the post-retirement years as being inevitably an extended period of poor health, disability, dependence and social isolation does not seem to be the perceived experience of the majority of older people in Britain.

As average life expectancy starts to reach its inevitable plateau, so improving the quality of life of the elderly population has become an increasing priority for health education in industrialized coun-

tries. One of the major goals listed in *Healthy people 2000* in the USA (DHHS, 1992) is to 'increase the span of healthy life for Americans'. In *The health of the nation* in England (DH, 1992), the aim is not only to reduce premature deaths and 'add years to life' but also to 'add life to years' by increasing the years that are free from ill-health and minimizing the adverse effects of illness and disability. Bebbington (1988) calculated 'the expectation of life without disability' from self-reported rates of disability and long-standing illness recorded in the British General Household Survey, together with estimates of the numbers of people living in institutions for the 'disabled'. Although total life expectancy in the UK increased over the period 1976–1988 (from 70 to 72.4 years in men and from 76.1 to 78.1 years in women), the 'expectation of life without disability' remained essentially unchanged. Over this same period, the years of illness and disability increased by 2.1 in men and 2.5 in women. The main effect of the recent increases in life expectancy has been to increase the number of years spent suffering from illness and disability. A similar study in the USA (Crimmins *et al.*, 1989) concluded that, between 1970 and 1980,

- Improved life expectancy during the twentieth century caused large increases in the proportion of elderly and very elderly people in the populations of industrialized countries.
- By the year 2000, 16% of British people were aged over 65 years and around 2% over 85 years.
- Not only are more people surviving to 65 years, but elderly people are also living longer.
- Women greatly outnumber men in the older age groups.
- Many elderly people in Britain live alone or in care accommodation and the proportions increase sharply with age.
- Elderly people are disproportionately represented in the lowest income groups in the UK.
- Elderly people spend less on food than younger adults, but it accounts for a higher proportion of their spending.
- The food expenditure of many elderly British people is at or below that considered necessary to purchase a 'modest but adequate' diet that is broadly in line with the current UK diet.
- Many elderly people are affected by some long-standing illness, but this does not prevent most of them perceiving their general health as good or fairly good.
- Improving the quality of life in the later years, rather than simply increasing life expectancy, is now being given a higher priority in health promotion.
- Some studies suggest that most of the recent increase in life expectancy in the USA and UK has been achieved by increasing the years spent suffering from illness and disability.
- A good diet and an active lifestyle should help to improve life expectancy and compress the period of chronic morbidity at the end of life.

increases in life expectancy were largely a result of more years spent with a chronic disabling illness. A good diet and an active lifestyle would contribute both to increasing total life span and to compressing the years of morbidity and disability at the end of the life span.

The effects of ageing

Ageing is characterized by a gradual decline in the ability of an organism to adapt to environmental stresses and an inability to maintain homeostasis. This loss of adaptability or capacity to maintain homeostasis results in increasing mortality rates, increased morbidity and increased rates of disability. There is measurable age-related deterioration in the functioning of most systems of the body. For example, ageing of the immune system leads to a decline in the efficiency of the immune surveillance and defensive mechanisms, which leads to an increased incidence of infection, autoimmune disease and cancer. The absorptive area of the small intestine decreases with age and it is likely that the absorption of several nutrients decreases with age. Deterioration is usually earliest and fastest in those systems in which there is no replacement of dead cells (e.g. brain, muscle and heart) than in those in which continual replacement occurs (e.g. red cells and intestinal epithelium). Generally, complex functions involving co-ordination are more affected than simple ones, e.g. reaction time increases more rapidly than the slowing in nerve conduction velocity.

There is a marked change in body composition with age: the proportions of lean tissue and bone decline and the proportion of fat increases. There is also a loss of height with age from about 30 years

onwards. There is a decline in the basal metabolic rate (BMR) with age, which is probably a function of the decline in lean body mass in the elderly, because when the basal metabolic rate is expressed per kilogram of lean body mass, there is no decline with age (COMA, 1992).

The speed of homeostatic regulation is reduced as people become elderly. Some examples are given below.

- In response to cold stress, old people tend to start shivering later and, in response to heat stress, they sweat later. Old people are also more susceptible to hypothermia because of their reduced BMR and their reduced levels of physical activity.
- There is slower restoration of acid-base balance if it is disturbed. Elderly people have reduced buffering capacity, a diminished respiratory response to acid-base disturbance and a reduced ability to eliminate excess acid or base via the kidney.
- There is a decline in the efficiency of the mechanisms regulating salt and water balance. For example, there is an age-related decline in the acuity of the thirst mechanisms. Rolls and Phillips (1990) found that, immediately after a period of water deprivation, elderly men drank much less than younger men and experienced much less thirst even though their levels of dehydration were similar.

Nutritional requirements of the elderly

Energy expenditure and therefore energy requirements fall as people become elderly. The two factors listed below are thought to be responsible for this decline in energy expenditure.

- All body systems deteriorate during ageing and there are increasing rates of mortality, morbidity and disability in the elderly.
- The proportions of muscle and bone in the adult body decrease with age and the proportion of fat increases.
- The loss of lean tissue in the elderly leads to a fall in basal metabolic rate.
- Ageing is characterized by a reduced speed of homeostatic regulation and a reduced ability to adapt to environmental change.

1. Levels of physical activity decline with increasing age: this may be accelerated by retirement and is inevitable in those who are housebound or bedridden.
2. Basal metabolic rate decreases in the elderly: this is due to the decline in lean body mass and may itself be largely a consequence of reduced levels of physical activity.

Table 14.12 shows the UK EARs and the US RDAs for energy for different age categories of men and women. These EARs and RDAs represent quite substantial reductions in the estimated energy requirements and, therefore, likely total food intake of elderly people, particularly of elderly men. The UK values for younger adults are set by multiplying the BMR by 1.4 (the **physical activity level, PAL**), whereas those for older adults are set using a PAL multiple of 1.5. This is despite clear evidence that physical activity decreases in the elderly. The rationale for this decision is discussed in Chapter 7.

Table 14.13 summarizes the differences between the dietary standards for elderly people and those for younger adults. For nutrients not listed in Table 14.13, the values for younger and older adults are the same. Although energy requirements and total food intakes tend to fall in the elderly, Table 14.13 suggests that there is no corresponding decrease in the requirement for most other nutrients. The UK RNIs for the elderly do indicate a few reductions in the RNI for older people, i.e.

Table 14.12 UK and US Dietary Reference Values for energy in older adults

Age (years)	Men kcal	MJ	Women kcal	MJ
<i>UK (EAR)</i>				
19–50	2550	10.60	1940	8.10
51–59	2550	10.60	1900	8.00
60–64	2380	9.93	1900	7.99
65–74	2330	9.71	1900	7.96
75+	2100	8.77	1810	7.61
<i>USA (RDA)</i>				
25–50	2900	12.12	2200	9.20
51+	2300	9.61	1900	7.94

Data sources COMA (1991) and NRC (1989). From Webb and Copeman (1996).

- protein in men – due entirely to differences in the assumed body weight of young and elderly men
- thiamin and niacin, which are set according to the assumed energy expenditure
- a substantial reduction in the female RNI for iron, which reflects the cessation of menstrual blood losses.

Similarly, in the US RDAs, there are reductions only in the RDAs for the three B vitamins, thiamin, riboflavin and niacin, whose RDA is based upon the expected energy intake, and a reduction in the RDA for iron in elderly women.

In young adults, endogenous production of vitamin D in the skin is the primary source of the

Table 14.13 Differences between Dietary Reference Values for elderly and younger adults

Nutrient	Male 19–50 years	50+ years	Female 19–50 years	50+ years
<i>UK (RNIs)</i>				
Protein (g)	55.5	53.3	45.0	46.5
Thiamin (mg)	1.0	0.9	—	—
Niacin (mg)	17	16	13	12
Vitamin D (μg) ^a	0	10	0	10
Iron (mg)	—	—	14.8	8.7
<i>USA (RDAs)</i>				
Thiamin (mg)	1.5	1.2	1.1	1.0
Riboflavin (mg)	1.7	1.4	1.3	1.2
Niacin (mg)	19	15	15	13
Iron (mg)	—	—	15	10

^a RNI applies after 65 years.
For other nutrients, the values are the same.
From Webb and Copeman (1996).

vitamin. The UK RNI for vitamin D in older adults shown in Table 14.13 reflects the view that endogenous production can no longer be relied upon in elderly people, who may be housebound and inadequately exposed to sunlight. Poor vitamin D status is identified in Chapter 13 as a major aetiological factor for osteoporosis in the very elderly.

Thus, elderly people are perceived as requiring intakes of most nutrients that are similar to those for other adults, but requiring average intakes of energy up to 20% lower than those of other adults. For many elderly people, the real fall in energy expenditure and food intake may be even greater than the figures in Table 14.2 suggest. COMA (1992) suggested that, for many inactive elderly people, who only spend an hour a day on their feet, a value of $1.3 \times \text{BMR}$ would represent their balance, rather than the $1.5 \times \text{BMR}$ used to set the EARs in COMA (1991). Of course, if very elderly people are losing weight, they are not even eating enough to maintain energy balance. This increases the risk that energy intakes can be met without fulfilling the requirements for all other nutrients. This risk increases if the nutrient density of the diet is low, i.e. if a substantial proportion of the energy is obtained from nutrient-depleted sources such as fatty or sugary foods or from alcohol. Energy intakes may be so low in many elderly people that it becomes difficult to obtain a good intake of the other nutrients without substantial alterations in the nature of the diet. Elderly people, or those responsible for providing meals for the elderly, were advised in COMA (1992) to ensure that their diet is of high nutrient density so

that adequate intakes of the other nutrients can be maintained despite a considerable decline in total food intake. COMA (1992) also suggested that the intake of sugars in the elderly tend to be towards the top end of the UK range (i.e. 10–20% of total energy). They concluded that the general dietary guidelines suggesting limiting non-milk extrinsic sugars to 10% of energy might be especially appropriate for older people to ensure a high nutrient density of their diets.

Note that the dietary standards for older adults have often been set at the same value as younger adults because of a lack of data on how nutrient requirements change with ageing. It is quite probable that requirements for some nutrients (e.g. calcium and vitamin B₆) may increase in the elderly and that reduced efficiency of absorption in the gut may affect requirements more generally.

The diets and nutritional status of elderly people

Longitudinal studies of the food intakes of groups of elderly people in the UK and in Sweden indicate that the average energy intakes of elderly people do indeed fall as they get older (see COMA, 1992). Gregory *et al.* (1990), in a survey of 16–64-year-old British adults, found a distinct trend towards decreasing energy intakes in the older age groups. The average intakes of 50–64-year-old men were 5% less than those of 35–49-year-old men (7% less in women). A Danish cross-sectional study of 1000 men and 1000 women in 1982–4 found a marked

- Energy requirements fall in the elderly due to the combined effects of reduced activity and lower BMR.
- In elderly immobile people, energy expenditure may be as low as 1.3 times BMR.
- There is no fall in the estimated requirement for most nutrients to compensate for the fall in energy needs and food intake.
- The iron needs of elderly women are reduced substantially because menstruation has ceased.
- Elderly housebound people are unlikely to obtain sufficient vitamin D from their diet and should either take supplements or ensure exposure to summer sunlight.
- Very low energy requirements mean that the nutrient density of the diet needs to be good.
- There are few data about how ageing affects nutrient requirements, and the need for some nutrients may increase in the elderly, e.g. due to reduced absorption.

decline in energy intakes across the age range 30–85 years. The average intake of 85 year olds was only 72% of that of 30 year olds in men and 79% in women (see Schroll *et al.*, 1993).

A survey of 750 elderly people living in the UK in the late 1960s (DHSS, 1972) indicated that only around 3% of the surveyed population were suffering from malnutrition and that, in the great majority of cases, malnutrition was associated with some precipitating medical condition. Around half of this population were re-assessed in 1972, by which time they were all aged over 70 years (DHSS, 1979). In this second survey, the prevalence of malnutrition was found to be around 7%, but twice as high as this in the over-80s. Once again, most of the nutritional deficiencies were related to some underlying disease; certain social factors were also identified as being associated with the risk of malnutrition in the elderly, e.g. being housebound and having a low income. According to this survey, the diets of elderly British people were in their general nature very similar to those of other adults.

The general impression created by the results of this survey were that, if elderly people have good general health, are reasonably mobile and affluent, they tend to have few specific nutritional problems. However, as people get older, they are more likely to suffer from a number of medical conditions that may

precipitate nutritional problems. The elderly are also more likely to be affected by a range of social factors that were associated with higher risk of nutritional inadequacies, such as:

- immobility and being housebound
- social isolation
- recent bereavement and depression
- low income
- living alone
- low social class
- low mental test score.

Finch *et al.* (1998) conducted a dietary and nutritional survey of a representative sample of British people aged over 65 years that included both free-living elderly people and those living in care accommodation. Table 14.14 compares the sources of energy for free-living elderly adults recorded in this survey with an earlier survey of adults aged 16–64 years (Gregory *et al.*, 1990). The time gap of 8 years between the two surveys must be borne in mind when interpreting Table 14.14. In terms of macronutrient composition, the diets of elderly people seem closer to current nutrition education guidelines than those recorded 8 years earlier for younger adults. In terms of food sources of energy, the elderly sample obtained less of their energy from drinks, sugar, confectionery and vegetables, but more from cereals and milk.

Table 14.14 A comparison of the macronutrient and food sources of energy recorded in samples of British adults aged 16–64 years^a and free-living adults aged over 65 years^b

	Percentage contribution to total energy intake	
	16–64 years	65+ years
Protein	14.6	15.9
Fat	38.4	35.0
Carbohydrate	42.3	46.7
Alcohol	4.9	2.4
Non-milk extrinsic sugars	16.2	12.7
Saturated fatty acids	16.0	14.7
Cereal foods	30	34
Milk and milk products	11	13
Fat spreads	6	7
Meat, fish, eggs	20	20
Vegetables (including potatoes)	12	10
Fruit and nuts	2	3
Sugar and preserves	9	6
Drinks	9	4

^a Gregory *et al.* (1990).

^b Finch *et al.* (1998).

Some other observations from this survey of elderly Britons (Finch *et al.*, 1998) are listed below.

- The energy intakes of free-living men aged over 85 years were 12% less than for those aged 65–74 years and 28% less than those reported for men aged 50–64 years in Gregory *et al.* (1990).
- The average energy intakes of free-living women changed little with age, but the average value of 1422 kcal (5.98 MJ) was low, about 12% less than that recorded by Gregory *et al.* (1990), for women aged 50–64 years.
- The average intakes of most major vitamins and minerals were above or close to the RNI and in many cases well above it.
- About 10% of the total sample had serum ferritin levels that were indicative of poor iron status.
- About 8% of the free-living sample and 37% of those living in institutions had biochemical indications of poor vitamin D status. Vitamin D intakes were only around a third of the RNI, but it is difficult to achieve this RNI without the use of supplements. This reinforces the recommendation of COMA (1992) that elderly people should either get regular exposure to summer sunlight or take vitamin D supplements.
- Intakes of zinc were at or below the RNI and 15% of men and 7% of women in institutions had biochemical indications of zinc deficiency.
- About two-thirds of the free-living sample and just under half of the institutional sample were overweight or obese ($BMI < 25$). About one in six of the institution sample were underweight

($BMI > 20$), as were 3% of the free-living men and 6% of the women.

Diet and disease risk in the elderly

There is an age-related deterioration in the immune system, especially in cell mediated immunity. Malnutrition has similar deleterious effects upon the immune system (see Chapter 15). Malnutrition could compound with the effects of ageing in depressing the immune function of elderly, malnourished people. Indicators of nutritional deficiency have been found to be associated with reduced responses in immune function tests in disease-free, elderly people. Nutritional supplements given to these elderly subjects were associated with improvements in the measures of both nutritional status and immunocompetence (Chandra, 1985). Chandra (1992) randomly assigned 96 free-living elderly people to receive either a micronutrient supplement or a placebo. After a year, several measures of immune function were higher in the supplemented group than in the controls and the supplemented group had less than half the number of days affected by infective illness compared to the controls. Woo *et al.* (1994) showed that nutritional supplements given to elderly people recovering from chest infections were effective in helping them to recover from their illness.

In the Allied Dunbar National Fitness Survey (1992), ageing was associated with a substantial

- Recorded energy intakes do decline with age in middle-aged and elderly people.
- Only a few per cent of elderly Britons in the 1960s and 1970s were found to be malnourished, but the prevalence did rise with age.
- Nutritional inadequacy in elderly Britons is often precipitated by some illness or is associated with adverse social circumstances.
- Elderly British people who were mobile, in good health and reasonably affluent had few specific nutritional problems and had diets that were in their general character similar to those of other Britons.
- Table 14.4 shows a comparison of the dietary energy sources of elderly and other British adults as determined by separate surveys using weighed inventory.
- Around 10% of elderly Britons show biochemical indications of iron and zinc deficiency.
- Many elderly people living in care accommodation show evidence of vitamin D deficiency.
- A sixth of elderly people living in care accommodation are underweight.

decline in aerobic fitness, muscle strength and flexibility. Only a relatively minor proportion of this decline seems to be due to age *per se* and much of it represents a form of disuse atrophy. The average aerobic fitness of the top 10% of men aged 65–74 years was higher than the average of the bottom 10% of men aged 25–34 years. The report's authors concluded that 'much functional disability among older people could be reversed or avoided through continued regular exercise as people grow older'. Fiatarone *et al.* (1994) found that resistance training led to significant increases in muscle strength, walking speed, stair-climbing power and spontaneous physical activity, even in very elderly nursing home residents (mean age 87 years).

Maintenance of a reasonable level of physical activity would seem to be strongly advisable for the elderly as a complement to sound nutrition in maximizing good health and quality of life. In addition to the general beneficial effects upon the cardiovascular system, increased activity will maintain energy expenditure and help to maintain lean body mass in the elderly. Exercise and improved fitness, strength and flexibility will help elderly people to continue with the everyday activities that are essential to independent living. The general benefits of exercise are reviewed in Chapter 16.

Overweight and obesity are generally associated with excess morbidity and reduced life expectancy (see Chapter 8). However, this conclusion is less secure in older people. Lehmann (1991), in a review of nutrition in old age, concluded that underweight is a more important indicator of increased mortality and morbidity risk than obesity. She suggested that, in the elderly, being moderately overweight is not associated with any excess mortality risk, whereas being underweight is associated with the following adverse outcomes:

- increased mortality
- increased risk of hip fracture
- increased risk of infections
- increased risk specific nutrient deficiencies.

Being underweight and having low reserves of energy and nutrients may leave elderly people less able to survive periods of illness or injury. The combined effects of low body weight, reduced lean tissue to fat ratio and extreme inactivity may also reduce the maintenance energy requirements of these elderly individuals very substantially and threaten their

ability to obtain adequate intakes of the other essential nutrients.

At least in the short term, low body weight rather than high body weight may be the better predictor of mortality in the elderly. Two studies, described below, illustrate this point.

1. Campbell *et al.* (1990) found that anthropometric indicators of low body weight, low body fat and low muscle bulk were associated with an increased risk of death in the 4 years following measurement in those aged over 70 years. In contrast, those with a high BMI had no increased risk of death over this period.
2. Mattila *et al.* (1986) found that, in Finns aged over 85 years, there was a progressive increase in the 5-year survival rate as BMI increased.

An expert working party in the UK (COMA, 1992) recommended that there should be more research into the prognostic significance of BMI in the elderly. It reviewed data from a 10-year cohort study of elderly people (65–74 years), which found that the relationship between mortality and BMI was U-shaped, with a tendency for mortality to rise at the extremes of the range, i.e. at very low or very high BMI.

The general dietary guidelines reviewed in Chapter 4 are still regarded as being appropriate for the bulk of elderly people. Despite this, the nutritional priorities for older people are different from those of younger adults. Dietary adequacy assumes a greater priority in the face of declining energy intakes. In those elderly people suffering weight loss, loss of taste perception and reduced ability to chew, maintaining high palatability and good food intake may be more important than conforming to current nutritional education guidelines. These guidelines are intended to reduce the long-term risk of degenerative diseases. It seems inevitable that, even if these preventative benefits still occur in the elderly, they are likely to be reduced. The application of several of these guidelines to the elderly is briefly reviewed below.

- The association between increased serum cholesterol concentration and increased risk of coronary heart disease has often been assumed to be less convincing in elderly people than in young and middle-aged adults. According to COMA (1992), more recent data from both the USA and the UK indicate that the predictive value of a raised serum cholesterol for increased coronary heart disease

risk is maintained in later years and thus that the general cholesterol-lowering guidelines should also apply to the elderly. Table 14.14 suggests that elderly people's diets are closer to the guideline values for fat than those of younger adults, although the proportion of energy from saturated fat is still well above the target value of 10% of total energy.

- Increased intakes of non-starch polysaccharides (fibre) are likely to reduce constipation and beneficially affect existing minor bowel problems such as haemorrhoids. Increased activity and prevention of dehydration would also improve bowel function. Dehydration is common amongst elderly people because of reduced acuity of the thirst mechanism, reduced ability to concentrate urine and physical problems with getting drinks or even voluntary suppression of fluid intake because of fears of incontinence or difficulties in using the toilet. Dehydration is also a cause of confusion in the elderly. Elderly people are advised to drink the equivalent of six to eight glasses of water per day, and those charged with the care of the elderly should make sure that immobile patients are regularly offered palatable drinks.

- Hypertension is very prevalent amongst the elderly population (see Chapter 13) and COMA (1992) recommended that an average intake of 6 g/day of salt is a reasonable target for elderly people as well as for other adults.
- COMA (1992) recommended that the targets for non-milk extrinsic sugars for elderly people should be the same as for the rest of the population. Moderating the intake of added sugar should help to ensure a high nutrient density of the diet. The majority of elderly people now still have some natural teeth and so the effects of sugars upon the teeth and gums are still relevant to older people. According to Finch *et al.* (1998), the sugar intakes of elderly people are fairly close to current guidelines, and these results are at variance with the expectations of COMA (1992), which assumed that the sugar intakes of elderly people tended to be towards the top of the UK range.
- Like the rest of the population, elderly people are advised to eat more fruit and vegetables (five portions per day). This should increase the intake of many vitamins, antioxidants, non-starch polysaccharide and some minerals, including potassium.

- Immune function deteriorates with age and is also depressed by malnutrition.
- Nutrient supplements can improve immune function in some elderly people, and malnutrition may contribute to the decline in immune function with age.
- Aerobic fitness, strength and flexibility all decline sharply with age in Britain, largely because of decreased activity rather than as an inevitable consequence of ageing.
- Resistance training can increase strength and performance, even in people in their 80s and 90s.
- Increased activity and fitness in elderly people would increase life expectancy and improve the quality of life in the later years and so enable more elderly people to remain independent for longer.
- Underweight is a much more important indicator of mortality risk in the elderly than overweight or obesity.
- Overweight elderly and very elderly people tend to have a reduced mortality risk in the subsequent few years.
- Maintaining dietary adequacy becomes an increasingly important goal as people become elderly and very elderly.
- The dietary guidelines for other adults are still generally appropriate for elderly people, providing that they do not compromise dietary adequacy, i.e. lower intakes of fat, saturated fat, added sugar and salt, but more non-starch polysaccharide, starch, fruits and vegetables.
- Dehydration is common in the elderly and it causes confusion and constipation; the equivalent of six to eight glasses of water per day is recommended for elderly people.
- Good intakes of non-starch polysaccharide, increased activity and ample fluid intakes should all contribute to improved bowel function in the elderly.

Nutrition as treatment

Diet as a complete therapy	371	Malnutrition in hospital patients	379
Diet as a specific component of therapy	374		

DIET AS A COMPLETE THERAPY

Overview and general principles

There are a small number of diseases or conditions for which dietary change can be an effective and complete therapy. In many such diseases, the patient is intolerant of a specific nutrient or component of food. **Food intolerance** has been defined as 'a reproducible, unpleasant (i.e. adverse) reaction to a specific food or ingredient, which is not psychologically based'. Where there is 'evidence of an abnormal immunological reaction to the food', this is termed a **food allergy** (Mitchell, 1988).

The therapeutic principles in cases of food intolerance are very simple:

- to identify the nutrient or foodstuff to which the patient is intolerant
- to devise a dietary regimen that provides a healthful and acceptable diet whilst either keeping intakes of the problem nutrient within tolerable limits or completely excluding the offending food-stuff.

In practice, it may prove difficult to pinpoint the food or foods causing the symptoms or even to establish with certainty that the symptoms are really due to food intolerance. For example, a range of diverse symptoms have been attributed to food allergy, e.g. abdominal pain and diarrhoea, urticaria (hives), eczema, asthma, allergic rhinitis, migraine and behaviour problems in children. Food allergy is

not the only cause of such symptoms and is not even a usual cause, so other, more probable, causes need to be eliminated before a diagnosis of probable food allergy can be made.

Once food allergy (or other intolerance) is suspected as a probable cause, the food or foods responsible need to be identified. Exclusion diets, comprised of foods that rarely cause allergy, may be used and, once symptoms disappear, individual foods added back. If symptoms reappear upon the re-introduction of a particular food, this suggests that this food may be responsible for the allergy. To be confident that symptoms are really due to sensitivity to a particular food, one really needs to **blind challenge** with the suspect food. Patients may develop symptoms in response to a placebo if they believe it contains the food they suspect of causing their symptoms or may not develop symptoms when they unknowingly consume it. A procedure analogous to the double-blind clinical trials discussed in Chapter 3 is the only way of being sure that symptoms are a physiological rather than a psychological response to the suspect food. If symptoms do not disappear with the exclusion diet, an **elemental diet** composed of purified nutrients may need to be used.

Skin sensitivity tests are sometimes used to identify causes of food allergy; small amounts of potential allergens are injected into the skin and the extent of any inflammatory skin reaction is then used to assess sensitivity. There is still doubt about how valid these skin reactions are as predictors of responses to substances taken by mouth.

- There are a few conditions for which diet can be an effective and complete therapy and this often involves restricting the intake of a food or ingredient to which the person is intolerant.
- A few conditions, such as pernicious anaemia, respond well to nutrient supplements.
- Food intolerance is an adverse reaction to a food component that is not psychological.
- It may be difficult to pinpoint the precise cause of a food allergy or other intolerance and a blind challenge may be needed to confirm that the response is not psychological.
- If someone is intolerant to a staple food or an essential nutrient, it may be difficult to devise a diet that alleviates the symptoms but is also adequate and acceptable.

In addition to diseases for which exclusion is the primary dietary aim, there are a few diseases that are not of primary dietary origin but can be alleviated by nutrient supplements. **Pernicious anaemia** is probably the best-known example; autoimmune destruction of the parietal cells of the stomach results in a failure of vitamin B₁₂ absorption. Most of the symptoms of this potentially fatal disease can be alleviated by regular injections of B₁₂. **Hartnup disease** is a rare but fatal inherited disorder with symptoms that resemble those of pellagra (niacin deficiency). The symptoms of this disease result from a failure to absorb the amino acid tryptophan (the precursor of niacin), which leads to niacin deficiency. The condition may respond to niacin treatment and, if plenty of protein is given, tryptophan may be absorbed in dipeptides and tripeptides.

Phenylketonuria (PKU) and **coeliac disease** have been selected to illustrate the problems of devising diets that need to severely restrict the intake of an essential nutrient or to exclude a common food ingredient. Coeliac disease also serves as an example of food allergy.

Phenylketonuria

PKU is an inherited disease that affects around 1 : 10 000 babies in the UK. Since the 1960s, all babies born in the UK have been tested for this condition within the first 2 weeks of life. The blood phenylalanine concentration is measured using the Guthrie test. Classical PKU is caused by a genetic defect in the enzyme phenylalanine hydroxylase that converts the essential amino acid phenylalanine to tyrosine:



Normally, excess phenylalanine, either from the diet or from endogenous protein breakdown, is metabolized via its conversion to tyrosine. In normal adults, 90% of the phenylalanine consumed is converted to tyrosine and only about 10% is used for protein synthesis. In PKU, this route is blocked and so excess phenylalanine and unusual metabolites of phenylalanine accumulate in the blood, where they seriously impair mental development. In PKU, tyrosine becomes an essential amino acid and some of the symptoms of PKU are probably associated with tyrosine deficiency. Untreated children are severely mentally retarded and prone to epilepsy, but they grow normally and have a normal life expectancy.

The objective of dietary management in PKU is to restrict phenylalanine intake so that mental development is not impaired, but to provide enough of this essential amino acid and all other nutrients to allow growth and physical development. In particular, adequate amounts of tyrosine must be provided. These objectives cannot be achieved with normal foods because all foods with good amounts of protein and tyrosine also contain high levels of phenylalanine. The tolerance to phenylalanine depends upon the precise nature of the biochemical lesion and varies from child to child. The amount of phenylalanine that the child can consume whilst keeping the blood levels within the acceptable range for his or her age is first determined. The phenylalanine allowance that has been defined in this way is taken in the form of normal foods, usually milk, cereals and potatoes. Meat, fish, eggs, cheese and pulses will be totally excluded and only sugars, fats, fruits and non-leguminous vegeta-

bles can be eaten relatively freely, and even for these there will be some restrictions. The resultant diet is, by design, very low in protein and tyrosine and it is likely to provide inadequate amounts of energy. Children are therefore given medical supplements that provide amino acids, energy and other nutrients, such as vitamin B₁₂, that are likely to be deficient. Infant formulae that are low in phenylalanine are available prior to weaning.

The prognosis for children with PKU depends upon how strictly the dietary regimen is followed, but even in families where control is poor, the children should still be educable. Prior to the 1960s, and the introduction of universal screening, most PKU children would have been very severely impaired (the average IQ of untreated children with PKU is about 50) and many would have spent much of their lives in institutions for the mentally handicapped. Children whose blood phenylalanine levels have been well controlled have normal IQs at 9 years old (see Elsas and Acosta, 1999). Current consensus suggests that some dietary regulation needs to be maintained throughout life. Neurological changes occur in adults with PKU when their blood phenylalanine concentration is allowed to rise, and these symptoms are reversed by the re-introduction of a low-phenylalanine diet. Special care needs to be taken with the diet of women with PKU intending to become pregnant, because the developing foetus is vulnerable to high phenylalanine levels. Babies of mothers with uncontrolled PKU often have congenital defects that are incompatible with life, and those who do survive fail to grow and develop normally. The dietary management of PKU and other inborn errors of metabolism is reviewed by Elsas and Acosta (1999).

Coeliac disease (gluten-induced enteropathy)

Coeliac disease results from a hypersensitivity or allergy to the gliadin component of the wheat protein **gluten**; similar proteins are found in barley, oats and rye and these also cause symptoms. The disease is classified as a food allergy because it involves an 'abnormal immunological response to food', but it differs from classical allergy, e.g. hay fever, in that it does not involve the production of antibodies to gluten of the type (**immunoglobulin E, IgE**) that is usually associated with allergic reactions. IgA and IgG antibodies to gliadin are present in the blood and there are increased numbers of immune cells in the affected intestine. Coeliac disease is also associated with other autoimmune diseases such as thyroiditis and diabetes.

In untreated coeliacs, the villi of the small intestine atrophy and there is excessive secretion of mucus. The diagnosis is confirmed by a biopsy of the jejunal mucosa using an endoscope. These changes in the small intestine result in severe malabsorption of fat and other nutrients, leading to fatty diarrhoea, distended abdomen, weight loss or growth failure and other symptoms that are the result of nutrient deficiencies such as anaemia, rickets/osteomalacia and bleeding due to vitamin K deficiency. In the long term, there may be cancerous changes in the intestine induced by continued exposure to gluten, leading to small bowel lymphoma. Some patients diagnosed with this condition can tolerate some dietary gluten without overt symptoms, but this may increase their later risk of small bowel lymphoma. Gluten sensitivity may also present as a chronic skin disorder, **dermatitis herpetiformis**.

- Coeliac disease is a hypersensitivity to the cereal protein gluten.
- Gluten causes atrophy of the villi in the small intestine, which leads to diarrhoea, malabsorption and a variety of abdominal symptoms.
- If untreated, the condition leads to growth failure, weight loss, a range of nutritional deficiencies and eventually may result in cancerous changes in the small bowel.
- The treatment is to exclude from the diet the wheat, barley, rye and oats that contain the gluten.
- Dietary compliance is made difficult by the widespread use of wheat flour in manufactured foods.
- Rice and maize do not contain gluten and a range of gluten-free wheat products is available, some of them on prescription in the UK.

The aim of dietary management in coeliac disease is to avoid all gluten-containing foods. The achievement of this goal is complicated because not only are bread and cereals staple dietary items, but flour is also an ingredient of many prepared foods. Gluten-free flour and bread are available on prescription in the UK and gluten-free cakes and biscuits (cookies) are commercially available. Many food manufacturers use a 'crossed grain' symbol to indicate that a particular product is free of gluten. Debenham (1992) has reviewed the dietary management of coeliac disease.

The symptoms of coeliac disease may present in early childhood or in adulthood. It has been suggested that the increased use of rice-based weaning foods and thus delayed exposure to wheat have caused a delay in the onset of the disease, with more adult and fewer childhood presentations.

DIET AS A SPECIFIC COMPONENT OF THERAPY

There are a number of conditions for which diet does not give complete relief from symptoms, nor is it the sole treatment, but for which there is an apparently sound theoretical and empirical basis indicating a specific role for a particular dietary regimen in its management. Three examples, **diabetes mellitus**, **cystic fibrosis** and **chronic renal failure**, are briefly overviewed.

Diabetes mellitus

Diet has long been regarded as a key element of the treatment of diabetes. Diabetes mellitus results from insufficient production of the pancreatic hormone **insulin**. In the severe **type I diabetes** (insulin-dependent diabetes mellitus, IDDM), which usually develops during childhood, there is destruction of the insulin-producing cells in the pancreas that results in an almost total failure of insulin supply. It is envisaged that some environmental factor such as a viral infection triggers the autoimmune destruction of the cells in the pancreas that produce insulin, the β -cells in the islets of Langerhans. This type of diabetes accounts for around 5% of all cases of diabetes. Type I diabetes is always treated by a combination of diet

and insulin injections. Patients cannot survive long without insulin therapy and before insulin therapy was available they usually died within a few weeks of diagnosis.

In the much more common and milder **type II diabetes** (non-insulin-dependent diabetes mellitus, NIDDM), there is a progressive inadequacy of insulin production. This inadequacy seems to stem from a failure of the pancreas to be able to compensate adequately for a progressive decline in sensitivity to insulin in the target tissues. Insulin sensitivity decreases as body fat content increases, and excessive weight gain is the most important trigger for the onset of type II diabetes in those who are genetically susceptible to it (see Chapter 9). A high-fat, low-fibre diet and inactivity are other lifestyle characteristics that decrease insulin sensitivity and predispose to this type of diabetes.

Type II diabetes is usually treated either by diet alone or by a combination of diet and oral hypoglycaemic drugs. Insulin injections may be used as a last-resort treatment for this type of diabetes. Around 5% of the UK population aged over 65 years are diagnosed diabetics, but there may be an equal number who are undiagnosed and untreated but still at risk of the long-term complications of diabetes. The risk of this form of diabetes increases with age and so the number of diabetics will continue to increase as the numbers of very elderly people in the population continue to grow.

All diabetics have a high blood glucose concentration and as a consequence they excrete glucose in their urine. This glucose in the urine acts as an osmotic diuretic and causes increased water loss, increased thirst and propensity to dehydration. In untreated type I diabetes, there are severe life-threatening symptoms:

- rapid weight loss (weight loss also occurs in type II diabetes and so tends to obscure the strength of the association between obesity and diabetes)
- very high blood glucose concentration and severe dehydration
- excessive production and excretion of ketone bodies, leading to acidosis, electrolyte imbalance and eventually to coma and death.

The symptoms of the disease suggest that the diabetic is carbohydrate intolerant, and so severe restriction of dietary carbohydrate and almost total exclusion of dietary sugar have been the rule for dia-

betic diets for the last 100 years or so. Prior to the 1970s, most diabetic diets would have contained no more than 40% of the calories as carbohydrates, and often much less than this. These low-carbohydrate diets were inevitably high in fat. According to Leeds (1979), when insulin treatment was first used in the 1920s, diabetic diets typically contained less than 10% of calories as carbohydrate and more than 70% as fat.

The historical aim of diabetic therapy was to alleviate the immediate symptoms of the disease, which are at best unpleasant and incapacitating and at worst acutely life threatening. Despite the availability for many years of therapies that are effective in this limited aim, the life expectancy of diabetics was and is considerably less than that of non-diabetics and diabetics have continued to suffer a high level of disability in their later years. A major additional objective of modern diabetic therapy is to increase the life expectancy and to reduce the long-term morbidity of diabetics, and so some discussion of the causes of the high morbidity and mortality of diabetics is necessary. Two major considerations seem to be important in this respect:

1. the hyperlipidaemia that is associated with diabetes and the consequent increase in risk of atherosclerosis
2. the persistent hyperglycaemia of diabetes.

Diabetics in Western countries have traditionally suffered much higher mortality from cardiovascular diseases than non-diabetics. This high rate of cardiovascular disease is not an inevitable consequence of diabetes, because Japanese and black East African diabetics have been relatively free of cardiovascular disease, unlike either Japanese American or black American diabetics (Keen and Thomas, 1988). A high-fat diet is widely accepted to be a risk factor for cardiovascular diseases and so the very high-fat diet that was in the past prescribed for diabetics almost certainly increased their propensity to cardiovascular disease still further.

Diabetics also suffer from a range of conditions that are attributed to degenerative changes in their microvasculature, i.e. capillaries. These changes in the microvasculature are responsible for the **diabetic retinopathy** that causes many diabetics to go blind and also for the high levels of renal disease (**diabetic nephropathy**) seen in diabetics. Changes in the capillaries of diabetics are thought to stem from a chem-

ical change in the proteins of the basement membrane of capillary cells that results from their continued exposure to high glucose levels; the proteins react abnormally with glucose and become **glycosylated**. Glycosylation probably also plays a role in the development of cataracts and in the degeneration of peripheral nerves (neuropathy) and gangrene often seen in older diabetics. Glycosylation of **low-density lipoproteins (LDLs)** may increase their atherogenicity. The proportion of glycosylated haemoglobin is used as an indicator of average blood glucose concentration over the previous 2–3 months. In a healthy person, glycosylated haemoglobin is less than 6% of the total, but it can rise to over 25% in some diabetics.

Modern diabetic management seeks to reduce these longer-term complications by:

- achieving and maintaining a more normal body weight
- minimizing hyperlipidaemia and hyperglycaemia.

The modern diabetic diet is almost the complete opposite of that recommended prior to 1970. It is a low-fat diet that is high in complex carbohydrate and fibre. More than half of the energy should come from carbohydrate, with the emphasis on unrefined foods high in complex carbohydrate and fibre. It has moderate levels of fat and saturated fat, ideally less than 35% of energy as fat and less than 10% as saturated fat. The very strict avoidance of sugar has been relaxed provided it is a component of a meal. The general avoidance of isolated sugary foods except for hypoglycaemic emergency is still recommended. A diet considered ideal for diabetics with type II diabetes is very similar to that recommended for the population as a whole. The matching of carbohydrate to insulin dose in type I diabetes is a specialist topic beyond the scope of this text, but the overall dietary strategy is as outlined here. Note that overdose of insulin causes the blood glucose level to fall below normal (hypoglycaemia), and this fall in blood glucose can lead to a variety of symptoms, ranging from confusion and disorientation through to coma and even death. Hypoglycaemia is treated by consuming a sugary snack or by the infusion of glucose if the patient is unconscious. It occurs, for example, if the patient misses or unduly delays a meal or takes too much insulin.

As seen in Chapter 8, obesity is a major risk factor for type II diabetes, and normalization of body weight and regulation of caloric intake to match expenditure improve the symptoms of type II diabetes. Much of the success of traditional low-carbohydrate diabetic diets may have been due to their regulation of energy intake rather than to carbohydrate restriction *per se*. Provided the diabetic is in energy balance, the proportions of energy that come from fat and carbohydrate do not really affect short-term diabetic control. A raised carbohydrate intake increases peripheral sensitivity to insulin and thus does not increase the need for insulin. This adaptation cannot occur in the absence of insulin and so

carbohydrate does have adverse effects in untreated severe diabetes, thus explaining the past assumption that carbohydrate was inevitably bad for diabetics. The well-documented effects of fibre in improving glucose tolerance are discussed in Chapter 8. An increase in the proportion of energy from carbohydrate, together with an increase in dietary fibre, have a generally favourable effect upon blood glucose control in both type I and type II diabetes. The moderate levels of fat and saturated fat in diabetic diets would be expected to reduce blood lipid levels and contribute to a reduction in atherosclerosis.

There is now firm evidence from the Diabetes Control and Complications Trial (DCCT, 1993) that

- Diabetes mellitus is caused by insulin deficiency.
- In type I diabetes, there is β -cell destruction during childhood and absolute insulin deficiency.
- In type II diabetes, the primary change is reduced sensitivity to insulin with age, which can precipitate a relative insulin deficiency.
- Type II diabetes is triggered by excessive weight gain, a high-fat low-fibre diet and inactivity.
- The acute symptoms of type II diabetes are relatively mild, whereas those of type I diabetes are acutely life threatening and insulin therapy is an absolute requirement.
- 95% of diagnosed diabetics are type II, many more remain undiagnosed and the numbers are rising with the growth in the very elderly population.
- Both types of diabetes lead to hyperglycaemia, glucose in the urine, increased urine flow and increased thirst.
- In untreated type I diabetes, there is rapid weight loss, severe dehydration and ketoacidosis, which will eventually result in coma and death.
- Diet is a key part of the treatment of both types of diabetes.
- Diabetics are particularly prone to heart disease, strokes, renal failure, blindness, peripheral nerve degeneration and gangrene.
- Good control of body weight, blood lipids and blood glucose is important for reducing the long-term complications of diabetes.
- Obesity reduces insulin sensitivity, high blood lipids increase atherosclerosis, and hyperglycaemia induces the excessive glycosylation of proteins that is responsible for many of the long-term complications.
- Modern diabetic diets are low in fat and saturated fat but high in fibre and complex carbohydrate.
- High-carbohydrate diets improve glycaemic control because they increase insulin sensitivity.
- Dietary fibre slows carbohydrate digestion and absorption and so improves glycaemic control.
- Low saturated fat intake reduces atherosclerosis.
- Better glycaemic control has been shown to reduce the renal, retinal and peripheral nerve damage associated with diabetes.
- The glycaemic index is a measure of the rise in blood glucose produced by individual carbohydrate foods.
- The increased use of foods with high glycaemic index (such as pulses) should facilitate better glycaemic control in diabetics.

type I diabetics who achieved good long-term glycaemic control had a reduced risk of nephropathy, retinopathy and neuropathy compared to those with poor glycaemic control.

Jenkins and Wolever (1981) coined the term **glycaemic index** to describe in a quantitative way the rise in blood glucose that different carbohydrate foods produce. The glycaemic index is the rise in blood glucose induced by the test food expressed as a percentage of that induced by the same amount of pure glucose (the areas under the blood glucose-time curves are compared). For example, bread and cornflakes have a high glycaemic index, porridge oats and All Bran® have moderate values, and pulses have low values. It is generally accepted that foods with a low glycaemic index would be beneficial for glycaemic control in diabetes management. The following factors are some that will affect the glycaemic index of an isolated food:

- the speed of digestion of the carbohydrate
- the amount and type of dietary fibre
- the presence of monosaccharides, such as fructose, which are slowly absorbed from the gut
- the ripeness of a fruit.

The presence of other foods in a meal will also modify the glycaemic response to a food. Manipulation of the glycaemic index has been generally regarded as too complex to be usefully incorporated into formal individual dietary advice for diabetics. However, a popular book (Leeds *et al.*, 1998) has sought to increase popular awareness and to encourage the use of this concept in food selection for better weight control, diabetic management and general healthy eating.

Cystic fibrosis

Cystic fibrosis is an inherited disease caused by an autosomal recessive gene and it affects as many as 1 in 2000 babies in the UK. There are in excess of 6000 patients with cystic fibrosis in the UK. The genetic lesion results in the production of a sticky mucus, which blocks pancreatic ducts and small airways in the lungs. In the pancreas, this blockage leads to the formation of cysts and progressive fibrosis and loss of function as these cysts are repaired. Similar changes occur in the lungs; there are repeated chest infections and progressive fibrotic damage to the

lungs. In the past, affected children would have been unlikely to survive into adulthood, but improved therapy is steadily increasing the life expectancy of affected patients. The increasing prevalence (in the UK about 130 extra patients per year) is largely accounted for by the increased survival of affected patients into early adulthood.

Regular physiotherapy to clear the lungs of the thick secretions and the availability of antibiotics to treat lung infections have been key factors in the improved survival and quality of life of sufferers, and lung transplants may further extend the life expectancy of cystic fibrosis patients. Improved dietetic management has also been an important factor in improving the prognosis for cystic fibrosis patients. The fibrosis of the pancreas leads to a failure to produce pancreatic juice and, as a consequence, there is poor digestion and absorption of fat and protein and poor absorption of fat-soluble vitamins. Untreated cystic fibrosis patients are at high risk of both general malnutrition and fat-soluble vitamin deficiencies. Evidence of vitamin A deficiency has frequently been found in untreated cystic fibrosis patients, and in some cases **xerophthalmia** (due to vitamin A deficiency) has been the presenting symptom. There is often anorexia in cystic fibrosis patients, especially associated with infection, and there is also increased energy expenditure that is believed to be a primary consequence of the disease.

The principal dietetic objective in this condition is to maintain the nutritional adequacy of the patient. Dietetic management should prevent the wasting, the deficiency diseases and the fatty diarrhoea of untreated cystic fibrosis. There is also evidence of an interaction between nutritional status and the pulmonary manifestations of cystic fibrosis, which ultimately determine survival. Malnourished patients are more prone to respiratory infections, whereas well-nourished patients suffer fewer episodes of pneumonia. The experiences of other sick and injured patients suggest that weak, malnourished people have an impaired ability to cough and expectorate (KFC, 1992).

The strategies employed to meet this objective of maintaining dietary adequacy are summarized below.

- Pancreatic supplements containing the missing digestive enzymes are given with food. These are now in the form of coated microspheres that are

- Cystic fibrosis is an inherited condition that results in progressive fibrotic damage to the lungs and pancreas.
- A combination of effective physiotherapy, antibiotic treatment of lung infections and improved dietary management now allows many cystic fibrosis patients to survive into adulthood.
- Failure of pancreatic secretion leads to malabsorption of fat, protein and fat-soluble vitamins.
- Anorexia and malabsorption reduce energy and nutrient supply and raised energy expenditure may increase requirements.
- The aim of dietary management is to prevent the wasting and deficiency diseases that cystic fibrosis can precipitate.
- Improved nutrition may slow pulmonary deterioration and certainly maintains muscle strength and the ability to cough and expectorate.
- Oral supplements of pancreatic enzymes ensure the proper digestion of food and enable patients to eat a normal diet.
- Vitamin supplements are a precautionary measure.
- Supplements or artificial feeding should be used during bouts of infection and anorexia.

protected from destruction in the stomach but disintegrate in the small intestine (duodenum).

