1.1 The lifespan approach to nutrition

The principal aim of this book is to explore relationships between nutrition and health and the contribution of nutrition-related factors to disease. In tackling this subject, there are many different approaches that could be taken, for example, considering diet and cardiovascular disease, nutrition and diabetes, obesity or immune function as separate and discrete entities, each worthy of their own chapter. The view of this author is that the final stages of life, that is, the elderly years, are effectively the products of events that occur through the full lifespan of an individual. Ageing is in actuality a continual, lifelong process of ongoing change and development from the moment of conception until the point of death. It is therefore inappropriate to consider how diet relates to chronic diseases that affect adults without allowance for how the earlier life experiences have shaped physiology. The lifespan approach that is used to organize the material in this book essentially asserts three main points:

1. All stages of life from the moment of conception through to the elderly years are associated with a series of specific requirements for nutrition.
2. The consequences of less than optimal nutrition at each stage of life will vary, according to the life stage affected.
3. The nature of nutrition-related factors at earlier stages of life will determine how individuals grow and develop. As a result, the relationship between diet and health in later stages of adult life, to some extent, depends upon events earlier in life. As a result, the nature of this relationship may be highly individual. Although we tend to divide the lifespan into a series of distinct stages, such as infancy, adolescence, early...
adulthood, middle age and older adulthood, few of these divisions have any real biological significance, and they are therefore simply markers of particular periods within a continuum. There are, however, key events within these life stages, such as weaning, the achievement of puberty or the menopause, which are significant milestones that mark profound physiological and endocrine changes and have implications for the nature of the nutrition and health relationship. On a continual basis, at each stage of life, individuals experience a series of biological challenges, such as infection, a change in the diet or exposure to carcinogens that threaten to disturb normal physiology and compromise health. Within a lifespan approach, it is implicit that the response of the system to each challenge will influence how the body responds at later life stages. Variation in the quality and quantity of nutrition is one of the major challenges to the maintenance of optimal physiological function and is also one of the main determinants of how the body responds to other insults.

In considering the contribution of nutrition-related factors to health and disease across the lifespan, it is necessary to evaluate the full range of influences upon quality and quantity of nutrition and upon physiological processes. This book therefore takes a broad approach and includes consideration of social or cultural influences on nutrition and health, the metabolic and biochemical basis of the diet–disease relationships and the influence of genetics and, where necessary, provides overviews of the main physiological and cellular processes that operate at each life stage. While the arbitrary distinctions of childhood, adolescence and adulthood have been used to divide the chapters, it is hoped that the reader will consider this work as a whole. For those requiring a primer in nutrition before engaging with specific chapters, the Appendix to the book describes the nutrients in simple terms. In this opening chapter, we consider some of the basic terms and definitions used in nutrition and lay the foundations for understanding more complex material in the following chapters.

1.2 The concept of balance

Balance is a term frequently used in nutrition, and, unfortunately, the precise meaning of the term may differ according to the context and the individual using it. It is common to hear the phrase ‘a balanced diet’, and, indeed, most health education literature that goes out to the general public urges the consumption of a diet that is ‘balanced’. In this context, we refer to a diet that provides neither too much nor too little of the nutrients and other components of food that are required for normal functioning of the body. A balanced diet may also be viewed as a diet providing foods of a varied nature, in proportions such that consumption of foods rich in some nutrients does not limit intakes of foods rich in others.

1.2.1 A supply and demand model

There is another way of viewing the meaning of balance or a balanced diet, whereby the relationship between nutrient intake and function is the main consideration. A diet that is in balance is one where the supply of nutrients is equal to the requirement of the body for those nutrients. Essentially, balance could be viewed as equivalent to an economic market, in which supply of goods or services needs to be sufficient to meet demands for those goods or services. Figure 1.1 summarizes the supply and demand model of nutritional balance.

Whether or not the diet is in balance will be a key determinant of the nutritional status of an individual. Nutritional status describes the state of a person’s health in relation to the nutrients in their diet and subsequently within their body. Good nutritional status would generally be associated with a dietary pattern that supplies nutrients at a level sufficient to meet requirements, without excessive storage. Poor nutritional status would generally (though not always) be associated with intakes that are insufficient to meet requirements.