- Cystic fibrosis patients are prescribed a diet that is high in energy and protein and, with the more effective pancreatic supplements, this need not now be low in fat.
- Vitamin supplements (particularly fat-soluble vitamins) are given.
- Dietary supplements or artificial feeding are given when patients are unable to take adequate food by mouth.

Dodge (1992) has reviewed the dietary management of cystic fibrosis.

Chronic renal failure

The weight of the kidney and the number of functional units (nephrons) both decline with age. There is also an age-related decline in measures of renal function such as glomerular filtration rate. In large numbers of elderly people, renal function declines to the point at which the **chronic renal failure (CRF)** becomes symptomatic. The incidence of CRF is five times higher in the over-60s than in those aged 20–49 years and ten times higher in the over-80s.

As the kidney fails, there is reduced excretion of the urea produced by the breakdown of excess dietary protein (or body protein in wasting sub-

jects). Levels of creatinine and urea in the blood rise and increased blood urea (**uraemia**) produces unpleasant symptoms such as headache, drowsiness, nausea, itching, spontaneous bruising, vomiting and anorexia. As the condition progresses, patients will eventually lapse into a coma and die unless they receive either dialysis or a kidney transplant. A diet that minimizes urea production would be expected to reduce the symptoms of uraemia, and low-protein diets have long been used in the symptomatic treatment of CRF.

There is a widespread belief that a low-protein diet is not only palliative but, if started in the early stages of renal failure, actually slows the pathological degeneration of the kidney. It thus extends the period during which conservative management can be used before either dialysis or transplantation becomes essential. It has even been argued that chronic over-consumption of protein in early adult life may be a contributory factor in the aetiology of CRF (e.g. Rudman, 1988).

This benefit of low-protein diets in slowing the progression of CRF is still a matter of dispute. Many of the early studies that claimed a benefit for low-protein diets were small, improperly controlled or short term. Even if very restrictive diets have some effect, this may be outweighed by the effects they have on the patient's quality of life and general nutritional status. Locatelli *et al.* (1991) reported only marginal benefits from a low-protein diet and a

controlled but 'normal' protein diet on the progression of CRF in a large sample of Italian patients with renal disease. They concluded that any extra benefits of the low-protein diet did not justify the restrictions imposed upon the patients. Compliance with the low-protein diet was acknowledged to be poor in this study. A large and long-term American cohort study suggested that moderate protein restriction slowed the progression of CRF in patients with moderate renal insufficiency. More severe restriction of protein intake in those with more advanced renal failure did not appear to offer any additional benefits as compared to moderate restriction (Klahr *et al.*, 1994).

Low-protein diets for chronic renal disease aim to restrict protein intake to around 40 g/day (and phosphorus intake to around 600 mg/day) and to give a high proportion of this protein in the form of high-quality animal protein. The typical protein intake of a UK male is 84 g/day. Intakes of lower quality cereal proteins, with their attendant relatively high nitrogen losses, are restricted and the diet also aims to be high in energy in order to prevent the breakdown of endogenous protein as an energy source. The overall aim is a high-energy, low-protein but high protein-quality diet. A range of high-energy but protein-reduced foods such as protein-reduced flour, bread, pasta and crackers is available, not only as direct sources of low-protein calories, but also to act as vehicles to carry fats and sugars, e.g. butter and jam (jelly). High-energy, low-protein supplements may also be used. In the past, dietary restriction was often initiated in asymptomatic patients based on biochemical evidence of mild insufficiency, but there was little evidence to support the use of this early intervention.

MALNUTRITION IN HOSPITAL PATIENTS

Overview

Every careful observer of the sick will agree in this, that thousands of patients are annually starved in the midst of plenty, from want of attention to the ways which alone make it possible for them to take food.

Florence Nightingale (1859) *Notes on nursing*.
Reprinted 1980, Edinburgh: Churchill Livingstone.

Many people with severe illness are at risk from an unrecognised complication – malnutrition... Doctors and nurses frequently fail to recognise under-nourishment because they are not trained to look for it.

Kings Fund Centre report (1992) *A positive approach to nutrition as treatment*. (KFC, 1992).

These two quotations are separated by almost a century and a half, but both suggested a high prevalence of malnutrition amongst hospital patients that is partly the result of inadequate care.

Several examples of conditions for which diet is a specific part of the therapy have been discussed earlier in the chapter and in such patients there is likely to be careful monitoring of their intake and early recognition of nutritional problems. However, good nutrition is no less important to all long-stay hospital patients and is a vital, if non-specific, complement to their medical or surgical treatment. General medical and nursing staff are likely to be less vigilant in monitoring these patients' food intake and nutritional status.

- Renal function declines with age and in many, usually elderly, people, chronic renal failure (CRF) becomes symptomatic due to high blood urea concentrations.
- Blood levels of urea and creatinine increase as the capacity of the kidney to excrete them declines.
- Low-protein diets reduce the production of urea and give symptomatic relief in CRF.
- Maintenance of good nutrient and energy intakes is necessary as wasting will increase urea production from endogenous protein breakdown.
- Moderate restriction of protein probably also slows the progress of CRF and so extends the time that the patient can be maintained without resort to dialysis or transplantation.

- Malnutrition is still prevalent amongst hospital patients nearly 150 years after Florence Nightingale first highlighted this problem.
- Nutrition has traditionally been inadequately covered in medical and nursing education and so hospital staff are neither trained to monitor nutritional status nor aware of its importance to successful treatment.
- Improving the nutritional support for hospital patients could generate considerable financial savings.

Historically, nutrition has been given low priority in medical and nursing education. Many doctors and nurses are neither trained to recognize signs of inadequate nutrition nor educated about the key importance of good nutrition in facilitating recovery. The key to improving the nutritional status of hospital or community patients is heightened awareness amongst all medical staff of the importance of sound nutrition to prognosis and their increased vigilance in recognizing indications of poor or deteriorating nutritional status. A specialist nutrition team can only personally supervise the nutritional care of a small number of patients. This team must rely upon other medical and nursing staff to identify nutritionally 'at-risk' patients and refer them quickly to the team. The general nutritional care of those patients not specifically at risk is also dependent upon non-nutrition specialists.

Budgetary pressures may encourage hospital managers to economize on the 'hotel' component of hospital costs and to concentrate resources upon the direct medical aspects of care. However, pruning the catering budget is likely to be a false economy if it leads to a deterioration in the nutritional status of patients. Any deterioration in nutritional status will lead to higher complication rates, longer hospital stays and increased costs despite of maintained or even improved surgical, medical and nursing care. In the report of an expert working party, it was estimated that the potential financial savings in the UK from a nationwide introduction of nutritional sup-

port for undernourished patients amounted to over £250 million at 1992 prices (KFC, 1992).

Prevalence of hospital malnutrition

Two papers published in the 1970s are widely credited with focusing attention onto the problem of hospital malnutrition. Bistrian *et al.* (1974) in the USA and Hill *et al.* (1977) in the UK reported that up to 50% of surgical patients in some hospitals showed indications of malnutrition, both general protein energy malnutrition and vitamin deficiencies. Since these landmark publications, there have been reports of a high prevalence of suboptimal nutrition in children's wards, medical and geriatric wards and in hospitals in other countries. For example, almost one in six children admitted to a Birmingham (UK) hospital were stunted or severely wasted – at least a quarter of children admitted with chronic respiratory, cardiac and digestive problems were short for their age (see KFC, 1992, and Powell-Tuck, 1997, for further examples and references).

Consequences of hospital malnutrition

Following a general review of the relationship between nutritional status and outcome in hospital patients, an expert committee in the UK concluded

- A high prevalence of malnutrition has been reported in most hospital specialties from medical to surgical and from paediatrics to geriatrics.

Table 15.1 Some of the likely consequences of undernutrition in hospital patients

Higher rates of wound infections because of slow healing and reduced immunocompetence
Increased risk of general infections, particularly respiratory infections and pneumonia
Muscle weakness which means that patients take longer to remobilize and weakness of the respiratory muscles will impair the ability to cough and expectorate and predispose to respiratory infections
Increased risk of pressure sores because of wasting and immobility
Increased risk of thromboembolism because of immobility
Malnutrition will reduce the digestive functions of the gut and probably also increase its permeability to bacteria and their toxins
Increased liability to heart failure because of wasting and reduced function of cardiac muscle
Apathy and depression

that malnutrition is associated with increased duration of stay, increased hospital charges, increased rates of complications and increased mortality (KFC, 1992). Some of the specific consequences of undernutrition in hospital patients are summarized in Table 15.1 (see KFC, 1992; Powell-Tuck, 1997).

Wasting of muscles and fat is one of the most obvious outward signs of starvation, but there is also wasting of vital internal organs such as the heart and intestines. In the classical starvation studies of Keys *et al.* (1950), chronic undernutrition in healthy subjects was found to cause apathy and depression. This would clearly hinder the recovery of malnourished hospital patients.

It is a general observation that, in famine areas, malnutrition is associated with high rates of infectious disease, with increased severity and duration of illness and ultimately with higher mortality. Providing medicines and medical personnel has limited impact if the underlying problem of malnutrition is not addressed.

Chandra (1993) has reviewed the effects of nutritional deficiencies upon the immune system and concluded that there were a number of demonstrable changes in immune responses as a consequence of malnutrition (see below).

- **Delayed-type cutaneous hypersensitivity** responses to a variety of injected antigens are markedly depressed, even in moderate nutritional deficiency.

In this type of test, the antigen is injected into the skin and the inflammatory reactions occurs some hours after injection – hence ‘delayed-type hypersensitivity’ (DTH). The best known of these reactions is the **Mantoux reaction**, in which individuals immune to the organism that causes tuberculosis respond to a cutaneous tuberculin injection with a DTH reaction. These DTH reactions are a measure of the cell-mediated immune response, which deals with pathogens that have the capacity to live and multiply within cells, e.g. the bacilli that cause tuberculosis and leprosy, some viruses such as the smallpox virus and parasites such as that responsible for the disease toxoplasmosis.

- Secretion of **immunoglobulin A (IgA)** – the anti body fraction that protects epithelial surfaces such as those in the gut and respiratory tract – is markedly depressed. This would make malnourished patients more prone to respiratory, gut and genito-urinary infections.
- Circulating antibody responses (**IgG**) are relatively unaffected in protein energy malnutrition, although the response may be delayed.
- White cells kill bacteria after they have ingested them by generating an oxidative pulse of superoxide radicals. Their capacity to kill ingested bacteria in this way is reduced in malnutrition.

Note that trauma also has an immunosuppressive effect and the degree of immunosuppression is proportional to the amount of trauma (Lennard and Browell, 1993). This means that, for example, the immune systems of poorly nourished patients undergoing major and traumatic surgery will be doubly depressed. Minimizing surgical trauma with modern techniques should lessen the immunosuppressive effect of surgery.

The causes of hospital malnutrition

Hospital malnutrition is not solely or even primarily a consequence of things that happen in the hospital environment. Many patients are already malnourished when they are admitted. For example, McWhirter and Pennington (1994) surveyed the nutritional status of 100 consecutive admissions to each of five areas of a major Scottish acute teaching hospital. They found that about 40% of patients were malnourished at the time of admission. A study by

- Malnutrition of hospital patients increases their duration of stay, their risk of complications and their mortality risk.
- Malnutrition decreases immune function and predisposes hospital patients to infection.
- In malnutrition, the cell-mediated immune response is reduced, there is reduced secretion of IgA, and white cells have a reduced capacity to generate an oxidative pulse to kill ingested bacteria.
- Some specific consequences of malnutrition for hospital patients are summarized in Table 15.1.

Mowe *et al.* (1994) suggests that, at least in the elderly, malnutrition may be one factor in precipitating the illnesses that necessitate hospital admission. They made a nutritional assessment of a large sample of elderly patients admitted to an Oslo hospital because of an acute illness (e.g. stroke, myocardial infarction and pneumonia). They also assessed the nutritional status of a matched sample of older people living within the hospital catchment area. Fifty to sixty percent of patients were undernourished at the time of their admission to hospital, and the nutritional status of the hospital group was much worse than that of the home group. Using several measures of nutritional status, they found that 86% of the home subjects showed no sign of malnutrition, compared to only 43% of the hospital group. They found that poor food intake in the month prior to admission was much more common in the hospital group than in the home group. In the period prior to admission, many more of the hospital group had:

- inadequate energy intakes (in the month before admission)
- intakes of vitamins and trace elements below 66% of the US Recommended Dietary Allowance (RDA)
- problems with buying and preparing foods
- eating difficulties and reduced enjoyment from eating.

They suggested that reduced nutrient intakes and deteriorating nutritional status in the period prior to admission could cause an increased risk of hospitalization.

Although many patients are already malnourished when they reach hospital, there is a tendency for any undernutrition to worsen during hospitalization (Powell-Tuck, 1997). In their survey of 500 Scottish hospital admissions, McWhirter and Pen-

nington (1994) reported that two-thirds of patients lost weight during their stay and that this weight loss was greatest amongst those patients identified as most malnourished on admission. Very few patients were given any nutritional support, but those who were showed a substantial mean weight gain. Those selected for nutritional support obviously tended to be the most severely malnourished on admission. The deterioration in nutritional status that often accompanies hospitalization can usually be avoided with appropriate care. Less than half of the malnourished patients in this survey had any nutritional information documented in their notes. The authors concluded that 'malnutrition remains a largely unrecognised problem in hospital and highlights the need for education on clinical nutrition'. In their earlier study, Hill *et al.* (1977) also found that only 20% of patients' notes contained any reference to their nutritional status, and then only a brief note such as 'looks wasted'. Only around 15% of the patients they surveyed had been weighed at any stage during their hospital stay.

Depressed intake is usually the main cause of nutritional inadequacy during illness or after injury. Several of the factors that may depress food intake in the sick and injured are listed below.

- Physical difficulty in eating: e.g. unconsciousness, facial injury, difficulties in swallowing, oral or throat infection, lack of teeth, arthritis and diseases affecting co-ordination and motor functions.
- Anorexia induced by disease or treatment: severe anorexia is a frequent consequence of malignant disease, but anorexia is a symptom of most serious illnesses and a side-effect of many treatments. It is likely in anyone experiencing pain, fever or nausea.
- Anorexia resulting from a psychological response to illness, hospitalization or diagnosis: starvation itself

- may lead to depression and apathy and may thus reduce the will to eat in order to assist recovery.
- Unacceptability of hospital food: this is most obvious in patients offered food that is not acceptable on religious or cultural grounds but, equally, low-prestige, unfamiliar and just plain unappetizing food may severely depress intake in people whose appetite may already be impaired.
 - Lack of availability of food: patients may be starved prior to surgery or for diagnostic purposes; they may be absent from wards at meal times. This may help to depress overall nutrient intake.

In general, the more specific and more obvious these influences are, the more likely they are to be addressed. It will be obvious that particular measures are needed to ensure that someone with a broken jaw can eat, but a patient who lacks teeth and has a sore mouth and throat is more likely to be overlooked. More allowance is likely to be made for a Jewish patient known to require Kosher food than for a patient who simply finds hospital food strange and unappetizing. Some service provision factors that may help to depress the food intake of hospital patients are listed in Table 15.2.

The factors listed below may increase the nutrient requirements of many sick and injured patients.

Table 15.2 Some service provision factors that may depress energy and nutrient intakes in hospital patients

Timing of meals: meals may be bunched together during the working day, with long enforced fasts from early evening to morning
Prolonged holding of food prior to serving leads to deterioration of both nutritional quality and palatability
Inherently unappetizing food and limited choice
Failure to allow choice, e.g. patients may initially be given meals selected by the previous occupant of the bed
Providing patients with portions of food that are insufficient for their needs, perhaps because staff underestimate the needs of bedridden patients
Plate wastage is not monitored or recorded by staff and so very low intakes are not recognized early
Inadequate amount of time allowed for slow feeders to finish their meals
Lack of staff help for those who need help with eating: staff shortages may mean that, by the time help is provided, the food is cold and unappetizing

Adapted from Webb and Copeman (1996).

- Increased nutrient losses: e.g. loss of blood, the protein lost in the exudate from burned surfaces, loss of glucose and ketones in the urine of diabetics and protein in the urine of patients with renal disease.
- Malabsorption of nutrients: e.g. due to diseases of the alimentary tract.
- Increased nutrient turnover: illness and injury lead to hypermetabolism, the **metabolic response to injury**.

In a classical series of studies that began in the 1930s, Cuthbertson demonstrated that, in patients with traumatic injury, the initial period of shock after injury was followed by a period of hypermetabolism. He coined the term **ebb** to describe the period of depressed metabolism or shock in the first 12–24 hours after injury and the term **flow** to describe the state of hypermetabolism that occurs once the initial period of shock is over. The flow phase is characterized by increased resting metabolism and oxygen consumption and an increased urinary loss of nitrogen associated with increased muscle protein breakdown. Cuthbertson found that the magnitude of this flow response was greater the more severe the injury and the better nourished the patient. Long-bone fractures may increase resting metabolic rate by 15–25% and extensive burn injuries can double the metabolic rate. In severe trauma, sepsis or advanced disease, nutritional support can only partly ameliorate the severe depletion of body fat and protein reserves that this hypermetabolic state produces (see Cuthbertson, 1980).

Following a traumatic injury, Cuthbertson found that the flow response generally reached its peak 4–8 days after injury and that the total loss of body protein could exceed 7% of total body protein within the first 10 days after injury. The increased metabolism was a consequence of the increased rate of body protein turnover and the use of body protein as an energy-yielding substrate. Cuthbertson regarded this catabolic response as an adaptive mechanism to provide a source of energy when incapacitated by injury (traumatic or infective) and a means of supplying amino acids for the synthesis of new tissue and for the functioning of the immune system. This catabolic response is partly mediated through hormonal responses to the stress of injury or illness. Levels of cortisol and catecholamine hormones from the adrenal gland rise in the flow phase, but induced hormone changes in healthy subjects are not able to

- The high prevalence of hospital malnutrition is largely explained by a high prevalence of malnutrition amongst new admissions.
- In some cases, malnutrition may even be a contributory factor to the illness that necessitates hospitalization.
- However, many patients lose weight during their hospital stay and any undernutrition at the time of admission tends to worsen.
- Much of the deterioration in nutritional status during hospitalization can be avoided if there is effective nutritional support.
- There is inadequate monitoring and recording of nutritional status in many hospitals.
- Weight loss is due to depressed food intake and sometimes to increased requirements.
- Intake of hospital patients is depressed for many reasons, including:
 - eating difficulties
 - anorexia due to illness, treatment or anxiety
 - unacceptability of hospital food
 - enforced periods of fasting
 - some more specific service provision factors, as listed in Table 15.2.
- Nutritional requirements may be increased if there are increased nutrient losses in urine or from the site of injury or if absorption of nutrients is impaired.
- Major trauma, sepsis and serious illness induce a hypermetabolic state in which there is increased metabolic rate and protein catabolism, leading to rapid wasting of muscle and adipose tissue.

reproduce the magnitude of protein catabolism seen in severe injury. A patient who is well nourished and able to mount a bigger flow response is enabled to recover more quickly (see Cuthbertson, 1980; Richards, 1980). This would suggest that improving nutritional status prior to planned major surgery should improve the acute recovery from surgery, although studies designed to test this hypothesis have produced variable results (KFC, 1992). A more recent review of the hypercatabolic state can be found in Shils *et al.* (1999).

Improving the nutritional care of hospital patients

AIMS OF DIETETIC MANAGEMENT OF GENERAL HOSPITAL PATIENTS

The two main aims of dietetic care of hospital patients are listed below.

1. To assess and monitor the nutritional status of all patients and, where necessary, to take measures to

correct nutritional inadequacy. Where feasible, the nutritional status of poorly nourished patients should be improved prior to planned surgery.

2. To ensure that, after injury or surgery or during illness, the input of energy, protein and other nutrients is sufficient to maintain body reserves or even to increase them in undernourished patients. In grossly hypercatabolic patients, it may only be possible to partly compensate for losses caused by this hypercatabolism.

The basal energy requirements of patients can be estimated from their body weight (see Chapter 3). The following additional allowances can then be added to this basal figure to estimate their total requirements.

- An amount to allow for the hypermetabolism of illness and injury, which will depend upon the exact nature of the condition. It might range from +5–25% in post-operative patients, those with long-bone fractures or mild to moderate infections, up to as much as +90% in patients with extensive burns.

- An amount to allow for the mobility of the patient, ranging from +20% in the immobile, 30% in the bed bound but mobile, to 40% in those able to move around the ward.
- An increment to allow replenishment of stores if the patient is already depleted.

Protein allowances will also need to be increased in line with the extent of the calculated hypermetabolism.

AIDS TO MEETING NUTRITIONAL NEEDS

The simplest and cheapest way of satisfying the nutritional requirements of sick and injured patients is to encourage and facilitate the consumption of appetizing and nutritious food and drink. For some patients, additional support measures may be needed (see list below).

- **Nutritional supplements:** concentrated energy and nutrient sources that are usually consumed in liquid form and are readily digested and absorbed.
- **Enteral feeding:** the introduction of nutrients directly into the stomach or small intestine by use of a tube that is introduced through the nose and into the stomach or intestine via the oesophagus. Tubes may be introduced directly into the gut by surgical means, particularly if there is likely to be an extended period of enteral feeding.
- **Parenteral feeding:** the patient may be supplied with nutrients intravenously, either as a supplement to the oral route or as the sole means of nutrition – **total parenteral nutrition (TPN)**. In TPN, the nutrients must be infused via an indwelling catheter into a large vein. In some cases of extreme damage to the gut, patients may be fed for years by TPN. A major factor that has enabled patients to be maintained indefinitely with TPN was the development of a means of safely infusing a source of fat intravenously.

Ideally, a specialist nutrition team should supervise these artificial feeding methods. TPN, in particular, requires specialist management and is usually only used when the oral route is ruled out, e.g. in patients who have had a large part of the intestine removed or have severe inflammatory disease of the intestine. Infection of the intravenous lines can be a major problem in TPN but, with skilled specialist management, can be largely eliminated (KFC, 1992).

KFC (1992) estimated that around 2% of hospitalized patients in Britain received some form of artificial feeding and that, in three-quarters of cases, an enteral tube was the method used. Their analysis of the workload of a nutrition team at one general hospital suggested that the number of patients receiving dietary supplements was similar to that of patients being fed artificially. A minority of patients may continue to receive enteral or parenteral nutrition at home. KFC (1992) estimated that, at that time, well in excess of 1000 patients were receiving enteral nutrition at home and 100–150 parenteral nutrition.

The use of oral supplements costs just a few pounds per week, but artificial feeding is much more expensive, and total parenteral nutrition costs several hundred pounds per week. It is therefore important that patients receive a level of intervention that is appropriate for their needs. Figure 15.1 shows a scheme that could be used to select the appropriate level of nutritional support for a given patient.

MEASURES THAT COULD IMPROVE THE NUTRITIONAL STATUS OF HOSPITAL PATIENTS

- Improved nutrition education for doctors and nurses and heightened awareness of the importance of adequate nutrition for patient recovery. Staff must be aware of the effect of illness and injury upon nutritional needs and must be able to carry out simple nutritional monitoring and to interpret the results.
- Inclusion of a nutritional assessment as part of the standard admissions procedure. KFC (1992) suggested that this should include information about appetite, recent weight changes, oral and dental health, social characteristics and physical indicators of nutritional status such as body mass index (BMI) or arm circumference. Many hospitals do now include such information in their admission forms and have developed simple scoring schemes to screen for those people who are nutritionally 'at risk'.
- Regular monitoring of patients' food intake and weight or other simple anthropometric indicator. This would facilitate the early identification of patients with depressed food intake and deteriorating nutritional status.
- Provision of appetizing, acceptable and nutritious food that reaches patients in good condition and

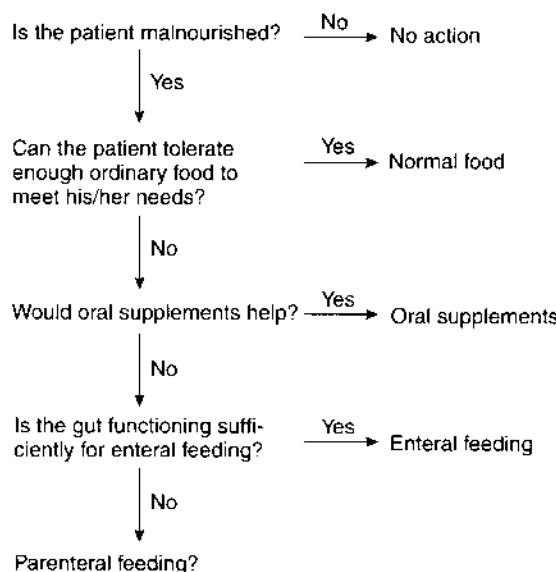


Figure 15.1 A scheme for deciding upon the appropriate level of nutritional support for a patient (based on a scheme in RCN, 1993).

ensuring that they receive the correct level of assistance with feeding. Ways of minimizing the adverse impact of the service provision factors listed in Table 15.2 should be sought and proper provision should be made for patients with special cultural needs. There should be formal procedures to record absences from wards at meal times and active measures to provide food for patients who are absent for meals and after a period of enforced fasting.

- Rapid referral of patients with poor or deteriorating nutritional status to a nutrition specialist. These patients can then be given the appropriate level of nutritional support by a specialist nutrition team. Several of the authors who have conducted surveys that have reported a high prevalence of malnutrition in hospitals have also noted that very few of the malnourished patients had received any nutritional support, e.g. Hill *et al.* (1977) and McWhirter and Pennington (1994).

In many acute conditions, nutritional support can substantially reduce mortality. A number of controlled studies on the effects of supplementary feeding in hospital patients have been conducted. Reduced complication rates, decreased length of hospital stay and reduced mortality rates are general findings from such studies (KFC, 1992; Pennington, 1997; Powell-Tuck, 1997). Larsson *et al.* (1990) carried out a randomized study of the effects of energy supplements using 500 patients admitted to a long-stay geriatric ward of a Swedish hospital. The supplements reduced the deterioration in nutritional status following admission and reduced the mortality rate. Two studies on the effects of supplementation in patients with hip fractures suggest that oral or enteral supplements improve both the nutritional status of the patients and measures of outcome. Supplementation was associated with reduced hospital stay, more rapid mobilization, reduced rates of complications and lower mortality (Bastow *et al.*, 1983; Delmi *et al.*, 1990; Pennington, 1997).

- The nutritional status of all patients should be routinely monitored and all appropriate measures taken to minimize any deterioration in nutritional status.
- Where necessary, special nutritional supplements can be given orally or patients can be fed by an enteral tube or parenterally (intravenously).
- Feeding costs increase steeply with increasing sophistication of nutritional support and so the minimum effective level of support should be used (see Figure 15.1).
- Controlled trials have shown that nutritional support can lead to significant patient benefits.
- Measures that could improve the nutritional status of hospital patients include:
 - improved nutrition education for medical and nursing staff
 - nutritional screening of patients at admission and regular monitoring during hospitalization
 - improvements in catering and the delivery of food to patients
 - rapid referral of 'at-risk' patients to a nutrition specialist.

Some other groups and situations

Vegetarianism	387	Nutrition and physical activity	395
Racial minorities	392		

VEGETARIANISM

Introduction

A strict vegetarian, or **vegan**, avoids consuming any animal products, i.e. meat, fish, eggs, dairy produce and perhaps even honey. Others are less strict in their avoidance and, although they do not eat meat, they may eat dairy produce, eggs or fish or any combination of these three. The prefixes **lacto**, **ovo** and **pesco** are used alone or in combination to describe these degrees of vegetarianism, e.g. one who avoids meat and fish but eats eggs and dairy produce is an ovolactovegetarian.

Johnston (1999) suggested that, at the start of the new millennium, around 10% of all Americans classified themselves as vegetarian. The numbers of Americans reporting that they are vegetarians has roughly trebled since the middle of the 1980s. Only a tiny fraction (about 4%) of these self-styled vegetarians are vegans and many of them occasionally eat meat and many more eat poultry or fish. The term vegetarian clearly has a variety of different meanings to different consumers. People who consume little meat and poultry but are not strictly vegetarian are sometimes referred to as *demi*vegetarians or *semi*vegetarians.

Surveys conducted by Gallup for the Vegetarian Society in Britain show that, in 1999, 5% of those sampled (about 4000 adults) reported being vegetarian, and this was twice as many as in a similar survey conducted in 1985 (www.vegsoc.org/info/realeat.html). Almost half of those surveyed in 1999 reported that

they were eating less meat and almost 9% reported that they were avoiding red meat. Vegetarianism was twice as common in women as in men and was particularly prevalent amongst women in the under-35 age groups.

The following are the major reasons cited for adopting vegetarianism in the USA and UK are.

- Health reasons: the vegetarian lifestyle may reduce the risk of several chronic diseases and there have been several major alarms about the safety of animal foods such as the bovine spongiform encephalopathy (BSE) crisis in Britain (see Chapter 17).
- Ecological and environmental reasons: in Chapter 2, it was noted that it requires between 3 kg and 10 kg of grain to produce 1 kg of meat.
- Animal welfare and cruelty concerns.

Several major religions restrict or discourage the consumption of meat, e.g. Buddhism, Hinduism and Seventh Day Adventism. In the UK, many members of the large south Asian community are lactovegetarians and these people probably represent the biggest group of vegetarians in the UK. Religious considerations and lack of meat availability mean that many people in developing countries eat little or no meat.

Traditionally, nutritionists have been concerned about the adequacy implications of vegetarian and particularly vegan diets. From an adequacy viewpoint, any major restriction of the categories from which food may be selected cannot be regarded as ideal. The greater the degree of restriction, the greater

are the chances of the diet being inadequate and also of any toxicants in food being consumed in hazardous amounts. Paradoxically, most recent interest has been shown in vegetarian diets as a possible means of increasing health. There have been numerous reports of a low incidence of chronic disease amongst vegetarian groups (see Thorogood, 1995). Indeed, as indicated above, this is one of the major reasons cited for adopting vegetarianism in the UK and USA.

A recurring theme of the dietary guidance offered in this book has been to encourage diversity of food choice. The avoidance of whole groups of foods runs contrary to that theme and is therefore regarded by the author as suboptimal. Paradoxically, another recurring theme has been to encourage adults in industrialized countries to reduce their consumption of meat and other animal foods and to increase their consumption of cereals, fruits and vegetables, i.e. to move towards a more vegetable-based diet. It is clear that a thoughtfully constructed vegetarian or even vegan diet is compatible with nutritional adequacy and may well be more diverse and adequate than the current diets of many omnivores. When people decide for cultural, religious, ethical or ecological reasons to adopt some degree of vegetarianism, the role of the nutrition educator should be to facilitate the healthful implementation of that personal choice. Nutritionists should only try to discourage practices that are irrevocably dysfunctional.

Adequacy of vegetarian diets

Animal-derived foods are the only, or the major, sources of some nutrients in typical UK and USA diets. The possibility that supplies of such nutrients might be inadequate in vegetarian or vegan diets needs to be considered and, where necessary, remedial measures identified that are consistent with maintaining the chosen diet. The degree to which less strict vegetarians are threatened by these potential inadequacies will depend upon the extent of individual restrictions. Many vegetarians make liberal use of dairy produce and, in some cases, also of eggs and fish, and so many of the theoretical deficiencies of totally vegetable-based diets will not apply to them.

Vegan diets theoretically contain no vitamin B₁₂, no vitamin D and no retinol. They are also likely to contain less total energy, protein, calcium, zinc, riboflavin and available iron.

Vegan and probably other vegetarian diets are likely to be less energy dense than omnivorous diets because they have less fat and more starch and fibre. Total energy intakes are, as a consequence, also likely to be lower than those of omnivores. Indeed, the reduced energy density of vegetarian diets is perceived as one of the major advantages of such a diet for affluent adult populations in which overweight and obesity are common. There is a general consensus that Caucasian vegans have lower energy intakes than omnivores and are lighter and have a lower pro-

- Only a tiny fraction of vegetarians are vegans who avoid all animal products.
- Most vegetarians consume milk and dairy produce (lactovegetarian); many consume eggs (ovo) and/or fish (pesco).
- Some people classify themselves as vegetarian despite eating some poultry or meat and could be termed demivegetarian.
- Several major religions encourage vegetarianism, but many affluent Western vegetarians choose to be vegetarian for health, ecological or ethical reasons.
- Vegetarianism restricts dietary choice and so theoretically increases the risks of dietary inadequacy.
- Most dietary guidelines in Western countries imply that a move towards a more vegetable-based diet would have health benefits.
- Vegetarians in Western countries often have lower rates of chronic disease and lower early mortality rates than omnivores.

portion of body fat. Most studies also suggest that other vegetarians are lighter than omnivores, although the difference is smaller than that between vegans and omnivores.

A less positive view of the low energy density of vegetarian diets is usually taken when children are the consumers. A diet that has too low an energy density could impair their growth. Studies in the USA, Holland and the UK have found that younger vegan children tend to be lighter and smaller in stature than omnivorous children. The differences are, however, generally small and the growth patterns and growth curves of vegan children are usually within normal limits and they tend to catch up this growth later in childhood. In a sample of British vegan children, Sanders (1988) found that energy intakes were below the Recommended Dietary Allowances (RDAs) at the time and that fat intakes were, on average, low but very variable (ranging from only 16% to 39% of calories). Sanders concluded that low fat intake and the low energy density were probably the major determinants of the low energy intakes. In turn, the low energy intakes were probably responsible for the anthropometric differences between vegan and omnivore children, rather than any differences in dietary quality. American Seventh Day Adventist children who are ovolactovegetarians grow similarly to other children and have no greater evidence of nutritional deficiencies (see Johnston, 1999).

Vegan diets are likely to be lower in protein than omnivorous or lactovegetarian diets. Individual plant proteins are also generally of lower quality than individual animal proteins. Thus, one of the traditional concerns of vegetarians has been about the protein adequacy of their diets. The priority attached to protein deficiency as a likely cause of human malnutrition has declined very sharply in recent decades (see Chapter 10). Mutual supplementation also means that the overall protein quality of a good, mixed vegetable meal or diet may not be substantially different from that of an omnivorous one. This means that, according to current estimates of requirements, vegetarian and even vegan diets are likely to be adequate, both in terms of their overall protein content and in their ability to supply the essential amino acids.

Vitamin D (cholecalciferol) is naturally present only in foods of animal origin (see Chapter 12). Vitamin D is added to soya milk, many infant foods, some breakfast cereals and to margarine and other

spreading fats in the UK. American milk is also often supplemented with vitamin D. Average intakes of vitamin D in the UK are generally well below estimated requirements, but even intakes of British vegans do not differ substantially from those of omnivores, provided that they eat vitamin-D-supplemented foods. Endogenous production (via the action of sunlight upon the skin) rather than dietary intake is regarded as the principal source of vitamin D for most adults and children. Endogenous production should therefore ensure adequacy of vitamin D status in strict vegetarians, even if they avoid supplemented foods. In some groups, current UK Dietary Reference Values (DRVs) suggest that endogenous production cannot be relied upon and so supplements of vitamin D may be necessary. In such groups and in anyone who has limited exposure to summer sunlight, the case for supplements is even stronger if they are vegan, especially if they avoid supplemented foods.

Retinol (vitamin A) is only present in animal foods, but ample supplies of carotene in fruits and vegetable will make the total vitamin A content of many affluent vegetarian diets higher than those of typical omnivorous diets.

In strictly vegetarian diets, there is no apparent vitamin B₁₂. Symptoms of megaloblastic anaemia and the neurological manifestations of vitamin B₁₂ might thus be expected to be prevalent amongst vegans. However, the human requirement for vitamin B₁₂ is extremely small (UK Reference Nutrient Intake (RNI) 1.5 g/day) and stores of the vitamin in the liver are large and could amount to several years' supply in an omnivore. In the UK, meat substitutes must be supplemented with B₁₂ and many vegans take B₁₂ supplements. Even in the absence of such alternative dietary sources, cases of clinical B₁₂ deficiency amongst Caucasian vegans are rare. There are indirect sources of B₁₂, even for those who consume no animal products or supplemented foods, such as from:

- micro-organisms and moulds contaminating plant foods
- insects or insect remains consumed with plant foods
- absorption of vitamin produced endogenously by gut bacteria, although most of this is produced below the point of absorption in the ileum and is excreted in faeces
- fermented foods or yeast extracts
- faecal contamination of foods such as seaweed.

A combination of the large stores, extremely low requirement and incidental consumption from the sources listed above means that the practical hazard of B_{12} deficiency is less than might be expected, even for vegans who take no active measures to ensure the adequacy of their B_{12} supply. High intakes of folic acid in vegans tend to mask the haematological consequences of B_{12} deficiency and, thus, if it does occur, it is the neurological symptoms that are more likely to be manifested. Johnston (1999) has suggested that the low incidence of B_{12} deficiency in some developing countries where a largely vegan diet is the norm may be because of higher levels of contamination of food and water with bacteria. She suggests that, in some products used as B_{12} supplements by vegans, the vitamin may be present in an inactive form.

Dairy produce and eggs are major sources of riboflavin in UK and US diets, and thus intakes of riboflavin are likely to be low in vegan, although not lacto-vegetarian, diets. Carlson *et al.* (1985) found average intakes of riboflavin in vegans of only 75% of the then RDA, compared with 140% of the RDA in ovo-lacto-vegetarians. There is no evidence of overt riboflavin deficiency amongst strict vegetarians in affluent countries, although it is generally one of the more prevalent micronutrient deficiencies. Plant sources of riboflavin are leafy vegetables, nuts and pulses.

Milk and dairy produce, including soya milk, are major sources of dietary calcium. Strict vegetarians who avoid these foods are likely to have lower calcium intakes than omnivores. High fibre and phytate intakes might also be expected to hinder the absorption of calcium from the gut. However, vegans apparently adapt to low calcium intakes and have reduced faecal losses of calcium. A low prevalence of osteoporosis has frequently been reported in vegetarian groups, such as Seventh Day Adventists, although very few studies have looked specifically at vegans. In general, many largely vegetarian populations in developing countries have very low calcium intakes by American and British standards, but also have very low rates of osteoporosis (see Chapter 13).

Haem iron is the most readily absorbed form of dietary iron. Hazel and Southgate (1985) found that between 12% and 25% of the iron in meat and fish was absorbed, compared with 2–7% from a variety of vegetable sources. Rice and spinach were at the bottom of the vegetable range and soya beans at the top. The higher fibre and phytate content of vegetarian diets would also tend to reduce iron absorption,

although high levels of vitamin C in vegetarian diets are an important promoter of iron absorption. Several groups have reported lower iron stores in lacto-vegetarians, although there is little evidence of increased clinical anaemia. Some studies have reported lower haemoglobin levels in lactovegetarian children and adults despite their having higher iron intakes than omnivores. This would suggest that poor iron absorption may be a particular problem with some vegetarian diets (see Johnston, 1999).

Over 40% of the zinc in omnivore diets comes from meat and fish and a further 20% comes from dairy foods. There is no persuasive evidence of unsatisfactory zinc status amongst vegetarians.

Vegetarian diets and nutritional guidelines

A vegetarian diet has been reported by a variety of authors, and using a variety of vegetarian groups, to be associated with:

- lower body weight and reduced prevalence of obesity
- lower blood pressure
- lower prevalence of osteoporosis
- reduced risk of coronary heart disease
- reduced incidence of certain cancers, particularly bowel cancer
- lower all-cause mortality.

It is very difficult to determine to what extent these associations are wholly or partly due to the lack of meat consumption *per se*. Caucasian vegetarians in the UK tend not to be from the lower social groups, many of them adopt vegetarianism for health reasons and thus they are likely to pay particular attention to the healthiness of their diet and their lifestyle in general. Those who are vegetarian for cultural or religious reasons also tend to adopt other practices that are considered healthful, e.g. the avoidance of alcohol and tobacco. American Seventh Day Adventists are a frequently studied group because many are vegetarians. They do, indeed, have lower rates of the diseases listed above, but they also tend to avoid alcohol and tobacco and are perhaps different in many other lifestyle characteristics from Americans in general. Some studies have compared the health of meat-eating and non-meat-eating Seventh Day Adventists and these have found that the vegetarians have lower rates

- The risk of dietary inadequacy in vegetarians is partly dependent upon the level of dietary restriction.
- Well-planned vegetarian diets are more varied and have higher levels of adequacy than the diets of many omnivores.
- Vegetarian diets are likely to be less energy dense than omnivorous diets and Western vegetarians tend to be lighter and leaner than omnivores.
- A vegan diet can restrict the growth of children, although the effect is usually small and temporary.
- Vegan diets reduce growth rates in children because they are low in fat and of low energy density rather than because of protein or other nutrient deficiency.
- If energy intakes are adequate, most diets, even vegan diets, should have ample supplies of protein and essential amino acids.
- If plenty of coloured fruits and green vegetables are consumed, vegetarian diets should have ample vitamin A in the form of carotenoids, even though retinol is only found in animal foods.
- Plant foods theoretically contain no vitamin B₁₂.
- Dietary deficiency of B₁₂ is uncommon because:

requirements are extremely small

omnivores have stores sufficient for several years

many vegetarians take supplements or eat supplemented foods

many plant foods contain some B₁₂ as a contaminant, e.g. from microbes.

- Vitamin D is not naturally present in vegetable foods, although some are supplemented.
- For most people, the principal source of vitamin D is endogenous synthesis in the skin when it is exposed to summer sunlight.
- Vegetarian diets lack haem iron, which is the most readily available form of dietary iron.
- The iron in vegetarian diets is likely to be less well absorbed than that in omnivorous diets and several surveys suggest that vegetarians have lower body iron stores than omnivores.
- Vitamin C is an important promoter of non-haem iron absorption.
- Animal foods are important dietary sources of riboflavin, calcium and zinc, but there is no persuasive evidence of inadequate status for these nutrients in vegetarians or vegans.

of coronary heart disease, cancer at several sites and a reduced risk of diabetes.

Compared to the UK population in general, Caucasian vegetarians in the UK would probably tend to:

- be more health conscious
- drink less alcohol
- make less use of tobacco
- take more exercise
- be lighter and less likely to be obese
- be less likely to come from the lower manual social classes
- be more likely to be female.

Compared to omnivores, the diets of vegetarians, and vegans in particular, tend to be lower in fat, satu-

rated fat, cholesterol and alcohol, but higher in fruits and vegetables and their associated antioxidants, complex carbohydrates, fibre and polyunsaturated fatty acids.

To assume that any differences in mortality or morbidity patterns of vegetarians are largely due to the avoidance of meat or animal foods *per se* is probably premature. Non-dietary risk factors may be very different in vegetarian and non-vegetarian groups. There are also numerous compositional differences between the diets of vegetarians and omnivores. Some of these are almost inevitable, but a thoughtfully constructed omnivorous diet could match even a good vegan diet in many respects. The key question is not whether vegetarians are healthier than omnivores, but how would the health records of vegetarians compare with

- Affluent Western vegetarians tend to:
 - be leaner than omnivores
 - have reduced levels of several risk factors and a lower prevalence of several chronic diseases
 - have reduced all-cause mortality.
- There are numerous social, lifestyle and dietary differences between affluent vegetarians and the rest of the population that may contribute to the observed health differences between them.
- Adopting a more vegetarian diet and lifestyle should produce more health benefits than simple meat avoidance.

omnivores of similar racial and social backgrounds who have similar smoking, drinking and activity patterns and who take equal care in the selection and preparation of a healthy diet? The point of this discussion is to suggest to those reluctantly contemplating vegetarianism on health grounds alone that giving up meat or 'red meat' may well, by itself, be ineffective, but that adopting some of the lifestyle and dietary practices characteristic of vegetarians will almost certainly be beneficial. This point is even more important for nutrition educators seeking to influence the behaviour of others. The discussion in the next section on the health of British Asians makes the point that vegetarianism and good health statistics are not invariably linked. Johnston (1999) and Hallas and Walker (1992) have written useful scientific reviews of vegetarianism; Thorogood (1995) has reviewed epidemiological studies of the relationship between vegetarianism and chronic disease.

- economic influences upon food choices
- cultural and religious factors that influence food selection and dietary practices.

Those offering dietary or other health education advice to people from a different cultural background need to carefully identify the particular needs and priorities of the target group and the reasons for them. This should enable them to determine what advice is appropriate from the scientific viewpoint. However, before they attempt to translate these scientific goals into practical advice and recommendations, they need to assure themselves that they are aware of and understand the cultural and social determinants of the group's practices. Only by doing this can they ensure that their advice is compatible with the beliefs and normal practices of the group.

There is a general tendency for the health of racial minority groups living in industrialized countries to be inferior to that of the majority population. Indeed, one of the health promotion goals for the American nation is to 'reduce the health disparity among Americans' (DHHS, 1992). If improved nutrition is credited with being a major factor in improving the health and longevity of the majority population, one would have to assume that some of this disparity is due to nutritional improvements lagging behind in minority groups. Racial minorities often tend to be economically and socially disadvantaged as compared to the majority population and this may also be a factor in some of these ethnic disparities in health. One of the most striking inequalities in the health of Americans is that between low-income groups and all other groups (DHHS, 1992).

RACIAL MINORITIES

Introduction and overview

Williams and Qureshi (1988) have reviewed the nutritional aspects of the dietary practices of racial minorities in the UK. Several topics and issues discussed in Chapter 2 are also of relevance when considering the nutrition of racial minorities:

- the effects of migration upon dietary habits and nutrition

The health and nutrition of particular minority groups

The health record of Native Americans is particularly poor, with two nutrition-related risk factors, alcohol abuse and obesity, clearly identified as important. Death rates from accident, homicide, suicide and cirrhosis of the liver are all much higher in Native Americans than in other Americans and are all associated with alcohol abuse. Rates of diabetes are also very high amongst Native Americans, with more than 20% of the members of many tribes affected by the disease. The increasing rates of diabetes in American Indians have paralleled the increasing rates of obesity: the proportion of overweight adults varies between a third and three-quarters, depending upon the tribe. Clearly, nutritional factors, including alcohol abuse, are major determinants of the poor health statistics of this minority group (DHHS, 1992).

In 1987, American blacks had a life expectancy of 69.4 years, compared with 75 years for the American population as a whole, and this gap widened during the first half of the 1990s. Mortality rates for heart disease are considerably higher for American blacks than for whites. The poverty rate amongst blacks is also three times that of whites and, when heart disease rates of blacks and whites are compared within income levels, the black rates are actually lower than those for whites. A good proportion of the poor health record of some American racial groups seems to result from their generally lower socioeconomic status (DHHS, 1992).

In the UK, the two largest racial minority groups are Afro-Caribbeans and British Asians. The Afro-Caribbeans originated from the former British colonies in the West Indies and large-scale immigration of this group started in the 1950s. British Asians are a racially and culturally mixed group and most of the immigration has occurred since 1960. They originate from the countries of Southern Asia (India, Pakistan, Bangladesh and Sri Lanka) or from the Asian communities of East Africa.

British Asians are made up of three major religious groups: Hindus, Sikhs and Muslims. Rates of vegetarianism are high. Hindus are usually lactovegetarians and they do not usually eat Western cheese. Sikhs do not eat beef, rarely eat pork and many are lactovegetarian. Muslims should not eat pork or consume alcohol, and any meat or animal fat should

have come from animals slaughtered and prepared according to Islamic ritual.

The potential health problems of vegetarianism, discussed earlier in the chapter, would thus apply to many within this group. For example, Robertson *et al.* (1982) suggested that nutritional deficiencies of folic acid and vitamin B₁₂ were virtually confined to the Asian population in the UK. They suggested that the practice of boiling milk with tea leaves when making tea destroys much of the B₁₂ present in the milk, and strict Hindus would have no other apparent source of B₁₂ in their lactovegetarian diet. They also suggested that prolonged gentle heating of finely cut-up foods also destroys much of the folic acid initially present in some Asian diets.

Mini-epidemics of rickets and osteomalacia occurred in the 1960s and 1970s amongst Asian children and women in the UK. For example, a 5-year-old Asian child living in Glasgow between 1968 and 1978 had a 3.5% chance of being admitted to hospital with rickets before the age of 16 years, whereas the equivalent risk for a white child was negligible (Robertson *et al.*, 1982). Vitamin D deficiency has also been reported as common amongst pregnant Asian women in the UK and this has serious implications for the health of their babies. Several factors probably contribute to this high prevalence of vitamin D deficiency (see list below).

- Avoidance of those few foods that contain good amounts of vitamin D, e.g. fatty fish, liver, fortified margarine and infant foods and eggs.
- A cultural tendency not to expose the skin to sunlight, perhaps coupled with pigmentation of the skin, would be expected to reduce endogenous vitamin D synthesis. This endogenous synthesis is normally regarded as the primary source of vitamin D.
- High levels of fibre and phytate in chapattis (a traditional unleavened bread) might impair calcium absorption and perhaps also prevent enterohepatic recycling of vitamin D (Robertson *et al.*, 1982).

The problem appears to be most prevalent amongst recent immigrants, with peaks of rickets cases associated with new waves of immigration. Robertson *et al.* (1982) suggested that the problem diminishes because of acculturation and partial adoption of a Western diet. They concluded that the most

practical way of addressing the problem was by increased awareness through health education campaigns and by the provision of vitamin D supplements to those 'at risk'.

The relative prominence of the different causes of mortality also varies quite considerably between different ethnic groups in the UK. The incidence of hypertension and strokes is high in the Afro-Caribbean population of the UK. It is also high amongst black Americans, with stroke mortality amongst black American men being almost twice that in the total male population. Africans and Afro-Caribbeans may have an ethnic predisposition to hypertension due to a hypersensitivity to salt (Law *et al.*, 1991a).

The rates of coronary heart disease amongst the Asian groups in the UK are higher than in the white population, despite the high prevalence of vegetarianism amongst British Asians. Age-standardized mortality is 40% higher for Asians in England and Wales than for the general population. A similar trend has been reported for Asians living in other industrialized countries. McKeigue *et al.* (1985) compared the dietary and

other risk factors for coronary heart disease amongst Asians and non-Asians living in London. They confirmed that the rates of coronary heart disease were much higher amongst the Asians and that the excess morbidity and mortality was common to the three major religious groups, despite considerable differences in dietary and lifestyle characteristics amongst the three. They found that the diets and lifestyles of Asians differed from those of their white neighbours in the following ways.

- The Asians had lower intakes of saturated fat and cholesterol and had lower serum cholesterol concentrations.
- The Asians smoked fewer cigarettes and drank less alcohol.
- Vegetarianism was much more common amongst the Asians.
- The Asians had lower rates of bowel cancer than the white population, even though high rates of bowel cancer and coronary heart disease tend to go together in populations.

- The health statistics of racial minorities are often worse than those of the majority population.
- Native Americans are much more likely to be obese and to develop type II diabetes than other Americans.
- Obesity and alcohol-related problems are major contributors to the generally poor health record of Native Americans.
- Black Americans have shorter life expectancy and higher heart disease mortality than white Americans.
- The relatively poor health statistics of black Americans can be largely attributed to their relative poverty, and this probably applies to other minority groups in industrialized countries.
- Afro-Caribbeans and South Asians are the largest racial minority groups in the UK.
- The potential problems of a lactovegetarian diet are relevant to many British Asians who are lactovegetarian for religious reasons.
- Rickets has been prevalent amongst British Asians due to the combined effects of low dietary availability and low endogenous synthesis because of inadequate exposure of the skin to sunlight.
- People of African and Caribbean origin have a high prevalence of hypertension, perhaps because they are genetically sensitive to the hypertensive effects of salt.
- Coronary heart disease is very prevalent amongst British Asians, even amongst those who are lactovegetarian and even though they generally have lower exposure to several established risk factors than the white population.
- High rates of type II diabetes and abdominal obesity amongst British Asians probably account for their susceptibility to coronary heart disease.
- Inactivity and excessive energy intake probably trigger the insulin resistance and abdominal obesity of British Asians.

McKeigue *et al.* (1991) reported high levels of insulin resistance amongst British Asians, and the prevalence of diabetes was more than four times higher than in the population as a whole. They also reported a very high incidence of central obesity in this Asian population (i.e. with a high waist : hip ratio). Vegetarianism amongst the Asian population does not seem to protect against overweight and obesity, as it usually does for Caucasian vegetarians. McKeigue *et al.* (1991) suggested that the only known environmental influences associated with insulin resistance are high caloric intake and inactivity. They therefore proposed that these Asian groups are adapted to life in environments in which high levels of physical activity and frugal use of food resources have been traditionally required. When these groups migrate to countries where levels of physical activity are low and caloric intakes high, this leads to insulin resistance, central obesity and other associated metabolic disturbances, including maturity-onset diabetes. This, in turn, predisposes them to coronary heart disease.

NUTRITION AND PHYSICAL ACTIVITY

Fitness

It does not require scientific measurement to convince most of us that regular exercise leads to increased levels of physical fitness. Any sedentary person who undertakes a programme of regular, vigorous exercise will soon notice that he or she can perform everyday activities like garden and household chores, climbing stairs, walking uphill or running for a bus with much less breathlessness and pounding of the heart than previously. Simple personal monitoring would show a reduction in resting heart rate and smaller increases in pulse rate in response to an exercise load.

Scientific appraisal of the benefits of physical activity upon long-term health or the effects of different training schedules upon fitness requires a precise definition of fitness and an objective method of measuring it. This allows changes in fitness levels to be monitored during training and allows the fitness levels of different individuals or populations to be compared. The scientific criterion normally used to define fitness is the **aerobic capacity**, or the rate of

oxygen uptake of the subject when exercising maximally – the **$\text{VO}_2 \text{ max}$** . This $\text{VO}_2 \text{ max}$ is a measure of the functional capacity of the cardiopulmonary system to deliver oxygen to the body muscles and other tissues. Direct measurement of oxygen uptake in maximally exercising subjects is feasible, but it requires relatively sophisticated laboratory equipment and trained personnel. It also requires subjects to perform sustained exercise to the limit of their capacity and this may be undesirable, and sometimes dangerous, in untrained people. The heart rate measured at several submaximal exercise levels is therefore often used to derive a measure of aerobic capacity. In one widely used test, subjects cycle on a static bicycle ergometer at defined work loads. The external work performed by the subject is calculated from the speed of cycling and the resistance applied to the bicycle fly-wheel. In the steady state, heart rate is assumed to equate with oxygen uptake. Three graded levels of work are performed and the steady-state heart rate is measured at each. The calculated work rate is then plotted against the heart rate on a graph and the best straight line drawn through the three points (Figure 16.1). By extending this straight line, the graph can be used in either of the ways listed below.

- To predict the work rate at the predicted maximum heart rate ($220 - \text{age in years}$) for the subject, the **maximum physical work capacity**, **PWC_{max}**.
- To predict the work rate at a defined heart rate, e.g. the physical work capacity at 170 beats per minute, the **PWC₁₇₀**.

The rate of 170 beats per minute is approximately 85% of the maximum heart rate in young adults and using this value involves less extrapolation than extrapolating to maximum heart rate. In older subjects, a lower defined heart rate would be used. Similar tests are available that involve walking or running on a treadmill rather than the use of a static bicycle.

Conditioning of the cardiopulmonary system to produce measurable improvements in aerobic capacity requires that the system be significantly exerted for sustained periods. In *Healthy people 2000* (DHHS, 1992), it has been assumed that, in order for exercise to increase or maintain aerobic fitness, it must comprise three sessions each week in which heart rate is raised to at least 60% of maximum for at

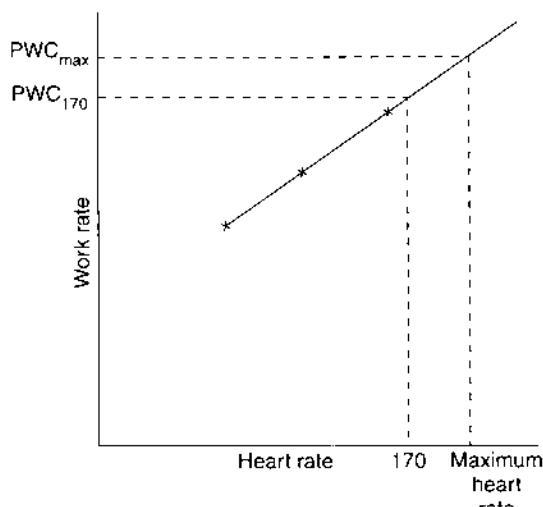


Figure 16.1 A graph of work rate against heart rate. The straight line drawn through the three measured points can be extrapolated to predict the work rate at 170 beats per minute (the PWC₁₇₀) or the work rate at the estimated maximum heart rate (the maximum physical work capacity, PWC_{max}). These are both used as indices of aerobic fitness. Maximum heart rate = 220 – age in years.

least 20 minutes' duration. Sometimes, a figure of 70% of maximum heart rate is used.

Regular exercise not only produces measurable increases in aerobic capacity, but also produces other rapid and readily demonstrable benefits in the short term, such as those listed below.

- Training increases muscle strength.
- Training increases endurance, i.e. the maximum duration for which one is able to sustain a particular activity. Both the endurance of individual trained muscle groups and the endurance of the cardiopulmonary system are increased.
- Training improves or maintains joint mobility or flexibility, i.e. the range of movements that subjects are able to safely and comfortably perform is increased.
- Training increases, or at least maintains, muscle mass. During bouts of sustainable activity such as aerobics, jogging or walking, the supply of oxygen to the muscles is sufficient for the muscle to produce energy by the normal **aerobic** (oxygen-requiring) processes – hence **aerobic exercise**. The most effective exercises for increasing muscle mass are high-intensity, short-duration activities for which the effort involved is close to the max-

imum, e.g. lifting a heavy weight, pushing against a resistance, press ups, jumping, climbing stairs or sprinting. Such exercises are so intense that the oxygen supply to the muscle is insufficient to produce all the energy needed for the intense activity and so the muscle temporarily 'tops up' the energy supply by switching to energy-producing processes that are **anaerobic** (not requiring oxygen) – hence **anaerobic exercise**.

Several of these latter benefits of exercise accrue at intensities and durations of activity that are considerably below those necessary to produce measurable changes in aerobic capacity. These latter effects of increased strength, endurance and flexibility may be of even greater importance to some subjects than increases in aerobic fitness, e.g. in the elderly. Note that in Chapter 14 it was suggested that much of the loss of physical capacity and functional disability associated with ageing is not a result of the ageing process *per se*, but is largely the result of inactivity. Even in extreme old age, resistance training resulted in measurable increases in strength and functional ability (Fiatarone *et al.*, 1994). A 40-minute walk three times a week did increase the aerobic fitness of a group of Texans in their 70s and, after 6 months, they had aerobic capacities that were typical of 50-year-old Texans.

There is also a substantial body of evidence suggesting that participation in regular physical activity produces psychological as well as physiological benefits. It improves the general sense of well-being and self-confidence, improves the ability to cope with stress and may even reduce the frequency and severity of personality disorders. Some of these psychological benefits may well be a consequence of the physiological conditioning (see Hayden and Allen, 1984; Fentem, 1992).

CURRENT LEVELS OF PHYSICAL ACTIVITY AND FITNESS

The average levels of physical activity of the populations of the wealthy industrialized countries have, by general agreement, declined in recent decades. Car ownership, workplace automation and labour-saving aids in the home and garden have drastically reduced the amount of obligatory physical exertion that most people are now required to undertake. Increased voluntary leisure-time activities have not been enough to offset this reduction in obligatory activity.

The average Briton now watches twice as much television as in the 1960s and many people spend more time watching television than they do at work or school. Between 1975 and 1995, the average distance walked by British adults fell by about 20%.

The increased levels of overweight and obesity in the UK and USA in recent decades have occurred despite a sharp decline in average energy intakes and this gives indirect evidence of the scale of this reduction in activity (discussed in Chapter 8). Energy expenditure (i.e. activity) has been declining faster than the decline in energy intake.

There are estimates of activity levels of the American population in *Healthy people 2000* (DHHS, 1992). The examples listed below were estimates for 1985 and so these are likely to be even lower now.

- In 1985, 12% of American over the age of 18 and 66% of 10–17 year olds engaged in vigorous exercise of sufficient frequency, intensity and duration to promote the development of aerobic fitness. In those over 18 with low incomes, this assumed rate was only 7%.
- 22% of over-18s were assumed to engage in light to moderate physical activity (i.e. reaching less than 60% of maximum heart rate) for at least 30 minutes on at least five occasions each week.

- 24% of over-18s were assumed to engage in no leisure-time physical activity, with figures of 32% for low-income people and 43% for the over-65s.

Table 16.1 gives a summary of activity levels determined by interview in a health survey of English adults (White *et al.*, 1993). More than half of men and women reported engaging in fewer than three periods of even a moderate level of activity each week. Around 20% reported engaging in no periods of moderate or vigorous activity in a 4-week period. The activity levels of women were, on average, less than those of men and there was a marked tendency in both sexes for activity level to decline with age. More than a third of 65–74 year olds had no period of even moderate activity in a 4-week period. Subjects with higher educational qualifications were more likely to be at the highest activity levels (levels 4 and 5 in Table 16.1).

The Allied Dunbar National Fitness Survey (1992) also found very low levels of reported physical activity and low levels of aerobic fitness amongst a representative sample of 4000 English adults. This survey report suggested that 70% of men and 80% of women had lower levels of physical activity than would be necessary to achieve aerobic fitness. For around one-third of men and two-thirds of the women, walking at 3 m.p.h. up a gradient of 1:20 would be difficult, with

Table 16.1 Percentage of English adults at various levels of reported activity

Activity level	Sex	Age group (years)				All
		16–24	35–44	45–64	65–74	
4 and 5 (highest)	M	37	24	9	2	20
	F	22	11	8	1	12
3	M	23	29	38	31	29
	F	28	34	42	21	30
1 and 2	M	31	33	33	36	32
	F	39	48	34	35	37
0 (lowest)	M	8	15	20	31	19
	F	10	8	16	43	21

Activity levels:

1 and 5 = 12 or more occasions in 4 weeks of vigorous or mixed vigorous/moderate activity.

3 = 12 or more occasions of moderate activity in 4 weeks.

1 and 2 = 1–11 occasions of at least moderate activity in 4 weeks.

0 = No occasions of moderate or vigorous activity in 4 weeks.

Moderate activity means requiring an energy expenditure of 5+ kcal (20 kJ) per minute and vigorous activity 7.5+ kcal (30 kJ) per minute.

Moderate activities: walking a mile at brisk or fast pace; heavy gardening work such as digging or cutting large areas of grass by hand; heavy DIY work such as knocking down walls or breaking up concrete; heavy housework such as walking with heavy loads of shopping or scrubbing/polishing floors by hand.

Vigorous activities: running or jogging; playing squash; several other sporting activities that cause the subject to breathe heavily or sweat a lot, such as aerobics, tennis, weight training and swimming.

Data source White *et al.* (1993).

even higher proportions in the over-65s. A large number of women over the age of 35 years would find even walking at 3 m.p.h. on the flat severely exerting and would need to rest or slow down after a few minutes. Many elderly women do not have enough strength in their legs to rise from a chair without using their arms. Many elderly and even middle-aged people do not have enough shoulder flexibility to allow them to wash their own hair in comfort.

There is also convincing evidence that the activity levels of children in Britain have decreased sharply in recent decades. As previously noted in Chapter 8, Durbin (1992) has suggested that the average energy intakes of teenagers were up to a third higher in the 1930s than they were in the early 1990s. Several other studies have also found a reduction in recorded energy intakes of various age groups of children in recent decades. Over these same decades, British children have got taller and fatter and this must indicate a decrease in their energy expenditure via physical activity. It was argued in Chapter 8 that the rising prevalence of childhood obesity is primarily a result of declining activity rather than of dietary changes *per se*. There was a 20% decrease in walking amongst British children in the 10 years up to 1995 and a 25% decrease in the amount of cycling. Between 1975 and 1995, the annual distance walked by children in the

11–15-year category dropped a third, from 248 miles in 1975 to 199 miles in 1994. Armstrong *et al.* (1990) used continuous heart-rate monitoring to assess the activity levels of a group of English teenagers. Very few of these children recorded three or more periods of 20 minutes' duration when their heart rate rose above 139 beats per minute, and more than three-quarters had no such 20-minute periods.

In the USA, *Healthy people 2000* set a series of very specific targets for increases in the physical activity levels of different sections of the American population for the year 2000 (DHHS, 1992). Some of these targets are listed below, although it is highly improbable that any of them have been realized.

- 20% of over-18s and 75% of 6–17 year olds should engage in vigorous physical activity that promotes aerobic fitness, i.e. three or more 20-minute periods each week of activity intense enough to raise the heart rate to 60% of maximum.
- At least 30% of those aged 6 and over should engage in light to moderate physical activity for at least 30 minutes each day (i.e. activities that do not raise heart rate to 60% of maximum, such as walking).
- No more than 15% of people aged over 6 years, 22% of over-65s and 17% of low-income people should fail to engage in any leisure-time physical activity.

- Scientific assessment of fitness uses some measure of aerobic capacity such as the maximum oxygen uptake ($\text{VO}_2 \text{ max}$).
- Aerobic capacity is usually extrapolated from pulse rates measured at submaximal exercise loads.
- Measurable increases in $\text{VO}_2 \text{ max}$ require regular and significant exertion for sustained periods, typically three weekly sessions of 20 minutes' duration when the heart rate is raised to 60% of maximum.
- Training also maintains muscle mass, increases strength, endurance and flexibility and produces beneficial psychological effects.
- These other benefits of exercise are discernible at lower intensities than that required to produce increases in $\text{VO}_2 \text{ max}$.
- The overall activity levels of affluent populations have decreased substantially in recent decades.
- Increasing mechanization has greatly reduced occupational energy expenditure and the energy required to perform the tasks of everyday living.
- The Allied Dunbar National Fitness Survey has confirmed the extreme inactivity and low levels of aerobic fitness, flexibility and strength of many adult Britons.
- Other evidence suggests that children are also much less active than they were even 25 years ago.

In *The health of the nation* (DH, 1992), the stated intention of the British Government is 'to develop detailed strategies to increase physical activity' in the light of the activity and fitness levels found in the Allied Dunbar National Fitness Survey (1992).

In *Healthy People 2000*, it was also seen as particularly important to encourage children to participate in physical activities that are likely to be carried through into adulthood. Participation in team sports and group games is rarely carried through into adulthood, whereas some other activities are much more likely to be continued, e.g. swimming, cycling, jogging, walking, dancing and games requiring only one or two participants. The increased provision of facilities for participating in such activities was also seen as a necessary component of this encouragement.

Long-term health benefits of physical activity

In the short term, increased physical activity improves aerobic capacity, endurance, strength, flexibility and various aspects of psychological well-being. Over a period of weeks or months, it also leads to changes in appearance and body composition, increased lean to fat ratio and improved muscle tone. It is inevitable that low levels of physical activity will lead to reduced energy requirements and predispose people to becoming overweight and obese. In the longer term, regular exercise contributes to reduced morbidity and mortality from several of the diseases of industrialization and therefore increases both life expectancy and years of healthy life. Some of the health benefits of regular exercise and increased fitness are summarized in Table 16.2.

Regular, sustained and moderate (aerobic) activity should aid in the long-term control of body weight. Inactivity undoubtedly predisposes to becoming overweight and obese. Regular physical activity is more likely to be effective as a preventative measure rather than as a relatively short-term measure to cure established obesity. An exercise programme is nonetheless an important adjunct to the dietary management of obesity. It increases the chances that any short-term weight losses will be sustained in the longer term. (See Chapter 8 for a discussion of the role of inactivity in causing obesity and the role of exercise in its treatment.)

Table 16.2 Some of the demonstrable benefits of regular exercise and increased physical fitness

Increased activity and fitness is associated with increased bone density in children and younger adults (e.g. Pocock <i>et al.</i> , 1986) and reduced levels of osteoporotic fracture in elderly people (e.g. Wickham <i>et al.</i> , 1989)
Increased physical activity (Paffenbarger <i>et al.</i> , 1986) and high level of aerobic fitness (Blair <i>et al.</i> , 1989) are associated with reduced all-cause mortality in men and women; these benefits are detectable even in the elderly and may even become more pronounced with age
Middle-aged male civil servants who participated in vigorous leisure-time activities had reduced risk of coronary heart disease as compared to those who did not (Morris <i>et al.</i> , 1980)
Men with active jobs have fewer heart attacks than those in similar but less active jobs (see Leon, 1985)
There are psychological benefits resulting from increased physical activity (Fentem, 1992); these may be direct effects of the physical activity, or indirect effects caused by the improved physical capabilities and conditioning that increased activity bring with it
Exercise maintains the ability to perform the activities of daily living as people become elderly and so maintains the morale, self-esteem and quality of life of older people (see Bennett and Morgan, 1992)
Increased activity leads to increased energy expenditure, both directly and as a result of the associated higher lean body (muscle) mass (particularly important in the elderly)
Exercise increases the anti-atherogenic high-density lipoprotein cholesterol concentration in blood (Macnair, 1994)
Those who are physically fit seem to be protected against the consequences of being overweight or even mildly obese (Lee <i>et al.</i> , 1998; discussed in Chapter 8)

Exercise increases and maintains bone density and therefore protects against osteoporosis. For example, Pocock *et al.* (1986) found physical fitness to be a major determinant of bone density in two regions frequently associated with osteoporotic fracture, the neck of the femur and the lumbar spine. Note, however, that extremes of physical exercise in women that are accompanied by very low body weight and amenorrhoea almost certainly accelerate bone loss and predispose to osteoporosis (see Chapter 13).

Regular physical activity is protective against cardiovascular disease. This may be a direct effect, e.g. by strengthening heart muscle, or an indirect

consequence of its effects on other risk factors. Physical activity reduces the tendency to overweight, insulin resistance and maturity-onset diabetes, it lowers blood pressure and has beneficial effects on blood lipid profiles (e.g. Blair *et al.*, 1989). As far back as 1980, Morris *et al.* found that middle-aged men who participated in vigorous leisure-time activity had lower rates of both fatal and non-fatal coronary heart disease than those who did not. This was based on an 8.5-year cohort study of 18 000 middle-aged, male, British civil servants. This apparent protective effect of exercise was maintained in all age groups and in subgroups classified according to smoking behaviour, family history of coronary heart disease and even in those with existing hypertension and subclinical angina. They concluded that vigorous exercise has a protective effect in the ageing heart against ischaemia and its consequences.

Several studies have looked at the comparative mortality of different occupational groups with differing levels of work-related activity (reviewed by Leon, 1985). Farmers and farm labourers have been the subject of several such studies because of the perception that their work requires considerable activity. Studies in California, North Dakota, Georgia and Iowa have all found lower levels of heart disease amongst farm workers than amongst other inhabitants of these regions (Pomrehn *et al.*, 1982). Less smoking and higher levels of activity were identified as differences between farmers and non-farmers in these studies. Pomrehn *et al.* (1982) analysed all death certificates in the state of Iowa over the period 1964–1978 for men aged 20–64 years and compared the death rates of farmers and non-farmers in Iowa over that period. They found that farmers had lower all-cause mortality and mortality from heart disease than non-farmers. They also compared the lifestyles and dietary characteristics of a sample of farmers and non-farmers living in Iowa in 1973. They found that the farmers were significantly less likely to smoke (19% versus 46% of non-farmers), were more active, were fitter and drank less alcohol. The farmers in their sample also had significantly *higher* total serum cholesterol levels. They concluded that the reduced total and heart disease

mortality of farmers could be attributed to a lifestyle that included regular physical activity and much less use of alcohol and tobacco than non-farmers.

In a 16-year cohort study of 17 000 Harvard graduates, ranging in age from 35 to 74 years, Paffenbarger *et al.* (1986) found that reported exercise level was strongly and inversely related to total mortality. They used reported levels of walking, stair climbing and sports play to calculate a relative activity index. The relative risk of death of those with an index of around 3500 kcal (14 MJ) per week was around half that of those with an index of less than 500 kcal (2 MJ) per week. They found that mortality rates amongst the physically active were lower, with or without consideration of other confounding factors, such as hypertension, cigarette smoking, body weight or weight gain since leaving college or early parental death.

In an 8-year cohort study of over 13 000 men and women, Blair *et al.* (1989) found a strong inverse correlation in both sexes between measured fitness level and relative risk of death from all causes. The apparent protective effect of fitness was not affected by correcting for likely confounding variables. In Chapter 8 there is a discussion of the data of Blair and his colleagues (Lee *et al.*, 1998), which suggests that people who are overweight or mildly obese but have high levels of aerobic fitness do not have excess mortality as compared to those who have similar fitness but are lean. Aerobic fitness seems to offset the usual health risks associated with being overweight or obese.

Diet as a means of improving physical performance

BODY WEIGHT, ATHLETIC PERFORMANCE AND ENERGY INTAKE

Athletes in training expend considerably more energy than the average person for whom the DRVs and RDAs are intended. Their total energy expenditure during training and competition may be well over twice their basal metabolic rate (BMR), as compared to the 1.4 or 1.5 times BMR assumed when

- Some of the long-term benefits of regular exercise and increased fitness are summarized in Table 16.2.

DRVs are set. During the 10 days of the Tour de France cycle race, athletes may expend 3.5–5.5 times their BMR (see Wilson, 1994). One would also expect, therefore, that the energy and nutrient intakes of those in training would be considerably higher than those of the average person. Even though the requirements for some nutrients may be increased by heavy exercise, one would therefore expect that athletes should still be unlikely to have problems of dietary inadequacy.

In some events, in which performance is judged upon aesthetic appeal (e.g. gymnastics, ice skating and dancing), thinness may be seen to increase the chances of a favourable verdict by judges. In some sports (e.g. middle and long distance running, jumping and cycling), being light may improve performance. In some sports, competitors are divided up into weight categories and competitors of both sexes may starve themselves to make the weight or even dehydrate themselves. These pressures may cause some athletes to consume much less than one would expect. Many female athletes, in particular, have much lower energy intakes than might be expected from their activity levels. According to Wilson (1994), the average recorded intakes of female runners in several studies were only about 100 kcal (420 kJ) per day higher than those recorded in non-athletes. This means that, unless their diets are nutrient dense, even small increases in nutrient requirements due to exercise may become significant in athletes with such unexpectedly low energy intakes.

Many female athletes have such low body weights that they cease to menstruate and have low levels of circulating sex hormones. This may adversely affect their bone health. It may make them more prone to stress fractures and increase their risk of developing osteoporosis in later life. Several studies have suggested that there is a high prevalence of eating disorders amongst female athletes competing in events in which low body weight might be considered advantageous (Wilson, 1994).

ENERGY AND NUTRIENT REQUIREMENTS

Adequate intakes of dietary energy and of micronutrients are prerequisites for maximum athletic performance. There is, however, no convincing evidence that any particular vitamin or mineral supplements boost performance in already well-nourished individuals.

There may be small increases in the requirements for some water-soluble vitamins in those undertaking vigorous training schedules, e.g. vitamin C, thiamin, niacin and riboflavin. Nevertheless, a well-balanced diet should contain sufficient of these vitamins, especially as total food intake will be higher during training. Some vitamins are toxic in excess (e.g. vitamins A and D) and excesses of some others may actually impair performance (e.g. niacin).

Iron status has frequently been reported to be low and iron requirements to be increased in athletes, particularly endurance athletes. This situation with regard to iron is complicated by the fact that endurance training produces a haemodilution effect similar to that described in pregnancy in Chapter 14. Training increases plasma volume and total circulating red cell mass, but the relatively larger increase in plasma volume causes blood haemoglobin concentration to fall. True anaemia (low circulating haemoglobin) and iron deficiency (low amounts of iron stored as ferritin) in athletes are usually due to low dietary intake. This would be particularly likely in athletes, often female athletes, who restrict their energy intake to keep their body weight low. Exercise may increase iron losses and it may impair the increase in iron absorption that normally occurs when iron stores are low. Training may thus slightly increase iron requirements, and poor iron intakes during training may precipitate iron deficiency or even anaemia in some athletes, especially female ones. Anaemia would certainly impair athletic performance, particularly in endurance events. Indeed, artificially boosting blood haemoglobin levels is used as an unfair way of trying to boost performance.

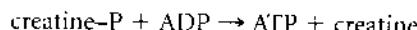
The protein requirements and optimal protein intakes of athletes and bodybuilders have been the subject of hundreds of studies stretching back over several decades. Whereas a detailed analysis of these studies is beyond the scope of this book, a few general conclusions are listed below.

- Prolonged heavy exercise does increase protein losses slightly as amino acids start to be used as substrates.
- Habitual protein intakes are almost always well in excess of requirements.
- If athletes consume sufficient energy and carbohydrate, any specific measures to boost protein intakes are unnecessary, even in ultra-endurance events.

- The large protein supplements taken by many athletes are almost certainly ineffective, both in increasing muscle mass and in boosting athletic performance.
- It was noted in Chapter 10 that excess protein intakes have been speculatively suggested to have adverse long-term consequences, e.g. an increased risk of chronic renal failure.

SUBSTRATE UTILIZATION DURING EXERCISE AND ITS IMPLICATIONS

In explosive events such as short sprints, jumping or throwing, the energy supply of the muscle during the event is mainly from **creatine phosphate** plus a contribution from the anaerobic metabolism of glycogen. Creatine phosphate is the muscle's short-term energy reserve and it can be used to convert adenosine diphosphate (ADP) to adenosine triphosphate (ATP) by transfer of a phosphate moiety:



Creatine is present in the diet in meat and fish and can be synthesized in sufficient quantities for normal circumstances from other amino acids, even if a vegetarian diet is eaten. Some athletes do take supplements of creatine, which do elevate the stores of creatine in muscles.

In the longer sprints and other events when there is more sustained and very intense activity, energy supply comes from creatine phosphate, anaerobic metabolism of glycogen plus a contribution from the aerobic metabolism of glycogen. The build up of lactate from the anaerobic metabolism of glycogen will produce fatigue and limit performance.

In endurance events, the aerobic metabolism of glucose plus a contribution from the metabolism of fat are the major sources of energy. In any endurance event or competitive sport (or even a long training session for a more explosive event), the depletion of muscle glycogen is a factor that potentially limits sustained high performance. One would expect glycogen stores in liver and muscle to be depleted after 1.5–2 hours of vigorous exercise, and depletion of muscle glycogen leads to fatigue – the so-called 'wall' that affects marathon runners. Training for endurance events increases muscle glycogen stores. Training also increases the aerobic capacity of muscles, enabling them to use fatty acids more effect-

ively. As the duration of exercise increases, so the contribution made by fatty acids to the total energy supply of the muscle increases. Once glycogen reserves have been used up, running speed is limited by the speed at which fatty acids can be metabolized aerobically.

Maximizing the glycogen stores in muscle and liver is now seen as a key element in the preparation for competition in endurance sports (i.e. longer than 1 hour of sustained heavy exercise). In some endurance sports, it may also be possible to take in extra carbohydrate during the event. In experimental studies, the maximum duration for which heavy exercise (75% of VO_2 max) can be sustained is directly proportional to the glycogen content of the muscle at the start. Exhaustion occurs when the muscle glycogen reserves are completely depleted. High-carbohydrate diets (70% of energy) during training increase the muscle glycogen content and so increase the maximum duration for which heavy exercise can be sustained. Many endurance athletes manipulate their diets and exercise schedules prior to competition in an attempt to increase their body glycogen stores – **carbohydrate loading**. This extra carbohydrate also increases muscle water content, which may not be advantageous for those in explosive events for which the size of muscle glycogen stores are not critical.

There are several theories as to the best way to increase muscle glycogen stores. One regime that is not associated with any obvious side-effects is for the athlete to eat a normal, high-carbohydrate diet during heavy training up to a week before the event. Then, in the week before competition, a mixed diet (50% carbohydrate) is eaten for the first 3 days, followed by a very high-carbohydrate diet (75% carbohydrate) during the rest of the week diet. Training intensity is wound down in the second half of the week, with a complete rest on the day before the competition. A high-carbohydrate diet should be continued during the week following competition to replenish carbohydrate stores. Carbohydrate taken in the first hour or two following the completion of the event seems to be most effective in replenishing muscle glycogen stores.

THE NEED TO PREVENT DEHYDRATION

A marathon runner competing in a warm environment may expect to lose up to 6 L of sweat during

the run and, even in a more temperate climate, sweat losses will often be in the range of 3–4 L. Prolonged exercise, particularly in a hot environment, can lead to dehydration, which produces fatigue and reduces work capacity. Profuse sweating and an inability to dissipate excess body heat in a warm, humid environment may lead to a combination of dehydration and hyperthermia, **heat stroke**, which can have serious, even fatal, consequences. Ideally, marathon runners should drink 0.5 L of fluid 15–30 minutes before competition and then take small, regular drinks (say every 3 miles), although competitive

rules may prevent drink stops in the early part of the race. Ideally, the athlete should take in enough water to replace that which is lost, rather than simply drink enough to satisfy thirst. These principles apply not just to marathon runners, but to anyone participating in events that involve long periods of heavy exertion and sweating, e.g. tennis players, long distance cyclists and soccer players.

The following referenced reviews of nutrition for exercise and sport may be useful for further reading: Eastwood and Eastwood (1988); Powers and Howley (1990); Hultman *et al.* (1999).

- Athletes in training expend much more energy than the average person for whom DRVs are intended.
- Low body weight and consequent amenorrhoea are common amongst female athletes who restrict their energy intake to enhance performance or because judges favour lean competitors.
- Amenorrhoea reduces bone density in female athletes and can make them more susceptible to stress fractures and to osteoporosis in later life.
- Requirements for some nutrients may be increased slightly by prolonged training.
- A balanced diet should provide sufficient nutrients, even during training, unless total energy intake is severely restricted.
- There is no convincing evidence that micronutrient supplements improve performance in those who are not deficient, but large excesses of some nutrients can be toxic and hinder performance.
- The iron stores of some athletes are low and this is largely due to poor iron intake.
- Endurance training produces a haemodilution effect, which may exaggerate the apparent prevalence of iron-deficiency anaemia amongst athletes.
- Prolonged heavy exercise may increase protein requirements slightly, but habitual intakes are almost always well in excess of even these increased requirements.
- There is no evidence that the large protein supplements taken by many athletes either increase performance or increase muscle mass.
- In explosive events, creatine phosphate is the principal source of muscle energy.
- Creatine is present in meat and fish and can be made from amino acids; nevertheless, creatine supplements can increase creatine stores in muscle.
- In endurance events, the depletion of muscle glycogen limits performance and so maximizing glycogen stores prior to competition is an important goal for endurance athletes.
- A high-carbohydrate diet increases muscle glycogen stores.
- Manipulating diet and training schedules during the week before competition can significantly increase glycogen stores – carbohydrate loading.
- Enormous amounts of fluid can be lost from the body during sustained heavy work or exercise.
- Dehydration produces fatigue, reduces work performance and can lead to potentially fatal heat stroke.
- Maintaining hydration during prolonged heavy exercise enhances performance, but thirst is an insensitive indicator of dehydration in these circumstances.

Part 5

The safety and quality of food

The safety and quality of food

Aims of the chapter	407	Bovine Spongiform Encephalopathy	425
Consumer protection	407	Food processing	430
Food poisoning and the microbiological safety of food	412	The chemical safety of food	434
		Functional foods	441

AIMS OF THE CHAPTER

This chapter provides an overview of factors that influence the safety and quality of food in industrialized countries and of the legal framework that is designed to try to ensure food safety and quality. The two major aims of this chapter are:

1. to give readers enough understanding of the sources of potential hazard in their food to enable them to take practical measures to improve the safety and quality of their own food
2. to help readers to make a realistic evaluation of the real hazards posed by potential threats to food safety, i.e. to identify priorities for improved food safety.

CONSUMER PROTECTION

Food law

The web pages on UK and European food law prepared by Dr David Jukes of Reading University are recommended for anyone seeking further information on this topic (www.fst.rdg.ac/foodlaw). A summary of US food legislation can be found in Tompson *et al.* (1991).

Food law has traditionally had the three major aims that are listed below.

1. *To protect the health of consumers:* to ensure that food offered for human consumption is safe and fit for people to eat.
2. *To prevent the consumer being cheated:* food laws try to ensure that verbal descriptions, labels and advertizements for food describe it honestly and that manufacturers and retailers cannot dishonestly pass off some substitute or inferior variant as the desired product.
3. *To ensure fair competition between traders:* to prevent unscrupulous traders gaining a competitive advantage by dishonestly passing off inferior and cheaper products as more expensive ones or by making dishonest claims about the merits of their product.

A further aim could be added for some current legislation governing the labelling of foods:

4. *To facilitate healthy consumer choices:* to give consumers enough information about the nutritional content of foods to allow them to make informed food choices that increase their compliance with nutrition education guidelines.

In the UK, the first comprehensive food legislation was passed in the middle of the nineteenth century and there have been regular revisions and additions to this legislation ever since. The 1990 Food Safety Act is the latest version. The first food legislation was

introduced in an attempt to combat widespread adulteration and misrepresentation of foods designed to cheat the consumer but also on occasion resulting in a product that was directly injurious to health. The sweetening of cider with sugar of lead (lead acetate) and the addition of red lead to cayenne pepper are examples of practices with serious health repercussions for the consumer. The watering of milk (often with dirty water), the addition of alum and bone ash to flour, the dilution of pepper with brick dust and the bulking out of tea with floor sweepings are all examples of practices designed to cheat the consumer. The sale of meat from diseased animals slaughtered in knackers' yards and the passing off of horseflesh as beef have also occurred.

Below are some of the main provisions of the UK 1990 Food Safety Act that illustrate a formal legal framework designed to achieve the aims listed above.

- It is illegal to add anything to food, take any constituent away, subject it to any treatment or process, or use any ingredient that would render the food injurious to health.
- It is forbidden to sell food not complying with food safety regulations. Food does not comply with food safety regulations if:

it has been rendered injurious to health by any of the above

it is unfit for human consumption (e.g. food that contains high levels of food-poisoning bacteria)

it is contaminated, e.g. by dirt, insect bodies, metal or glass fragments.

- It is illegal to sell any food not of the nature, substance or quality demanded by the purchaser. Customers are given a legal right to expect to be given what they ask for and to expect certain minimum standards for any food product.
- It is an offence to describe, label or advertise a food in a way that is calculated to mislead as to its nature, substance or quality. This is an additional requirement on top of any specific food-labelling requirements.

The 1990 Food Safety act also empowers government ministers to issue, with parliamentary approval, specific regulations to ensure food safety, e.g. setting maximum permitted levels of contaminants. There are many such specific regulations, or Statutory Instruments. For a successful prosecution

under this act, it is not necessary to prove intent upon the part of the manufacturer or retailer. The main defence against prosecution is for the defendant to show that they took all reasonable precautions and 'exercised due diligence' in trying to prevent the defect in the food.

Over the last decade or so there have been a series of highly publicized alarms in Britain about the safety of food, most notably the bovine spongiform encephalopathy (BSE) crisis discussed later in the chapter. This has undermined public confidence in the safety of British food and in the ability of the **Ministry of Agriculture, Fisheries and Food (MAFF)** to balance the interests of farmers and food producers against those of consumers when dealing with food safety issues. As a consequence of these concerns, an independent **Food Standards Agency** was established and began operating in 2000. This agency is charged with ensuring public health in relation to food. The agency is expected to consult widely, assess scientific information impartially, and give advice to ministers on legislation and controls that should ensure and improve food safety. The agency is expected to take into account the costs and consequences of any proposed actions so that they are proportionate to the risks involved.

Enforcement of food legislation is, in the UK, primarily the responsibility of Trading Standards Officers (TSOs) and Environmental Health Officers (EHOs) employed by local authorities. The EHOs are mainly concerned with issues relating to food hygiene and the microbiological safety of food; they are involved in investigating food-poisoning outbreaks. The TSOs are primarily concerned with food labelling, advertising and the chemical safety of food. These officials do not just police the food safety laws, but also give guidance and advice to local businesses relating to food safety issues.

The **Codex Alimentarius Commission** is an international body that was set up by the United Nations in 1962 to try to establish international food standards. More than 160 countries are members of this commission. Membership of the commission does not oblige members to accept the international standards and most industrialized countries have retained their own national standards. This means that, so far, the commission has been a useful forum for international debate of food standards and safety issues, but that the direct impact of its standards has been very limited.

- Food law exists to protect consumers' health and to try to ensure high standards of honesty and responsibility in food manufacturing and retailing.
- The 1990 UK Food Safety Act makes it illegal to:
 - do anything to food that would make it harmful to health
 - sell food that is unfit to eat or does not comply with food safety regulations
 - dishonestly describe a food
 - sell food that is of unacceptable quality.
- UK ministers are also empowered by the 1990 Act to make specific food regulations, subject to approval by parliament.
- An offence can be committed under the 1990 Act even if there is no intention, unless the defendant took reasonable measures and care to try to avoid the defect in the food or its description.
- An independent Food Standards Agency has recently been set up whose brief is to ensure public health in relation to food.
- Enforcement of food safety legislation in the UK is usually the responsibility of local government.
- The Codex Alimentarius Commission is an international body that sets international food standards.

Food labelling

LABELLING IN THE UK

Most packaged foods in Britain must contain the following information on the label:

- the name of the food
- a list of ingredients
- an indication of its minimum durability
- an indication of any special storage conditions or conditions of use that are required (e.g. refrigerate, eat within 3 days of opening, etc.)
- a 'lot' number so that they can be identified in the case of a problem that requires them to be withdrawn from sale
- the name and address of the manufacturer or packer of the food.

If a food is perishable and likely to become microbiologically unsafe after a relatively short storage period, it carries a '**use by**' date and it is illegal to offer for sale any food that is passed its '**use by**' date. Such foods usually need to be stored in a refrigerator or chill cabinet and should not be eaten after their '**use by**' date, e.g. cooked meats, ready meals, sandwiches, meat pies etc. Other foods that are unlikely to become hazardous on storage are labelled with a '**best before**' date. Although unlikely to represent a hazard after that date, their eating quality may be

impaired, e.g. biscuits (cookies), cakes, bread and chocolate. Foods may be legally offered for sale after their '**best before**' date provided they have not become unfit for human consumption.

In Britain and the European Union (EU), the ingredients of packaged foods must be listed in order of their prominence, by weight, in the final product. All additives must be listed and their function given. Certain foods are exempted from the requirement to carry a list of ingredients, such as fresh fruit and vegetables, cheese, butter, alcoholic drinks and any food composed of a single ingredient.

In Britain, the great majority of packaged foods also carry nutritional information on the labels. Most manufacturers voluntarily choose to include some nutritional information on their labels and such nutrition labelling is compulsory if the manufacturer makes any nutritional claim for the product, e.g. low fat or high fibre. The minimum nutritional information the food must then carry is the energy, protein, carbohydrate and fat content per 100 g, together with the amount of any nutrient for which a claim is made.

LABELS IN THE USA

American food labelling regulations underwent a major revision in the Nutrition Labelling and Education Act of 1990 (NLEA). Many food 'activists'

would regard this very formal and detailed set of regulations, amounting to 4000 pages, as an ideal model for other countries to follow. This short review is offered to readers from outside the USA as an illustration of the complexity involved when the statutory regulation of full nutritional labelling of food is attempted.

The stated aims of the US legislation were to produce a simple, understandable label that clears up confusion and helps consumers to make healthy choices and also:

To encourage product innovation, so that companies are more interested in tinkering with the food in the package, not the word on the label.

(D.A. Kessler, FDA Commissioner, 1991).

Ironically, one of the original purposes of food law was to stop producers 'tinkering with the food in the package'! This brief summary of these American food labelling regulations illustrates the complexity of trying to frame food labelling rules that cover every eventuality. The regulations have tried to ensure that all labels contain all the useful information, in a clear, comprehensible and truly informative form, and are free from misleading information and claims. Shortly after these regulations were introduced in 1992, it was suggested in the influential journal *Nature* (vol. 360, p. 499) that the resulting labels were 'all but incomprehensible' and 'leaving the consumer in need of a mainframe computer to translate it all into practical terms'.

Listed below is a sample of the important changes introduced in NLEA (source FDA, 1992).

- *Range of foods.* Nutrition labelling is mandatory on almost all packaged foods. Food stores are also asked to display nutrition information on the most frequently consumed raw fruits, vegetables, fish, meat and poultry.
- *List of ingredients.* A full list of ingredients is required on all packaged foods, even those made according to the standard recipes set by the Food and Drug Administration (FDA) such as mayonnaise, macaroni and bread. The ingredients must be listed in descending order of weight in the food, and a number of additional specific requirements are listed, e.g. all colours must be listed by name, caseinate (used in non-dairy coffee whitener) must be identified as being from milk, and the source of any protein hydrolysate must be indicated (e.g.

hydrolysed milk protein). One purpose of these specific rules is to assist people with food allergies.

- *Serving sizes.* These have to be stated and, for many products, the serving size is laid down so as to properly reflect what an adult actually eats, e.g. the 'reference amount customarily consumed' for a soft drink is 8 ounces (about 250 g). There are detailed regulations about what happens if packages contain less than two statutory servings, e.g. if the carton also contains less than 200 g/200 ml, the whole package should be treated as one serving. This measure has been introduced to prevent producers misleading consumers as to, for example, the fat, sugar or energy content of their product by using unrealistically small portion sizes.
- *Nutritional information.* The label must contain the amounts of the following per serving: calories, calories from fat, total fat, saturated fat, cholesterol, sodium, total carbohydrate, complex carbohydrate, dietary fibre, sugars, protein, vitamins A and C, calcium and iron. Other nutritional information that can be voluntarily listed (e.g. B vitamins) is also specified. The label must not only give the amount of the nutrient in absolute terms, but must also put it into the context of an average ideal diet, e.g. if a serving contains 13 g of fat, this is equivalent to 20% of the total fat in the ideal 2000-kcal diet. Not only the content of the label but also its format are laid down. A nutrition information panel from a sample label is shown in Figure 17.1.
- *Descriptive terms.* Several of the descriptive terms that are frequently used on food labels are formally defined and the food must comply with this statutory definition, e.g. 'free' as in, say, 'fat free' or 'sugar free'; 'low' as in 'low calorie' or 'low sodium'; and 'less' as in 'less fat'. Other descriptive terms are light, lite, reduced and high. Low-fat versions of products such as butter, sour cream and cheese need no longer be called imitation or substitute, but may be called simply low fat, provided they comply with certain specified criteria. This is aimed at improving the image of low-fat foods.
- *Health claims.* Five relationships between nutrients and disease are regarded as well enough established to be allowed to be used in health claims on food labels, they are:

calcium and osteoporosis

sodium and hypertension

Serving sizes are now more consistent across product lines, stated in both household and metric measures, and reflect the amounts people actually eat.

The list of nutrients covers those most important to the health of today's consumers, most of whom need to worry about getting too much of certain items (fat for example), rather than too few vitamins or minerals, as in the past.

The label will now tell the number of calories per gram of fat, carbohydrates, and protein.

Nutrition Facts

Serving Size $\frac{1}{2}$ cup (114 g)

Servings Per Container 4

Amount Per Serving

Calories 90	Calories from Fat 30
-------------	----------------------

% Daily Value*

Total Fat 3g	5%
--------------	----

Saturated Fat 0g	0%
------------------	----

Cholesterol 0mg	0%
-----------------	----

Sodium 300mg	13%
--------------	-----

Total Carbohydrates 13g	4%
-------------------------	----

Dietary Fiber 3g	12%
------------------	-----

Sugars 3g	
-----------	--

Protein 3g

Vitamin A 80%	• Vitamin C 60%
---------------	-----------------

Calcium 4%	• iron 4%
------------	-----------

* Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

	Calories	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300 mg	300 mg
Sodium	Less than	2,400 mg	2,400 mg
Total Carbohydrate		300g	375g
Fiber		25g	30g

Calories per gram

Fat 9	• Carbohydrates 4	• Protein 4
-------	-------------------	-------------

Calories from fat are now shown on the label to help consumers meet dietary guidelines that recommend people get no more than 30% of their calories from fat.

% Daily Value shows how a food fits into the overall daily diet.

Daily values are also something new. Some are maximums, as with fat (65g or less); others are minimums, as with carbohydrates (300g or more). The daily values on the label are based on a daily diet of 2000 and 2500 calories. Individuals should adjust the values to fit their own calorie intake.

Figure 17.1 A typical nutrition information panel from an American food label. From the Food and Drug Administration (FDA, 1992).

saturated fat and cholesterol and coronary heart disease
fat and cancer
folic acid and neural tube defects.

Most other claims are not yet permitted and only foods that meet the compositional criteria specified for each claim can carry the claim. The suggested wordings for such claims are also laid down. Several other, more general, claims are permitted, e.g. it is not permitted to

claim that antioxidant vitamins prevent cancer but it is authorized for substances low in fat and high in fruits and vegetables. Note that, in Britain, foods may not carry a claim that they cure or prevent a specific disease.

The US FDA estimated the cost of these labelling changes at \$1.4–2.3 billion (£1–1.5 billion) over 20 years. They assumed that these costs would be more than compensated for by the improvements in the diet that they facilitate and the resultant reduction in health care costs!

- In the UK, most packaged foods must have on their label the food's name, ingredients, shelf life, storage requirements, batch number and the name and address of the packer.
- A 'use by' date on a food in the UK indicates that the food is likely to be microbiologically unsafe after that date.
- A 'best before' date indicates that a UK food's quality may have deteriorated after that date, but it is unlikely to have become unsafe.
- Most British foods carry some nutritional information on the label and they must do so if they make any nutritional claim for the product.
- If nutritional information is used on a British label, it must include the energy and macronutrient content per 100 g and the amount of any nutrient for which a claim is made.
- There are now very detailed and complex regulations controlling food labelling in the USA.
- Almost all packaged foods in the USA must have a label, which must:
 - contain a full list of ingredients
 - use prescribed serving sizes for nutritional information
 - give detailed nutritional information according to prescribed rules
 - only use descriptive terms such as high or low if the food meets certain prescribed criteria
 - only make permitted health claims and then only if the food meets the prescribed criteria.

FOOD POISONING AND THE MICROBIOLOGICAL SAFETY OF FOOD

Introduction

There cannot be many readers of this book who have not experienced several bouts of microbial food poisoning. There seems to be a general acceptance that the incidence of microbial food poisoning has risen sharply in recent decades in many industrialized countries. For example, the number of officially recorded and verified cases of food poisoning in England and Wales rose throughout the 1980s and 1990s from under 15 000 in 1982 to more than 100 000 by the end of the 1990s. Food poisoning is rarely serious enough to cause death, but it can be fatal in the very young, very old and those people whose immune system is suppressed by drugs or disease. One to two hundred people die each year in the UK as a consequence of food poisoning. To put this figure for deaths from food poisoning into perspective, in the 1940s, prior to the introduction of general pasteurization of milk, 1500 people died in the UK each year from bovine tuberculosis contracted from drinking contaminated milk. Bovine tuberculosis has now been all but eradicated in the UK.

The number of officially recorded and verified cases of food poisoning will, almost inevitably, only represent the 'tip of the iceberg' because only a minority of sufferers will even consult a physician. Reported cases have been variously estimated as representing between 1% and 10% of all cases, and a recent UK report suggested that as many as one in six of the population might be made ill by food-poisoning bacteria each year. Even though recorded cases represent only a small and indeterminate fraction of actual cases, it still seems almost certain that real cases have risen in line with the number of recorded ones. It is also likely that infections that produce a relatively mild or relatively short-lasting illness will probably be under-represented in the recorded data as compared to their true incidence. The relatively high incidence in children may again be because parents are more likely to consult a doctor about food poisoning in a child.

The causes of foodborne diseases

THE CAUSATIVE ORGANISMS

Textbooks of food microbiology usually list more than ten organisms or groups of organisms as significant

- Recorded cases of food poisoning rose sharply during the 1980s and 1990s to a current total of around 100 000 cases annually.
- Recorded food-poisoning cases only represent a small fraction of total cases and as many as one in six of the British population may be affected each year.
- Food poisoning is seldom fatal and most of the 100–200 annual deaths in the UK are in vulnerable groups such as the elderly, infants and those with impaired immune function.

causes of food poisoning. In classical food poisoning, large numbers of organisms generally need to grow in the food to cause illness, and this normally presents as acute gastro-enteritis. There are a number of other foodborne diseases, e.g. typhoid and cholera, which usually require relatively few organisms to cause illness, and in these cases there is usually no need for the organism to grow in food, as there is in classical food poisoning. Several organisms that are now important causes of food poisoning require only a small number of bacteria to cause illness, and so there is no need for them to multiply in the food to cause illness (e.g. *Campylobacter* and *Escherichia (E.) coli* 0157).

The relative importance of the various organisms in causing food poisoning varies in different countries and communities and depends upon the diet and food practices of the population. For example, food poisoning caused by the organism *Vibrio (V.) parahaemolyticus* is associated with the consumption of undercooked fish and shellfish. In Japan, where the consumption of raw fish (sushi) is widespread, this organism is the most common cause of food poisoning. It is, however, a relatively uncommon causative agent in the UK and USA, where sushi bars and Japanese restaurants are usually the source of infection.

Roberts (1982) analysed more than 1000 outbreaks of food poisoning that occurred in England and Wales and she found that four organisms or groups of organisms accounted for almost all of these outbreaks: the *Salmonellae*, *Clostridium (C.) perfringens*, *Staphylococcus (S.) aureus* and *Bacillus (B.) cereus*. Similar analyses of US outbreaks had also identified the first three on this list as the most common causative agents (Bryan, 1978).

During the 1980s, the incidence of food poisoning in the UK caused by *Salmonella* increased dramatically, mainly due to increases in infection

with *S. enteriditis* from poultry and eggs. As a group, the *Salmonella* now account for over a third of all recorded cases of bacterial food poisoning in England and Wales. *Salmonellae* are the second most common cause of food poisoning. *Campylobacter jejuni* was only recognized as a general cause of food poisoning in the late 1970s, when techniques for its isolation and identification became available. The organism is difficult to grow on standard laboratory media and does not usually multiply in food. The *Campylobacter* are now the organisms most frequently causing acute bacterial gastro-enteritis in the UK and account for over half of all recorded cases. *Campylobacter* were not mentioned in the Roberts' report of 1982. It requires only a few of these organisms to cause illness, which can be spread from person to person, and in this respect is more like the other group of foodborne diseases such as typhoid and cholera. *C. perfringens* is now the third most common cause of bacterial food poisoning, whereas *B. cereus* and *S. aureus* each represents less than 1% of recorded food-poisoning cases.

Listeriosis was also not mentioned in the earlier report of Roberts (1982), nor is it mentioned in textbooks of food microbiology published well into the 1980s. It has, however, been a cause of great public concern in the UK and USA in recent years, with the demonstration of the causative agent (*Listeria (L.) monocytogenes*) in many supermarket chilled foods. It has been of particular concern because it may cause miscarriage, birth defects and the death of newborn infants if pregnant women are infected.

E. coli 0157 is another food-poisoning organism that has attracted much media and public attention in the last few years. This bacterium was associated with an outbreak of food poisoning in Scotland that affected hundreds of people and killed 20 elderly people. This is one of a group of bacteria that produces a

verocytotoxin that damages cells in the gut and the endothelium of blood vessels. These verocytotoxin-producing *E. coli* (VTEC) only emerged as a cause of food poisoning in the early 1980s, but now account for over 1000 cases per year. They have attracted a lot of attention because a proportion of those affected develop renal complications and even acute renal failure, which can be fatal. It has significant mortality, especially in children and the elderly.

HOW BACTERIA MAKE US ILL

Food-poisoning organisms produce their ill-effects in one of the three ways listed below.

1. Some organisms cause illness by colonizing the gut when they are ingested, e.g. the *Salmonellae*, the *Campylobacter* and *V. parahaemolyticus*. They produce an infection.
2. Some organisms produce a toxin when they grow in the food, and it is the ingestion of this toxin that is responsible for the illness associated with the organism, e.g. *S. aureus*, *B. cereus* and *Clostridium (C.) botulinum*. The illness is caused by chemical intoxication.
3. Some organisms, such as *C. perfringens*, *E. coli* 0157 and the other VTEC infect the gut and produce a toxin after ingestion that is responsible for the food-poisoning symptoms.

Some cases of gastro-enteritis cannot be attributed to bacteria or bacterial toxins and many are caused by foodborne viruses.

CIRCUMSTANCES THAT LEAD TO FOODBORNE ILLNESS

In order for foodborne disease to occur, the food must have been contaminated with the causative organism at some stage prior to ingestion. There are many possible means by which the food can be contaminated with micro-organisms and some examples are listed below.

- Fruits and vegetables may become contaminated by being washed or watered with dirty water.
- The food handler can be a source of contamination.
- Insects, such as flies, may be a source of contamination
- Food may be cross-contaminated from contact with another food – particularly important if cooked food that is to undergo no further heat treatment is contaminated by raw meat or poultry.

With some bacteria (e.g. those causing food-borne diseases such as *Shigella* dysentery, typhoid, paratyphoid and cholera), illness can result from the consumption of small numbers of organisms (say, 10–10 000) and thus contamination of any food with these organisms may, in itself, be sufficient to cause illness. Such organisms are said to have a low **infective dose** and in these cases infection may well result from drinking contaminated water or, in some cases, from spread from person to person. Of the major food-poisoning bacteria, both the *Campylobacter* and *E. coli* 0157 also have a low infective dose.

Foreign travel and gastro-enteritis (**travellers' diarrhoea**) seem to be very frequently associated. Some of these cases may be due to a wide variety of food-poisoning organisms, including those listed earlier, but a high proportion are probably due to strains of *E. coli* to which the traveller has not acquired any immunity. Cook (1998) has suggested that up to 80% of cases of acute diarrhoea amongst travellers may be due to strains of *E. coli* that are much less virulent than the potentially fatal infection of *E. coli* 0157. Faecal contamination of drinking water, bathing water or perhaps contamination of water used to irrigate fruits or vegetables are the likely sources of the organisms. The lack of immunity means that foreigners may have a much lower infective dose than the indigenous population and so foreigners become ill but the local population is untouched. Low infective doses from water, ice in drinks, or raw fruit and salad vegetables washed in contaminated water may trigger the infection. Those on short visits to foreign countries, especially where the water supply may be suspect, would be well-advised to consume only drinks made with boiled or purified water, to avoid ice in drinks, to peel or wash in purified water any fruits or vegetables that are to be eaten raw.

- The organism may be widespread in the environment, e.g. in the soil, water, dust or air.
- The organism may be present in the animal carcass at slaughter.
- Eggs and meat may become contaminated with animal faeces.
- Fish and shellfish may become contaminated by sewage in seawater – particularly important in filter-feeding molluscs such as mussels.

Many common food-poisoning organisms (including *Salmonellae*) need to be ingested in higher numbers (often in excess of 100 000) to cause illness in a healthy adult. These organisms have a high infective dose and they do not usually cause illness as a result of the consumption of contaminated water, because of dilution and because water does not have the nutrients to support bacterial proliferation. Similarly, with many organisms that cause illness by producing a toxin in the food, once again a large bacterial population usually needs to have accumulated before sufficient toxin is produced to cause illness. For food poisoning to occur in healthy adults, not only must food be contaminated with these organisms, but the following conditions also usually need to be met:

- the food must support growth of the bacteria (and allow the production of toxin if relevant)

- the food must at some stage have been stored under conditions that permit the bacteria to grow (and produce toxin).

Note that, even within a group of healthy young adults, there will be variation in the susceptibility to any particular infective organism. In highly vulnerable groups, infective doses are likely to be at least an order of magnitude smaller. Babies are vulnerable because their immune systems are not fully developed, and the functioning of the immune system also declines in the elderly (see Chapter 14). Others who are at increased risk of food poisoning are those people whose immune systems have been compromised by illness, starvation, immunosuppressive drugs or radiotherapy. Not only is the infective dose lower in such vulnerable groups, but the severity of the illness is likely to be greater, perhaps even life threatening.

- At least ten groups of bacteria are important potential causes of bacterial food poisoning.
- Classical food-poisoning bacteria require large numbers to be present in the food to produce illness, i.e. they have a high infective dose.
- Some organisms produce illness when only a few are present in the food, i.e. they have a low infective dose.
- A group's diets and dietary practices determine the relative importance of different food-poisoning bacteria.
- In the early 1980s, the *Salmonellae*, followed by *C. perfringens*, *S. aureus* and *B. cereus*, accounted for almost all recorded outbreaks of food poisoning.
- In recent years, the *Campylobacter* have risen from being unrecognized as a cause of food poisoning to being by far the most common cause in Britain.
- Other organisms that have attracted particular attention in recent years are:

S. enteriditis, because it became the most common cause of *Salmonella* food poisoning and its presence in British poultry and eggs was very widespread

L. monocytogenes, because it is present in many chilled foods and it can cause miscarriage, birth defects and stillbirth if consumed by pregnant women

E. coli 0157, because it may lead to acute and potentially fatal renal complications in children and elderly people.

- Food-poisoning bacteria produce illness by infecting the gut or by producing a toxin, either in the food or in the gut, after ingestion.
- Initial contamination of food with the organism is a prerequisite for all foodborne diseases.
- If organisms have a high infective dose, the bacteria must have multiplied in the food prior to ingestion and so the food and storage conditions must have been suitable for bacterial growth.
- Travellers' diarrhoea is often caused by strains of *E. coli* to which the traveller is more sensitive than the local population.
- Contaminated water may be directly or indirectly responsible for many cases of travellers' diarrhoea.

ing, and the speed of recovery slower. Most food-poisoning deaths occur amongst these vulnerable groups.

Principles of safe food preparation

The safe preparation of food involves measures and practices that attempt to do some or all of the following:

- minimize the chances of contamination of food with bacteria
- kill any contaminating bacteria, e.g. by heat or irradiation
- ensure that food is stored under conditions, or in a form, that prevent bacterial proliferation.

Some bacteria produce changes in the flavour, appearance and smell of the food that reduce its palatability – i.e. they cause **spoilage**. The growth of spoilage organisms serves as a warning that the food is stale or has been the subject of poor handling procedures, inadequate processing or poor storage conditions. Food-poisoning organisms, on the other hand, can grow in a food without producing changes in palatability. Thus, the absence of spoilage is no guarantee that the food is safe to eat, otherwise food poisoning would be a much rarer occurrence.

REQUIREMENTS FOR BACTERIAL GROWTH

Many bacteria must actively grow in food before they are numerous enough or have produced enough toxin to cause illness in healthy people. In order to grow in food, bacteria need the following conditions to be met.

- **Nutrients.** Many foods are ideal culture media for the growth of micro-organisms. The same nutrients that people obtain from them can also be used to support bacterial growth. One reason why meat, meat products, eggs, seafood and milk products are so often the causes of food poisoning is because they are nutrient rich and generally support bacterial growth.
- **A suitable temperature.** Bacterial growth slows at low temperatures and the growth of most pathogens ceases at temperatures of less than 5 °C. At the temperatures maintained in domestic

freezers (−18 °C), all bacterial growth will be arrested, even though organisms can survive extended periods at such freezing temperatures. Growth of food-poisoning organisms will also not occur at temperatures greater than 60 °C and most will be slowly killed. At temperatures of over 73 °C, non-spore-forming bacteria will be very rapidly killed, although those capable of forming spores may survive considerable periods even at boiling point. Note also that, for example, *S. aureus* produces a heat-stable toxin and so foods heavily contaminated with this organism remain capable of causing food poisoning even if heat treatment or irradiation kills all of the bacteria.

- **Moisture.** Bacteria require moisture for growth. Drying has been a traditional method of extending the storage life of many foods. High concentrations of salt and sugar preserve foods by increasing their osmotic pressure and thus making the water unavailable to bacteria, i.e. they reduce the **water activity** of the food. Moulds generally tolerate much lower water activities than bacteria and grow, for example, on the surface of bread, cheese and jam (jelly).
- **Favourable chemical environment.** A number of chemical agents (preservatives), by a variety of means, inhibit bacterial growth. They create a chemical environment unfavourable to bacterial growth. Many pathogenic and spoilage organisms only grow within a relatively narrow pH range; they will not, for example, grow in a very acid environment. Acidifying foods with acetic acid (pickling with vinegar) is one traditional preservation method. Fermentation of food using acid-producing bacteria (e.g. the *Lactobacilli*) has also been traditionally used to preserve and flavour foods such as yoghurt, cheese and sauerkraut.
- **Time.** The contaminated food needs to be kept under conditions that favour bacterial growth for long enough to allow the bacteria to multiply to the point at which they represent a potentially infective dose. Note that, under favourable conditions, bacteria may double their numbers in less than 20 minutes and so 1000 organisms could become 1 million in 3 hours.
- **Oxygen.** Food-poisoning organisms do not require oxygen for growth, although it may hastens their growth. *C. botulinum* is an anaerobe that is killed by oxygen.

Table 17.1 Factors most often associated with 1000 outbreaks of food poisoning in England and Wales

	%
Preparation of food in advance of needs	61
Storage at room temperature	40
Inadequate cooling	32
Inadequate heating	29
Contaminated processed food	19
Undercooking	15
Inadequate thawing	}
Cross-contamination	4–6
Improper warm holding	}
Infected food handler	}
Use of leftovers	}
Consumption of raw food	}

After Roberts (1982).

SOME SPECIFIC CAUSES OF FOOD-POISONING OUTBREAKS

Roberts (1982) identified the factors that most often contributed to outbreaks of food poisoning in England and Wales (Table 17.1). Bryan (1978) had earlier concluded that the factors contributing to food-poisoning outbreaks in the USA and UK were similar. Hobbs and Roberts (1993) suggest that factors contributing to outbreaks of food poisoning are unlikely to have changed significantly since that time. Such analyses give a useful indication of practices that are likely to increase the risk of food poisoning and so indicate some of the practical measures that can be taken to reduce the risk. It must be borne in mind that Roberts' (1982) analysis would not have included cases of infection with those bacteria that have relatively recently attracted attention as important causes of food poisoning and have a low infective dose, e.g. *Campylobacter* and *E. coli* 0157.

In her analysis of food-poisoning outbreaks, Roberts (1982) found that two-thirds of outbreaks could be traced to foods prepared in restaurants, hotels, clubs, institutions, hospitals, schools and canteens. Less than 20% of outbreaks were linked to food prepared in the home. In a later review of these figures, Hobbs and Roberts (1993) suggested that they were seriously distorted due to the low rate of inclusion of home-related incidents because of the lack of epidemiological data on these outbreaks. Nevertheless, outbreaks of food poisoning that orig-

inate from home-prepared food will usually generate far fewer individual cases than outbreaks associated with commercial and institutional catering:

Situations where food is prepared in quantity for a large number of people are most likely to give rise to most food poisoning.

Hobbs and Roberts (1993).

For example, a flaw in the pasteurization process of a major US dairy resulted in over 16 000 confirmed cases and around 200 000 suspected cases of food poisoning in Illinois in 1985 (Todd, 1991).

In the vast majority of the outbreaks investigated by Roberts (1982), meat and poultry were identified as the source of the infection. Eggs, as well as poultry, are now also recognized as a source of the organism causing the most common type of *Salmonella* food poisoning (*S. enteritidis*).

Certain foods, such as those listed below, are associated with a low risk of food poisoning.

- Dry foods such as bread and biscuits or dried foods – the low water activity prevents bacterial growth.
- High-sugar or salted foods such as jam, because of the low water activity.
- Fats and fatty foods such as vegetable oil lack nutrients and provide an unfavourable environment for bacterial growth.
- Commercially canned foods, because they have been subject to vigorous heat treatment and maintained in airtight containers.
- Pasteurized foods such as milk, because they have been subjected to heat treatment designed to kill likely pathogens.
- High-acid foods such as pickles.
- Vegetables generally do not support bacterial growth. They can, however, be a source of contamination for other foods, e.g. if used in prepared salads containing meat, eggs or mayonnaise. Fruit and vegetables contaminated by dirty water may be a direct source of infection if the infecting organism has a low infective dose.

Canned or pasteurized foods cannot be assumed to be safe if they may have been contaminated after opening or, in the case of dried foods, after rehydration.

This information may be particularly useful if one is obliged to select food in places where food hygiene seems less than satisfactory.

- Spoilage organisms adversely affect the taste, smell and appearance of food, but food-poisoning organisms generally do not.
- Safe handling and preparation of food require that the risks of contamination and bacterial proliferation are minimized and/or that the killing of bacteria during preparation is maximized.
- In order to grow, bacteria require nutrients, moisture, warmth, time and a suitable pH.
- The safe storage of food involves depriving likely pathogens and spoilage bacteria of one or more of their requirements for growth.
- Table 17.1 lists some of the factors most often linked to outbreaks of food poisoning due to mainly high infective-dose bacteria in the early 1980s.
- Institutional and commercial catering is responsible for most large outbreaks of food poisoning.
- Meat, poultry and eggs are the sources of most food-poisoning outbreaks in the UK.
- Some foods are seldom associated with food poisoning, e.g.:

foods with low water activity, i.e. dry, salted or high-sugar foods
 fats and oils
 newly opened canned or pasteurized foods
 pickled foods
 fruits and vegetables, unless they act as vehicles for contaminated water.

Some practical guidelines to avoid food poisoning

MINIMIZE THE RISKS OF BACTERIAL CONTAMINATION OF FOOD

Some specific steps for reducing the risk of food contamination are listed in Table 17.2. Much of the meat, poultry and eggs that we purchase must be assumed to be already contaminated when they are purchased. It will require improvements in farming and slaughtering practices to reduce the initial contamination of these foods. For example, there has been a sharp fall in the number of infections with *S. enteriditis* in the UK following the introduction of stringent new safety measures for egg production and handling, including a massive hen vaccination programme.

Most of the tips for reducing bacterial contamination in Table 17.2 are commonsense measures that require little amplification or explanation, but carelessness can cause us to overlook them, even though they seem obvious when written down. Most people would require no prompting to avoid a butcher who kept his meat in a warm place where flies could crawl over it or to avoid buying food that is obviously spoiled or past its 'use by' date. Lax hygiene can allow the cook's hands to become a source of faecal organisms or of *S. aureus*, which is frequently present in

skin lesions and in the nasal passages of many people. It is vital to avoid cross-contamination of ready-to-eat foods with raw foods such as meat, which are usually contaminated with bacteria. The bacteria in the raw food will probably be killed during cooking, but if the bacteria multiply in the ready-to-eat food, there is no further heat treatment to eliminate them. There are very strict food hygiene regulations that are intended to ensure that raw and cooked meats are separated in supermarkets and butchers' shops and staff should be trained to avoid cross-contaminating cooked meats.

MAXIMIZE THE KILLING OF BACTERIA DURING HOME PREPARATION OF FOOD

Some tips for maximizing the killing of bacteria during home preparation are listed in Table 17.2. Even though many raw foods may be contaminated when bought, if they are thoroughly cooked, almost all of them should be made safe by this cooking. Most of the common food-poisoning organisms are killed rapidly at temperatures over 73 °C. Meat that is eaten 'rare' will not have undergone sufficient heat treatment to kill all bacteria and represents a hazard if contaminated prior to cooking; poultry should always be well cooked. Note that this is especially important for hamburgers and other minced meat. With a joint or

Table 17.2 Some tips for reducing the risk of food poisoning

Reduce bacterial contamination of food:

- Buy food that is fresh and from suppliers who store and prepare it hygienically
- Keep food covered and protected from insect contamination
- Wash hands thoroughly before preparing food and after using the toilet or blowing your nose
- Cover any infected lesions on hands
- Ensure that all utensils and surfaces that come into contact with food are clean
- Ensure that any cloth used to clean utensils and surfaces or to dry dishes is clean and changed regularly
- Avoid any contact between raw food (especially meat) and cooked food
- Do not store raw meat in a position in the refrigerator where drips can fall onto other foods that will not be heat-treated before being eaten
- Wash utensils, hands and surfaces in contact with raw food before they are used with cooked food

Maximize killing of bacteria during processing:

- Cook meat, poultry and eggs thoroughly so that all parts of the food reach a minimum of 75 °C
- Defrost frozen meat and poultry thoroughly before cooking
- Re-heat food thoroughly so that all parts of it reach 75 °C

Minimize the time that food is kept in conditions that allow bacteria to grow:

- Prepare food as close to the time of consumption as is practical (prolonged holding of food also leads to loss of nutrients)
- When food is prepared well in advance of preparation, it should be protected from contamination and should not be stored at room temperature
- Discard foods that are past their 'use by' dates or show evidence of spoilage (e.g. 'off' flavours and smells or discolouration)
- Food that is kept hot should be kept at 65 °C or above
- Food that is kept cool should be kept below 5 °C
- When food is cooked for storage it should be cooled quickly
- Leftovers or other prepared food that are stored in a refrigerator should be covered and used promptly
- Store eggs in a refrigerator and use them within three weeks of purchase

slice of meat, most of the contaminating bacteria will be on the outside and these will be killed even if the centre of the meat is pink and underdone. However, when meat is minced, the bacteria are spread throughout the meat, so that the centre of a hamburger contains as many bacteria as the outside.

Raw eggs (e.g. in homemade mayonnaise) and undercooked eggs represent a potential hazard, especially to the elderly, the sick and the very young. People in Britain have been officially advised to only eat eggs that are thoroughly cooked, e.g. not soft boiled eggs.

MINIMIZE THE TIME THAT FOOD IS STORED UNDER CONDITIONS THAT PERMIT BACTERIAL MULTIPLICATION

Some tips for minimizing the time that food is stored under conditions that allow bacteria to multiply are listed in Table 17.2. The time between food preparation and consumption should be as short as is practically reasonable. Where food is prepared well in advance of serving, it should be protected from contamination and should not be stored at room temperature. Buffets, where prepared food may be left uncovered at room temperature for some hours before consumption, are particularly hazardous.

Most bacteria only grow between 5 °C and 65 °C and are slowly killed at temperatures of over 65 °C. This means that food that is kept hot should be kept above 65 °C and any food that is to be reheated should be thoroughly reheated so that all parts of it reach 75 °C (the temperature at which non-spore-forming bacteria are rapidly killed). Food that is to be kept cool should be kept below 5 °C because most food-poisoning bacteria will not grow below this temperature. Food that has been cooked but is to be stored cool should be cooled rapidly so that it spends the minimum time within the growth range of bacteria. Leftovers, or other prepared foods, that are stored in a refrigerator should be covered and used promptly; care should be taken to ensure that they do not come into contact with raw food, especially flesh foods.

The principal aim of this chapter is to encourage greater care and awareness in the home preparation of food. In order to effect reductions in the rates of food poisoning produced outside the home, there needs to be improved training of catering workers and regulatory changes that lead to improvements to commercial and institutional catering practices. Food hygiene laws and regulations are complex, often extremely detailed, and vary from nation to nation (recent attempts to harmonize food hygiene laws within the EU have highlighted this latter point). Food hygiene laws must remain the province of specialist texts of food microbiology and food technology (e.g. see Jacob, 1993). Note, however, that, in

- Many practical measures that will reduce the risk of food poisoning are listed in Table 17.2.
- It should be assumed that meat, poultry, eggs and raw milk are likely to be contaminated when purchased.
- Heat will kill bacteria, but some heat-resistant spores may survive boiling and some bacterial toxins may be heat resistant and remain in food even when all the bacteria have been killed.
- Most cases of food poisoning can be treated by simple supportive measures aimed at body fluid balance.
- Antibiotics do not generally affect the course of a bout of food poisoning.
- Compounds that stop diarrhoea by inhibiting gut motility are generally contraindicated because they may prolong the infection.

principle, food hygiene laws represent attempts by governments to enshrine the safety principles and practices outlined for home food preparation into a formal legal framework so as to ensure that they are carried through into mass catering situations.

A NOTE ABOUT THE TREATMENT OF FOODBORNE DISEASE

Most cases of food poisoning go unrecorded because the symptoms are not severe enough or prolonged enough to prompt people to consult their doctor. The best way of treating simple food poisoning is to maintain the patient's fluid and electrolyte levels with oral rehydration fluids containing glucose and salts. In severe cases, the intravenous infusion of fluid may be necessary. In simple food poisoning, antibiotics are not used because they do not affect the course of the disease and their use may simply encourage the development of antibiotic resistance in bacteria. As a general rule, self-medication with preparations that stop diarrhoea by inhibiting gut motility should not be used, especially with children, because they increase the retention of the bacteria and increase exposure to any toxin. These antimotility agents may, for example, increase the risk of renal damage in cases of infection with *E. coli* 0157.

Pinpointing the cause of a food-poisoning outbreak

Identifying the cause of food-poisoning outbreaks should highlight high-risk practices and thus may be useful for improving food handling. If a single event

(e.g. serving a contaminated dish at a buffet) is the cause of an outbreak, a time distribution of cases like that shown in Figure 17.2 would be expected. Each food-poisoning organism has a characteristic average incubation period from the time of ingestion to the onset of symptoms. Incubation times in individuals will vary around this average: a few people develop symptoms early, most towards the middle of the range and a few towards the end of the range, as in Figure 17.2. The variation in the speed of onset of symptoms is influenced by all sorts of factors, such as:

- the dose of organism or toxin consumed
- the amount of other food consumed and thus the dilution and speed of passage through the gut
- individual susceptibility.

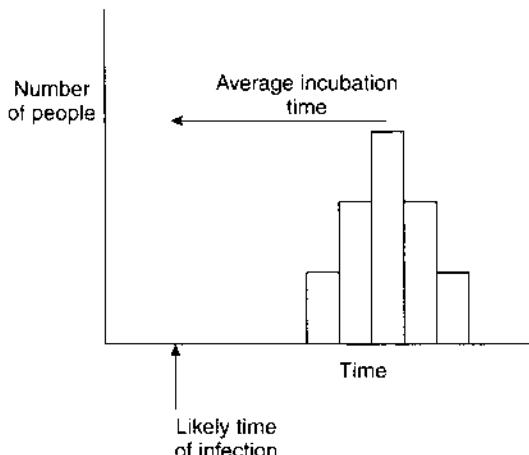


Figure 17.2 Theoretical time distribution of people first experiencing food-poisoning symptoms after infection.

The organism causing the illness can be identified from the clinical picture and/or laboratory samples and then, knowing the incubation period, the approximate time at which all of the victims were infected and the meal at which the organisms were consumed can be pinpointed (see Figure 17.2). Sampling of leftovers or circumstantial evidence can then identify the responsible food and possibly the unsafe practice that led to the outbreak.

Board (1983) gave an illustrative case history of an outbreak of food poisoning amongst the passengers and crew of a charter flight. More than half of the passengers on this flight developed food poisoning, a few at the end of the flight and large numbers at the airport within 2 hours of landing. The clinical picture (the symptoms, the rapid onset and short duration) indicated that the illness was probably due to ingestion of toxin from *S. aureus*. The time distribution of cases indicated that the breakfast served on the plane 90 minutes before landing was the likely source of the illness, and samples of uneaten breakfasts confirmed the presence of this organism in ham omelette. The food had been contaminated by infected sores on the hands of one of the cooks at the catering company that had supplied the breakfasts to the airline. A long time (3 days) had elapsed between the initial preparation of the food and its eventual serving to the passengers, and the food had not been kept cool for much of the time between preparation and serving.

If no food samples are available for analysis, the responsible food can often be implicated by calcula-

Table 17.3 Food-specific attack rates for people present at a social function identified as the point of infection for an outbreak of food poisoning

Food	All ill		Attack rate (%)	Did not eat		Attack rate (%)
	Ill	Not ill		Ill	Not ill	
A	30	131	19	0	51	0 ^a
B	0	27	0	30	154	16 ^a
C	8	49	14	21	125	14
D	21	95	18	9	84	10
E	8	69	10	22	107	17
F	13	60	18	17	121	12

^a Statistically significant difference between the attack rates of those who did and did not consume that particular food.

After Hobbs and Roberts (1993).

tion of **attack rates**. All of those participating in the meal identified as the point of infection are asked to complete a questionnaire about which foods they consumed and whether they became ill. The attack rate is the percentage of people becoming ill. For each food, the percentage of those who became ill is calculated for those who did and those who did not consume that food, i.e. the **food-specific attack rates**. A statistically significantly higher attack rate for those consuming a particular food as compared to those not consuming it indicates that it was the likely source of the infection. In the example in Table 17.3, eating food A significantly increased the attack rate, whereas eating food B significantly reduced it. Food B

- Food-poisoning bacteria each produce a characteristic set of symptoms and have characteristic incubation times.
- Incubation times will vary between individuals, depending upon dose consumed, its dilution with other food and drink and the susceptibility of the person.
- If the time distribution of cases in an outbreak is 'normal', as in Figure 17.2, this indicates a single point of infection and one can predict when this occurred from the average incubation period of the organism.
- One can identify which specific food was responsible because there will be a higher 'attack rate' amongst those who ate this food than amongst those who did not.
- If there is a continuing source of infection or person-to-person spread, the time distribution of cases will show multiple peaks or a tendency to plateau, rather than the single peak seen in Figure 17.2.
- Even with a continuing source of infection, it should still be possible to identify a factor common to all victims during the window of time when they were likely to have been infected.

appears protective because A and B were alternative choices on the menu and so people who ate food B did not eat food A. It would thus seem almost certain that food A was the source of the infection.

The time distribution of people reporting symptoms of food poisoning may not follow the bell-shaped model shown in Figure 17.2. For example, there may be multiple peaks or the number of new cases may tend to plateau rather than fall away. Such distributions would indicate either person-to-person spread or a continuing source of infection, e.g. a processing fault, an infected food handler, continuing use of a contaminated ingredient, or continued sale of a contaminated processed food. If the incubation period of the causative organism is known, a questionnaire should establish what all of the victims were doing (and eating) within the window of time when infection is predicted to have occurred. It should then be possible to identify a common factor, e.g. all victims ate a particular processed food, ate at a particular place or bought a processed food from a particular source during the period when they were likely to have become infected.

A review of some common food poisoning organisms and foodborne illnesses

THE CAMPYLOBACTER

The *Campylobacter* were only recognized as a general source of foodborne disease in 1977. The organisms do not grow in food at room temperature and do not grow readily upon normal laboratory media. This means that suitable methods needed to be developed for their identification. From being almost totally ignored as a cause of food poisoning 20 years ago, this organism is now easily the most common cause of acute bacterial gastro-enteritis in the UK, with around 50 000–60 000 cases recognized each year. The illness has a relatively long incubation period (2–10 days) and it produces a quite severe and prolonged illness. It initially produces flu-like symptoms and abdominal pain, which last for 1–2 days, and this is followed by profuse, sometimes bloody, diarrhoea, lasting for 1–3 days. It may take a week or more for complete recovery. The organism may contaminate water, unpasteurized milk, meat and poultry and it

may reside in the intestines of domestic animals. The organism has a low infective dose and thus is potentially transmissible through water, through the hands of children playing with or near domestic animals, or even from person to person. Undercooked poultry is the most common cause. Poultry or meat that has been inadequately cooked on garden barbecues has been blamed for many cases in the UK. The organism has a high optimum growth temperature but is killed rapidly at temperatures of 75 °C.

THE SALMONELLA

Many different types of the *Salmonella* group are responsible not only for common food poisoning, but also for the more serious foodborne and waterborne illnesses typhoid and paratyphoid. The common food-poisoning organisms usually need in excess of 100 000 bacteria to represent an infective dose. These latter organisms produce diarrhoea, stomach pain, vomiting and fever about 12–48 hours after ingestion, and the illness lasts for between 1 and 7 days. Intensive farming methods and slaughtering practices mean that raw meat, poultry, eggs and raw milk should be regarded as potentially contaminated with these organisms. These organisms grow at temperatures between 6 °C and 46 °C. They are relatively heat sensitive and are killed slowly at 60 °C and very rapidly at 75 °C. They are not tolerant of very acid conditions, i.e. pH less than 4.5.

Meat and especially poultry have long been regarded as foods with a high potential to cause *Salmonella* food poisoning, but in recent years the contamination of eggs with *S. enteriditidis* has been a major cause for concern in the UK. Many thousands of hens were slaughtered in an apparently unsuccessful attempt to eradicate these organisms from the laying flock. The incidence of *S. enteriditidis* infection has fallen sharply in the UK in the last few years because of changes in egg production and handling and a mass vaccination programme for hens.

Eggs have a number of natural barriers that tend to reduce the risk of bacterial entry, but they are sold in millions, so even if only a very small proportion are contaminated, this can represent, in absolute terms, a large number of potentially infected meals. Eggs also have a number of systems that inhibit bacterial growth (bacteriostatic systems) by making nutrients unavailable to the bacteria. The likelihood of a single fresh egg containing an infective dose for a healthy adult is thus

low. However, if eggs are stored at room temperature and kept for long periods, a small number of original contaminating bacteria can multiply to the point at which they represent an infective dose. Eggs should be stored in a refrigerator and they should not be kept for more than 3 weeks. Fresh and well-cooked eggs should represent no hazard at all and Britons are currently advised to eat only well-cooked eggs. Any potential hazard from eggs is increased very significantly if uncooked egg is added to other foods, e.g. mayonnaise or cheesecake, and then the mixture is stored under conditions that allow bacterial proliferation, e.g. at room temperature.

C. PERFRINGENS

This organism has a high infective dose of between 100 and 1000 million organisms. Illness is due to a toxin produced by bacteria within the gut. The disease has an incubation period of 8–12 hours and lasts for up to 24 hours. The symptoms are diarrhoea and abdominal pain, but not usually either vomiting or fever. The organism grows at temperatures of up to 53 °C and the spores need temperatures of 100 °C to kill them. Food poisoning from this organism is usually associated with meat and poultry consumption and the organism is of widespread distribution in the gut of animals and people and also in soil and dust. When food is inadequately cooked, these organisms may continue to grow during the initial heating of the food; some will survive at high temperatures by forming spores. Surviving spores may start to grow as the food cools and during storage of the cooked food. The organism grows very rapidly at temperatures towards the top end of its tolerable range.

E. COLI 0157 AND THE VTEC BACTERIA

Media and public awareness of *E. coli* 0157 was increased sharply in the UK by an outbreak of food poisoning in Lanarkshire, Scotland, in November 1996 that affected around 500 people and caused the deaths of 20 elderly people. Interviews with the initial victims established that all of them had eaten cooked meat products from a single butcher or had attended a lunch at which a steak pie supplied by this butcher had been served.

Cattle are almost always the source of infection with *E. coli* 0157. Undercooked beef and, more especially, beef products such as burgers, raw milk and cheese made from unpasteurized milk are the poten-

tially infective foods. The organism has a low infective dose and so cases have occurred due to contact with farm animals (e.g. children visiting farms), contamination of drinking water and even person-to-person spread.

E. coli 0157 is one of a group of bacteria that produce toxins that have potentially fatal effects, the verocytotoxin-producing *E. coli* (VTEC). These toxins damage gut cells and cells in the vascular endothelium. The onset time is typically 3–4 days after exposure, but can range from 1 to 14 days. The symptoms are abdominal cramps, vomiting and diarrhoea, but a proportion of sufferers will go on to develop more serious complications, including kidney failure, which will be fatal in some children and elderly people. It is one of the major causes of acute renal failure in children.

S. AUREUS

The illness caused by this organism is due to a toxin present in food in which this organism has grown. More than 5 million bacteria are usually regarded as necessary to produce enough toxin to evoke illness. The symptoms are vomiting, diarrhoea and abdominal pain but no fever. Symptoms are rapid in onset (2–6 hours) and of short duration (6–24 hours). The organism is heat sensitive. It grows within the range 8–46 °C and is killed at temperatures in excess of 60 °C, but the toxin, once formed in food, is very heat stable and will not be eliminated by further heat treatment or irradiation of the food.

Food handlers are an important source of contamination. Many people carry the organism in their nasal passages and this can lead to contamination of hands and thence food; infected lesions on the hands may contain this organism. The organism needs foods rich in protein to produce toxin, but it is comparatively tolerant of high salt and sugar concentrations, e.g. as found in ham or custard. *S. aureus* does not compete well with other bacteria present in raw food, but if essentially sterile, cooked food is then contaminated by the handler, then, without any competition, the organism is able to grow even in foods with relatively high salt or sugar contents.

B. CEREUS

This organism is commonly found in soil and thus upon vegetables. It grows in starchy foods and, in the UK and USA, poisoning due to this organism is

usually associated with the consumption of rice. The organism produces heat-resistant spores, which may survive cooking of rice. If the cooked rice, containing viable spores, is stored at temperatures that permit growth of the organism (7–49 °C), an infective dose will accumulate. Boiling rice in bulk and then storing it at room temperature prior to flash frying, e.g. in Chinese restaurants, is a common scenario for food poisoning due to this organism. The resultant illness is acute in onset (1–16 hours) and of short duration (6–24 hours) and it is caused by toxins produced by the organism in food. An infective dose is in excess of 100 000 organisms in the food. The symptoms depend upon the particular toxin involved, but may include diarrhoea, abdominal pain, nausea or vomiting.

C. BOTULINUM

This organism grows under anaerobic conditions in low-acid foods. Canned or bottled meats, fish or vegetables (but not acidic fruit) are potential sources of illness. The organism is killed by oxygen and thus the storage of these foods in airtight containers allows growth. The organism may also occasionally grow deep within non-canned flesh foods or meat products. Commercial canning processes involve a heat treatment that is specifically designed to ensure that all of the heat-stable spores of this organism are killed, the so-called **botulinum cook**. Cured meats and meat products contain nitrate, nitrite and salt, which are effective in preventing the growth of this organism, and the addition of these preservatives is legally required in the UK for meat products such as sausages. Home canning or bottling of vegetables results in significant numbers of cases of botulism in the USA each year because the heat treatment is insufficient to kill the *Clostridia* spores. The very rare cases in the UK have usually been traced to occasional rogue cans of imported meat or fish, e.g. after heat treatment, contaminated water might be sucked into the cooling can through a damaged seam. The most recent UK outbreak in 1989 was traced to a can of nut purée that was subsequently used to flavour yoghurt. In the USA, honey is regarded as a potential source of small amounts of toxin and so honey is not recommended for young children.

The organism produces an extremely potent toxin within the food. This toxin is destroyed by heat so danger foods are those eaten unheated (e.g. canned

salmon or corned beef) or those that are only warmed. The toxin, although a protein, is absorbed from the gut and affects the nervous system, blocking the release of the important nerve transmitter acetylcholine. A dry mouth, blurred vision and gradual paralysis are the symptoms, and death will occur due to paralysis of the respiratory muscles. With rapid treatment, including artificial ventilation, about 75% of those affected survive. In those victims who survive, recovery may take weeks or months. The incubation period is normally 18–36 hours. This organism is included in this review even though it rarely produces illness in the UK because the symptoms are frequently fatal and because the imperative to avoid contamination of food with this organism has had a great influence upon food-processing practices.

L. MONOCYTOGENES

Chilling of food has been a traditional method of storing food; it reduces both microbial growth and the rate of chemical deterioration. An enormous range of chilled foods is now available in supermarkets, e.g. fresh meat and fish, cheese, cooked meals, sandwiches, pies etc. Such foods are susceptible to the development of infective doses of pathogenic bacteria that continue to grow slowly at low temperatures. Small variations in temperature within the low range (0–8 °C), can have marked effects on the rate at which these low-temperature organisms grow. Strict temperature control is therefore of great importance in ensuring the safe storage of chilled foods. For foods that are to be eaten without further cooking, or just rewarming, current UK legislation requires that they be stored below 5 °C, e.g. cook-chill ready meals, sandwiches, cured meat and fish.

The organism *L. monocytogenes* is one of these low-temperature organisms and it has attracted much media attention and public concern in the UK and USA in the last few years. This organism will grow slowly at temperatures as low as 3 °C and it will tolerate a salt concentration of 10% and a pH down to 5.2. Soft cheeses and pâté, which have a relatively long shelf life, are considered to be at particular risk from this organism. Its presence has been reported in many cook-chill ready meals and it is likely to survive the mild reheating in a microwave oven recommended for many of these foods.

Table 17.4 Some characteristics of important food-poisoning bacteria

Bacteria	Effect of 75 °C	Incubation	Infective dose	Mode of pathology	Likely source	Other points
<i>Campylobacter</i>	Killed	2–10 days	Low	Infection	Meat, poultry	
<i>Salmonella</i>	Killed	12–48 hours	High	Infection	Meat, poultry, eggs	
<i>C. perfringens</i>	Spores resist	8–12 hours	High	Toxin in food	Meat, poultry	
<i>E. coli</i> 0157	Killed	1–14 days	Low	Toxin in gut	Cattle	Can be lethal
<i>S. aureus</i>	Killed	2–6 hours	High	Toxin in food	Food handler	Heat-resistant toxin
<i>B. cereus</i>	Spores resist	1–16 hours	High	Toxin in food	Rice	
<i>C. botulinum</i>	Spores resist	18–36 hours	High	Lethal toxin in food	Improperly canned or bottled food	Rare; often fatal; anaerobic
<i>L. monocytogenes</i>	Killed	2 days–7 weeks	Low in pregnancy and vulnerable groups	Infection	Chilled foods, pâté, soft cheese	Grows at 3 °C; tolerates salt and sugar
<i>V. parahaemolyticus</i>	Killed	2–48 hours	High	Infection	Fish, shellfish	

The organism causes a disease known as listeriosis in vulnerable groups such as the elderly, infants, pregnant women and the immunodeficient, although it does not usually cause illness in healthy people. Symptoms range from those resembling mild flu to severe septicaemia and, occasionally, meningitis. If infection with this organism occurs during pregnancy, abortion, foetal abnormality or neonatal death may result. In the UK, pregnant women are currently advised to avoid those foods that are liable to harbour *Listeria* (e.g. pâté and soft cheese made with unpasteurized milk) and to thoroughly reheat cook-chill foods. An increased use of chilling led to an increase in recorded cases of listeriosis from around 25 per year in the early 1970s to around 300 per year by 1988. In 1988, there were 26 deaths of newborn babies, 52 total deaths and 11 abortions caused by listeriosis. By the late 1990s, recorded cases of listeriosis had fallen to just over 100 cases per year as a result of increased awareness and measures to reduce it. The rise in the prominence of listeriosis with the increased use of food chilling is a good illustration of how changing eating habits and processing methods can lead to changes in the prominence of different types of food-poisoning organisms. The characteristics of some important food-poisoning bacteria are summarized in Table 17.4.

BOVINE SPONGIFORM ENCEPHALOPATHY

Introduction

When I completed the first edition of this book in 1994, there was no evidence that BSE would directly affect human beings. It is now clear that the human form of this disease has affected people (with about 80 deaths by November 2000) and that humans almost certainly acquired it by eating contaminated beef products. It is still unclear what the likely scale of the epidemic of the human form of this disease will be. Current estimates are that anywhere between a few hundred and 135 000 people in Britain will eventually die as a result of this foodborne disease. There is a wealth of reliable information on this disease available through the internet, and three recommended sites are listed below. Prusiner (1995) has written a review of the **prion** diseases of which BSE is an example.

1. The MAFF website about BSE – www.maff.gov.uk/animalh/bse/index.html
2. The CJD surveillance unit website – www.cjd.ed.ac.uk
3. The website of the official UK government inquiry into BSE – www.bse.org.uk

This disease has already had a devastating effect upon British farming and consumer confidence in food safety. Its eventual impact upon human health cannot yet be accurately predicted. It is thus given much greater weighting here than in the previous edition.

Overview

BSE is an apparently new disease of cattle that first appeared in British cattle in 1985. The affected animals are excitable, aggressive, easily panicked, have an abnormal gait and a tendency to fall down. These symptoms gave rise to the popular name 'mad cow disease'. Up to November 2000, around 175 000 British cattle had developed this disease and millions more had been slaughtered and burnt in order to try to eradicate it. Over 80% of cases of BSE have been in dairy cows and most of these in older cows. Only a handful of symptomatic cases have been recorded in cattle under 3 years old. The cattle epidemic peaked in 1992, when around 37 000 cases were recorded. Table 17.5 charts the rise and fall of the BSE epidemic in cattle, which now appears to be almost over.

The symptoms of BSE and the spongiform lesions seen in the brains of affected animals are similar to those seen in other transmissible spongiform encephalopathies (TSEs), such as those listed below. These diseases are all inevitably fatal.

Table 17.5 Annual numbers of cases of BSE in British cattle

Year	Number of cases
1986	17
1987	500
1988	3000
1989	7500
1990	15000
1991	26000
1992	37000
1993	34000
1994	30000
1995	18000
1996	10500
1997	5500
1998	4300
1999	3000
2000	1500 (to 30/09/00)

- **Scrapie** of sheep and goats: a disease that has been endemic in British sheep since the eighteenth century without any apparent effect upon human health.
- **Creutzfeld–Jakob disease (CJD)**: a rare disease that kills around 30–50 elderly people in the UK each year.
- **Kuru**: a disease that was prevalent amongst the Fore tribe in Papua New Guinea.

All of these diseases are potentially transmissible by injecting extracts of diseased brain into healthy animals. The 'infective agent' is resistant to heat and radiation and, unlike any other infective agents or any living organism, appears to contain no nucleic acid. Stanley Prusiner proposed that transmission was caused by an infective protein or prion and he was awarded a Nobel Prize for this work.

TSEs can be acquired in the three ways listed below.

1. *Spontaneous sporadic cases*: most cases of conventional CJD fall into this category. CJD affects around one in a million people in most populations. It almost always occurs in elderly people. It is unrelated to the presence of scrapie in the sheep flock.
2. *Genetic*: a few cases of CJD are familial.
3. *Transmission*:

experimentally induced by injecting infected brain into healthy animals
iatrogenically induced (caused by medical treatment), e.g. by injections of contaminated growth hormone or gonadotrophins extracted from the pituitary glands of human cadavers, by the use of contaminated surgical instruments or by transplants of infected corneal tissue
kuru is thought to have been caused by eating infected human brains in cannibalistic rituals.

The incubation period of all of these transmissible encephalopathies is long. In kuru and iatrogenic CJD, the average time from exposure to the onset of clinical symptoms is thought to be around 20 years.

In 1996, it was formally announced that a new variant of CJD (*vCJD*) had been identified in Britain. It now seems almost certain that it has been caused by eating contaminated beef products. The symptoms of *vCJD* are different from those of sporadic CJD, it also has different histological characteristics, a different time course, a different electroencephalogram (EEG)

pattern and, most significantly, it affects much younger people. Sporadic CJD rarely affects people under 60 years, but all cases of vCJD have been in younger age groups, including teenagers as young as 14 years old.

WHAT CAUSED THE EPIDEMIC OF BSE IN BRITISH DAIRY CATTLE?

Since the 1920s, cattle, and especially dairy cows, have been fed a protein supplement made by rendering animal remains, including cattle remains. It now seems almost certain that this feed was the cause of the epidemic, but it is unclear how it initially became contaminated with the BSE agent. The most popular theory, and one still favoured on the MAFF web-site in early 2001, was that a change in the rendering process of this feed in the early 1980s allowed large amounts of sheep scrapie prion to remain in the feed. This scrapie prion then triggered some cows to develop BSE. The disease was then spread by the recycling of BSE-infected cattle material in the feed. This theory offered much reassurance, because scrapie had been around for hundreds of years and had not been transmitted to people by eating sheep products. Although the scrapie prion had apparently crossed the species barrier, it was to another ruminant.

The official inquiry into BSE concludes that this theory is incorrect and that BSE was present in cattle several years before 1985 and that the recycling of infective cattle material in feed spread the disease. It was suggested in this report that the original source of the epidemic may have been a single animal with a mutation in the prion protein gene.

In 1988, the inclusion of ruminant protein in feeds for ruminants was banned and this should have eliminated the source of infection in cattle. However, animals born long after this feed ban came into force still developed BSE, and there are several reasons for this:

- Existing stocks of banned feed were used up on some farms.
- The use of ruminant protein remained legal for pig and poultry feed and this allowed cross-contamination of ruminant feeds to occur during production. The feed ban was extended to pigs and poultry in 1996.
- There is some vertical transmission (from cow to calf) of the disease. It is estimated that a calf born within 1 year of the mother developing clinical symptoms of BSE has a 10% chance of being infected by the mother. Infection occurs either in

the womb or during delivery. There is no transmission from paternal semen. There is no evidence of any significant horizontal transmission, i.e. from cow to cow.

THE INFECTIVE AGENT IN TSES

The dominant theory is that the infective agent in these TSEs is an infective protein or prion (Prusiner, 1995). This prion protein (PrP) is a normal protein found in the brains and other tissues of animals. It is suggested that, in normal, healthy animals, the normal PrP is present largely in α -helical form and this is easily broken down by proteases. In TSEs, large amounts of this PrP are present in a different configuration (β -sheet form) and this is extremely resistant to proteases. It is also very resistant to heat and radiation. When the abnormal PrP comes into contact with the normal form, it causes the normal form to change shape, a chain reaction is established and large amounts of abnormal protein accumulate and destroy brain cells, leading to the clinical symptoms and spongiform lesions in the brain. According to Prusiner, there are three likely ways in which TSEs can develop.

1. *Spontaneous*: some spontaneous production of abnormal PrP can occur and this will occasionally trigger the chain reaction as in spontaneous CJD in elderly people.
2. *Genetic*: a mutation in the PrP gene may occur, which makes the protein more likely to flip to its abnormal structure. There are rare examples of a familial form of CJD in people. It may be that brain tissue from a cow with such a PrP mutation, when fed back to other cows, triggered off the BSE epidemic.
3. *Transmission*: if abnormal PrP is introduced into the brain via the diet or via direct injection of infected brain material, this can trigger the chain reaction.

TRANSMISSION OF BSE TO HUMANS

Scrapie has long been prevalent in British sheep, but it is not transmitted to humans via the diet and it is hard to transmit sheep scrapie to rodents. There was a general belief in the 1980s and early 1990s that BSE would probably not jump the species barrier and affect humans. This view was encouraged by the widespread belief that BSE was scrapie that had been transmitted

to cows. In 1990, cases of BSE-like diseases were reported in several exotic species in zoos and these had been given feed containing ruminant protein. More disturbingly, the first of several cases of a TSE in domestic cats was identified. These observations suggested that BSE had jumped the species barrier and affected species other than ruminants and even carnivores. It is still hoped that the 'species barrier' will have limited the spread of BSE to humans and the numbers of people currently incubating vCJD.

In general, the more differences there are in the amino acid sequences of the PrPs of two species, the more difficult it is to transmit the disease experimentally. Cow prion protein has about 30 amino acid differences compared to the human protein, and it is already clear that normal genetic variation in the human PrP affects susceptibility to vCJD. It is also clear that transmission of TSEs is most efficient when infective material is injected directly into the brain and less efficient when it is given orally.

In experimental transmission studies, only some tissues in kuru and BSE contain infective quantities of abnormal PrP, i.e. brain, spinal cord and lymphatic tissues. Milk and muscle meat do not contain infective quantities of PrP in such studies. Steaks and joints of beef from young beef cattle should have been unlikely sources of human infection, even at the peak of the BSE epidemic. Concern has focused upon meat products, including baby foods, which often included the potentially infected tissues and were often made from meat obtained from elderly dairy cows. A number of measures have been introduced by the British government in order to protect the human population from BSE. The most important of these measures are listed below.

- In 1988, it was ordered that all clearly infected animals should be destroyed and their carcasses incinerated. Milk from infected animals was also removed from the food chain. All animals are now inspected before slaughter to make sure they do not show symptoms of BSE.
- In 1989, a ban on specified bovine offal (SBO) entering the human food chain was introduced. It was required that all of the likely infective tissues should be removed from all cattle at slaughter, i.e. the brain, spinal cord, spleen, thymus, tonsils and intestines. This measure, if properly enforced, should have afforded almost complete protection to people eating beef. It now seems that, up until 1996, the enforcement of this measure was lax and

that significant amounts of these tissues almost certainly did enter the human food chain between 1989 and 1996. In 1997, these controls were extended to cosmetics, pharmaceuticals and medical products.

- Since 1996, only cattle under 30 months of age have been allowed to be sold for human consumption. Very few young animals have developed BSE and so this measure should mean that little infective material actually reaches the slaughterhouse. Animals up to 42 months old can be sold if they come from specialist beef producers whose animals have only been fed upon grass and who meet very strict criteria under the Beef Assurance Scheme.
- For a time, it was forbidden to sell beef with bone in it, largely because of concerns that it might have nerve tissue adhering to it.

The measures listed above, taken together with the low and still declining incidence of BSE in cattle, should afford complete consumer protection at this point in time. In 1996, the EU instituted a worldwide ban on British beef exports. This ban has now been lifted, but many individual countries still refuse to allow British beef to be imported and the recovery in British beef exports has been extremely slow to gather any momentum. As at November 2000, no case of BSE has been confirmed in any British cow born after September 1996.

THE HUMAN 'EPIDEMIC' OF vCJD

It was in the spring of 1996 that the British government formally announced that a new variant of CJD (vCJD) had been discovered and that this was probably linked to eating contaminated beef before the bovine offal ban was introduced in 1989. Table 17.6 shows the annual numbers of deaths from vCJD since 1995. Experience

Table 17.6 Annual numbers of deaths from vCJD in the UK

Year	Deaths ^a
1995	3
1996	10
1997	10
1998	18
1999	14
2000	22 ^b

^a Includes probable cases up to 28 September 2000

- Bovine spongiform encephalopathy (BSE), or 'mad cow disease', is a new and inevitably fatal neurological disease of cattle.
- BSE produces spongiform lesions in the brain similar to those seen in sheep scrapie and human Creutzfeld–Jakob disease (CJD) and kuru.
- The British cattle epidemic appears to be almost over, but it has affected more than 175 000 cattle, mostly in dairy herds and especially older cows.
- More than 80 people have died from a new form of CJD (vCJD), which seems to be the human form of BSE, probably contracted through eating contaminated beef products.
- BSE and vCJD belong to a group of so-called prion diseases that can be transmitted by injecting or eating infected tissue.
- The infective agent (prion) in these prion diseases is thought to be a normal protein that has undergone a conformational change from α -helix to β -pleated sheet.
- When abnormal prion protein is introduced into the brain, it causes other normal prion protein molecules to change shape and this begins a chain reaction, which ultimately destroys large areas of brain tissue.
- The abnormal prion protein is very resistant to protease enzymes and to heat and radiation, so it is not destroyed by cooking or by the body's natural defences.
- Contaminated animal feed made from rendered ruminant remains was almost certainly the source of the cattle epidemic.
- It was at first thought that the initial trigger for BSE was sheep scrapie protein that had survived the rendering process.
- The belief that BSE was essentially scrapie in cows was reassuring because scrapie has been endemic in British sheep for centuries without affecting human health.
- It is now thought likely that BSE is a completely new disease that may have been initiated by the recycling of the infected remains of a single animal with a mutant prion protein.
- The British government banned the use of ruminant protein in ruminant feed in 1988, but some contaminated feed continued to be fed to cattle until 1996 and this prolonged the cattle epidemic.
- There is occasional transmission of BSE from cow to calf, but not apparently from cow to cow or bull to calf.
- Milk and meat from symptomatic animals have been destroyed since 1988, and all cattle are inspected at slaughter for signs of the disease.
- All beef sold after 1989 in the UK should have had all of the infective tissues (the specified bovine offal, SBO) removed, but this measure was not effectively enforced until 1996.
- Since 1996, only cattle under 30 months old and free from SBO have been permitted to be sold in the UK (with a few specified exceptions).
- Current measures, coupled with the low and declining incidence of BSE in cattle, should ensure the safety of current beef sales.
- A great deal of infected cattle material probably entered the human food chain prior to 1996 and millions of people may have been exposed to it.
- It is impossible to predict the likely scale of the human epidemic of vCJD, partly because of the lengthy but indeterminate incubation period and partly because there is no way of testing how many healthy people are incubating the infection.
- It can only be hoped that the 'species barrier' will have limited the spread of the cattle disease to humans who have eaten contaminated beef products.

with kuru and iatrogenic CJD suggests that the average incubation period for vCJD may be as high as 20 years. This makes it impossible as yet, and probably for several years to come, to make any reliable estimate of the total number of cases of vCJD that will occur.

FOOD PROCESSING

Some general pros and cons of food processing

The term food processing covers a multitude of processes to which food may be subjected. These processes may be traditional or modern and they may occur in the home, catering unit or factory. Unless one eats only raw, home-grown foods, it is impossible to avoid 'processed foods'. It would be impossible to feed a large, urban, industrial society without some commercial processing of foods. For example, around 50% of all food purchases in the UK is frozen or chilled. Processing includes cooking, smoking, drying, freezing, pasteurizing, canning, irradiation etc.

The processing of foods, particularly the preparation of ready-to-eat foods, 'adds value' to the ingredients and increases their commercial potential for retailers and processors. Food processors and retailers represent some of the most commercially successful businesses in developed countries. This high profitability has tended to politicize food-processing issues, sometimes to the detriment of constructive critical debate. The term 'processed food' is often used to convey very negative connotations about modern food and food suppliers. However, it is such an all-embracing term, covering foods across the whole spectrum of composition, that its use in this way is unhelpful and probably warns of the prejudice or political motivation of the critic. Each processed food should be judged on its own merits and should be considered within the context of the whole diet rather than purely in isolation.

Commercial processing of food has a number of objectives, in addition to the obvious commercial ones, such as those listed below.

- Processing reduces or prevents the chemical deterioration and microbial spoilage of foods. It thus increases the shelf-life of foods and increases food availability, reduces waste and may reduce the cost of food.

- By destroying pathogens or preventing their growth, processing lowers the risk of foodborne diseases.
- It can increase the palatability of foods.
- Processing enables new varieties of foods to be created.
- Commercially prepared food products reduce the time that individuals need to spend on food preparation. Even those unable or unwilling to spend much time on food preparation can eat an interesting and varied diet if they can afford to buy high-quality prepared foods.

Commercial processing usually, but not always, has some negative effects on the nutrient content of the unprepared food. These losses may be very similar to, or even less than, those involved in the home preparation of food. They are often not very significant in the diets of North Americans and Western Europeans, who, in the main, have ample intakes of the essential nutrients. In many cases, the nutrient content of processed foods (e.g. vitamins in frozen vegetables) may actually be higher than in stale versions of the same food bought 'fresh'. It is often argued that food processors encourage people to consume foods high in fat, sugar and salt but low in complex carbohydrate, and thus are partly responsible for the increased prevalence of the diseases of industrialization. It seems probable, however, that even without the encouragement of food processors, our natural inclination to consume such a diet would have prevailed. It is true, however, that, for example, most of the salt in the UK diet has been added by food manufacturers and reductions in the salt of commercially processed foods would have a major impact on total salt intakes. Although sales of sugar have dropped considerably in recent years in the UK, there has been a smaller fall in total sugar intake because more sugar is consumed within commercially prepared foods. Commercial processing tends to reduce the consumers' ability to monitor and regulate their intakes of sugar, salt, fat, etc.

Commercial processors and food retailers provide what they think they can profitably sell. They have, for example, been quick to recognize the commercial opportunities offered by the demand for 'healthy foods'. Many highly processed foods have a very healthy image, e.g. some margarine and low-fat spreads, some breakfast cereals, low-fat salad dressings, low-fat yoghurts, low-calorie drinks and certain meat-substi-

- The term 'processed food' covers a huge and diverse range of food products and so general discussion of their merits and flaws is of limited usefulness.
- Commercial processing often extends the shelf-life of foods and may increase their microbiological and chemical safety.
- Commercial processing can improve the palatability of individual foods and increase variety by making foods more accessible and even allowing the creation of new foods.
- The availability of commercially prepared foods decreases the time individuals need to spend on food preparation, although this may increase food costs.
- Nutrient losses during the commercial processing of foods may be comparable to losses during home preparation and generally do not threaten the adequacy of Western diets.
- Many commercially prepared foods are high in fat, salt and/or sugar, but low in complex carbohydrate.
- Food manufacturers will, at least superficially, respond to consumer demands for healthier processed foods if it is profitable for them to do so.
- Many highly processed foods are marketed on their image as healthier options.

tutes of vegetable or microbial origin. Some of the modern **functional foods** that are marketed on their proposed health benefits are the products of modern food processing, e.g. margarine with high levels of plant sterols that lower blood cholesterol.

Specific processing methods

CANNING

This is a very good method for the long-term preservation of food. The food is maintained in sealed containers and thus protected from oxidative deterioration and the growth of aerobic micro-organisms. The food is subjected to a vigorous heat treatment after canning that is designed to ensure that there is no practical possibility that even heat-resistant bacterial spores will survive. The so-called botulinum cook ensures the destruction of all spores of the potentially lethal toxin-producing, anaerobic organism *C. botulinum*. If commercially canned food is consumed soon after opening, there is negligible risk of it being associated with food poisoning.

PASTEURIZATION

Mild heat is applied to a food that is sufficient to destroy likely pathogenic bacteria and to reduce the number of spoilage organisms without impairing the

food itself. It is usually associated with milk, but other foods, such as liquid egg, liquid ice cream mix and honey, may be pasteurized. The pasteurization of milk traditionally involved holding it at 63 °C for 30 minutes. This was designed to ensure the destruction of the organism responsible for tuberculosis, but it also kills *Salmonellae* and the other heat-sensitive pathogens usually found in milk. The same result can be achieved by the modern **high-temperature short-time** method in which the milk is raised to 72 °C for 15 seconds.

ULTRA HIGH TEMPERATURE (UHT) TREATMENT

Once again, this is usually associated with milk but is also applied to other foods such as fruit juice. Traditionally, milk was sterilized by heating it to temperatures of 105–110 °C for 20–40 minutes. Such severe heating causes marked chemical changes in the milk that impairs its flavour, appearance, smell and nutrient content. The **ultra high temperature (UHT)** method relies upon the principle that the rate of chemical reactions only doubles with a 10 °C rise in temperature, whereas the rate of bacterial killing increases about tenfold. Thus, full sterilization can be achieved by holding the milk at 135 °C for 2 seconds, with little chemical change in the milk. Two seconds at 135 °C has approximately the same bacterial killing effect as 33 minutes at 105 °C, but results in chemical changes equivalent to only 16 seconds at

this temperature. After UHT treatment, the milk is placed aseptically into sterile containers and will keep for up to 6 months. Chemical deterioration rather than microbial spoilage limits the duration of storage of UHT products.

COOK CHILL PROCESSING

With the cook-chill process, foods are cooked separately as they would be in the home kitchen. Bacterial cells will be killed by this process but spores will survive. The food is then chilled rapidly to temperatures below 5 °C to minimize the growth of the surviving spore-forming bacteria. The food is then divided into portions and packaged. Rigorous standards of hygiene are required during this portioning and packaging stage to prevent re-contamination with spoilage and food-poisoning organisms. Most cook-chill foods have a maximum permitted shelf-life of 6–8 days, partly because of the difficulty of preventing contamination by the ubiquitous *Pseudomonas* group of spoilage organisms. The introduction of cook-chill foods into hospitals is an area of particular concern. Many groups that are very vulnerable to organisms such as *Listeria* are concentrated in hospitals, e.g. pregnant women and infants in maternity units, people whose immune systems have been suppressed by disease (e.g. autoimmune deficiency syndrome, AIDS) or by treatment (e.g. immunosuppressant drugs), and those weakened by injury, disease or old age.

FOOD IRRADIATION

Irradiation involves exposing food to ionizing radiation, which can either be X-rays or, more usually, gamma-rays emitted from a radioactive source such as cobalt-60 (reviewed by Hawthorn, 1989). There is no contact between the radioactive source and the food and so the irradiated food does not itself become radioactive. The irradiation of food is not a new idea, its potential was recognized and demonstrated around a 100 years ago. It is only in recent years, however, that the widespread application of irradiation in food processing has become economically viable because of the widespread availability of gamma-emitters like cobalt-60. Many countries have now approved the use of irradiation for some categories of foods.

In the USA, irradiation is classified with the food additives because it causes chemical changes in food.

Its use is permitted for a wide range of foods, including onions, potatoes, herbs and spices, nuts and seeds, wheat, fish, poultry and meat. Irradiated food must be clearly labelled as having been irradiated.

In the UK, prior to 1990, irradiation was forbidden for all foods with the exception of herbs and spices. At the beginning of 1991, this ban was withdrawn and irradiation of a wide range of foods is now permitted in properly licensed centres, e.g. fruit, vegetables, cereals, fish, shellfish and poultry may all now be legally irradiated in the UK. There is currently only one site in the UK that is licensed to irradiate human food and the licence only covers the irradiation of spices and condiments. Irradiated foods must be labelled as such. New EU regulations mean that, even when a prepared food contains only a small amount of irradiated ingredient, this must be declared on the label.

The effects of irradiation on the food vary according to the dose used (Figure 17.3).

Some of these changes are used to extend the shelf-life of foods, such as by inhibiting sprouting, delaying ripening and reducing the numbers of spoilage organisms.

The killing of insect pests reduces losses and damage to food from this cause and reduces the need to fumigate with chemical insecticides.

The elimination of food-poisoning organisms should reduce the risks of food poisoning, and sterilization of food by irradiation may be particularly useful in ensuring the safety of food intended for immunosuppressed patients.

Lowest dose	Inhibits sprouting, e.g. potatoes Delays ripening Kills insects pests Eliminates parasites, e.g. <i>Trichinella</i> worms in pork Reduces microbial load and delays spoilage Eliminates most pathogens (but not <i>C. botulinum</i>)
Highest dose	Full sterilization

Figure 17.3 The effects of varying doses of ionizing radiation on food.

Despite these apparent advantages of food irradiation, its introduction has been vigorously opposed by certain pressure groups within the UK. It has been suggested that the full effects of irradiation are not yet well enough understood to be totally confident of its safety.

Some potential problems of irradiated foods are listed below.

- The ionizing radiation induces certain chemical changes within the food; it increases free-radical production and certain chemical species known as radiolytes. (See Chapter 12 for a discussion of free radicals and antioxidants.) There is natural concern that these chemical changes may have detrimental effects on health, although it must be added that all processing methods produce some chemical changes in food, e.g. smoking and barbecuing lead to the production of small amounts of potentially carcinogenic chemicals in food. The chemical changes induced in food by irradiation are less than those produced by heat processing. The low level of chemical change induced by irradiation is illustrated by the extreme difficulty in

chemically identifying whether or not a food sample has been irradiated.

- Some bacterial and fungal toxins will not be destroyed by irradiation so, if irradiation is used to compensate for earlier poor hygiene, food poisoning from these toxins may result.
- At the doses usually used on foods, spores will survive and may later grow in the food. Irradiation is known to increase mutation rates and so, conceivably, new and dangerous mutants of micro-organisms might be produced in food.
- At high doses, e.g. those required to achieve sterilization, there may be distinct adverse effects on the palatability of the food and major losses of some nutrients such as thiamin and vitamin E. At dose levels likely to be used for 'pasteurization' of food, irradiation produces almost no change in the flavour or nutrient content, unlike other methods of food processing and preservation.
- There is concern about the safety of workers at establishments where food irradiation takes place. There is thus a clear need to regulate and licence facilities for food irradiation.

- Rigorous heat treatment and sealed containers ensure that properly canned food will often remain in good condition and microbiologically safe for many years.
- The mild heat treatment of pasteurization should eliminate likely pathogens and reduce levels of spoilage organisms without adversely affecting the flavour or nutrient value of the food.
- Sterilization using ultra high temperature (UHT) for very short periods results in far less chemical change than traditional sterilization techniques.
- Cook-chill processing does not kill all spore-forming bacteria, and food may also be contaminated with other spoilage and pathogenic organisms during packaging.
- *Listeria monocytogenes* can still grow in chilled foods and this can have serious consequences for pregnant women and other vulnerable groups.
- Irradiating foods with varying doses of X-rays or γ -rays has a variety of potential uses that are summarized in Figure 17.3.
- Consumer resistance has limited the marketing of irradiated foods, although many countries permit the irradiation of many foods.
- Irradiation can extend the shelf-life of many foods and/or improve their microbiological safety.
- High doses of radiation do induce potentially adverse chemical changes in food and loss of nutrients.
- Mild irradiation produces imperceptible changes in flavour or loss of nutrients and it is difficult to detect chemically whether or not a food has been irradiated.
- Some microbial toxins will survive irradiation, and radiation may increase mutation rates in organisms that survive the irradiation process.

THE CHEMICAL SAFETY OF FOOD

Overview of chemical hazards in food

There are three potential sources of chemical hazard in food:

1. *food additives*: chemicals that are deliberately added to food during processing
2. *natural toxicants*: compounds naturally present in the food that may have toxic effects
3. *contaminants*: substances that are incidentally or inadvertently added to foods during agricultural production, storage or processing, e.g. residues of drugs given to food animals, pesticide or fertilizer residues, contaminants leaching into food from packaging or containers, fungal toxins.

One dilemma that increasingly faces those trying to ensure the chemical safety of food is to decide at what level in the food the presence of any particular chemical substance represents a real hazard to consumers. Total absence of all traces of potentially hazardous chemicals from all foods is, and always has been, an impossible ideal. Almost every chemical, including most nutrients, is toxic if the dose is high enough. More substances are being subjected to rigorous safety tests, which usually involve exposing animals to high doses of the chemical for prolonged periods. Many chemicals that are ubiquitous in our environment can be shown to have toxic potential by such tests. Analytical procedures are also becoming increasingly sophisticated and sensitive, making it possible to detect infinitesimally small amounts of potentially hazardous chemicals in foods. This combination of factors, especially when distorted by popular journalists, can lead to the impression that the chemical hazards in our food are increasing at an alarming rate. At least some of this apparent increase in chemical danger from food is artefact caused by increased awareness of the potential hazards of chemicals and increased ability to detect small amounts of these chemicals.

In a report by the Institute of Food Technologists (IFT, 1975), the panel distinguish between **toxic**, which they define as 'being inherently capable of producing injury when tested by itself', and **hazard**, which is defined as 'being likely to produce injury under the circumstances of exposure' as in food. Food naturally contains thousands of toxic substances, including many nutrients, but very few of these represent real hazard. For example, the average US consumer ingests around 10 000 mg of **solanine** from potatoes each year. This is enough of this atropine-like alkaloid to kill someone if consumed in a single dose. The same average US consumer ingests about 40 mg of lethal hydrogen cyanide in his or her annual kilogram of lima beans and 14 mg of arsenic from seafood.

Most natural toxins are present in low concentrations in natural foods and they do not usually accumulate. They are usually metabolized or excreted by a variety of mechanisms that are also involved in the disposal of ingested synthetically produced chemicals such as food additives, drugs and residues of agricultural chemicals. There is a variety of mechanisms and sites through which the different toxicants induce their effect and so, generally, the effects of small amounts of individual toxicants are unlikely to be additive. Thus, 1% of a toxic dose of 100 different toxins almost certainly will not produce ill-effects (IFT, 1975).

Natural toxicants and contaminants

CIRCUMSTANCES THAT MAY INCREASE CHEMICAL HAZARD

The chances of toxic potential becoming real hazard are likely to be increased if there is exaggerated consumption of one food for a prolonged period. This is true whether the toxin is naturally present, a contaminant or a deliberate additive. This is one reason for encouraging diversity in the diet. Examples of natural toxicants in food producing serious illness may occur when abnormally high amounts of a

- Natural toxicants, contaminants and food additives must all be considered potential sources of chemical hazard in food.
- Sensitive analytical techniques can detect numerous substances in food that have toxic potential, but few of them represent a real hazard to consumers who eat a varied diet.

particular food are consumed. The plant *Lathyrus sativa* has been widely grown in Asia and North Africa and the seeds (chickling peas or khesari dhal) regularly consumed. However, during very dry seasons, it was consumed in very large quantities because the plant is very drought resistant. When consumed in very large quantities, it can produce a severe disease of the spinal cord, **lathyrism**, which can lead to permanent paralysis. When only small quantities are consumed as part of a mixed diet, they do not constitute a hazard and are nutritious.

Chemical hazard from food may arise if individual susceptibility causes increased sensitivity to a particular toxin. The common broad bean (*Vicia faba*) contains a substance that causes haemolysis (red cell breakdown). This can lead to a severe anaemia called **favism** in those who are genetically susceptible because they are deficient in a particular enzyme (glucose 6-phosphate dehydrogenase). As many as 35% of some Mediterranean peoples and 10% of American blacks have this particular genetic deficiency. Vomiting, abdominal pain and fever are the acute symptoms; jaundice and dark-coloured urine may occur as a result of the haemolysis, with, eventually, severe anaemia a possibility.

Traditional methods of preparing, processing and selecting foods often minimize any potential hazard they may represent. Cassava is one of the most important staple foods for millions of people in the tropics; it may, for example, provide up to 60% of the calorific intake in Nigeria. Cassava contains certain alkaloids that release cyanide when acted upon by an enzyme in the cassava. The traditional method of preparing cassava involves peeling and soaking for several days and most of the cyanide is lost due to fermentation. Cases of sometimes fatal cyanide poisoning are associated with inadequate processing of the cassava (particularly shortening of the fermentation time) and the increased use of lower-quality 'bitter' cassava, which has a higher cyanide content (see Akintonwa and Tunawashe, 1992). As another example, polar bear liver contains toxic concentrations of retinol. Eskimos avoided eating the liver, but unwary polar explorers have been poisoned by eating it.

SOME NATURAL TOXICANTS IN 'WESTERN' DIETS

Few natural toxicants are thought to represent significant hazards to Western consumers. In her analysis of more than 1000 recorded outbreaks of food poi-

soning in England and Wales, Roberts (1982) found that 54 outbreaks (about 5%) were due to chemical toxicity – 47 due to **scrombotoxin** from fish and seven to a **haemagglutinin** in red kidney beans.

Some potential natural chemical hazards in UK food are listed below.

- Scrombotoxic poisoning is caused by heat-stable toxins liberated by the action of bacteria upon fish protein during spoilage. The toxin is thus produced by bacterial action but is not in itself a bacterial toxin. Symptoms occur shortly after eating the contaminated fish and include a peppery taste in the mouth, flushing of the face and neck, sweating and sometimes nausea and diarrhoea.
- Eating raw red kidney beans leads to short-lasting symptoms of nausea, vomiting and diarrhoea. These symptoms are thought to be due to a toxin that causes red cells to agglutinate (stick together) – a haemagglutinin. The symptoms are probably due to damage to intestinal cells caused by this toxin. The haemagglutinin is destroyed by vigorous boiling, but it may persist if the beans are cooked by prolonged gentle heating in a slow cooker. These beans should always be subjected to ten minutes of vigorous boiling before being eaten. Roberts (1982) found that all of the outbreaks she investigated were due to eating the beans in an uncooked or undercooked state. As a general rule, raw or undercooked beans should not be eaten as they contain several mildly toxic factors and factors that interfere with the proper digestion of protein.
- Some cheeses contain the substance **tyramine**, which can cause a rise in blood pressure. This may be dangerous for those taking certain antidepressant drugs because they sensitize the individual to the effects of tyramine.
- Solanine is an atropine-like substance found in potatoes. In high enough doses, it will cause headache, vomiting and diarrhoea and may perhaps even result in circulatory collapse and neurological disturbances. Levels of solanine in potatoes are rarely enough to produce illness, and established outbreaks of potato poisoning are rare.
- Many species of *Brassicae* (cabbage family) contain goitrogens – chemicals that induce goitre. These do not represent a real hazard at levels of consumption normally found in industrialized countries.

- Mussels that have ingested the plankton species *Gonyaulax tamarensis* may contain hazardous amounts of a heat-stable neurotoxin – **saxitoxin**. At certain times, the sea may turn red ('red tides') due to large numbers *Gonyaulax* in the water. These 'red tides' may occasionally occur even off the coasts of Britain and the US and at such times it is dangerous to eat mussels.
- Some fungi produce toxic chemicals – **mycotoxins**. Some mycotoxins represent a hazard to the inexperienced gatherer of wild fungi and some fungal toxins are deliberately consumed because they contain hallucinogenic agents. Moulds also grow upon many foods that will not support the growth of bacteria because they are more tolerant of low water activity and low pH than bacteria. Dry foods such as nuts and bread, sugary foods such as jam (jelly) and salty foods such as cheese may all go mouldy. Several of the toxins produced by moulds are potent carcinogens. **Aflatoxins** produced by *Aspergillus flavus* have been responsible for outbreaks of fatal poisoning in Taiwan, India and Uganda. They cause gastrointestinal bleeding, liver damage and pul-

monary oedema. They have been shown to cause liver cancer in animal studies. Mouldy nuts would be a likely source of aflatoxin in the USA and UK. Mouldy grain is likely to contain the toxin **ergot** and this has caused serious outbreaks of poisoning in the past, and still does in some countries, a condition referred to as St Vitas dance. Mouldy food should therefore be regarded as potentially hazardous and should be discarded. In the early 1990s, the detection of small amounts of the mycotoxin patulin in some brands of apple juice received great publicity in the UK.

Mould growth is deliberately encouraged in the production of some foods, such as blue cheese and mould-ripened cheese. This is not thought to represent a hazard at usual levels of consumption.

RESIDUES OF AGRICULTURAL CHEMICALS

Expert opinion is that the residues of agricultural chemicals in foods represent no significant hazard to consumers when current regulations are adhered to. These assurances have been insufficient to convince a

- Food naturally contains small amounts of many toxic substances that, because they are eaten in such small doses, do not cause illness.
- The risk of a natural toxicant causing illness is increased by:
 - abnormally high consumption of the toxin-containing food
 - increased susceptibility to the toxin, e.g. genetic
 - inadequate removal of the toxin during processing.
- Some toxins present in many Western diets are:
 - scrombotoxin produced by bacterial action upon fish protein
 - haemagglutinins and other toxins in raw or undercooked beans
 - tyramine in cheese, which interacts with antidepressant drugs
 - an atropine-like alkaloid, solanine, normally present in minute amounts in potatoes
 - goitrogens present in minute amounts in plants of the cabbage family
 - mussels contaminated with a neurotoxin produced by the red plankton *Gonyaulax tamarensis*
 - mycotoxins produced by fungi and by mould growth upon food.
- Most, but not all, scientists believe that, when used in permitted amounts, residues of agricultural chemicals do not represent a significant hazard to consumers.
- Most scientists therefore do not believe that more expensive organic food produced without agricultural chemicals is healthier.
- Many consumers do not accept the scientific consensus about organic food and there may well be stronger ecological arguments for less chemical-dependent food production methods.

significant minority of the population who are willing to pay considerably higher prices for organic produce that has been grown without the aid of modern agricultural chemicals. These consumers are also willing to accept the less than perfect appearance of the fruits and vegetables that organic farming practices sometimes produce. Many producers and suppliers have been quick to recognize the commercial opportunities afforded by this new market. People who can afford it clearly have the right to choose food grown in this way, but the practicability of supplying the whole population with organic food is doubtful and there are no scientifically quantifiable health benefits. There may well be more convincing environmental arguments for farming in this way.

- They are dangerous chemicals *per se*; in particular, that chronic exposure to them will lead to an increase in cancer risk.
- They can be used to disguise faulty or inferior products and thus deceive the consumer. Colourings, flavouring and emulsifiers can disguise the high-fat and low-grade meat used in some meat products. Colourings and flavourings can disguise the lack of fruit in 'fruit' drinks and yoghurts. Polyphosphates can be used to artificially increase the weight of meat and poultry by increasing its water content.
- Even generally safe additives may trigger adverse reactions in some individuals, e.g. allergy to the yellow food colourant tartrazine and to sulphites (sulphur dioxide) which are used to preserve many foods.
- Preservatives can be used by manufacturers to compensate for poor hygiene standards.
- Many of the cosmetic additives are unnecessary or perhaps even imposed upon the consumer by food producers and retailers.

Food additives

Around 300 food additives are used in the UK, together with another 3000 flavourings. The average British person probably consumes several kilograms of these in a year.

USES

Chemicals are deliberately added to food for a variety of purposes, such as:

- as processing aids, e.g. emulsifiers, flour improvers and anti-caking agents
- to improve the sensory appeal of foods, e.g. colours and flavours
- to improve the nutritional value of foods, e.g. vitamins and/or minerals added to bread, breakfast cereals and drinks
- to prevent the growth of food-poisoning and spoilage organisms, e.g. nitrites and nitrates in meat products
- to inhibit the growth of moulds, e.g. propionic acid or vinegar in bread
- to inhibit the chemical deterioration of foods, e.g. antioxidants such as vitamins C and E.

SOME ARGUMENTS AGAINST THE USE OF FOOD ADDITIVES

Those who are opposed to the use of food additives or the scale of their current usage have used all of the arguments below to criticize them.

SOME COUNTER-ARGUMENTS

Additives with food-preserving functions are a necessity if large urban populations are to be supplied with a variety of safe, nutritious and affordable foods. Traditional preservatives have long been used for these purposes, e.g. salt, sugar, woodsmoke, vinegar, alcohol, nitrites and nitrates. When evaluating the safety of modern preservatives or when judging the merits of foods claiming to be 'free from all artificial preservatives', it should be borne in mind that most of these traditional preservatives have been implicated in disease. For example, salt has been linked to hypertension, nitrites lead to the generation of potentially carcinogenic nitrosamines, and small amounts of potential carcinogens are found in smoked foods.

Some additives are essential as processing aids or otherwise necessary for the manufacture of a considerable number of supermarket foods. Some of these foods are considered to be 'healthy foods' and are seen as important in helping consumers to comply with current nutrition education guidelines. Emulsifiers and stabilizers are essential for the production of many 'reduced-fat' products such as low-fat spread and polyunsaturated margarine. Anti-caking agents are needed for the manufacture of many powders that

are to be instantly rehydrated, such as coffee whitener. Artificial sweeteners or **sugar replacers** are necessary for the production of many 'low calorie' and 'reduced-sugar' foods.

The additives that are most vulnerable to criticism are those that serve cosmetic purposes, i.e. that are there to enhance the appearance or palatability of the food. These may even be claimed to be doing a positive disservice to Western consumers by encouraging overconsumption and obesity. If a purely scientific model of food function is used, appearance and palatability could almost be regarded as a decadent irrelevance. However, few people would really want to regard the positive sensory appeal of food as an optional extra, although, as already noted, flavourings and colourings do have the potential to allow manufacturers to dupe the public into buying prepared foods made from poor-quality ingredients.

FOOD ADDITIVE REGULATION

Different countries have different regulations governing the use of food additives, but the common purposes of such regulations are to ensure that the additives are used safely, effectively, honestly and in the minimum amounts necessary to achieve their objectives.

The **Food and Drug Administration (FDA)** regulates the use of food additives in the USA. In order to obtain approval to use a new additive, a manufacturer is required to provide evidence that the additive is effective, safe and can be detected and measured in the final product. The approval of the additive is then considered after a public hearing at which experts testify for and against its use. Additives that were in use before this procedure was adopted were put onto a list of substances **generally recognized as safe (GRAS)**. The GRAS list is subject to a continuing process of review and revision. More than 35 years ago, Congress approved the so-called **Delaney clause**: 'no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal'. This requirement is now regarded as unreasonably absolute in its prohibition, and the FDA deems additives to be safe if the risk of human cancer is less than one in a million. Saccharin is a permitted sweetener in the USA, despite being reported to cause cancer when administered in very large doses to animals, i.e. although shown to be toxic, it is not thought to represent a hazard to the US consumer.

In the UK, the responsibility for the regulation of the use of food additives lies with the **Ministry of Agriculture Fisheries and Food (MAFF)**. A **Food Advisory Committee** decides whether any proposed new additive or new usage of an existing additive is necessary. This committee is comprised of a wide variety of interested persons, including academics, consumer representatives, food industry representatives and public analysts. Grounds for necessity would be factors such as:

- increased shelf-life of a product
- reduced cost
- an improved product for the consumer
- necessity for the manufacture of a new product or for the introduction of a new manufacturing process.

Once need has been established, evidence must be presented to an expert **Committee on Toxicology**, which assesses the safety of the additive. Existing additives may be referred to this committee for safety review.

E numbers have been used to designate food additives in Britain and the rest of the EU. They were introduced as part of efforts to harmonize legislation within the EU and to overcome language barriers. They were originally envisaged as something that would help to engender consumer confidence. If an additive had an E number, the consumer could be totally assured of its safety. In reality, however, the effect was quite the opposite: E numbers on food labels evoke suspicion and have been used to epitomize everything that is unwholesome about 'modern, adulterated and artificial food'. A full listing of E numbers can be found in the web pages of Dr David Jukes (www.fst.rdg.ac.uk/foodlaw). A quick perusal of this extensive list shows that several nutrients and compounds that would be familiar (and perhaps less threatening by name) are included in this list:

- E101 is riboflavin
- E300 is vitamin C
- E302 is calcium carbonate
- E307 is vitamin E
- E120 is cochineal
- E140 is chlorophyll
- E150 is caramel
- E160a is carotene
- E260 is acetic acid (vinegar)
- E553b is talc
- E901 is beeswax.

TESTING THE SAFETY OF FOOD ADDITIVES

Listed below are the four potential sources of information that might be used to evaluate the safety of food additives:

1. human experiments
2. human epidemiology
3. *in vitro* tests
4. animal experiments.

One of the major concerns relating to the safety of food additives is that life-time exposure to an additive may increase the risk of chronic disease and especially of human cancer. Human experiments are inevitably either short term or small scale; such experiments can be of no real use in identifying even relatively large increases in cancer risk due to particular additives. Also, epidemiology cannot be expected to pinpoint harmful effects of individual additives. Epidemiology involves relating exposure to changes in disease rates, yet we are all exposed to hundreds of additives in amounts that would be difficult to quantify on an individual basis and, even on a 'worse-case scenario', the increased cancer rate due to the additive is likely to be small. It proved difficult to show convincingly with epidemiological methods that smoking causes lung cancer, even though the association is thought to be strong, the level of exposure is relatively easy to establish and there is an identifiable matched but unexposed population.

In vitro tests involve the use of isolated mammalian cells or micro-organisms. The best established of these tests are those, such as the Ames test, that use the ability of compounds to cause mutation in bacteria as an indicator of their likely carcinogenicity in mammals. Mutagenesis, carcinogenesis and teratogenesis (causing foetal abnormalities) may all be caused by damage to the genetic material DNA. Such tests are a useful way of screening out some potential carcinogens or teratogens and thus reducing the number of compounds that need to undergo full animal testing. *In vitro* tests are, however, not able to provide positive assurance that a compound or one of its metabolites is going to be safe in higher animals. Such tests are, in any case, validated by comparison of the mutagenic effects of chemicals with their carcinogenic effect when used in animal tests.

Animal tests upon food additives can be categorized under four main headings: acute and subacute

toxicity tests; absorption, distribution, metabolism and excretion tests; teratogenicity tests; and, long-term carcinogenicity tests.

- *Acute and subacute toxicity tests.* These tests seek to establish just how much of the additive is required to produce acutely toxic and fatal effects. In the subacute tests, animals are exposed to very high doses for 3 months and tissues are examined microscopically at autopsy for signs of damage.
- *Absorption, distribution, metabolism and excretion tests.* These seek to establish how much of the additive is absorbed, where it goes after absorption, how it is chemically processed after absorption and how quickly it is excreted. Such studies can give important pointers as to the dangers likely to be associated with the consumption of the additive (see examples below).

If the compound is not absorbed from the gut, any adverse effects are likely to be confined to the gut.

If a water-soluble substance is absorbed and then excreted unchanged in the urine, the bladder is the likely danger organ if the substance is carcinogenic.

If the substance is absorbed but only slowly excreted or metabolized, it will tend to accumulate. Chronic low intakes may lead to the build up of high levels.

If the substance is metabolized, the metabolites need to be identified and their potential toxicity also needs to be assessed.

If substances are absorbed and detoxified in the liver and then excreted in the bile or urine, the liver may be likely to be affected if the compound is carcinogenic.

- *Teratogenicity tests.* These involve feeding large amounts of the compound to pregnant animals to see whether the compound causes birth defects or in any way harms the developing foetuses.
- *Long-term carcinogenicity tests.* These usually involve exposing test animals to relatively high doses of the additive over their whole life span. The tumour rates in exposed animals are then compared to those of control animals who have been maintained under identical conditions but have not been exposed to the additive. Such controlled experiments mean that experimenters are able to confidently attribute any increased tumour rate in the test group to the carcinogenic

effect of the additive. There are a number of potential criticisms of these tests, such as those discussed below.

- Substances that are not carcinogenic in laboratory animals may be carcinogenic in people.
- The controlled conditions of testing are very different from those of use. In the tests, genetically homogeneous animals are used, animals are usually fed single additives and are not usually exposed to other chemicals (e.g. drugs, cigarette smoke or alcohol). Animals are also maintained on defined laboratory diets. In use, the additive will be fed to genetically diverse people - perhaps only certain

genotypes will be sensitive to its toxic effects. In use, there is the possibility that the additive may only become toxic if it interacts with other chemicals or perhaps it becomes harmful under particular dietary or environmental conditions, e.g. deficiency of antioxidant vitamins or minerals.

- Laboratory animals have relatively high background tumour rates. This means that the additive needs to produce relatively large increases in tumour rate to be detected with the comparatively small numbers of animals used in the tests. The signal (increase due to additive) must be large to detect it against the background noise (spontaneous tumours not due to additive and occurring

- Hundreds of food additives and thousands of flavourings are in general use.
- Additives are used:
 - as processing aids
 - to preserve food and increase its microbiological safety
 - to improve the taste, texture or appearance of food
 - to increase the nutrient content of food.
- Some criticisms applied to food additives include:
 - they may be toxic or have carcinogenic potential
 - they can be used to disguise low-quality ingredients or poor standards of hygiene
 - idiosyncratic reactions may occur even with additives that are generally safe
 - flavourings and colourings are unnecessary.
- However:
 - preservatives are necessary for safe mass production of food and even traditional preservatives have their problems
 - food additives are necessary to produce some modern foods that are marketed as healthy options.
- In the USA, the use of food additives is regulated by the FDA and, in the UK, by MAFF.
- Regulatory authorities try to ensure that additives are necessary, used in minimum amounts to achieve their purposes, and safe when used in permitted amounts.
- In the EU, all additives have a designated E number to be used on food labels.
- Even if huge doses of a chemical produce adverse effects in laboratory animals, it does not necessarily mean that small amounts represent a hazard to people.
- Safety testing of additives is inevitably very dependent upon animal tests.
- Differences between species, the defined conditions of laboratory tests and the high spontaneous tumour rates in laboratory animals all tend to reduce confidence in the validity of these as tests of human safety.
- A large safety margin is used when the dose that is demonstrably safe in animals is translated into an acceptable intake for people.

randomly in both groups). Thus, these tests may be insensitive indicators of carcinogenicity, and even small increases in risk can be significant if hundreds of millions of people are to be exposed.

Safety testing of additives depends very much upon animal experiments because they are the only practical methods available. In order to try to ensure safety in use, despite the potential flaws discussed, wide safety margins are used. A **no-effect level** is identified in animal experiments, i.e. a dose of lifetime exposure that produces no detectable effects in any species used. This no-effect level is then divided by a safety factor (usually 100) to determine an **acceptable daily intake** for people. It is, of course, difficult to control the additive intakes of individuals by regulating the maximum amounts permitted in particular foods. The difficulty of converting dosages between species of different sizes has already been discussed in Chapter 3.

Most food additives have been subjected to extensive safety testing. Many substances naturally present in foods would probably fail such tests and, of course, a number of additives are naturally present in some foods. Nevertheless, the inevitable flaws in the testing procedures do suggest the need to use the minimum effective amounts and for there to be a continuing critical review of which additives are to continue to be recognized as safe. Millstone (1985) critically reviewed the testing and regulation of food additives from a British perspective.

the 'active' ingredients of functional foods, are found in ordinary foods, such as the **phytochemicals** found in fruits and vegetables.

Functional foods are legally classified as foods rather than medicines. In Britain, they are not permitted to carry claims that they are able to prevent or cure specific diseases – only substances legally classified as medicines are permitted to carry such claims. According to Dowden (see the first web-site listed at the end of this section), it would be acceptable to say that a food 'provides calcium, which is important for strong bones', but not that it 'provides calcium which helps prevent osteoporosis'. Permitted health claims on US food labels are discussed earlier in the chapter.

In 1999, the European and American markets for functional foods that made a specific health claim were estimated at well over £1 billion each and growing strongly. In Europe, **probiotics** account for over 70% of the functional food market. Probiotics are fermented dairy products that contain living cultures of *Lactobacilli* or *Bifidobacteria*. These bacteria are said to improve the microbial balance in the intestine and they may reduce the incidence or severity of gastrointestinal and perhaps vaginal infections. **Prebiotics** are non-digestible food ingredients that selectively promote the growth of some bacteria within the gut and thus have similar effects to probiotics. The plant polysaccharide inulin may have such an action.

Some other examples of actual or potential functional foods are briefly discussed below.

- **Plant sterols.** Some plant sterols, such as sitosterol, ergosterol and, more recently, sitostanol, inhibit the absorption and reabsorption of cholesterol in the gut. They can thus lower blood cholesterol concentration. Margarine and other dairy products containing such plant sterols are currently being heavily marketed on this basis (see Chapter 11).
- **Olestra.** This is a synthetic fat that is not digested in the gut and so passes out through the faeces. It has the palatability benefits of fat but yields no calories. Its use is permitted in the USA in savoury snacks such as potato crisps (chips), but not in the EU. There are a number of short-term problems with such products and many scientists have serious concerns about the long-term effects of regularly eating foods containing fat substitutes like olestra. In the short

FUNCTIONAL FOODS

Functional foods can be defined as those that have components or ingredients incorporated into them in order to give a specific medical or physiological benefit in addition to their purely nutritional value. These foods often carry some form of health claim on the packaging or in their advertising. They are also sometimes referred to as neutraceuticals, a term that implies that they have both nutritional and pharmaceutical functions. The term functional food could encompass some ordinary foods that have been fortified with a nutrient, such as bread or breakfast cereals fortified with high levels of folic acid. Folic acid supplements taken by pregnant women reduce the risk of their babies being affected by a **neural tube defect** (see Chapter 14). Many of

term, they can cause problems because of the entry of indigestible fatty material into the large bowel, e.g. diarrhoea, bowel urgency, anal 'leakage', abdominal cramps and greasy staining of toilets and underwear. They also reduce the absorption of fat-soluble substances, including fat-soluble vitamins and cholesterol. In the USA, foods containing olestra have to be fortified with fat-soluble vitamins to compensate for their reduced absorption, although there may well be other, unknown but beneficial fat-soluble substances in foods. Olestra is currently restricted to use in savoury snacks in the USA on the grounds that these foods are not usually eaten with meals and so the olestra will not interfere with the absorption of vitamins in normal meals. Compounds like olestra are being promoted as potentially able to replace up to 30 g of fat per day in the diet.

- **Phytochemicals.** This is a term applied to a whole range of plant chemicals that may have beneficial

effects upon health. The possible health benefits of carotenoids and other antioxidants are discussed in Chapter 12. There are some substances in plants, the phyto-oestrogens, that have some oestrogen activity. These bind with the human oestrogen receptor and exert some oestrogen-like effect. They may increase total oestrogen activity when endogenous levels are low (e.g. after the menopause) and thus help protect against osteoporosis, or they may decrease oestrogen activity when it is high and possibly protect against breast cancer.

A supplement to the *British Journal of Nutrition* is recommended for further reading (BJN, 1998).

The following web-sites contain useful information on functional foods:

- www.dotpharmacy.co.uk/upneutra.html
- www.just-food.com/features/detail.asp?art=194
- www.nutrition.org.uk (search for functional foods 20 key facts)

- Functional foods or neutraceuticals have components that are intended to confer specific health benefits.
- Functional foods usually carry some health claim on their packaging, although in the UK they cannot claim to cure or prevent a specific disease.
- The worldwide market for functional foods is growing rapidly and 1999 sales in Europe and the USA each amounted to over £1 billion.
- Probiotics are fermented dairy products containing live bacteria.
- Probiotics accounted for 70% of the European market for functional foods in 1999.
- Probiotics are intended to alter the microbial balance in the intestine and hinder the growth of pathogens.
- Prebiotics are indigestible food components that are intended to enhance the growth of certain bacteria within the gut.
- Other examples of functional foods are those containing:
 - plant sterols that can lower blood cholesterol
 - calorie-free fat substitutes like olestra
 - a diverse range of phytochemicals that may act as antioxidants or have weak oestrogen activity

Glossary

Acceptable daily intake – the daily intake of a food additive judged to be safe for lifetime consumption.

Acculturation – the cultural changes that occur when two cultures interact, e.g. when immigrants adopt the cultural practices of the indigenous majority.

Activity diary – a system for measuring the activity level. A detailed diary is kept of activities in each time block (say, 5 minutes) of the measurement period. Energy expenditure can be crudely estimated from the assumed energy costs of the individual activities.

Adaptive thermogenesis – an increase in heat production whose supposed function is to burn off surplus calories and prevent excessive weight gain. It is suggested to occur in **brown fat**.

Adenosine triphosphate (ATP) – an important short-term intracellular energy store. Energy released during the metabolism of foodstuffs is ‘trapped’ as ATP. ATP energy drives synthetic and other energy-requiring cellular processes.

β_3 -adrenoreceptor – a receptor found specifically on **brown fat**. Drugs that stimulate this receptor should increase heat production and burn off surplus calories.

Aerobic capacity – the capacity of the cardiopulmonary system to supply oxygen to the tissues – one definition of fitness. It can be measured by the $vO_2 \text{ max}$, the PWC_{max} or the PWC_{170} .

Aerobic exercise – (aerobic – using oxygen) moderate exercise in which the oxygen supply to the muscles is sufficient to allow it to generate the required energy by normal aerobic means.

Ames test – a test that assesses the carcinogenic potential of a chemical from its tendency to cause mutation in bacteria.

Amino acids – the units that make up proteins. They have an **amino group** (NH_2) and a **carboxyl** or acid group ($COOH$).

α -amylase – a starch-digesting enzyme present in saliva and pancreatic juice.

Amylopectin – the greater part of dietary starch with the sugar residues in branched chains.

Amylose – a starch with the sugar residues in largely unbranched chains.

Anaemia – a low concentration of haemoglobin in blood. **Iron-deficiency anaemia** is characterized by small red blood cells (**microcytic anaemia**). Anaemia due to folic acid or vitamin B_{12} deficiency is characterized by large, unstable red cells (**macrocytic anaemia**); also called **megaloblastic anaemia** after the immature red cell precursors, megaloblasts, in the bone marrow.

Anaerobic exercise – (anaerobic – not using oxygen) intense exercise in which the supply of oxygen is insufficient to generate all of the required energy and so some is generated by other non-oxygen-requiring methods (e.g. weight lifting).

Anorexia nervosa – a potentially fatal, psychological disease principally seen in adolescent girls. It is characterized by an obsessive desire to be thin and self-starvation.

Anthropometry – measurement of the human body, e.g. measurements of weight and body dimensions.

Appetstat – a collective name for the brain centres that monitor and maintain body energy stores by regulating the feeding drive; analogous to a thermostat regulating temperature.

Appetite – the desire to eat.

Arachidonic acid, 20:4 ω -6 – a fatty acid of the ω -6 (n-6) series, the main precursor of the **eicosanoids**.

Artificial sweeteners – intensely sweet chemicals that can be used to sweeten foods but are not carbohydrate and are essentially calorie free, e.g. saccharin and aspartame.

Atherosclerosis – degeneration of arterial walls that is triggered by a build up of fatty deposits (hardening of the arteries).

Attack rates – in food poisoning, the proportion of people who become ill after exposure to a food-poisoning ‘incident’. **Food-specific attack rate** – the proportion of people eating a particular food who become ill.

Available carbohydrate – sugars and starches, the digestible carbohydrates. Non-starch polysaccharides are **unavailable carbohydrates** because they are indigestible by gut enzymes.

Balance – the difference between intake and losses of a nutrient. If intake exceeds losses – **positive balance**; if

losses exceed intake – **negative balance**. Note particularly **energy balance** and **nitrogen balance**.

Basal metabolic rate (BMR) – the rate of energy expenditure (metabolic rate) in the resting state, i.e. lying down in a warm room and some time after eating. The minimum metabolic rate in a conscious person.

Behaviour therapy – a psychological approach to the treatment of, for example, obesity. Subjects modify behaviour to avoid situations that trigger inappropriate eating behaviour and reward appropriate behaviours, e.g. taking exercise.

Beriberi – a deficiency disease due to lack of thiamin.

'**Best before**' – the date marked on UK foods that are unlikely to become microbiologically unsafe but whose eating qualities may have deteriorated by this time.

Beta (β)-oxidation pathway – the process by which fatty acids are metabolized to acetyl coenzyme A in the mitochondria.

Bioelectrical impedance – a method of estimating body fat content. It relies upon the principle that the resistance to an alternating current (impedance) passed between two electrodes on the body will depend upon the body composition.

Bitot's spots – white spots on the surface of the cornea caused by clumps of epithelial cells; a symptom of vitamin A deficiency.

Blind challenge – a diagnostic procedure used to confirm food intolerance. The subject is offered foods that do and do not contain the suspected cause but is unaware of which is which.

Body mass index (BMI) – the weight in kilograms divided by the height in metres squared. It is used as an indicator of overweight or obesity. Ideal range for adults is 20–25 kg/m².

Bomb calorimeter – a device for measuring the heat released when food is burned; a crude indication of its dietary energy value.

Botulinum cook – vigorous heat treatment applied to canned foods to ensure the destruction of spores of *Clostridium botulinum*.

Botulism – a rare but potentially fatal form of food poisoning caused by the toxin of *Clostridium botulinum*. It causes paralysis, including paralysis of the respiratory muscles.

Bovine spongiform encephalopathy (BSE) – a new disease of cattle. One of a group of fatal degenerative brain infections (**transmissible spongiform encephalopathies**, TSEs) that result in 'spongiform lesions' in the brain. Other spongiform encephalopathies are **scrapie**, in sheep and goats, and **kuru**, which is associated with cannibalism in New Guinea. **Creutzfeld–Jakob disease (CJD)** is a rare disease, normally confined to elderly people, but a new variant (**vCJD**) linked to eating BSE-contaminated beef has killed more than 80 young and middle-aged people in the UK.

Bran – the outer layer of the cereal grain containing most of the fibre (NSP); it may be removed during milling. White flour has had the bran removed, but in wholemeal flour it remains.

Brown adipose tissue (BAT) or brown fat – literally, adipose tissue that appears brown. It is prominent in small mammals and babies, where it is the principal site of non-shivering thermogenesis. Postulated as a site for **luxoskonsumption** or **adaptive thermogenesis**.

Bulimia nervosa – an eating disorder characterized by alternating periods of bingeing and compensatory behaviour such as fasting, intense exercise, induced vomiting or purging.

C-terminal – the end of a protein that has a free carboxyl group.

Cafeteria feeding – a method of increasing fatness in laboratory animals by allowing them access to a variety of tasty and 'fattening' foods; a model of human obesity.

Calcitriol – the active hormone produced from vitamin D (cholecalciferol) in the kidney – 1,25-dihydroxycholecalciferol.

Calorie – used to mean **kilocalorie (kcal)** in nutrition: a unit of energy defined as 'the heat energy required to raise the temperature of a litre of water by 1 °C'. 1 kcal = 4.2 kJ.

Cancer cachexia – weight loss associated with malignant disease.

Carbohydrate loading – training and dietary regimens that attempt to boost muscle glycogen stores.

β -carotene – a plant pigment found in many coloured fruits and vegetables that acts as a vitamin A precursor and an antioxidant.

Carotenoids – a group of plant pigments that includes **β -carotene**, **lycopene** and **lutein**. Some have vitamin A activity, but all are thought to have antioxidant potential.

Case-control study – a type of epidemiological study. Cases of a disease are matched with unaffected controls and the diet or some other exposure of the two groups compared. Differences may suggest the cause of the disease.

Casein – a milk protein; the dominant protein in cow milk.

Casein-dominant formula – infant formula with casein as the dominant protein.

Catalase – an iron-containing enzyme involved in free radical disposal.

Centile – hundredths of the population, e.g. the fifth centile for height would be the height below which would be found 5% of a population. Note also **quintile** (division into fifths), **quartile** (division into quarters) and **tertile** (division into thirds).

Chemical score – a measure of protein quality; the amount of the limiting amino acid in a gram of test protein as a percentage of that in a gram of reference protein, e.g. egg.

Cholecalciferol – vitamin D₃.

Cholecystokinin (CCK) – a hormone released from the gut after feeding that stimulates bile production. It may also act as a **satiety signal**. CCK is also found in the brain, where it acts as a nerve transmitter.

Chronic renal failure – progressive loss of renal function associated with uraemia. Low-protein diets give symptom-

atic relief of uraemia and may slow down progression of the disease.

Chylomicrons – triacylglycerol-rich lipoproteins found in plasma after feeding. The form in which absorbed fat is transported away from the gut.

Codex Alimentarius Commission – an international body set up by the UN in 1962 that seeks to set international food standards.

Coeliac disease – allergy to the wheat protein, gluten, causing damage to the absorptive surface of the gut. It is characterized by malabsorption, various gastrointestinal symptoms and malnutrition.

Coenzyme/cofactor – a non-protein substance needed for an enzyme to function; many are derivatives of vitamins.

Cohort study – a type of epidemiological study. Measurements are recorded on a sample or cohort of people and are subsequently related to mortality or morbidity.

Collagen – the most abundant protein in the human body. A key structural component of muscle, bone, heart muscle and connective tissue.

Colostrum – fluid rich in protein and antibodies that is secreted from the breast in very early lactation.

Committee on Toxicology (COT) – the UK committee that advises upon the safety of food additives.

Confounding variable – an epidemiological term: an association between a dietary variable and a disease may be an artefact caused because both test variables are influenced by a third 'confounding' variable, e.g. a relationship between alcohol intake and lung cancer might be because smoking causes lung cancer and smokers drink more alcohol than non-smokers.

Creatine phosphate – short-term energy store in muscles; the principal source for energy used in explosive activities such as short sprints.

Cretinism – a condition in young children caused by undersecretion of thyroid hormones, e.g. in iodine deficiency. Physical and mental development can be severely impaired.

Cultural relativism – the opposite of ethnocentrism – trying to regard all different cultural practices as normal.

Cultural superfood – a food that has acquired a cultural significance that goes beyond its purely nutritional or dietary importance, e.g. rice in Japan.

Cystic fibrosis – an inherited disease that results in progressive fibrosis and loss of function of the lungs and pancreas. Failure of pancreatic secretion reduces the digestion and absorption of fat and protein and leads poor fat-soluble vitamin absorption.

Cytokines – proteins produced as part of the immune response. They may be responsible for some of the anorexia and metabolic derangements seen in **cancer cachexia**.

db/db mouse – a genetically obese mouse that has a mutation in the leptin receptor gene and so fails to respond to leptin.

Death rate – crude death rate is the number of deaths per year per 1000 people in the population. **Age-specific death**

rate is the number of deaths of people within a specified age range per year per 1000 people in that age range. It may also be for a specified cause.

Decayed, missing and filled teeth (DMF) – a measure of dental health, usually of children.

Deficiency disease – a disease caused by lack of a nutrient.

Delaney clause – passed by the US Congress: 'no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal'. Nowadays, it is regarded as unreasonably absolute in its prohibitions.

Delayed-type cutaneous hypersensitivity (DTH) – a functional test of cell-mediated immune function. Antigen is injected into the skin and a reaction occurs after several hours delay, e.g. the **Mantoux reaction** to a cutaneous tuberculin injection.

Demi-span – the distance from the sternal notch to the web between the middle and ring fingers when the arm is stretched out horizontally. It is used as an alternative to height in the elderly and disabled.

Demiquet index – the body weight (kg) divided by the demi-span (m) squared, used as an alternative to body mass index (BMI) in elderly people.

Dermatitis herpetiformis – a chronic skin disorder caused by hypersensitivity to **gluten**.

Diabetes mellitus – a disease caused by lack of the pancreatic hormone insulin. It is characterized by high blood glucose, glucose loss in urine and, in severe cases, ketosis. **Severe type 1 diabetes** usually presents in childhood and requires insulin replacement; the milder **type 2 diabetes** usually presents after middle age, does not usually require insulin therapy and is triggered by a decline in insulin sensitivity.

Diabetic nephropathy – progressive loss of renal function associated with diabetes.

Diabetic retinopathy – progressive degeneration of the retina that may cause blindness in diabetics.

Diet-heart hypothesis – the suggestion that high dietary saturated fat increases low density lipoprotein (LDL) cholesterol, leading to increased atherosclerosis and ultimately increased risk of coronary heart disease.

Dietary fibre – plant material in the diet that resists digestion by human gut enzymes – 'roughage'. It is almost synonymous with **non-starch polysaccharide**, but includes resistant starch and lignin.

Dietary Reference Values (DRVs) – a general term to cover all UK dietary standards. It includes the RNI (listed separately), the **Estimated Average Requirement (EAR)** and the **Lower Reference Nutrient Intake (LRNI)** the estimated requirement of people with a particularly low need for the nutrient. Where data are limited, a **safe intake** is given.

Discretionary salt – salt that is added during home cooking of food or at the table.

Double-blind trial – a clinical trial using real and placebo treatments where neither the subject nor operator knows which is which until the trial is completed.

Doubly labelled water method – a method of estimating the long term energy expenditure of free living subjects. Subjects are given water labelled with both the heavy isotopes of oxygen and hydrogen. Labelled oxygen is lost as both CO_2 and water whereas hydrogen is lost only as water. The difference between the rate of loss of oxygen and hydrogen can thus be used to estimate CO_2 output and therefore energy expenditure.

Dual-centre hypothesis – the notion that feeding is regulated by two centres within the hypothalamus: a spontaneously active **feeding centre** in the lateral hypothalamus that is periodically inhibited by a **satiety centre** in the ventromedial hypothalamus.

Duplicate sample analysis – estimating nutrient intake by chemical analysis of duplicate samples of the food that has been eaten.

E numbers – the European system of designating food additives by number.

Eating disorders – a general term that encompasses **anorexia** and **bulimia nervosa** and the disorders of those people who are similarly affected but do not meet the strict diagnostic criteria for these conditions.

Eicosanoids – a group of locally acting, regulatory molecules synthesized from essential fatty acids, e.g. **prostaglandins, thromboxanes and prostacyclins**.

Eicosapentaenoic acid, 20:5 ω_3 (EPA) – fatty acid of the ω -3 or n-3 series. Together with **docosahexaenoic, 22:6 ω_3 (DHA)**, the 'active ingredients' of fish oils.

Elemental diet – a diet composed of purified nutrients, e.g. in the diagnosis of food allergy.

Endergonic reaction – a chemical reaction that absorbs energy.

Energy balance – see Balance.

Energy density – the amount of energy per unit weight of food, e.g. kcal/g

Energy equivalent of oxygen – the energy expended when 1 L of oxygen is consumed by a subject. It varies according to the mix of substrates being metabolized (around 4.8 kcal/L [20 kJ/L]).

Enteral feeding – feeding via the gut, e.g. using a nasogastric or a surgically placed tube.

Enteroto-hepatic circulation – re-absorption and return to the liver of substances secreted in the bile, e.g. bile acids.

Enzymes – proteins that speed up cellular reactions. All cellular reactions are 'catalysed' by specific enzymes.

Erythrocyte glutathione reductase activation coefficient (EGRAC) – a biochemical indicator of riboflavin status derived from one of the erythrocyte enzyme activation tests used to assess vitamin status. The extent to which the enzyme is activated by addition of an excess of vitamin-derived coenzyme is inversely related to the vitamin status of the donor.

Essential amino acids – those amino acids that are essential in the diet because they cannot be made by transamination.

Essential fatty acids – polyunsaturated fatty acids that must be provided in the diet. They serve structural functions and are the precursors of the **eicosanoids**.

Estimated Average Requirement (EAR) – see Dietary Reference Values.

Ethnocentrism – the tendency to regard one's own cultural practices as the norm and those of others as inferior or wrong.

Exergonic reaction – a chemical reaction that releases energy.

Factorial calculation – an apparently logical prediction of, say, nutrient needs by taking into account a series of factors, e.g. predicting the additional calcium needs of pregnant women from the rate of calcium accumulation during pregnancy and the assumed efficiency of calcium absorption in the gut.

Familial hypercholesterolaemia – an inherited condition characterized by very high serum LDL cholesterol (type IIa). The defect is in the **LDL receptor** and is associated with increased risk of heart disease.

Fat-cell theory – the proposal that overfeeding in childhood permanently increases susceptibility to obesity by increasing the number of fat cells.

Fat-soluble vitamins – the lipid-soluble vitamins, i.e. A, D, E and K.

Fatty acids – the major components of fats. They are made up of a hydrocarbon chain and a terminal **carboxyl** or acid group (COOH); If there are no double bonds in the hydrocarbon chain – **saturated**; one double bond – **monounsaturated**; more than one – **polyunsaturated**. In **ω -3/n-3 polyunsaturated fatty acids**, the first double bond is between carbons 3 and 4 (prominent in fish oils). In **ω -6/n-6 polyunsaturated fatty acids**, the first double bond is between carbons 6 and 7 (the predominant polyunsaturates in most vegetable oils).

Favism – a haemolytic anaemia that may result from eating broad beans. Around 30% of Mediterranean people and 10% of black Americans may be susceptible.

Fenfluramine and dextrofenfluramine – appetite-suppressant drugs that mimic the actions of 5HT in the brain. They were withdrawn from sale in 1996.

Ferritin – an iron–protein complex, the principal storage form of iron in the body.

Fetal alcohol syndrome (FAS) – a syndrome of children born to mothers who consume excess alcohol during pregnancy.

First-class proteins (complete proteins) – proteins with good amounts of all the essential amino acids, e.g. most animal and legume proteins.

Fitness – conditioning brought about by exercise. It is often defined as aerobic capacity, but could include all physical and mental effects of training.

Flavin adenine dinucleotide (FAD) and **flavin mononucleotide (FMN)** – prosthetic groups derived from riboflavin; important as hydrogen acceptors/ donators in several cellular oxidation/reduction reactions.

Food additives – chemicals deliberately added to food, e.g. preservatives, colours, flavourings etc.

Food and Drug Administration (FDA) – the US department responsible for food and drug regulation.

Food Advisory Committee (FAC) – the UK committee that advises upon the necessity for food additives.

Food allergy – intolerance to a food that involves an abnormal immunological reaction.

Food balance sheets – a population method of estimating food intake. The food available for human consumption is estimated and then expressed on a *per capita* basis.

Foodborne disease – a disease that can be transmitted via food, including both classical food poisoning and diseases such as typhoid and cholera, for which there is no need for the organism to grow in the food.

Food groups – dividing foods up into groups or categories according to their nutrient contents. Selecting from each group should ensure nutritional adequacy, e.g. the **four food group plan**: the meat group; the fruit and vegetable group; the milk group; and the cereal group.

Food guide plate – a visual guide to food selection used in the UK. A development of the four food group plan, which indicates the relative proportions of foods that should come from the four groups plus fats and sweets.

Food guide pyramid – a visual guide to food selection. A development of the four food group plan but it additionally indicates the ideal proportions from the different groups to meet current dietary guidelines.

Food intolerance – ‘a reproducible, unpleasant (i.e. adverse) reaction to a specific food or ingredient which is not psychologically based’.

Food poisoning – classical food poisoning is an illness, usually acute gastro-enteritis, caused by ingestion of food in which there has been active growth of bacteria. In current usage, it includes other foodborne illnesses caused by bacteria and viruses that have not actively grown in food.

Food Standards Agency – a new and independent body set up to oversee and advise on food safety issues in the UK.

Fractional catabolic rate – the proportion of the total body pool of a nutrient that is catabolized each day, e.g. around 3% of body vitamin C is broken down each day.

Free radicals – highly reactive chemical species produced as a by-product of oxidative processes in cells. Free-radical damage has been implicated in the aetiology of many diseases, including cancer and atherosclerosis.

Functional foods – foods that have components or ingredients incorporated into them in order to give a specific medical or physiological benefit in addition to their purely nutritional value.

Galactosaemia – an inherited condition in which galactose cannot be metabolized and accumulates in the blood. Sufferers must avoid lactose, milk sugar.

Gastric stapling – a surgical treatment for obesity. The capacity of the stomach is reduced by surgically stapling off a large part of it.

Gatekeeper – someone who effectively controls the food choices of others. Wives and mothers have traditionally been family gatekeepers, but it also describes catering managers in institutions.

Generally recognized as safe (GRAS) – a list of permitted food additives whose use became widespread before current regulatory procedures were established.

Gingivitis – an inflammation of the gums that is caused by plaque bacteria and is responsible for most tooth loss in older people.

Gluconeogenesis – the synthesis of glucose, e.g. from amino acids.

Glucostat theory – the notion that sensors (e.g. in the ventromedial hypothalamus) modulate the feeding drive in response to blood glucose concentration or rate of glucose utilization.

Glutathione peroxidase – a selenium-containing enzyme involved in free-radical disposal.

Glutathione reductase – a flavoprotein enzyme involved in free-radical disposal.

Gluten – a protein in wheat.

Glycaemic index – the rise in blood glucose induced by the test food expressed as a percentage of that induced by the same amount of pure glucose (the areas under the blood glucose-time curves are compared). Carbohydrate foods that are rapidly digested and absorbed have a high glycaemic index and cause steep rises in blood glucose and insulin release.

Glycogen – a form of starch stored in the liver and muscles of animals.

Glycosylated – (as in glycosylated haemoglobin) having reacted abnormally with glucose. Proteins become glycosylated due to hyperglycaemia in diabetics. Glycation may cause some of the long-term problems of diabetes, e.g. diabetic retinopathy and nephropathy.

Goitre – swelling of the thyroid gland, e.g. due to lack of dietary iodine.

Gold thioglucose – a chemical that produces permanent obesity in mice by damaging the ventromedial region of the hypothalamus.

Haemagglutinin – a substance that causes red blood cells to clump together. It is found in red kidney beans and causes vomiting and diarrhoea if undercooked beans are eaten.

Haemoglobin – an iron-containing pigment in blood that transports oxygen.

Hartnup disease – a rare inherited disorder with symptoms of pellagra. It is caused by failure to absorb tryptophan and responds to niacin therapy.

Hazard – ‘being likely to produce injury under the circumstances of exposure’, e.g. as in food (cf. **toxic**).

Heat stroke – a combination of hyperthermia and dehydration that may affect endurance athletes.

Hexose – a six-carbon sugar, e.g. glucose, fructose or galactose.

High-density lipoprotein (HDL) – a blood lipoprotein. It carries cholesterol from the tissues to the liver; anti-atherogenic.

High temperature–short time – a modern method of pasteurization that exposes food to a higher temperature than traditional pasteurization but for a much shorter time.

HMG CoA reductase – a rate-limiting enzyme in cholesterol biosynthesis that is inhibited by the statin group of cholesterol-lowering drugs.

Hormone replacement therapy (HRT) – oestrogen administered to postmenopausal women when the natural production of sex hormones ceases.

Hot and cold – a traditional food classification system, widely used in China, India and South America.

Hunger – the physiological drive or need to eat.

Hydrogenated vegetable oils – most margarine and solid vegetable-based shortening; vegetable oil that has been treated with hydrogen to solidify it.

Hydroxyapatite – the crystalline, calcium phosphate material of bones. In the presence of fluoride, **fluorapatite** is formed.

5-Hydroxytryptamine (5-HT) or serotonin – a nerve transmitter in the brain.

Hydroxyl radical – a free radical.

Hypernatraemic dehydration – literally, dehydration with high blood sodium concentration. It was associated with high-sodium infant formulae prior to 1976 in the UK.

Hypertension – high blood pressure.

Immunoglobulin (Ig) – antibodies. IgA protects mucosal surfaces, IgG is the main circulating antibody and IgE the fraction associated with allergic responses.

Incidence – the number of new cases of a disease occurring within a specified time period.

Infective dose – the number of micro-organisms that must be ingested to provoke illness.

Intestinal bypass – a surgical treatment in which a large proportion of the small intestine is bypassed. It is used in the treatment of obesity.

Irradiation – of food: exposing food to ionizing radiation in order to delay ripening, inhibit sprouting, kill insect pests or kill spoilage and pathogenic organisms.

Isomers – chemicals that have the same set of atoms but whose spatial arrangement is different.

Jaw wiring – a surgical procedure used to restrict feeding and treat obesity.

Joule – the Standard International unit of energy: 'the energy expended when a mass of 1 kg is moved through a distance of 1 m by a force of 1 newton'. **Kilojoule (kJ)** = 1000 joules; **megajoule (MJ)** = a million joules; 4.2 kJ = 1 kcal.

Keshan disease – a degeneration of heart muscle seen in parts of China and attributed to selenium deficiency.

Ketone bodies – substances produced from fatty acids and used as alternative brain substrates during starvation – β -hydroxybutyrate, acetoacetic acid and acetone.

Ketosis – Toxic accumulation of ketone bodies, e.g. in diabetes.

Keys' equation – one of a number of equations that seek to predict the change in plasma cholesterol in response to specified changes in the nature and amount of dietary fat. **Krebs' cycle** – the core pathway of oxidative metabolism in mitochondria. Activated acetate feeds in and carbon dioxide, ATP and reduced coenzymes (e.g. NADH₂) are the products.

Kwashiorkor – one manifestation of protein energy malnutrition in children. There is oedema and a swollen liver, and some subcutaneous fat may still remain. It was traditionally attributed to primary protein deficiency, but is now widely considered to be a manifestation of simple starvation.

Lactose – milk sugar, disaccharide of glucose and galactose.

Lactose intolerance – an inability to digest lactose, found in many adult populations; milk consumption causes diarrhoea and gas production.

Lathyrism – a severe disease of the spinal cord caused by eating very large amounts of chickling peas (*Lathyrus sativa*).

LDL receptor – a receptor that binds to LDL and removes it from the blood; 75% of LDL receptors are in the liver.

Leptin – a newly discovered protein hormone that is produced in adipose tissue. Its concentration in blood reflects the size of the body fat stores. It seems to have the characteristics of the 'satiety factor' that was postulated in the **lipostat theory** of body-weight regulation.

Limiting amino acid – the essential amino acid present in the lowest amount relative to requirement. It can limit the extent to which other amino acids can be used for protein synthesis.

Lipase – a fat-digesting enzyme, e.g. in pancreatic juice.

Lipid peroxyl radical – a free radical produced by oxidative damage to polyunsaturated fatty acid residues.

Lipoprotein lipase – an enzyme in capillaries that hydrolyses fat in lipoproteins prior to absorption into the cell.

Lipoproteins – lipid–protein complexes. Fats are attached to **apoproteins** and transported in blood as lipoproteins.

Lipostat theory – the notion that body fat level is regulated by a satiety factor (leptin?) released from adipose tissue in proportion to the amount of fat stored within it.

Liposuction – literally, sucking out adipose tissue. A surgical procedure used in the cosmetic treatment of obesity.

Listeriosis – foodborne disease caused by *Listeria monocytogenes*. It can cause miscarriage, stillbirth or neonatal death and is associated with chilled foods such as pâté, cheese made with raw milk and chilled ready meals.

Lower Reference Nutrient Intake (LRNI) – see Dietary Reference Values.

Low birth weight (LBW) – of babies: birth weight less than 2.5 kg. It is associated with increased risk of perinatal mortality and morbidity.

Low-density lipoprotein (LDL) – a blood lipoprotein that is rich in cholesterol and the form in which cholesterol is transported to the tissues; atherogenic.

Luxoskonsumption – an adaptive increase in metabolic rate in response to overfeeding – **adaptive thermogenesis**.
MAFF – Ministry of Agriculture, Fisheries and Food (UK).

Marasmus – one manifestation of protein energy malnutrition in children. It is attributed to simple energy deficit.
Metabolic rate – the rate of body energy expenditure. It may be measured directly as the heat output in a **whole-body calorimeter** or indirectly from the rate of oxygen consumption or carbon dioxide production.

Metabolic response to injury – changes in metabolism that follow mechanical or disease ‘injury’. A period of reduced metabolism, shock or ebb, is followed by a longer period of increased metabolism and protein breakdown, flow.

Metabolizable energy – the ‘available energy’ from a food after allowing for faecal loss of unabsorbed material and urinary loss of nitrogenous compounds from incomplete oxidation of protein.

Micelles – minute aggregates of fat-digestion products with bile acids that effectively solubilize them and facilitate their absorption.

Mid-arm circumference (MAC) – the circumference of the mid part of the upper arm. It is an anthropometric indicator of nutritional status, used in Third World children and bedridden hospital patients who cannot be weighed.

Mid-arm muscle circumference (MAMC) – a measure of lean body mass = MAC [$\pi \times$ triceps skinfold].

Mindex – body weight (kg) divided by demi-span. It is used as a measure of body fatness in the elderly.

Mitochondria – subcellular particles; the site of oxidative metabolism.

Mutual supplementation of proteins – compensation for deficit of an essential amino acid in one dietary protein by excess in another.

Mycotoxins – toxic substances produced by fungi or moulds, e.g. **aflatoxins** (mouldy nuts), **patulin** (apples or apple juice) and **ergot** (mouldy grain). Some are known carcinogens.

Myelin – the fatty sheath around many nerve fibres.

Myoglobin – a pigment similar to haemoglobin that is found in muscle.

N-terminal – the end of a protein with a free amino group.

NADP – a phosphorylated derivative of NAD that usually acts as a hydrogen donor, e.g. in fat synthesis.

National Food Survey (NFS) – an example of a **household survey**. It is an annual survey of the household food purchases of a representative sample of UK households. It provides information about regional and class differences in food purchasing and national dietary trends.

National Research Council (NRC) – a group that has produced several important reports relating to food and nutri-

tion in the US; under the auspices of the National Academy of Sciences.

Net protein utilization (NPU) – a measure of protein quality. The percentage of nitrogen retained when a protein (or diet) is fed at amounts that do not satisfy an animal’s total protein needs. Losses are corrected for those on a protein-free diet.

Neural tube defects (NTD) – a group of birth defects in which the brain, spinal cord, skull and/or vertebral column fail to develop properly in early embryonic life, e.g. spina bifida and anencephaly.

Niacin equivalents – a way of expressing the niacin content of food. It includes both pre-formed niacin in the diet and niacin that can be produced endogenously from dietary tryptophan.

Niacytin – the bound form of niacin in many cereals. In this form it is unavailable to humans.

Nicotine adenine dinucleotide (NAD) – a hydrogen acceptor molecule produced from niacin. A coenzyme in numerous oxidation/reduction reactions. Re-oxidation of reduced NAD in oxidative phosphorylation yields most of the ATP from the oxidation of foodstuffs.

Night blindness – impaired vision at low light intensity; an early sign of vitamin A deficiency.

Nitrogen balance – see **Balance**.

No-effect level – the intake of an additive that produces no adverse effects in animal tests.

Non-milk extrinsic sugars – added dietary sugars, i.e. all the sugars other than those naturally in fruits, vegetables and milk.

Non-starch polysaccharide (NSP) – dietary polysaccharides other than starch. They are indigestible by human gut enzymes and thus make up the bulk of **dietary fibre**.

Normal distribution – even distribution of individual values for a variable around the mean value. Most individual values are close to the mean and the further away from the mean, the lower the frequency of individuals becomes. It results in a bell-shaped curve when a frequency distribution is plotted.

Nutrient density – the amount of a nutrient per kcal or per kJ. It indicates the value of a food or diet as a source of the nutrient.

Nutritional supplements – concentrated nutrient sources, usually given in liquid form to boost the intakes of patients.

ob/ob mouse – a genetically obese mouse that has a mutation in its leptin gene and so fails to produce functional leptin.

Obligatory nitrogen loss – the loss of nitrogen that still occurs after a few days’ adaptation to a protein-free diet.

Odds ratio – an indirect measure of relative risk in case-control studies: ‘the odds of case exposure divided by the odds of control exposure’.

Oedema (US edema) – swelling of tissues due to excess water content. It is a symptom of several diseases, local injury and severe malnutrition.

Oil of evening primrose – a source of gamma-linolenic acid (18:3_{ω6}). It is widely taken as a dietary supplement.

Olestra – an artificial fat that is not digested. If added to food, it gives some of the palatability effects of fat without adding calories.

Orlistat – a drug that blocks fat digestion and so reduces its absorption from the gut.

Osteomalacia – a disease of adults caused by lack of vitamin D.

Osteoporosis – a progressive loss of bone matrix and mineral making bones liable to fracture, especially in post-menopausal women.

Oxidative phosphorylation – the re-oxidation of reduced coenzymes in the mitochondria with oxygen. This process produces the bulk of the ATP in aerobic metabolism.

Parentral feeding – intravenous feeding.

Pasteurization – mild heat treatment of foods that kills likely pathogens without impairing their palatability or nutrient content, e.g. pasteurized milk.

Peak bone mass (PBM) – the maximum bone mass reached by adults in their thirties. After this, bone mass starts to decline.

Pellagra – a deficiency disease due to a lack of niacin.

Pentose phosphate pathway – a metabolic pathway that produces ribose for nucleic acid synthesis and reduced NADP for processes such as fatty acid synthesis.

Peptidases – enzymes that digest proteins. Endopeptidases hydrolyse peptide bonds within the protein and break it up into small peptides. Exopeptidases hydrolyse the N-terminal (aminopeptidases) or C-terminal (carboxypeptidases) amino acid.

Peptide bond – a bond that links amino acids together in proteins. An amino group of one amino acid is linked to the carboxyl group of another.

Pernicious anaemia – an autoimmune disease; failure to produce **intrinsic factor** in the stomach necessary for vitamin B₁₂ absorption. Symptoms are severe **megaloblastic anaemia** and **combined subacute degeneration of the spinal cord** leading to progressive paralysis.

Phenylketonuria (PKU) – an inherited disease in which there is inability to metabolize the amino acid phenylalanine to tyrosine. Intake of phenylalanine must be strictly controlled to avoid severe mental retardation.

Phospholipids – lipids containing a phosphate moiety; important components of membranes.

Phylloquinone – the major dietary form of vitamin K.

Physical activity level (PAL) – when estimating energy expenditure, the number by which the **basal metabolic rate** is multiplied to allow for energy used in the day's activity. It ranges from around 1.3 (e.g. housebound, elderly person) to well over 2 (e.g. a serious athlete in training).

Phytochemicals – chemicals found in fruits and vegetables. Some may reduce the risk of chronic disease.

Placebo – a 'dummy' treatment that enables the psychological effects of treatment to be distinguished from the physiological effects.

Plaque – a sticky mixture of food residue, bacteria and bacterial polysaccharides that adheres to teeth.

Prebiotics – non-digestible food ingredients that selectively promote the growth of some bacteria within the gut and have effects similar to those of **probiotics**, e.g. the plant polysaccharide inulin.

Pre-eclampsia – hypertension of pregnancy.

Prevalence – the number of cases of a disease at any point in time; it depends upon both the incidence and the duration of the illness.

Prions – the transmissible agents in spongiform encephalopathies; abnormally shaped brain proteins that, on contact with the normal brain proteins, cause them also to change configuration. Prions are very resistant to proteases, heat and irradiation and their accumulation in brain causes tissue destruction.

Probiotics – fermented dairy products that contain living cultures of *Lactobacilli* that may improve the microbial balance in the intestine. They may reduce gastrointestinal and perhaps vaginal infections.

Proline hydroxylase – a vitamin C-dependent enzyme vital for collagen formation.

Prostaglandins – see **Eicosanoids**.

Prosthetic group – a non-protein moiety that is tightly bound to an enzyme and is necessary for enzyme function.

Protein-energy malnutrition (PEM) – the general term used to cover undernutrition due to lack of energy or protein, or both. It encompasses **kwashiorkor** and **marasmus**.

Protein gap – the notional gap between world protein requirements and supplies that disappeared as estimates of human protein requirements were revised downwards.

Protein turnover – the total amount of protein broken down and resynthesized in a day.

Prothrombin time – a functional test of the prothrombin level in blood that is used as a measure of vitamin K status.

P:S ratio – the ratio of polyunsaturated to saturated fatty acids in a fat or diet.

PWC₁₇₀ – the work load at which the pulse rate reaches 170 beats per minute; a measure of aerobic fitness.

PWC_{max} – the work load at maximum heart rate; a measure of aerobic fitness.

Recommended Dietary Allowance (RDA) (USA) – the suggested average daily intake of a nutrient for healthy people in a population. It is equivalent to the UK RNI, and represents the estimated requirement of those people with a particularly high need for the nutrient. The energy RDA is the best estimate of average requirement.

Reference Nutrient Intake (RNI) – the estimated need of those people with a particularly high need for the nutrient. People consuming the RNI or above are practically assured of adequacy.

Relative risk – an epidemiological term: the ratio of the number of cases per 1000 in an exposed population, to those in an unexposed population, e.g. the ratio of deaths per 1000 in smokers versus non-smokers.

Reliability – the technical accuracy or repeatability of a measurement (cf. **Validity**).

Resistant starch – starch that resists digestion by α -amylase in the gut and so passes into the large intestine undigested. It behaves like a component of **non-starch polysaccharide**.

Respiratory quotient (RQ) – ratio of carbon dioxide produced to oxygen consumed. It varies according to the substrates being oxidized and can indicate the mix of substrates being oxidized by a subject.

Retinol – vitamin A.

Retinol-binding protein – a protein in plasma that binds to and transports retinol.

Retinol equivalents – the way of expressing vitamin A content: 1 μg retinol or 6 μg carotene = 1 μg retinol equivalents.

Retinopathy of prematurity – blindness in newborn babies caused by exposure to high oxygen concentrations. It is believed to result from free-radical damage.

Rhodopsin – a light-sensitive pigment in the rods of the retina. It contains a derivative of vitamin A, **11 cis-retinal**. All visual pigments are based upon **11 cis-retinal**.

Rickets – a disease of children caused by lack of vitamin D.

Risk factors – factors such as high blood pressure, high blood cholesterol and low fruit and vegetable consumption that predict a higher risk of developing a particular disease.

Saccharide – a sugar. Note, monosaccharide (comprised of one sugar unit), disaccharide (two units), oligosaccharide (a few) and polysaccharide (many).

Satiety signals – physiological signals that induce satiation, e.g. high blood glucose and stomach fullness.

Saxitoxin – a plankton neurotoxin that may cause poisoning if mussels are eaten at times when this red plankton is abundant in the sea – ‘red tides’.

Scrombotoxin – a toxin produced in spoiled fish by the action of bacteria upon fish protein.

Scurvy – a deficiency disease due to lack of vitamin C.

Second-class (incomplete) protein – a dietary protein that is relatively deficient in one or more essential amino acids.

Sensory-specific satiety – the phenomenon whereby, during eating, one’s appetite for a previously consumed food diminishes rapidly but one’s appetite for other foods is much less affected. As a consequence, increased variety might lead to overeating.

‘Set point’ theory – the notion that body weight-control mechanisms operate to maintain a fixed level of body fatness; analogous to a thermostat set to maintain a fixed temperature.

Sitostanol – a plant sterol that is effective in lowering plasma cholesterol by inhibiting cholesterol absorption (and reabsorption) in the gut. Other plant sterols have a similar but lesser effect.

Skin sensitivity tests – used in the identification of (food) allergens: suspected allergens are injected into the skin and the extent of the skin reaction is used to indicate sensitivity.

Solanine – a poisonous alkaloid found in small amounts in potatoes.

Soluble fibre/non-starch polysaccharide (NSP) – that part of the dietary fibre that forms gums or gels when mixed with water; cf. **Insoluble fibre/NSP** – that part which is insoluble in water, e.g. cellulose.

Specific activity – the amount of activity per unit weight, e.g. the amount of radioactivity per milligram of labelled substance or the amount of enzyme activity per milligram of protein.

Spoilage – deterioration in the appearance or palatability of food caused by the action of spoilage bacteria.

Standard deviation (SD) – a statistical term that describes the distribution of individual values around the mean in a normal distribution. Approximately 95% of individual values lie within two SD either side of the mean.

Standard mortality ratio (SMR) – a way of comparing death rates in populations of differing age structures: ‘the ratio of actual deaths in a test population to those predicted assuming it had the same age-specific death rates as a reference population’.

Subjective norm – the behaviour that is perceived to be the normal option.

Sugar–fat seesaw – the tendency for fat and sugar intakes of affluent individuals to be inversely related, i.e. low-fat diets tend to be high-sugar diets, and vice versa. Part of a more general **carbohydrate–fat seesaw** that is almost inevitable as these are the two principal energy sources in most diets.

Superoxide dismutase – a zinc-containing enzyme that disposes of superoxide free radicals.

Superoxide radical – a free radical.

Taurine – an amino acid. It is not used in protein synthesis and can be made from cysteine. It is present in large amounts in breast-milk and may be essential for babies.

Teratogenic – causing birth defects.

Thermic effect of feeding/postprandial thermogenesis – a short-term increase in metabolic rate (thermogenesis) that follows feeding. It is due to energy expended in digestion, absorption etc.

Thermogenesis – literally, heat generation. Metabolic heat generation may be increased by exercise (**exercise-induced thermogenesis**), eating (**diet-induced thermogenesis**), drugs (**drug-induced thermogenesis**) and cold stress (either shivering or **non-shivering thermogenesis**).

Thiamin pyrophosphate – an important coenzyme produced from thiamin (vitamin B_1).

Thyrotrop(h)in – a pituitary hormone that stimulates the release of **thyroxine** and **triiodothyronine** from the thyroid gland.

α -tocopherol – a major dietary form of vitamin E.

Total parenteral nutrition (TPN) – feeding that is wholly by infusion into a large vein.

Toxic – ‘being inherently capable of producing injury when tested by itself’, cf. **Hazard**.

Trans fatty acids – isomeric forms of unsaturated fatty acids in which the hydrogen atoms around a double bond are on opposite sides, cf. most natural fatty acids where they are on the same side (*cis* isomer). The major sources are hydrogenated vegetable oils.

Transamination – the transfer of an amino group (NH_2) from one amino acid to produce another.

Transferrin – a protein in plasma that transports iron. The level of **transferrin saturation** with iron is a measure of iron status.

Transit time – the time taken for ingested material to pass through the gut. Fibre/NSP decreases the transit time.

Travellers' diarrhoea – gastro-enteritis experienced by travellers. It is often caused by *E. coli* bacteria from contaminated water to which the traveller has low immunity.

Triacylglycerol/triglyceride (TAG) – the principal form of fat in the diet and adipose tissue. It is composed of glycerol and three fatty acids.

Tumour necrosis factor – a **cytokine**.

Twenty-four-hour recall – a retrospective method of estimating food intake. Subjects recall and quantify everything eaten and drunk in the previous 24 hours. Note also **diet histories**, in which an interviewer tries to obtain a detailed assessment of the subject's habitual intake, and self-administered **food frequency questionnaires**.

Tyramine – a substance found in some foods (e.g. cheese) that causes a dangerous rise in blood pressure in people taking certain antidepressant drugs.

Ultra high temperature (UHT) – sterilization by exposing food (e.g. milk) to very high temperatures for very short times; it induces much less chemical change than traditional methods.

Uraemia – high blood urea concentration, seen in renal failure. It leads to symptoms that include nausea, anorexia, headache and drowsiness.

USDA – United States Department of Agriculture.

'Use by' – the date marked on foods in the UK to indicate when they are likely to become microbiologically unsafe.

Validity – the likelihood that a measurement is a true measure of what one is intending to measure, cf. **Reliability**, e.g. a biochemical index of nutritional status could be very precise (high reliability) but not truly indicate nutritional status (low validity).

Vegetarian – one who eats only food of vegetable origin. A **vegan** avoids all animal foods. The prefixes **lacto-** (milk), **ovo-** (eggs) and **pesco-** (fish) are used if some animal foods are eaten.

Very low-density lipoprotein (VLDL) – a triacylglycerol-rich blood lipoprotein. It transports endogenously produced fat to adipose tissue.

Very low-energy diets (VLEDs) – preparations designed to contain very few calories but adequate amounts of other nutrients; used in obesity treatment.

Verocytotoxin-producing *E. coli* (VTEC) – this emerged as a cause of food poisoning in the early 1980s and accounts for more than 1000 cases per year. Some of those affected develop renal complications and some die from acute renal failure.

vO₂ max – the oxygen uptake when exercising maximally; the capacity of the cardiopulmonary system to deliver oxygen tissues; one definition of fitness.

Waist-to-hip ratio – the ratio of waist circumference to that at the hip; an indicator of the level of health risk of obesity. A high value is considered undesirable.

(The) **'Wall'** – the state of exhaustion in endurance athletes when muscle glycogen reserves are depleted.

Water activity – the availability of water to bacteria in food. Drying and high solute concentration reduce water activity, i.e. reduce its availability to bacteria.

Water-soluble vitamins – vitamins that are soluble in water, i.e. the B and C vitamins.

Weaning – the process of transferring infants from breast-milk or infant formula onto a solid diet.

Weighed inventory – a prospective method of measuring food intake. Subjects weigh and record everything consumed. Household measures rather than weighing may be used.

Weight cycling or yo-yo dieting – cycles of weight loss and regain.

Wernicke–Korsakoff syndrome – a neurological manifestation of thiamin deficiency; often associated with alcoholism.

Whey – a milk protein; the dominant protein in human milk.

Whey-dominant formula – an infant formula with whey as the major protein.

World Health Organization (WHO) – the 'health department' of the United Nations.

Xerophthalmia – literally, dry eyes. It covers all the ocular manifestations of vitamin A deficiency, which range from drying and thickening of the conjunctiva through to ulceration/rupture of the cornea and permanent blindness.

References

- Ajzen, I and Fishbein, M 1980 *Understanding attitudes and predicting social behaviour*. Engelwood Cliffs, NJ: Prentice-Hall.
- Akintonwa, A and Tunwashe, OL 1992 Fatal cyanide poisoning from cassava-based meal. *Human and Experimental Toxicology* **11**, 47–9.
- Allied Dunbar National Fitness Survey 1992 *A report on activity patterns and fitness levels*. London: Sports Council.
- Altschul, AM 1965 *Proteins their chemistry and politics*. New York: Basic Books Inc.
- American Cancer Society 1997 New American Cancer Society guidelines on diet, nutrition and cancer prevention. *Journal of the National Cancer Institute* **89**, 198.
- Andres, R, Elahi, D, Tobin, JD, Muller, DC and Brait, L 1985 Impact of age on weight goals. *Annals of Internal Medicine* **103**, 1030–3.
- Anon 1980 Preventing iron deficiency. *Lancet* **i**, 1117–18.
- Anon 1984 Marine oils and platelet function in man. *Nutrition Reviews* **42**, 189–91.
- Armstrong, N, Balding, J, Gentle, P and Kirby, B 1990 Physical activity among 11 to 16 year old British children. *British Medical Journal* **301**, 203–5.
- Artaud-Wild, SM, Connor, SL, Sexton, G and Connor, W 1993 Differences in coronary mortality can be explained by differences in cholesterol and saturated fat intakes in 40 countries but not in France and Finland. *Circulation* **88**, 2771–9.
- Ashwell, M 1994 Obesity in men and women. *International Journal of Obesity* **18** (Supplement), S1–S7.
- Asp, N-G, van Amelsvoort, JMM and Hautvast, JGAJ 1996 Nutritional implications of resistant starch. *Nutrition Research Reviews* **9**, 1–31.
- Bado, A, Levasseur, S, Attoub, S, et al. 1998 The stomach is a source of leptin. *Nature* **394**, 790–3.
- Baker, EM, Hodges, RE, Hood, J, Saubelich, HE, March, SC and Canham, JE 1971 Metabolism of ^{14}C - and ^3H -labeled L-ascorbic acid in human scurvy. *The American Journal of Clinical Nutrition* **24**, 444–54.
- Barker, DJP, Cooper, C and Rose, G 1998 *Epidemiology in medical practice*, 5th edn. Edinburgh: Churchill Livingstone.
- Barker, M, Robinson, S, Osmond, C and Barker, DJ 1997 Birth weight and body fat distribution in adolescent girls. *Archives of Disease in Childhood* **77**, 381–3.
- Bastow, MD, Rawlings, J and Allison, SP 1983 Benefits of supplementary tube feeding after fractured neck of femur; a randomised controlled trial. *British Medical Journal* **287**, 1589–92.
- Bavly, S 1966 Changes in food habits in Israel. *The Journal of the American Dietetic Association* **48**, 488–95.
- Bebbington, AC 1988 The expectation of life without disability in England and Wales. *Social Science and Medicine* **27**, 321–6.
- Bennett, K and Morgan, K 1992 Activity and morale in later life: preliminary analyses from the Nottingham Longitudinal Study of Activity and Ageing. In *Nutrition and physical activity*. Ed. Norgan, NG. Cambridge: Cambridge University Press, pp. 129–42.
- Bennett, N, Dodd, T, Flately, J, Freeth, S and Bolling, K 1995 *Health Survey for England 1993*. London: HMSO.
- Becker, MH (ed.) 1984 *The health belief model and personal health behaviour*. Thorofare, NJ: Charles B. Slack.
- Bender, DA 1983 Effects of a dietary excess of leucine on tryptophan metabolism in the rat: a mechanism for the pellagraogenic action of leucine. *British Journal of Nutrition* **50**, 25–32.
- Bingham, SA 1990 Mechanisms and experimental and epidemiological evidence relating dietary fibre (non starch polysaccharides) and starch to protection against large bowel cancer. *Proceedings of the Nutrition Society* **49**, 153–71.
- Bingham, SA 1996 Epidemiology and mechanisms relating diet to risk of colorectal cancer. *Nutrition Research Reviews* **9**, 197–239.
- Binns, NM 1992 Sugar myths. In *Nutrition and the consumer*. Ed. Walker, AF and Rolls, BA. London: Elsevier Applied Science, pp. 161–81.

- Bistrian, BR, Blackburn, GL, Hallowell, E and Heddle, R 1974 Protein status of general surgical patients. *Journal of the American Medical Association* **230**, 858–60.
- BJN 1998 Functional Food Science in Europe. *British Journal of Nutrition* **80**, Supplement 1.
- Blair, SN, Kohl, HW, Paffenbarger, RS, Clark, DG, Cooper, KH and Gibbons, LW 1989 Physical fitness and all cause mortality. A prospective study of healthy men and women. *Journal of the American Medical Association* **262**, 2395–401.
- Blot, WJ, Li, J-Y, Tosteson, TD, et al. 1993 Nutrition intervention trials in Linxian, China: supplementation with specific vitamin, mineral combinations, cancer incidence and disease-specific mortality in the general population. *Journal of the National Cancer Institute* **85**, 1483–92.
- Blundell, JE and Burley, VJ 1987 Satiation, satiety and the action of dietary fibre on food intake. *International Journal of Obesity* **11** (Supplement 1), 9–25.
- Board, RG 1983 *A modern introduction to food microbiology*. Oxford: Blackwell Scientific Publications.
- Bolton-Smith, C and Woodward, M 1994 Dietary composition and fat and sugar ratios in relation to obesity. *International Journal of Obesity* **18**, 820–8.
- Bröbeck, JR 1974 Energy balance and food intake. In *Medical physiology*, 13th edn, Vol. 2. Ed. Mountcastle, VB. Saint Louis, MI: Mosby, pp. 1237–72.
- Brown, MS and Goldstein, JL 1984 How LDL receptors influence cholesterol and atherosclerosis. *Scientific American* **251** (5), 52–60.
- Bruckdorfer, R 1992 Sucrose revisited: is it atherogenic? *The Biochemist* June/July, 8–11.
- Bryan, FL 1978 Factors that contribute to outbreaks of foodborne disease. *Journal of Food Protection* **41**, 816–27.
- Burkitt, DP 1971 Epidemiology of cancer of the colon and rectum. *Cancer* **28**, 3–13.
- Burr, ML, Fehily, AM, Gilbert, JF, et al. 1991 Effects of changes in fat, fish and fibre intakes on death and myocardial infarction: Diet and Reinfarction Trial (DART). *Lancet* **ii**, 757–61.
- Buttriss, J 1999 *n-3 fatty acids and health*. British Nutrition Foundation briefing paper. London: BNF.
- Byers, T 1995 Body weight and mortality. *New England Journal of Medicine* **333**, 723–4.
- Campbell, AJ, Spears, GFS, Brown, JS, Busby, WJ and Borrie, MJ 1990 Anthropometric measurements as predictors of mortality in a community aged 70 years and over. *Age and Ageing* **19**, 131–5.
- Carlson, E, Kipps, M, Lockie, A and Thomson, J 1985 A comparative evaluation of vegan, vegetarian and omnivore diets. *Journal of Plant Foods* **6**, 89–100.
- CWT 1995 Caroline Walker Trust. *Eating well for older people*. London: The Caroline Walker Trust.
- Carpenter, KJ 1994 *Protein and energy: a study of changing ideas in nutrition*. Cambridge: Cambridge University Press.
- Chan, JM, Rimm, EB, Colditz, GA, Stamper, MJ and Willett, WC 1994 Obesity, fat distribution and weight gain as risk factors for clinical diabetes in men. *Diabetes Care* **17**, 961–9.
- Chandra, RK 1985 Nutrition–immunity–infection interactions in old age. In *Nutrition, immunity and illness in the elderly*. Ed. Chandra, RK. New York: Pergamon Press, pp. 87–96.
- Chandra, RK 1992 Effect of vitamin and trace-element supplementation on immune responses and infection in elderly subjects. *Lancet* **340**, 1124–6.
- Chandra, RK 1993 Nutrition and the immune system. *Proceedings of the Nutrition Society* **52**, 77–84.
- Chapuy, MC, Arlot, ME, Delmas, PD and Meunier, PJ 1994 Effects of calcium and cholecalciferol treatment for three years on hip fractures in elderly women. *British Medical Journal* **308**, 1081–2.
- Charlton, J and Quaife, K 1997 Trends in diet 1841–1993. In *The health of adult Britain 1841–1994*, Vol. 1. Ed. Charlton, J and Murphy, M. London: The Stationery Office, pp. 93–113.
- Chesher, 1990 Changes in the nutritional content of household food supplies during the 1980s. In *Household food consumption and expenditure*. Annual report of the National Food Survey Committee. London: HMSO.
- Church, M 1979 Dietary factors in malnutrition: quality and quantity of diet in relation to child development. *Proceedings of the Nutrition Society* **38**, 41–9.
- Clement, K, Vaisse, C, Lahlous, N, et al. 1998 A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398–401.
- Coleman, DL 1978 Obese and Diabetes: two mutant genes causing diabetes-obesity syndromes in mice. *Diabetologia* **14**, 141–8.
- COMA 1984 Committee on Medical Aspects of Food Policy. *Diet and cardiovascular disease*. Report on health and social subjects No. 28. London: HMSO.
- COMA 1988 Committee on Medical Aspects of Food Policy. *Present day practice in infant feeding: third report*. Report on health and social subjects No. 32. London: HMSO.
- COMA 1989a Committee on Medical Aspects of Food Policy. *Report of the panel on dietary sugars and human disease*. Report on health and social subjects No. 37. London: HMSO.
- COMA 1989b Committee on Medical Aspects of Food Policy. *The diets of British schoolchildren*. Report on health and social subjects No. 36. London: HMSO.
- COMA 1991 Committee on Medical Aspects of Food Policy. *Dietary reference values for food energy and nutrients*

- for the United Kingdom. Report on health and social subjects No. 41. London: HMSO.
- COMA 1992 Committee on Medical Aspects of Food Policy. *The nutrition of elderly people*. Report on health and social subjects No. 43. London: HMSO.
- COMA 1994a Committee on Medical Aspects of Food Policy. *Nutritional aspects of cardiovascular disease*. Report on health and social subjects No. 46. London: HMSO.
- COMA 1994b Committee on Medical Aspects of Food Policy. *Weaning and the weaning diet*. Report on health and social subjects No. 45. London: HMSO.
- COMA 1998 Committee on Medical Aspects of Food and Nutrition Policy. *Nutritional aspects of the development of cancer*. Report on health and social subjects No. 48. London: The Stationery Office.
- COMA (2000) Committee on Medical Aspects of Food and Nutrition Policy. *Folic acid and the prevention of disease*. London: The Stationery Office.
- Conning, D 1991 *Antioxidant nutrients in health and disease*. Briefing paper 25. London: The British Nutrition Foundation.
- Cook GC (1998) Diarrhoeal disease: a worldwide problem. *Journal of the Royal Society of Medicine* **91**, 192–4.
- Cooper, C and Eastell, R 1993 Bone gain and loss in premenopausal women. *British Medical Journal* **306**, 1357–8.
- Cramer, DM, Harlow, BL, Willett, WC, et al. 1989 Galactose consumption and metabolism in relation to risk of ovarian cancer. *Lancet* **ii**, 66–71.
- Crimmins, EM, Saito, Y and Ingegneri, D 1989 Changes in life expectancy and disability-free life expectancy in the United States. *Population and Development Review* **15**, 235–67.
- Crouse, JR 1989 Gender, lipoproteins, diet, and cardiovascular risk. *Lancet* **i**, 318–20.
- Cummings, SR, Kelsey, JL, Nevitt, MC and O'Dowd, KJ 1985 Epidemiology of osteoporosis and osteoporotic fractures. *Epidemiological Reviews* **7**, 178–208.
- Cuthbertson, DP 1980 Historical approach. (Introduction to a symposium on surgery and nutrition). *Proceedings of the Nutrition Society* **39**, 101–5.
- Czeizel, AF and Dudas, I 1992 Prevention of first occurrence of neural tube defects by periconceptual vitamin supplementation. *New England Journal of Medicine* **327**, 1832–5.
- Davies, JNP 1964 The decline of pellagra in the United States. *Lancet* **ii**, 195–6.
- Davies, J and Dickerson, JWT 1991 *Nutrient content of food portions*. Cambridge: Royal Society of Chemistry.
- DCCT 1993 The Diabetes 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. *New England Journal of Medicine* **329**, 977–86.
- Debenham, K 1992 Nutrition for specific disease conditions. In *Nutrition and the consumer*. Ed. Walker, AF and Rolls, BA. London: Elsevier Applied Science, pp. 249–70.
- Debons, AF, Krimsky, I, Maayan, MI, Fani, K and Jimenez, LA 1977 The goldthioglucose obesity syndrome. *Federation Proceedings of the Federation of American Societies for Experimental Biology* **36**, 143–7.
- Delmi, M, Rapin, C-H, Delmas, PD, Vasey, H and Bonjour, J-P 1990 Dietary supplementation in elderly patients with fractured neck of the femur. *Lancet* **335**, 1013–16.
- Denton, D, Weisinger, R, Mundy, NI, et al. 1995 The effect of increased salt intake on the blood pressure of chimpanzees. *Nature Medicine* **1** (10), 1009–16.
- Devine, A, Criddle, RA, Dick, IM, Kerr, DA and Prince, RL 1995 A longitudinal study of the effect of sodium and calcium intakes on regional bone density in post-menopausal women. *American Journal of Clinical Nutrition* **62**, 740–5.
- DH 1992 Department of Health. *The health of the nation. A strategy for health in England*. London: HMSO.
- DH 1993 Department of Health. *Report of the Chief Medical Officer's Expert Group on the sleeping position of infants and cot death*. London: HMSO.
- DHHS 1992 Department of Health and Human Services. *Healthy people 2000*. Boston: Jones and Bartlett Publishers Inc.
- DHSS 1972 Department of Health and Social Security. *A nutrition survey of the elderly*. Report on health and social subjects No. 3. London: HMSO.
- DHSS 1979 Department of Health and Social Security. *Nutrition and health in old age*. Report on health and social subjects No. 16. London: HMSO.
- DHSS 1988 Department of Health and Social Security. *Present day practice in infant feeding*. Report on health and social subjects No. 32. London: HMSO.
- Dietz, WH 1999 How to tackle the problem early? The role of education in the prevention of obesity. *International Journal of Obesity* **23** (Supplement 4), 57–9.
- Dietz, WH and Gortmaker, SL 1985 Do we fatten our children at the television set? Obesity and television viewing in children and adolescents. *Pediatrics* **75**, 807–12.
- Diplock, AT 1991 Antioxidant nutrients and disease prevention: an overview. *American Journal of Clinical Nutrition* **53**, 189S–93S.
- Dobson, B, Beardsworth, A, Keil, T and Walker, R 1994 *Diet, choice and poverty*. London: Family Policy Studies Centre.
- Dodge, JA 1992 Nutrition in cystic fibrosis: a historical overview. *Proceedings of the Nutrition Society* **51**, 225–35.
- Draper, A 1994 Energy density of weaning foods. In *Infant nutrition*. Ed. Walker AF and Rolls BA. London: Chapman and Hall, pp. 209–23.

- Drenick, EJ, Bale, GS, Seltzer, F and Johnson, DG 1980 Excessive mortality and causes of death in morbidly obese men. *Journal of the American Medical Association* **243**, 443–5.
- Dunnigan, MG 1993 The problem with cholesterol. *British Medical Journal* **306**, 1355–6.
- Durnin, JVGA 1992 Physical activity levels – past and present. In *Nutrition and physical activity*. Ed. Norgan, NG. Cambridge: Cambridge University Press. pp. 20–7.
- Durnin, JVGA and Womersley, J 1974 Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged 16 to 72 years. *British Journal of Nutrition* **32**, 77–97.
- Dyerberg, J and Bang, HO 1979 Haemostatic function and platelet polyunsaturated fatty acids in Eskimos. *Lancet* **ii**, 433–5.
- Easterbrook, PJ, Berlin, JA, Gopalan, R and Matthews, DR 1992 Publication bias in clinical research. *Lancet* **337**, 867–72.
- Eastwood, M and Eastwood, M 1988 Nutrition and diets for endurance runners. *British Nutrition Foundation Nutrition Bulletin* **13**, 93–100.
- Eklund, H, Finnstrom, O, Gunnarskog, J, Kallen, B and Larsson, Y 1993 Administration of vitamin K to newborn infants and childhood cancer. *British Medical Journal* **307**, 89–91.
- Elsas, LJ and Acosta PB 1999 Nutritional support of inherited metabolic disease. In *Modern nutrition in health and disease*, 9th edn. Ed. Shils, ME, Olson, JA, Shike, M and Ross AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 1003–56.
- Enerback, S, Jacobsson, A, Simpson, EM, et al. 1997 Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. *Nature* **387**, 90–7.
- Engelberg, H 1992 Low serum cholesterol and suicide. *Lancet* **339**, 727–9.
- Englyst, HN and Kingman SM 1993 Carbohydrates. In *Human nutrition and dietetics*, 9th edn. Ed. Garrow, JS and James, WPT. Edinburgh: Churchill Livingstone, pp. 38–55.
- Eskes, TKAB 1998 Neural tube defects, vitamins and homocysteine. *European Journal of Pediatrics* **157** (Supplement 2), S139–41.
- FAO 1996 *Food for all*. Book available to download free from (www.fao.org/wfs/wfsbook/e/ffa01-e.htm), pp. 21–31.
- FDA 1992 Food and Drug Administration. The new food label. *FDA Backgrounder* **92–4**, 1–9.
- Fentem, PH 1992 Exercise in prevention of disease. *British Medical Bulletin* **48**, 630–50.
- Fiatarone, MA, O'Neill, EF, Ryan, ND, et al. 1994 Exercise training and nutritional supplementation for physical frailty in very elderly people. *New England Journal of Medicine* **330**, 1769–75.
- Fieldhouse, P 1998 *Food and nutrition: customs and culture*, 2nd edn. London: Stanley Thornes.
- Filteau, S and Tomkins, A 1994 Infant feeding and infectious disease. In *Infant nutrition*. Ed. Walker, AF and Rolls, BA. London: Chapman and Hall, pp. 143–62.
- Finch, S, Doyle, W, Lowe, C, et al. 1998 *National Diet and Nutrition Survey: people aged 65 years and over*. London: The Stationery Office.
- Fleury, C, Neverova, M, Collins, S, et al. 1997 Uncoupling protein 2: a novel gene linked to obesity and hyperinsulinaemia. *Nature Genetics* **15**, 269–73.
- Forbes, GB 1999 Body composition: influence of nutrition, physical activity, growth and aging. In *Modern nutrition in health and disease*. Ed. Shils, ME, Olson, JA, Shike, M and Ross, AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 789–810.
- Forte, JG, Miguel, JM, Miguel, MJ, de Pyadua, F and Rose, G 1989 Salt and blood pressure: a community trial. *Journal of Human Hypertension* **3**, 179–84.
- Foster, K, Lader, D and Cheesbrough, S 1997 *Infant feeding* 1995. London: The Stationery Office.
- Frantz, ID and Moore, RB 1969 The sterol hypothesis in atherosclerosis. *American Journal of Medicine* **46**, 684–90.
- Fraser, GE 1988 Determinants of ischemic heart disease in Seventh-day Adventists: a review. *American Journal of Clinical Nutrition* **48**, 833–6.
- Friedman, J and Halaas, JL 1998 Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–70.
- Frost, CD, Law, MR and Wald, NJ 1991 By how much does dietary salt reduction lower blood pressure? II – Analysis of observational data within populations. *British Medical Journal* **302**, 815–18.
- Fujita, Y 1992 Nutritional requirements of the elderly: a Japanese view. *Nutrition Reviews* **50**, 449–53.
- Garrow, JS 1992 The management of obesity. Another view. *International Journal of Obesity* **16** (Supplement 2), S59–S63.
- Garrow, JS and James, WPT 1993 *Human nutrition and dietetics*, 9th edn. Edinburgh: Churchill Livingstone.
- Gey, KF, Puska, P, Jordan, P and Moser, UK 1991 Inverse correlation between plasma vitamin E and mortality from ischemic heart disease in cross-cultural epidemiology. *American Journal of Clinical Nutrition* **53**, 326S–34S.
- Gibbs, WW 1996 Gaining on fat. *Scientific American* August, 70–6.
- Gilbert, S 1986 *Pathology of eating. Psychology and treatment*. London: Routledge and Kegan Paul.
- GISSI 1999 GISSI-Prevenzione Investigators. Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. *Lancet* **354**, 447–55.
- Gleibermann, L 1973 Blood pressure and dietary salt in human populations. *Ecology of Food and Nutrition* **2**, 143–56.
- Golding, J, Greenwood, R, Birmingham, K and Mott, M 1992 Childhood cancer: intramuscular vitamin K, and

- pethidine given during labour. *British Medical Journal* **305**, 341–5.
- Gortmaker, SL, Must, A, Perrin, JM, Sobol, AM and Dietz, WH 1993 Social and economic consequences of overweight in adolescence. *New England Journal of Medicine*, **329**, 1008–12.
- Gounelle de Pontanel, H 1972 Chairman's opening address. In *Proteins from hydrocarbons. The proceedings of the 1972 symposium at Aix-en-Provence*. London: Academic Press, pp. 1–2.
- Gregory, JR, Collins, DL, Davies, PDW, Hughes, JM and Clarke, PC (1995) *National Diet and Nutrition Survey: children aged 1½ to 4½ years*, Vol. 1. London: HMSO.
- Gregory, J, Foster, K, Tyler, H and Wiseman, M 1990 *The Dietary and Nutritional Survey of British Adults*. London: HMSO.
- Gregory, J, Foster, K, Tyler, H and Wiseman, M 1994 *The Dietary and Nutritional Survey of British Adults – further analysis*. London: HMSO.
- Groff, JL, Gropper, SS and Hunt, SM 1995 *Advanced nutrition and human metabolism*, 2nd edn. Minneapolis/St Paul: West Publishing Company.
- Gronbaek, M, Deis, A, Sorensen, TIA, et al. 1994 Influence of sex, age, body mass index, and smoking on alcohol intake and mortality. *British Medical Journal* **308**, 302–6.
- Groom, H 1993 What price a healthy diet? *The British Nutrition Foundation Nutrition Bulletin* **18**, 104–9.
- Group 1994 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 cancer in male smokers. *New England Journal of Medicine* **330**, 1029–35.
- Grundy, SM 1987 Monounsaturated fatty acids, plasma cholesterol, and coronary heart disease. *American Journal of Clinical Nutrition* **45**, 1168–75.
- Grundy, SM 1998 Hypertriglyceridemia, atherogenic dyslipidemia and the metabolic syndrome. *American Journal of Cardiology* **81**(4A), 18B–25B.
- Hallas, TJ and Walker, AF 1992 Vegetarianism: the healthy alternative? In *Nutrition and the consumer*. Ed. Walker, AF and Rolls, BA. London: Elsevier Applied Science, pp. 211–47.
- Halliday, A and Ashwell, M 1991a *Non-starch polysaccharides*. Briefing paper 22. London: British Nutrition Foundation.
- Halliday, A and Ashwell, M 1991b *Calcium*. Briefing paper 24. London: British Nutrition Foundation.
- Hamilton, EMN, Whitney, EN and Sizer, FS 1991 *Nutrition. Concepts and controversies*, 5th edn. St Paul, MN: West Publishing Company.
- Harper, AE 1999 Defining the essentiality of nutrients. In *Modern nutrition in health and disease*, 9th edn. Ed. Shils, ME, Olson, JA, Shike, M and Ross, AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 3–10.
- Harris, MB 1983 Eating habits, restraint, knowledge and attitudes toward obesity. *International Journal of Obesity* **7**, 271–86.
- Harris, WS 1989 Fish oils and plasma lipid and lipoprotein metabolism in humans: a critical review. *Journal of Lipid Research* **30**, 785–807.
- Hawthorn, J 1989 Safety and wholesomeness of irradiated foods. *British Nutrition Foundation Nutrition Bulletin* **14**, 150–62.
- Hayden, RM and Allen, GJ 1984 Relationship between aerobic exercise, anxiety, and depression: convergent validation by knowledgeable informants. *Journal of Sports Medicine* **24**, 69–74.
- Hazel, T and Southgate, DAT 1985 Trends in the consumption of red meat and poultry – nutritional implications. *British Nutrition Foundation Nutrition Bulletin* **10**, 104–17.
- HEA 1994 Health Education Authority. *The balance of good health. The national food guide*. London: Health Education Authority.
- Hegsted, DM 1986 Calcium and osteoporosis. *Journal of Nutrition* **116**, 2316–19.
- Hennekens, CH, Buring, JF, Manson, JF, et al. 1996 Lack of effect of long-term supplementation with beta-carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine* **334**, 1145–9.
- Henson, S 1992 From high street to hypermarket. Food retailing in the 1990s. In *Your food: whose choice?* Ed. National Consumer Council. London: HMSO, pp. 95–115.
- Hetzel, BS 1993 Iodine deficiency disorders. In *Human nutrition and dietetics*, 9th edn. Ed. Garrow, JS and James, WPT. Edinburgh: Churchill Livingstone, pp. 534–55.
- Hetzel, BS and Clugston, GA 1999 Iodine. In *Modern nutrition in health and disease*, 9th edn. Ed. Shils, ME, Olson, JA, Shike, M and Ross, AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 253–64.
- Hill, GL, Blackett, RL, Pickford, I, et al. 1977 Malnutrition in surgical patients. An unrecognised problem. *Lancet* **i**, 689–92.
- Hirsch, J 1997 Some heat but not enough light. *Nature* **387**, 27–8.
- Hirsch, J and Han, PW 1969 Cellularity of rat adipose tissue: effects of growth, starvation and obesity. *Journal of Lipid Research* **10**, 77–82.
- Hobbs, BC and Roberts, D 1993 *Food poisoning and food hygiene*, 6th edn. London: Edward Arnold.
- Hodge, AM, Dowse, GK, Poelape, P, Collins, VR, Imo, T and Zimmet, PZ 1994 Dramatic increases in the prevalence of obesity in Western Samoa over the 13 year period 1978–1991. *International Journal of Obesity* **18**, 419–28.
- Hubley, J 1993 *Communicating health – an action guide to health education and health promotion*. London: Macmillan.

- Hulley, SB, Walsh, JMB and Newman, TB 1992 Health policy on blood cholesterol: time to change directions. *Circulation* **86**, 1026–9.
- Hultman, E, Harris, RC and Spriet, LL 1999 Diet in work and exercise performance. In *Modern nutrition in health and disease*. Ed. Shils, ME, Olson, JA, Shike, M and Ross, AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 761–82.
- Hunt, P, Gatenby, S and Rayner, M 1995 The format for the national food guide: performance and preference studies. *Journal of Human Nutrition and Dietetics* **8**, 335–51.
- Huse, DM and Lucas, AR 1999 Behavioural disorders affecting food intake: anorexia nervosa, bulimia nervosa and other psychiatric conditions. In *Modern nutrition in health and disease*, 9th edn. Ed. Shils, ME, Olson, JA, Shike, M and Ross, AC. Baltimore: Williams and Watkins, pp. 1513–21.
- IFPRI 1996 International Food Policy Research Institute. *Key trends in feeding the world*. www.cgiar.org/ifpri/2020/synth/trends.htm
- IFT 1975 A report by the Institute of Food Technologists' Expert Panel on Food Safety and Nutrition and the Committee on Public Information. Naturally occurring toxicants in foods. *Food Technology* **29** (3), 67–72.
- Iribarren, C, Sharp, DS, Burchfiel, CM and Petrovich, H 1995 Association of weight loss and weight fluctuation with mortality among Japanese American men. *New England Journal of Medicine* **333**, 686–92.
- Jacob, M 1993 Legislation. In *Food poisoning and food hygiene*, 6th edn. Ed. Hobbs, BC and Roberts, D. London: Edward Arnold, pp. 280–302.
- Jacobs, D, Blackburn, H, Higgins, M, et al. 1992 Report of the conference on low blood cholesterol: mortality associations. *Circulation* **86**, 1046–59.
- Jacobson, MS 1987 Cholesterol oxides in Indian ghee: possible cause of unexplained high risk of atherosclerosis in Indian immigrant populations. *Lancet* **ii**, 656–8.
- James, WPT, Ralph, A and Sanchez-Castillo, CP 1987 The dominance of salt in manufactured food in the sodium intake of affluent societies. *Lancet* **i**, 426–8.
- Jarjour, LMA, Prentice, A, Sawo, Y, et al. 1994 Changes in the diet of Mandinka women in The Gambia between 1978–79 and 1990–91: consequences for calcium intakes. *Proceedings of the Nutrition Society* **53**, 258A.
- Jebb, SA, Goldberg, GR, Coward, WA, Murgatroyd, PR and Prentice, AM 1991 Effects of weight cycling caused by intermittent dieting on metabolic rate and body composition in obese women. *International Journal of Obesity* **15**, 367–74.
- Jelliffe, DB 1967 Parallel food classifications in developing and industrialised countries. *American Journal of Clinical Nutrition* **20**, 279–81.
- Jenkins, DJA and Wolever, TMS 1981 Slow release carbohydrate and the treatment of diabetics. *Proceedings of the Nutrition Society* **40**, 227–36.
- Johnell, O and the MEDOS Study Group 1992 The apparent incidence of hip fracture in Europe: a study of national register sources. *Osteoporosis International* **2**, 298–302.
- Johnston, PK 1999 Nutritional implications of vegetarian diets. In *Modern nutrition in health and disease*. Ed. Shils ME, Olson JA, Shike M and Ross AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 1755–69.
- Jones, A 1974 *World protein resources*. Lancaster, England: Medical and Technical Publishing Company.
- Jones, E, Hughes, R and Davies, H 1988 Intake of vitamin C and other nutrients by elderly patients receiving a hospital diet. *Journal of Human Nutrition and Dietetics* **1**, 347–53.
- Joossens, JB, Hill, MJ, Elliott, P, et al. 1996 On behalf of the European Cancer Prevention (ECP) and the Intersalt Cooperative Research Group: dietary salt, nitrate and stomach cancer mortality in 24 countries. *International Journal of Epidemiology* **25**, 494–504.
- Joossens, JV and Geboers, J 1981 Nutrition and gastric cancer. *Proceedings of the Nutrition Society* **40**, 37–46.
- Kanarek, RB and Hirsch, E 1977 Dietary-induced overeating in experimental animals. *Federation Proceedings of the American Society for Experimental Biology* **36**, 154–8.
- Kanis, JA 1993 The incidence of hip fracture in Europe. *Osteoporosis International* **3** (Supplement 1), s10–s15.
- Keen, H and Thomas, B 1988 Diabetes mellitus. In *Nutrition in the clinical management of disease*. Ed. Dickerson, JWT and Lee, HA London: Edward Arnold, pp. 167–90.
- Kennedy, GC 1966 Food intake, energy balance and growth. *British Medical Bulletin* **22**, 216–20.
- Keys, A, Anderson, JT and Grande, F 1959 Serum cholesterol in man: diet fat and intrinsic responsiveness. *Circulation* **19**, 201–4.
- Keys, A, Anderson, JT and Grande, F 1965 Serum cholesterol responses to changes in the diet. II. The effect of cholesterol in the diet. *Metabolism* **14**, 759–60.
- Keys, A, Brozek, J, Henschel, A, Mickelson, O and Taylor, HL 1950 *The biology of human starvation*. Minneapolis: University of Minnesota Press.
- KFC 1992 King's Fund Centre. *A positive approach to nutrition as treatment*. London: King's Fund Centre.
- King, J and Ashworth, A 1994 Patterns and determinants of infant feeding practices worldwide. In *Infant nutrition*. Ed. Walker, AF and Bolls, BA. London: Chapman and Hall, pp. 61–91.
- Klahr, S, Levey, AS, Beck, GJ, et al. 1994 The effects of dietary protein restriction on the progression of chronic renal disease. *New England Journal of Medicine* **330**, 877–84.
- Klatsky, AL, Armstrong, MA and Friedman, JD 1992 Alcohol and mortality. *Annals of Internal Medicine* **117**, 646–54.

- Kremer, JM, Bigauette, J, Mickalek, AV, et al. 1985 Effects of manipulation of dietary fatty acids on clinical manifestation of rheumatoid arthritis. *Lancet* i, 184-7.
- Kromhout, D 1990 n-3 fatty acids and coronary heart disease: epidemiology from Eskimos to Western populations. *British Nutrition Foundation Nutrition Bulletin* 15, 93-102.
- Kuczmarski, RJ, Flegal, KM, Campbell, SM and Johnson, CL 1994 Increasing prevalence of overweight among US adults. *Journal of the American Medical Association* 272, 205-11.
- Kune, GA, Kune, S and Watson, LF 1993 Perceived religiousness is protective for colorectal cancer: data from the Melbourne Colorectal Cancer Study. *Journal of the Royal Society of Medicine* 86, 645-7.
- Larsson, J, Unosson, N, Ek, AC, Nilsson, L, Thorslund, S and Bjurulf, P 1990 Effect of dietary supplement on nutritional status and clinical outcome in 501 geriatric patients - a randomised study. *Clinical Nutrition* 9, 179-84.
- Laurent, C 1993 Private function? *Nursing Times* 89 (47), 14-15.
- Law, MR, Frost, CD and Wald, NJ 1991a By how much does dietary salt reduction lower blood pressure? I - Analysis of observational data among populations. *British Medical Journal* 302, 811-15.
- Law, M.R., Frost, CD and Wald, NJ 1991b By how much does dietary salt reduction lower blood pressure? III - Analysis of data from trials of salt reduction. *British Medical Journal* 302, 819-24.
- Leather, S 1992 Less money, less choice. Poverty and diet in the UK today. In *Your food: whose choice?* Ed. National Consumer Council. London: HMSO, pp. 72-94.
- Lee, CD, Jackson, AS and Blair, SN 1998 US weight guidelines, is it important to consider cardiopulmonary fitness. *International Journal of Obesity* 22 (Supplement 2), 52-7.
- Leeds, AR 1979 The dietary management of diabetes in adults. *Proceedings of the Nutrition Society* 38, 365-71.
- Leeds, AR, Brand Miller, J, Foster-Powell, K and Colagiuri, S 1998 *The glucose revolution*. Australia: Hodder & Stoughton.
- Lehmann, AB 1991 Nutrition in old age: an update and questions for future research: part 1. *Reviews in Clinical Gerontology* 1, 135-45.
- Leichter, I, Margulies, JY, Weinreb, A, et al. 1982 The relationship between bone density, mineral content, and mechanical strength in the femoral neck. *Clinical Orthopaedics and Related Research* 163, 272-81.
- Lennard, TWJ and Browell, DA 1993 The immunological effects of trauma. *Proceedings of the Nutrition Society* 52, 85-90.
- Leon, DA 1985 Physical activity levels and coronary heart disease: analysis of epidemiologic and supporting studies. *Medical Clinics of North America* 69, 3-19.
- Leon, DA 1991 Influence of birth weight on differences in infant mortality by social class and legitimacy. *British Medical Journal* 303, 964-7.
- Leon, DA 1993 Failed or misleading adjustment for confounding. *Lancet* 342, 479-81.
- Livingstone, B 1996 More ins than outs of energy balance. *British Nutrition Foundation Nutrition Bulletin* 21 (Supplement), 6-15.
- Locatelli, F, Alberti, D, Graziani, G, Buccianti, G, Redaelli, B, Giagrande, A and the Northern Italian Cooperative Study Group 1991 Prospective, randomised, multicentre trial of the effect of protein restriction on progression of chronic renal insufficiency. *Lancet* 337, 1299-304.
- Lock, S 1977 Iron deficiency anaemia in the UK. In *Getting the most out of food*, No. 12. London: Van den Berghs and Jurgens Ltd, pp. 111-36.
- McClaren, DS 1974 The great protein fiasco. *Lancet* ii, 93-6.
- McColl, K 1988 The sugar-fat 'seesaw'. *The British Nutrition Foundation Nutrition Bulletin* 13, 114-18.
- McCormick, J and Skrabaneck, P 1988 Coronary heart disease is not preventable by population interventions. *Lancet* i, 839-41.
- McGill, HC 1979 The relationship of dietary cholesterol to serum cholesterol concentration and to atherosclerosis in man. *American Journal of Clinical Nutrition* 32, 2664-702.
- MacGregor, GA and de Wardener, HE 1999 *Salt, diet and health*. Cambridge: Cambridge University Press.
- MacGregor, GA, Markandu, N, Best, F, et al. 1982 Double-blind random crossover of moderate sodium restriction in essential hypertension. *Lancet* i, 351-5.
- McKeigue, PM, Marmot, MG, Adelstein, AM, et al. 1985 Diet and risk factors for coronary heart disease in Asians in northwest London. *Lancet* ii, 1086-9.
- McKeigue, PM, Shah, B and Marmot, MG 1991 Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet* 337, 382-6.
- Macnair, AL 1994 Physical activity, not diet, should be the focus of measures for the primary prevention of cardiovascular disease. *Nutrition Research Reviews* 7, 43-65.
- McNutt, K 1998 Sugar replacers. A new consumer education challenge. *British Nutrition Foundation Nutrition Bulletin* 23, 216-23.
- McWhirter, JP and Pennington, CR 1994 Incidence and recognition of malnutrition in hospital. *British Medical Journal* 308, 945-8.
- MAFF 1991 Ministry of Agriculture Fisheries and Food. *Fifty years of the National Food Survey 1940-1990*. Ed. Slater, JM. London: HMSO.
- MAFF 1993 Ministry of Agriculture Fisheries and Food. *The National Food Survey 1992*. Annual report of the National Food Survey Committee. London: HMSO.

- MAFF 1997 Ministry of Agriculture Fisheries and Food. *The National Food Survey 1996*. Annual report of the National Food Survey Committee. London: HMSO.
- Mangiapani, EH and Salter, AM 1999 *Diet, lipoproteins and coronary heart disease. A biochemical perspective*. Nottingham: Nottingham University Press.
- Manson, JE, Colditz, GA, Stampfer, MJ, et al. 1990 A prospective study of obesity and coronary heart disease in women. *New England Journal of Medicine* **322**, 882–9.
- Manson, JE, Willett, WC, Stampfer, MJ, et al. 1995 Body weight and mortality among women. *New England Journal of Medicine* **333**, 677–85.
- Marsh, AG, Sanchez, JV, Mickelsen, O, Chaffee FL and Fagall, SM 1988 Vegetarian lifestyle and bone mineral density. *American Journal of Clinical Nutrition* **48**, 837–41.
- Maslow, AH 1943 A theory of human motivation. *Psychological Reviews* **50**, 370–96.
- Mathews, F (1996) Antioxidant nutrients in pregnancy; a systematic review of the literature. *Nutrition Research Reviews* **9**, 175–95.
- Mattila, K, Haavisto, M and Rajala, S 1986 Body mass index and mortality in the elderly. *British Medical Journal* **292**, 867–8.
- Mattson, FH and Grundy, SM 1985 Comparison of effects of dietary saturated, monounsaturated and polyunsaturated fatty acids on plasma lipids and lipoproteins in man. *Journal of Lipid Research* **26**, 194–202.
- Mayer, J 1956 Appetite and obesity. *Scientific American* **195** (5), 108–16.
- Mayer, J 1968 *Overweight: causes, cost and control*. Engelwood Cliffs, NJ: Prentice Hall.
- Mela DJ 1997 Fat and sugar substitutes: implications for dietary intakes and energy balance. *Proceedings of the Nutrition Society* **56**, 820–8.
- Mensink, RP and Katan, MJ 1990 Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. *New England Journal of Medicine* **323**, 439–45.
- Miettinen, TA, Puska, P, Gylling, H, Vanhanen, H and Vartiainen, E 1995 Reduction of serum cholesterol with sitostanol-ester margarine in a mildly hypercholesterolemic population. *New England Journal of Medicine* **333**, 1308–12.
- Miller, DS 1979 Non-genetic models of obesity. In *Animal models of obesity*. Ed. Festing, MWF. London: Macmillan, pp. 131–40.
- Miller, DS and Payne, PR 1969 Assessment of protein requirements by nitrogen balance. *Proceedings of the Nutrition Society* **28**, 225–34.
- Millstone, E 1985 Food additive regulation in the UK. *Food Policy* **10**, 237–52.
- Mitchell, EB 1988 Food intolerance. In *Nutrition in the clinical management of disease*. Ed. Dickerson, JWT and Lee, HA. London: Edward Arnold, pp. 374–91.
- Montague CT, Farooqi, IS, Whitehead, JP, et al. 1997 Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–8.
- Morgan, JB 1988 Nutrition for and during pregnancy. In *Nutrition in the clinical management of disease*, 2nd edn. Ed. Dickerson, JWT and Lee, HA. London: Edward Arnold, pp. 1–29.
- Morgan, JB 1998 Weaning: when and what. *British Nutrition Foundation Nutrition Bulletin* **23** (Supplement 1), 35–45.
- Morris, JN, Everitt, MG, Pollard, R, Chave, SPW and Semmence, AM 1980 Vigorous exercise in leisure-time: protection against coronary heart disease. *Lancet* **i**, 1207–10.
- Mowe, M, Bohmer, T and Kindt, E 1994 Reduced nutritional status in elderly people is probable before disease and probably contributes to the development of disease. *American Journal of Clinical Nutrition* **59**, 317–24.
- Moynihan PJ 1995 The relationship between diet, nutrition and dental health: an overview and update for the 90s. *Nutrition Research Reviews* **8**, 193–224.
- MRC 1991 The MRC Vitamin Study Group. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. *Lancet* **338**, 131–7.
- MRFIT 1982 Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial. *Journal of the American Medical Association* **248**, 1465–77.
- NACNE 1983 The National Advisory Committee on Nutrition Education. *A discussion paper on proposals for nutritional guidelines for health education in Britain*. London: Health Education Council.
- Naismith, DJ 1980 Maternal nutrition and the outcome of pregnancy – a critical appraisal. *Proceedings of the Nutrition Society* **39**, 1–11.
- Neaton, JD, Balckburn, H, Jacobs, D, et al. 1992 Serum cholesterol level and mortality findings for men screened for the Multiple Risk Factor Intervention Trial. *Archives of Internal Medicine* **152**, 1490–500.
- Nelson, M, Atkinson, M and Meyer, J 1997 *Food portion sizes. A photographic atlas*. London: MAFF Publications.
- Nicolaas, G 1995 *Cooking: attitudes and behaviour*. London: HMSO.
- NRC 1943 National Research Council. *Recommended dietary allowances*, 1st edn. Washington, DC: National Academy of Sciences.
- NRC 1989a National Research Council. *Recommended dietary allowances*, 10th edn. Washington, DC: National Academy of Sciences.
- NRC 1989b National Research Council. Diet and health: implications for reducing chronic disease risk. *Nutrition Reviews* **47**, 142–9.
- Oliver, MF 1981 Diet and coronary heart disease. *British Medical Bulletin* **37**, 49–58.
- Oliver, MF 1991 Might treatment of hypercholesterolaemia increase non-cardiac mortality. *Lancet* **337**, 1529–31.

- Omen, GS, Goodman, GE, Thornquist, MD, et al. 1996 Combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* **334**, 1150–5.
- Paffenbarger, RS, Hyde, RT, Wing, AL and Hsieh, C-C 1986 Physical activity, all-cause mortality, and longevity of college alumni. *New England Journal of Medicine* **314**, 605–13.
- Paolini, M, Cantelli-Forti, G, Perocco, P, Pedulli, GF, Abdel-Rahman, SZ and Legator, MS (1999) Co-carcinogenic effect of β -carotene. *Nature* **398**, 760–1.
- Passim, H and Bennett, JW 1943 *Social progress and dietary change*. National Research Council Bulletin 108. Washington, DC: National Research Council.
- Passmore, R and Eastwood, MA 1986 *Human nutrition and dietetics*, 8th edn. Edinburgh: Churchill Livingstone.
- Pauling, L 1972 *Vitamin C and the common cold*. London: Ballantine Books.
- Pennington, CR 1997 Disease and malnutrition in British hospitals. *Proceedings of the Nutrition Society* **56**, 393–407.
- Peto, R, Doll, R, Buckley, JD and Sporn, MB 1981 Can dietary beta-carotene materially reduce human cancer rates? *Nature, London* **290**, 201–8.
- Pocock, NA, Eisman, JA, Yeates, MG, Sambrook, PN and Ebert, S 1986 Physical fitness is a major determinant of femoral neck and lumbar spine bone mineral density. *Journal of Clinical Investigation* **78**, 618–21.
- Poleman, TI 1975 World food: a perspective. *Science* **188**, 510–18.
- Pomrehn, PR, Wallace, RB and Burmeister, LF 1982 Ischemic heart disease mortality in Iowa farmers. *Journal of the American Medical Association* **248**, 1073–6.
- Pond, C 1987 Fat and figures. *New Scientist* 4 June, 62–6.
- Poskitt, EME 1988 Childhood. In *Nutrition in the clinical management of disease*. Ed. Dickerson, JWT and Lee, HA. London: Edward Arnold, pp. 30–68.
- Poskitt, EME 1998 Infant nutrition: lessons from the third world. *British Nutrition Foundation Nutrition Bulletin* **23** (Supplement 1), 12–22.
- Powell-Tuck J 1997 Penalties of hospital undernutrition. *Journal of the Royal Society of Medicine* **90**, 8–11.
- Powers, SK and Howley ET (1990) *Exercise physiology*. Dubuque, Io: WC Brown.
- Prentice, A 1997 Is nutrition important in osteoporosis? *Proceedings of the Nutrition Society* **56**, 357–67.
- Prentice, AM 1989 Energy expenditure in human pregnancy. *British Nutrition Foundation Nutrition Bulletin* **14**, 9–22.
- Prentice, AM, Black, AE, Murgatroyd, PR, et al. 1989 Metabolism or appetite: questions of energy balance with particular reference to obesity. *Journal of Human Nutrition and Dietetics* **2**, 95–104.
- Prentice, AM, Goldberg, GR, Jebb, SA, Black, AE, Murgatroyd, PR and Diaz, EO 1991 Physiological responses to slimming. *Proceedings of the Nutrition Society* **50**, 441–58.
- Prentice, AM and Jebb, SA 1995 Obesity in Britain: gluttony or sloth? *British Medical Journal* **311**, 437–9.
- Prentice, A, Laskey, MA, Shaw, J, et al. 1993 The calcium and phosphorus intakes of rural Gambian women during pregnancy and lactation. *British Journal of Nutrition* **69**, 885–96.
- Prescott-Clarke, P and Primatesta, P 1998 *Health Survey for England 1996*. London: The Stationery Office.
- Prusiner, SB 1995 The prion diseases. *Scientific American* January, 30–7.
- Rapala, JM, Virtamo, J, Ripatti, S, et al. 1997 Randomised trial of alpha-tocopherol and beta-carotene supplements on incidence of major coronary events in men with previous myocardial infarction. *Lancet* **349**, 1715–20.
- Ravnskov, U 1992 Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *British Medical Journal* **305**, 15–19.
- RCN 1993 Royal College of Nursing. *Nutrition standards and the older adult*. Dynamic Quality Improvement Programme. London: RCN.
- Renaud, S and de Lorgeril, M 1992 Wine, alcohol, platelets and the French paradox for coronary heart disease. *Lancet* **339**, 1523–6.
- Report 1992 *Folic acid and the prevention of neural tube defects*. Report from an expert advisory group. London: Department of Health.
- Richards, JR 1980 Current concepts in the metabolic responses to injury, infection and starvation. *Proceedings of the Nutrition Society* **39**, 113–23.
- Riemersma, RA, Wood, DA, Macintyre, CCA, Elton, RA, Fey, KF and Oliver, MF 1991 Risk of angina pectoris and plasma concentrations of vitamins A, C, and E and carotene. *Lancet* **337**, 1–5.
- Rimm, EB, Giovannucci, EL, Willett, WC, et al. 1991 Prospective study of alcohol consumption and risk of coronary disease in men. *Lancet* **338**, 464–8.
- Rimm, EB, Stampfer, MJ, Ascheirio, A, Giovannucci, EL, Colditz, GA and Willett, WC 1993 Vitamin E consumption and the risk of coronary heart disease in men. *New England Journal of Medicine* **328**, 1450–5.
- Rissanen, A, Heliovaara, M, Knekt, P, Reunanen, A, Aromaa, A and Maatela, J 1990 Risk of disability and mortality due to overweight in a Finnish population. *British Medical Journal* **301**, 835–7.
- Rivers, JPW and Frankel, TL 1981 Essential fatty acid deficiency. *British Medical Bulletin* **37**, 59–64.
- Roberts, D 1982 Factors contributing to outbreaks of food poisoning in England and Wales 1970–1979. *Journal of Hygiene, Cambridge* **89**, 491–8.
- Robertson, I, Glekin, BM, Henderson, JB, Lakhani, A and Dunnigan, MG 1982 Nutritional deficiencies amongst ethnic minorities in the United Kingdom. *Proceedings of the Nutrition Society* **41**, 243–56.

- Rolls, BA 1992 Calcium nutrition. In *Nutrition and the consumer*. Ed. Walker, AF and Rolls, BA. London: Elsevier Applied Science, pp. 69–95.
- Rolls, BJ and Phillips, PA 1990 Aging and disturbances of thirst and fluid balance. *Nutrition Reviews* **48**, 137–44.
- Rolls, BJ, Rowe, EA and Rolls, ET 1982 How flavour and appearance affect human feeding. *Proceedings of the Nutrition Society* **41**, 109–17.
- Rosado JL 1997 Lactose digestion and maldigestion: implications for dietary habits in developing countries. *Nutrition Research Reviews* **10**, 137–49.
- Rothwell NJ and Stock MJ 1979 A role for brown adipose tissue in diet-induced thermogenesis. *Nature, London* **281**, 31–5.
- Rudman, D 1988 Kidney senescence; a model for ageing. *Nutrition Reviews* **46**, 209.
- Rutishauser, IHE 1992 Vitamin A and carotenoids in human nutrition. In *Nutrition and the consumer*. Ed. Walker, AF and Rolls, BA. London: Elsevier Applied Science, pp. 17–68.
- Ryan, YM, Gibney, MJ and Flynn, MAT 1998 The pursuit of thinness: a study of Dublin schoolgirls aged 15y. *International Journal of Obesity* **22**, 485–7.
- Sadler, M (1994) *Nutrition in pregnancy*. London: British Nutrition Foundation.
- Samaras, K and Campbell, IV 1997 The non-genetic determinants of central adiposity. *International Journal of Obesity* **21**, 839–45.
- Sanders, TAB 1985 Influence of fish-oil supplements on man. *Proceedings of the Nutrition Society* **44**, 391–7.
- Sanders, TAB 1988 Growth and development of British vegan children. *American Journal of Clinical Nutrition* **48**, 822–5.
- Sanders, TAB, Mistry, M and Naismith, DJ 1984 The influence of a maternal diet rich in linoleic acid on brain and retinal docosahexaenoic acid in the rat. *British Journal of Nutrition* **51**, 57–66.
- Sanderson, FH 1975 The great food fumble. *Science* **188**, 503–9.
- Schachter, S 1968 Obesity and eating. *Science* **161**, 751–6.
- Schachter, S and Rodin, J 1974 *Obese humans and rats*. New York: Halstead Press.
- Schrauwen PS, Walder, K and Ravussin, E 1999 Role of uncoupling proteins in energy balance. In *Regulation of food intake and energy expenditure*. Ed. Westerterp-Plantenga, MS, Steffens, AB and Tremblay, A. Milan: EDRA, pp. 415–28.
- Schroll, M, Jorgensen, L, Osler, M and Davidsen, M 1993 Chronic undernutrition and the aged. *Proceedings of the Nutrition Society* **52**, 29–37.
- Schutz, HG, Rucker, MH and Russell, GF 1975 Food and food-use classification systems. *Food Technology* **29** (3), 50–64.
- Schutz, Y, Flatt, JP and Jequier, E 1989 Failure of dietary fat intake to promote fat oxidation is a factor favouring the development of obesity. *American Journal of Clinical Nutrition* **50**, 307–14.
- Seidell, JC 1992 Regional obesity and health. *International Journal of Obesity* **16** (Supplement 2), S31–4.
- Shepherd, J, Cobbe, SM, Forde, I, et al. 1995 Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *New England Journal of Medicine* **333**, 130–7.
- Shils, ME, Olson, JA, Shike, M and Ross, AC 1999 *Modern nutrition in health and disease*, 9th edn. Philadelphia: Lippincott, Williams and Wilkins.
- Shipley, MJ, Pocock, SJ and Marmot, MG 1991 Does plasma cholesterol concentration predict mortality from coronary heart disease in elderly people? 18 year follow up in the Whitehall study. *British Medical Journal* **303**, 89–92.
- Simopoulos, AP 1991 Omega 3 fatty acids in health and disease and in growth and development. *American Journal of Clinical Nutrition* **54**, 438–63.
- Sims, EAH, Danforth, E Jr, Horton, ES, Bray, GA, Glennon, JA and Salans, LB 1973 Endocrine and metabolic effects of experimental obesity in man. *Recent Progress in Hormone Research* **29**, 457–96.
- Smith, GD, Song, F and Sheldon, IA 1993 Cholesterol lowering and mortality: the importance of considering initial level of risk. *British Medical Journal* **306**, 1367–73.
- Smith, GP, Jerome, C, Cushin, BJ, Eterno, T and Simansky, KJ 1981 Abdominal vagotomy blocks the satiety effect of cholecystokinin in the rat. *Science* **213**, 1036–7.
- Smith, R 1987 Osteoporosis: cause and management. *British Medical Journal* **294**, 329–32.
- Sobal, J and Stunkard, AJ 1989 Socioeconomic status and obesity: a review of the literature. *Psychological Bulletin* **105**, 260–75.
- Sorensen, TIA 1992 Genetic aspects of obesity. *International Journal of Obesity* **16** (Supplement 2), S27–9.
- Sorensen, TIA, Echwald, SM and Holm, J-C 1996 Leptin in obesity. *British Medical Journal* **313**, 953–4.
- Spector, TD, Cooper, C and Fenton Lewis, A 1990 Trends in admissions for hip fracture in England and Wales 1968–1985. *British Medical Journal* **300**, 1173–4.
- SSSS Group 1994 Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* **344**, 1383–9.
- Stampfer, MJ, Hennekens, CH, Manson, JE, Colditz, GA, Rosner, B and Willett, WC 1993 Vitamin E consumption and the risk of coronary disease in women. *New England Journal of Medicine* **328**, 1444–9.
- Steinberg, D, Parthasarathy, S, Carew, EE, Khoo, JC and Witztum, JC 1989 Beyond cholesterol: modifications of

- low-density lipoprotein that increase its atherogenicity. *New England Journal of Medicine* **320**, 915–24.
- Stephens, NG, Parsons, A, Schofield, PM, et al. 1996 Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). *Lancet* **347**, 781–6.
- Symonds, ME and Clarke, L 1996 Nutrition–environment interactions in pregnancy. *Nutrition Research Reviews* **9**, 135–48.
- Tartaglia, LA, Dembski, M, Weng, X, et al. 1995 Identification and expression cloning of a leptin receptor OB-R. *Cell* **83**, 1263–71.
- Thorogood, M 1995 The epidemiology of vegetarianism and health. *Nutrition Research Reviews* **8**, 179–92.
- Tisdale, MJ 1997 Isolation of a novel cachectic factor. *Proceedings of the Nutrition Society* **56**, 777–83.
- Tobian, L 1979 Dietary salt (sodium) and hypertension. *American Journal of Clinical Nutrition* **32**, 2659–62.
- Todd, E 1991 Epidemiology of foodborne illness: North America. In *Foodborne illness*. A *Lancet* review. Ed. Waites, WM and Arbuthnott, JP. London: Edward Arnold, pp. 9–15.
- Tompson, P, Salsbury, PA, Adams, C and Archer, DL 1991 US food legislation. In *Foodborne illness*. A *Lancet* review. Ed. Waites, WM and Arbuthnott, JP. London: Edward Arnold, pp. 38–43.
- Trowell, HC 1954 Kwashiorkor. *Scientific American* **191** (6), 46–50.
- Truswell, AS 1999 Dietary goals and guidelines: national and international perspectives. In *Modern nutrition in health and disease*. Ed. Shils ME, Olson, JA, Shike M and Ross, AC. Philadelphia: Lippincott, Williams and Wilkins, pp. 1727–42.
- UK National Case-Control Study Group 1993 Breast feeding and risk of breast cancer in young women. *British Medical Journal* **307**, 17–20.
- UN 1998 *United Nations 1998 revision of the world population estimates and projections*. www.popin.org/pop1998/4.htm
- USDA 1992 United States Department of Agriculture. *The food guide pyramid*. Home and Garden Bulletin Number 252. Washington, DC: United States Department of Agriculture.
- Wadden, TA and Stunkard, AJ 1985 Social and psychological consequences of obesity. *Annals of Internal Medicine* **103**, 1062–7.
- Wald, N, Idle, M, Boreham, J and Bailey, A 1980 Low serum-vitamin-A and subsequent risk of cancer. *Lancet* **ii**, 813–15.
- Walker, AF 1990 The contribution of weaning foods to protein-energy malnutrition. *Nutrition Research Reviews* **3**, 25–47.
- Waterlow, JC 1979 Childhood malnutrition – the global problem. *Proceedings of the Nutrition Society* **38**, 1–9.
- Waterlow, JC, Cravioto, J and Stephen, JML 1960 Protein malnutrition in man. *Advances in Protein Chemistry* **15**, 131–238.
- Waterlow, JC and Payne, PR 1975 The protein gap. *Nature* **258**, 113–17.
- Watts, G 1998 Avoiding action. *New Scientist*, 28 November, 53.
- Webb, GP 1989 The significance of protein in human nutrition. *Journal of Biological Education* **23** (2), 119–24.
- Webb, GP 1990 A selective critique of animal experiments in human-orientated biological research. *Journal of Biological Education* **24** (3), 191–7.
- Webb, GP 1992a A critical survey of methods used to investigate links between diet and disease. *Journal of Biological Education* **26** (4), 263–71.
- Webb, GP 1992b Viewpoint II: Small animals as models for studies on human nutrition. In *Nutrition and the consumer*. Ed. Walker, AF and Rolls, BA. London: Elsevier Applied Science, pp. 279–97.
- Webb, GP 1994 A survey of fifty years of dietary standards 1943–1993. *Journal of Biological Education* **28** (1), 101–8.
- Webb, GP 1995 Sleeping position and cot death: does health promotion always promote health? *Journal of Biological Education* **29** (4), 279–85.
- Webb, GP 1998 *Teach yourself weight control through diet and exercise*. London: Hodder and Stoughton.
- Webb, GP and Copeman, J 1996 *The nutrition of older adults*. London: Arnold.
- Webb, GP, Jagot, SA and Jakobson, ME 1982 Fasting induced torpor in *Mus musculus* and its implications in the use of murine models for human obesity studies. *Comparative Physiology and Biochemistry* **72A**, 211–19.
- Webb, GP and Jakobson, ME 1980 Body fat of mice and men: a class exercise in theory or practice. *Journal of Biological Education* **14** (4), 318–24.
- Wenkam, NS and Woolff RJ 1970 A half century of changing food habits among Japanese in Hawaii. *Journal of the American Dietetic Association* **57**, 29–32.
- Wharton, B (1998) Nutrition in infancy. *British Nutrition Foundation Nutrition Bulletin* **23**, Supplement 1.
- Wheeler, E 1992 What determines food choice, and what does food choice determine? *British Nutrition Foundation Nutrition Bulletin* **17** (Supplement 1), 65–73.
- Wheeler, E and Tan, SP 1983 From concept to practice: food behaviour of Chinese immigrants in London. *Ecology of Food and Nutrition* **13** (1), 51–7.
- White, A, Freeth, S and O'Brien, M 1992 *Infant feeding 1990*. London: HMSO.
- White, A, Nicolaas, G, Foster, K, Browne, F and Carey, S 1993 *Health Survey for England 1991*. London: HMSO.

- WHO 1990 *Diet, nutrition and the prevention of chronic diseases*. Geneva: World Health Organization.
- Wickham, CAC, Walsh, K, Barker, DJP, Margetts, BM, Morris, J and Bruce, SA 1989 Dietary calcium, physical activity, and risk of hip fracture: a prospective study. *British Medical Bulletin* **299**, 889–92.
- Willett, WC, Stampfer, MJ, Colditz, GA, Rosner, BA and Speizer, FE 1990 Relation of meat, fat and fiber intake to the risk of colon cancer in a prospective study among women. *New England Journal of Medicine* **323**, 1664–72.
- Willet, WC, Stampfer, MJ, Manson, JE, et al. 1993 Intake of trans fatty acids and risk of coronary heart disease among women. *Lancet* **341**, 581–5.
- Williams, CA and Qureshi, B 1988 Nutritional aspects of different dietary practices. In *Nutrition in the clinical management of disease*. Ed. Dickerson, JWT and Lee, HA. London: Edward Arnold, pp. 422–39.
- Wilson, JH 1994 Nutrition, physical activity and bone health in women. *Nutrition Research Reviews* **7**, 67–91.
- Woo, J, Ho, SC, Mak, YT, Law, LK and Cheung, A 1994 Nutritional status of elderly people during recovery from chest infection and the role of nutritional supplementation assessed by a prospective randomised single-blind control trial. *Age and Ageing* **23**, 40–8.
- Wortman, S 1976 Food and agriculture. *Scientific American* **235** (3), 31–9.
- WRI 1997 World Resources Institute. *Sustainable agriculture*. www.wri.org/rio-5/rio5Agri.html
- Yanovsky, SZ, Hubbard, VS, Lukaski, HC and Heymsfield, SB 1996 Bioelectrical impedance analysis in body composition measurement. Proceedings of a National Institutes of Health Technology Assessment conference in Bethesda, MD, December 12–14, 1994. *American Journal of Clinical Nutrition* **64** (Supplement), 387S–532S.
- Yeung, DL, Cheung, LWY and Sabrey, JH 1973 The hot–cold food concept in Chinese culture and its application in a Canadian-Chinese community. *Journal of the Canadian Dietetic Association* **34**, 197–203.
- Yu, S, Derr, J, Etherton, TD, and Kris Etherton, PM 1995 Plasma cholesterol-predictive equations demonstrate that stearic acid is neutral and monounsaturated fatty acids are hypocholesterolaemic. *American Journal of Clinical Nutrition* **61**, 1129–39.
- Zhang Y, Proenca, R, Maffei, M, Barone, M, Leopold, L and Friedman, JM 1994 Positional cloning of the mouse *obese* gene and its human homologue. *Nature* **372**, 425–32.
- Ziegler, RG 1991 Vegetables, fruits and carotenoids and the risk of cancer. *American Journal of Clinical Nutrition* **53**, 251S–9S.
- Zimmet, P, Hodge, A, Nicolson, M, et al. 1996 Serum leptin concentration, obesity, and insulin resistance in Western Samoa: cross sectional study. *British Medical Journal* **313**, 965–9.

Index

Page numbers in **bold** signify the major/most important reference to the topic.

- Acculturation 52, **55–8**, 100, 393–4
see also migration
- Activity diary 89
- Adequacy (dietary) 5–6, 129–30
- Adolescence *see childhood*
- Aerobic capacity *see fitness*
- Ageing, effects of 364
- Alcohol
- effects on health **122–4**, 347–8
 - intakes 123
 - recommendations 123
- Alcoholism
- nutritional consequences 142, 281, 290, 291
- Ames test 439
- Amino acids 231–5
- essential **234–5**, 238
 - limiting 235
 - see also protein*
- α -amylase *see carbohydrate, digestion*
- Amylose/amyllopectin *see starch*
- Anaemia
- macrocytic/megaloblastic *see folic acid/vitamin B₁₂ deficiency*
 - microcytic *see iron deficiency*
- Animal experiments
- as factor in protein gap theory 238, **241–2**
 - in determining human nutrient needs 75–6
 - in identifying dietary causes of disease 105–9
 - in food additive testing 107, 439–40
- Anorexia Nervosa *see eating disorders*
- Anthropometric assessment **75–86**, 385
- body mass index *see separate entry*
 - body density 83
 - children 84–5
 - mid arm circumferences 83–4
 - skinfold calipers 81–2
- Antioxidants 119, 226, 227, **298–306**
- see also beta-carotene, vitamins C and E, etc.*
- Appetite 190 et seq.
- Appetite centre/appetstat *see Hypothalamus and food intake regulation*
- Arachidonic acid *see fatty acids, essential*
- Artificial sweeteners 20, **214–15**, 438
- Ascorbic acid *see vitamin C*
- Asians (British) diet and health 98, 186, 225, 267, 393–5
- Assessment (nutritional) methods 60–91
- anthropometric 78–86
 - see also separate entry and body mass index*
 - biochemical 86–8
- clinical 76–8
- dietary 61–76
- Athletic performance and diet 400–3
- Attack rates *see food poisoning*
- ATP, adenosine triphosphate 136–8 et seq.
- Bacteria, growth requirements 416
- see also food poisoning*
- Balance (nutrient)
- use in setting dietary standards 73–4
 - see also under energy and protein*
- Behaviour therapy, obesity treatment 191
- Beriberi *see thiamin deficiency*
- 'Best before' date 409
- Bias (scientific) **14–15**, 110–11
- Bile acids 228, 352
 - entero-hepatic circulation 223, 229, 256
- Bioelectrical impedance 82–3
- Biotin 296
- Blind challenge 379
- Blood pressure *see hypertension and under salt*
- Body mass index, BMI **79–81**
- categories 79, 178–9
 - distribution in UK population 179–82
- Body water to measure fatness 83
- Bomb calorimeter 149
- Bone mass *see osteoporosis*
- Boutism 424
- see also Clostridium botulinum under food poisoning*
- Botulinum cook 424, 431
- Bovine spongiform encephalopathy, BSE/Creutzfeld Jacob Disease, CJD 425–9
- Bran, cereal **221**, 223
- see also non starch polysaccharide*
- Bread, as European cultural superfood **33–4**
- Breast milk composition 352–4
- see also infant feeding*
- Brown adipose tissue 138, **167–9**
- Bulimia (nervosa) *see eating disorders*
- Cafeteria feeding (rats) 167
- Calcium **126**, **317–27**, 333, 359–60
- balance, role of hormones/vitamin D **287**, **318**
 - pregnancy needs **319–20**, **344**
 - requirements/standards **74**, **318–20**
- Calorie *see energy, units*
- Cancer **98** et seq., 115, 122
- cachexia 162–3
 - dietary causes 103

- Cancer (contd)
see also salt/gastric cancer, non-starch polysaccharide/bowel cancer
- Canning 431
- Carbohydrate 209–29
 classification 210–11
 dietary sources 209, 211–12
 digestion 210, 219–21, 356
 loading (athletes) 402
 metabolism 138–40
 recommendations 116, 121, 210
see also individual fractions e.g. sugars
- β -carotene 15, 299, 302–4
 supplements, possible hazards 12–13, 302–4
 vitamin A precursor 67, 279, 302, 389
- Carotenoids 299, 302–4
see also β -carotene
- Casein:whey ratio (milk) 353
- Case-control studies *see under epidemiology*
- Cassava (toxins) 313, 435
- Catalase 299
- Centiles (for growth) 84–5
- Chemical hazards (in food) 434–40
 natural toxicants 434–6
 contaminants 436–7
see also food additives
- Chemical score *see protein, quality measurement*
- Childhood/adolescence 358–61
 effects of vegan diet 359
- Chloride *see salt*
- Cholecalciferol *see vitamin D*
- Cholecystokinin (CCK) 172, 175–6
- Cholesterol, dietary 9–10, 106, 107, 264–5
 oxides 265
- Cholesterol, serum/plasma levels
 distribution of UK values 261
 effects of dietary factors 9–10, 106, 261–5, 268–9
 high values and heart disease 7–8, 17–18, 256–7, 260–1, 265–72, 369–70
 risks of low values 19–20
- Chronic renal failure, dietary management 378–9
- Chylomicrons *see lipoproteins, plasma*
- Clinical trials 110–12
- Codex Alimentarius Commission 408
- Celiac disease 4, 357–8, 373–4
- Cohort studies *see under epidemiology*
- Cook chill processing 432
- Coenzymes 136
- Colostrum 354
- Confounding variables *see under epidemiology*
- Collagen 297
- Contaminants *sec chemical hazards*
- Copper 299
- Cori cycle 139
- Coronary heart disease *see diet heart hypothesis*
- Cost of 'healthy diet' 48–52
see also economics and food selection
- Cot death 10–11, 103
- Creatine in sport 402
- Cretinism *see iodine deficiency diseases*
- Creutzfeld Jacob Disease, CJD *see bovine spongiform encephalopathy*
- Cultural conservationist approach 20, 30
- Cultural relativism 30
- Cultural superfoods 33–5
- Culture as a learned phenomenon 55
- Cystic fibrosis 94, 281, 377–8
- db/db mouse 173 et seq.
- Death rate, calculations 93–6
- Deficiency diseases 3, 6
see also under minerals, vitamins and individual nutrients
- Dehydration 370, 402–3
- Delaney clause (food additives) 438
- Delayed-type cutaneous hypersensitivity (DTH) *see immunity and nutrition*
 effect of malnutrition
- Demiquet index *see anthropometric assessment*
- Dental health, effects of diet 8, 113, 215–19, 357
- Dermatitis Herpetiformis 373
see also coeliac disease
- Diabetes mellitus 7, 175, 184–5, 374–7
- Dietary change, effecting 20–6, 126–9
- Diet-heart hypothesis 109, 260–1, 265–72, 301–2
 supporting evidence 109–10, 266–72
- Dietary standards 4, 60, 68–76, 238
 comparison UK/US values 71
 estimating requirement (methods) 72–6
 UK DRVs and US RDAs 69–71
 Uses 71
see also under individual nutrient or population groups e.g. elderly
- Dietary Reference Values DRVs
see dietary standards UK
- 1,25 dihydroxycholecalciferol *see vitamin D*
- Diseases of industrialisation 7–8, 16, 115, 116, 130
 DMF, decayed, missing and filled teeth *see dental health*
- Double blind trial 111–12
- Doubly-labelled-water-method for estimating energy expenditure 90, 341
- Dual-centre hypothesis *see hypothalamus and food intake regulation*
- E-numbers (food additives) 438
- Eating disorders 159–62, 323
- Economics and food selection
see under food availability
- Eggs (as source of food poisoning) 413, 422
- Eicosanoids 245
 synthesis 254–5
 effect of fish oils on production 273
- Elderly 361–70
 demographic/social trends 361–4
 nutrition/health 364–70
- Elemental diets 371
- Energy in nutrition 143–205
 balance and its regulation 108–9, 164–77
 density 151–2, 388–9
 effect of fat 251–2
 role in obesity 9, 152
 weaning diets 152, 356–7
 dietary sources 149–51, 154–5
 expenditure, measurement 88–91
 falling intakes 6, 120, 194–5
 metabolisable 149
 requirements/dietary standards 145–8, 241, 365–6, 384
 units 145
 yields of macronutrients 149–50
- Endurance *see fitness measures*
- Enteral feeding 385
- Enterohepatic circulation *see bile acids*
- Enzymes 136–7

- Enzyme activation tests 74–5
EPA see fish oils
Epidemiology 91–105
 case-control studies 103–4
 cohort studies 101–3, 269–70
 confounding variables 15–16, 91–3
 cross cultural comparisons 96–8, 267
 migration studies 26, 55, 100–1
 time trends 98–100
Eskimos, diet and health 98, 272 et seq.
Essential nutrient, criteria for 4
Ethnocentrism and food 29–30
Evening primrose oil 255
Exercise and fitness 15, 268–9, 395–403
 current levels 120, 194–5, 396–8
 long-term benefits 102, 187, 267, 333, 399–400
 recommendations 120, 398–9
 role in obesity prevention/treatment 120, 187, 194 et seq., 399
 short-term benefits 395–6
Experiments, human 109–14, 268–9
see also animal experiments
- Factorial calculations**, to predict nutrient needs 74–5, 343 et seq.
- Familial hypercholesterolaemia** 105, 260, 269
- Family size**, effect on food spending 48
- Fat** 245–76
 as cause of bowel cancer 103, 226 et seq.
 as cause of obesity 152, 193–4
 intakes 16, 24, 125, 127, 251
 dietary functions 251–5
 dietary sources 250–1
 digestion and absorption 257–8
 effect on viscosity of weaning foods 253
 metabolism 140–1
 P:S ratio 9, 16, 109–10, 248–9, 251
 recommendations 9–10, 116, 120–1
 transport in blood *see lipoproteins plasma*
see also diet-heart hypothesis
- Fatty acids**
 types and distribution 247–9
- Fat cell theory (obesity)** 198–9
- Favism** 435
- FDA**, Food and Drug Administration (US) 438
- Ferritin** *see iron*
- Fibre** *see non starch polysaccharide*
- Fish oils** 113–14, 129, 272–6
- Fitness measures** of 395–6
see also exercise
- Flavoproteins**, FMN, FAD *see riboflavin*
- Fluoride** 113
 safe intakes/recommendations 126, 357
see also dental health
- Foetal alcohol syndrome FAS** 347–8
- Folic acid** 295–6
 deficiency and macrocytic anaemia 279, 295–6
 pregnancy, prevention of neural tube defects 112, 296, 345–7, 441
- Food additives** 437–41
 acceptable daily intake 441
 regulations 438
 safety testing 107, 439–41
 uses 437
- Food allergy** *see food intolerance*
- Food availability**, factors affecting 27–8, 38–59
 physical/geographical 39–40
 economic 6, 40–51
- cultural 51–8
 gatekeepers 58–9
- Food balance sheets** 61–2
- Food classification** 30–5
 anthropological 33–5
 biological 28–9, 30–1, 129–33
 consumer 31–3
- Food groups** 30–1, 129–30
- Food Guide Plate** 132–3
- Food Guide Pyramid** 130–2
- Food intake measurement** 61–5
- Food intolerance** 371–2
see also individual conditions e.g. coeliac disease
- Food labelling** 409–12
- Food poisoning, foodborne disease** 412–29
 causes 412–18
 incidence 412
 identifying source of outbreaks 420–2
 prevention 418–19
- Food poisoning organisms** 412–16, 422–5
Bacillus cereus 413 et seq., 423–4
Campylobacter 413 et seq., 422
Clostridium botulinum 424, 431
Clostridium perfringens 413 et seq., 423
Escherichia Coli 413 et seq., 423
See also verocytotoxin producing E. coli (VTEC)
Listeria monocytogenes 413, 424–5, 432
Salmonellae 413 et seq., 422–3
Staphylococcus aureus 413 et seq., 420–1, 423
Vibrio parahaemolyticus 413
- Food processing** 430–3
see also specific methods e.g. canning
- Food safety legislation** 407–9, 438
- Food selection** 20, 27–59
 model 36–9
see also food availability, factors affecting
- Food spoilage** 416, 432
- Food tables** 60, 65–8
- Fractional catabolic rate** 73 et seq.
- Free radicals and oxidative damage** 13, 267, 268, 297–301 et seq.
see also under individual antioxidants e.g. carotene and vitamin E
- Fruit and vegetables** 8, 15, 119, 302
- Functional foods** 265, 267, 430, 441–2
- Galactosaemia** 4, 212
- Galactose** *see lactose*
- Gatekeepers** *see under food availability*
- Gluconeogenesis** 141, 156
- Glucostat theory (food intake regulation)** 171, 172
- Glutathione peroxidase** 299
- Glutathione reductase** 75, 299
- Gluten** *see coeliac disease*
- Glycaemic index** 377
- Glycogen** 219, 402
- Glycolytic sequence** *see carbohydrate metabolism*
- Glycosylated proteins**, in diabetes 184, 375
- Goitre** *see iodine deficiency diseases*
- Gold thioglucose** 172
- GRAS**, generally recognised as safe (food additives) 438
- Guidelines (dietary)** 7, 115–135, 369–70
 economic and ecological implications 134
- Haemagglutinin** (in red kidney beans) 435
- Haemoglobin** 75, 86, 313 et seq.

- Hartnup disease 372
 HDL, high density lipoprotein *see lipoproteins, plasma*
 Health promotion 6–26
 criteria for intervention 13–20
 individual and population approach 17–19
 likelihood of inducing change 21–6, 27–8
 potential hazards 9, 13, 19–20, 23/7
 Heat stroke (in athletes) 403
 Hierarchy of human needs (Maslow) 36–8
 High temperature-short time (pasteurisation) 431
 Hormone replacement therapy, HRT 323, 325 et seq.
 Hospital malnutrition 379–86
 see also starvation
 Hot and cold classification of foods 31
 Hydroxyapatite (bone mineral) 320
 Hypernatraemic dehydration and infant formula 354
 Hypertension 7, 8, 184, 327–34
 see also under salt
 Hypothalamus and food intake regulation 169–77
 Immunity and nutrition 368, 381
 Incidence (disease) 94
 Infant feeding 349–58
 benefits of breastfeeding 352–5
 factors in choice of feeding method 53, 350–2
 milk composition, comparisons 76, 107, 241, 352–4
 prevalence of breastfeeding 349–50
 weaning 355–8
 Infective dose (of microorganisms) 414–15
 see also food poisoning/foodborne disease
 Internal/external hypothesis (of obesity) 190–1
 Intervention trials 112–14, 2/0–2
 Intrinsic factor *see vitamin B₁₂*
 Iodine (deficiency diseases/goitre) 78, 311–13
 goitrogens 313, 435
 Iron 126, 299, 313–17
 assessing iron status 75, 86, 314
 deficiency/microcytic anaemia 297, 314–17, 359, 360, 401
 pregnancy needs 12, 344–5
 toxicity 12, 314
 Irradiation of food 432–3
 Jewish dietary rules 36, 55
 Joule *see energy units*
 Kcal, kilocalorie *see energy units*
 Keshan disease *see selenium*
 Ketone bodies 136, 141, 157, 374
 Krebs' cycle 140–1
 Kwashiorkor *see Protein Energy Malnutrition*
 Lactation 348–9
 Lactic acid 139
Lactobacilli 353, 354, 441
 Lactose (intolerance) 212, 13
 Lathyrism 435
 Leptin 168, 173–7
 Life expectancy changes 7, 41–2, 361–2
 Lipase *see fat digestion*
 Lipoprotein lipase 257 et seq.
 Lipoproteins, serum/plasma 255–7, 258–60, 298
 LDL receptor 259–60
 See also cholesterol, plasma and diet heart hypothesis
 Lipostat theory (regulation of food intake) 171, 172–7
 Listeriosis *see food poisoning organisms, Listeria monocytogenes*
 Low birth weight 339 et seq.
 Low density lipoprotein (LDL) *see lipoproteins, plasma*
 Luxuskonsumption *see thermogenesis adaptive*
 MAFF (Ministry of agriculture, fisheries and food, UK) 408, 438
 Malnutrition *see starvation and hospital malnutrition*
 Marasmus *see Protein energy malnutrition*
 Meals on wheels 59
 Metabolic rate 88–91, 146–8
 comparison between individuals/species 90–1
 see also energy expenditure
 Metabolic response to injury 383–4
 Mid-arm circumferences *see anthropometry*
 Migration, nutritional aspects 55–8
 see also acculturation, epidemiology/migration studies
 Milk raw (as source of foodborne disease) 412, 423
 Minerals 299 et seq., 307–34
 intakes (UK) 310
 summary table 308–10
 see also under individual minerals
 Mutual supplementation (proteins) 236, 242
 Mycotoxins 436
 NAD, nicotinic adenine dinucleotide/NADP 138, 141, 277
 see also niacin/factions
 National Food Survey (UK) 44–6, 62–3, 150–1, 154–5, 211
 Net protein utilisation (NPU) *see protein quality measurement*
 Niacin vit. B₃ 292–3
 deficiency/pellagra 3, 41, 279, 280, 292–3
 dietary sources 279–80, 292
 functions 277, 292
 niacin equivalents 292
 requirements/dietary standards 292
 Night blindness *see vitamin A deficiency*
 Non milk extrinsic sugars 121, 213
 Non starch polysaccharide (NSP), dietary fibre 21, 119, 122–149,
 150, 221–9
 dietary sources 211, 221–2
 effects of 211, 222–3
 recommendations 116, 121, 211
 role in preventing bowel cancer/heart disease 225–9
 soluble/insoluble 221–2
 Normal distribution 69–70
 Nutrient density 152–4, 366
 Nutritional supplements 385
 ob/ob mouse 168–9, 173 et seq.
 Obesity and overweight 178–205, 388–9, 393, 395
 causes 9, 174–7, 188–96
 consequences 119–20, 182–8, 333, 369
 prevalence 179–82
 treatment/prevention 177, 196–205, 399
 Odds ratio 104–5
 Olestra 441–2
 Organic produce 436–7
 Osteomalacia *see vitamin D deficiency*
 Osteoporosis 302–6, 327
 β-oxidation *see fatty acid metabolism*
 Oxidative phosphorylation 137–8, 168–9
 Pantothenic acid 140, 296
 Parenteral feeding 254
 Pasteurisation 431
 Pellagra *see niacin deficiency*
 Pentose phosphate pathway 142, 290
 Pernicious anaemia *see vitamin B₁₂ deficiency*
 Phenylketonuria, PKU 4, 215, 372–3

- Phospholipids 245
Phyloquinone see vitamin K
 Physical activity level (PAL) 146–8, 201, 365
 Placebo 111
 Plant sterols 265, 441
 Population growth and world food supplies 41–4
 Potassium 126, 333
 Poverty, effects on food choices *see food availability/economic*
Pre-eclampsia (in pregnancy) 342
Pregnancy 338–48
 dietary standards 339–40 et seq.
 energy 341–3
 minerals 344–5
 protein 343–4
 vitamins 345–7
 folate supplements/neural tube defects 112, 296, 345–7
 iron supplements 12, 344–5
 species comparisons 338
Prevalence (disease) 94
Prions 425 et seq., 427
see also bovine spongiform encephalopathy
Prosthetic groups 137
Prostaglandins see eicosanoids
Protein 230–44, 339
 balance (nitrogen) 233–4
 content of foods 233–4
 digestion 230–1
 metabolism 141
 milk content, species differences 76, 241, 352
 quality 234–7
 requirements/standards 233, 235, 240–3, 343–4, 348
 significance in human nutrition, changes 11–12, 15, 237–44
Protein Energy Malnutrition 242–3
Prothrombin time 75
P:S ratio see under fatty acids
PWC_{max}, PWC₁₇₀ see aerobic capacity
Pyridoxine (vitamin B₆) 293–4
- Quality of life measurement 363–4**
- Racial minorities, diet/health 55–6, 392–5
see also Asians (British)
RDA recommended dietary allowances see dietary standards
Red kidney beans, toxicity 435
Red meat, masculine image 34
Reference nutrient intake (RNI) see dietary standards
Relative risk 102–3
Reliability 104–5
Resistant starch 233–5
Respiratory quotient (RQ) 89
Retinol see vitamin A
Retinol equivalents see vitamin A, units
Retinopathy of prematurity 300
Rhodopsin see vitamin A, functions
Riboflavin, vit. B₂ 75, 291–2, 299
Rice
 association with beriberi 290–1
 Japanese cultural superfood 33
Rickets see vitamin D deficiency
Risk factors 7–8, 22, 184–5
- Salt 20, 327–34
 association with gastric cancer 99–100, 327
 effects on blood pressure 16, 96–8, 111–12, 327 et seq., 330–4
 intake/requirements 328–30
- measuring intake 67–8, 329–30, 332
 recommendations 117, 121, 327–8
Satiety centre/signals see hypothalamus and regulation of feeding
Saxitoxin 436
Scrombotoxin 435
Scurvy see vitamin C deficiency
Selenium 299 et seq.
Sensory specific satiety 29, 192–3
Set point theory 166–7
Seventh Day Adventists, diet and health 98, 387 et seq.
Sitostanol see plant sterols
Skin sensitivity tests (allergy) 371
Sodium see salt
Solanine 434, 435
Soluble fibre see non starch polysaccharide soluble/insoluble
Specific activity 73
Standard mortality ratio (SMR) 94–6
Starches 219–21
Starvation 155–9
 adverse effects 5, 157–9, 161, 380–1, 382
 effects upon body composition 156
 responses/adaptations 156–7, 168–9
- Statins 271–2
Statistical significance 14
Subjective norm 24, 198
Sucrose 213–14
Sugar replacers 215
Sugars 212–14
 dietary sources 127, 131, 211
 recommendations 117, 121, 122
 relative sweetness 212
 sugar-fat-seesaw 128, 150, 193
see also dental health, effects of diet
Superoxide dismutase 299
Sushi (as cause of food poisoning) 413
- Taboos, dietary 52–5, 393**
- Taurine 352
Thermic effect of feeding 164
Thermodynamics, first law 164
Thermogenesis 109, 164
 adaptive 164, 167–9
Thiamin 289–91
deficiency/beriberi 3, 73, 279, 280, 290–1
functions 277, 290
requirements/dietary standards 290
sources 289–90
- Thyroid hormones see iodine*
α-tocopherol see vitamin E
Total parenteral nutrition (TPN) see parenteral nutrition
Transferrin 87, 314
Transamination 141, 235, 293
Transit time 222–3
Traveller's diarrhoea 414
Triacylglycerol (see fat)
Trichinosis 55
Triglyceride see fat
Tyramine 107, 435
- UHT, ultra high temperature 431**
'Use by' date 409
- Validity 106**
Vegetarianism 53, 98, 387–92
 adequacy considerations 359, 388–90
 health consequences 390–2, 393, 395

- Verocytotoxin-producing *E. coli* (VTEC) 413–14, 423
 Very low energy diets 204–5
 Vitamins 277–97
 circumstances that cause deficiencies 253, 280–1
 classification 278
 deficiency diseases 72, 73, 278–9
 human synthesis and precursors in food 67, 279–80
 summary tables 282, 283–4
 see also individual vitamins
 Vitamin A 282–7
 deficiency, night blindness, xerophthalmia 5, 73, 279, 285–6, 435
 functions 277, 284–5
 requirements/standards 74–5, 285
 role in cancer prevention/treatment 104
 sources 282–4
 toxicity and teratogenicity 286–7, 339, 347
 units, retinol equivalents 284
 see also β-carotene
 Vitamin B *see individual entries e.g. thiamin, niacin etc*
 Vitamin B₆ *see pyridoxine*
 Vitamin B₁₂ 279, 294–5
 absence from vegan diets 280–1, 294–5, 389–90, 393
 pernicious anaemia 4, 281
 Vitamin C 71 et seq. 280, 296–7
 deficiency/scurvy 3, 279, 288, 325–6
 high intakes benefits/dangers 305–6
- Vitamin D 287–8, 318 et seq.
 deficiency 3, 279, 288, 325–6
 rickets in British Asians 55, 288
 sunlight and skin production 280, 281, 287, 347, 365–6
 Vitamin E 288–9, 299 et seq.
 high intakes benefits/dangers 304–5
 Vitamin K 75, 289
 injection in newborn 279, 289
 warfarin as antimetabolite 75, 281, 289
- Waist:hip ratio 185–6
 Wall (endurance athletes) 402
 Warm holding, effects upon food 64–5, 67, 383
 Water activity (foods) 416
 Weaning *see infant feeding*
 Weight cycling 160, 186–7
 Wernicke-Korsakoff syndrome 142, 281, 290, 291
- Xerophthalmia *see vitamin A deficiency*
- Yoghurt 104, 267, 441
- Zinc 299 et seq.