The supply and demand model provides a useful framework for thinking about the relationship between diet and health. As shown in Figure 1.1, maintaining balance with respect to any given nutrient requires the supply of the nutrient to be equivalent to the overall demand for that nutrient. Demand comprises any physiological or metabolic process that utilizes the nutrient and may include use as an energy-releasing substrate, as an enzyme cofactor, as a structural component of tissues, as a substrate for the synthesis of macromolecules, as a transport element or as a component of cell–cell signalling apparatus. The supply side of the balance model comprises any means through which nutrients are made available to meet demand. This goes beyond delivery through food intake and includes stores of the nutrient that can be mobilized within the body and quantities of the nutrient that might be synthesized de novo (e.g. vitamin D is synthesized in the skin through the action of sunlight).

1.2.2 Overnutrition

When supply does not match demand for a nutrient, then the system is out of balance and this may have important consequences in terms of health and disease. Overnutrition (Figure 1.1) will generally arise because
the supply of a nutrient is excessive relative to demand. This is either because intake of foods containing that nutrient increases, because the individual consumes supplements of that nutrient or because demand for that nutrient declines with no equivalent adjustment occurring within the diet. The latter scenario particularly applies to the elderly, for whom energy requirements fall due to declining physical activity levels and resting metabolic rate (Rivlin, 2007). Commonly, intakes of energy that were appropriate in earlier adulthood will be maintained, resulting in excessive energy intake.

The consequences of overnutrition are generally not widely considered in the context of health and disease, unless the nutrient concerned is directly toxic or harmful when stored in high quantities. The obvious example here is, again, energy, where overnutrition will result in fat storage and obesity. For many nutrients, overnutrition within reasonable limits has no adverse effect as the excess material will either be stored or excreted. At megadoses, however, most nutrients have some capacity to cause harm. Accidental consumption of iron supplements or iron overload associated with inherited disorders is a cause of disease and death in children. At high doses, iron will impair oxidative phosphorylation and mitochondrial function, leading to cellular damage in the liver, heart, lungs and kidneys. Excess consumption of vitamin A has been linked to the development of birth defects in the unborn fetus (Martinez-Frias and Salvador, 1990). Vitamin D intoxication, for example, can arise due to overconsumption of supplements (Conti et al., 2014) and leads to the formation of kidney stones and neurological damage.

Overnutrition for one nutrient can also have effects upon nutritional status with respect to other nutrients and can impact on physiological processes involving a broader range of nutrients. For example, regular consumption of iron supplements can impact upon absorption of other metals such as zinc and copper by...
competing for gastrointestinal transporters and hence promote undernutrition with respect to those trace elements. Having an excess of a particular nutrient within the body can also promote undernutrition with respect to another by increasing the demand associated with processing the excess. For example, a diet rich in the amino acid methionine will tend to increase circulating and tissue concentrations of homocysteine. The processing of this damaging intermediate increases the demand for B vitamins, folic acid, vitamin B6 and vitamin B12, which are all involved in pathways that convert homocysteine to less harmful forms (Lonn et al., 2006; see Section “Folic acid and plasma homocysteine” of Chapter 8).

1.2.3 Undernutrition
Undernutrition arises when the supply of nutrient fails to meet demand. This can occur if intakes are poor or if demands are increased (Figure 1.1). In the short–medium term, low intakes are generally cushioned by the fact that the body has reserves of all nutrients that can be mobilized to meet demand. As such, for adults, it will usually require prolonged periods of low intake to have a significantly detrimental effect on nutritional status.

1.2.3.1 Increased demand
There are a number of situations that may arise to increase demand in such a way that undernutrition will arise if supply is not also increased accordingly. These include pregnancy, lactation and trauma. Trauma encompasses a wide range of physical insults to the body, including infection, bone fracture, burns, surgery and blood loss. Although diverse in nature, all of these physiological insults lead to the same metabolic response. This acute phase response (Table 1.1) is largely orchestrated by the cytokines including tumour necrosis factor-α (TNF-α), interleukin-6 (IL-6) and interleukin-1 (IL-1) (Grimble, 2001). Their net effect is to increase demand for protein and energy and yet paradoxically they have an anorectic effect. Thus, demand increases and supply will be impaired, together leading to protein–energy malnutrition. While in many developing countries, we associate protein–energy malnutrition with starvation in children, in developed countries such as the United Kingdom, protein–energy malnutrition is most commonly noted in surgical patients and patients recovering from major injuries (Wild et al., 2010).

1.2.3.2 The metabolic response to trauma
The human body is able to adapt rates of metabolism and the nature of metabolic processes to ensure survival in response to adverse circumstances. The metabolic response to adverse challenges will depend upon the nature of the challenge. Starvation leads to increased metabolic efficiency, which allows reserves of fat and protein to be utilized at a controlled rate that prolongs survival time and hence maximizes the chances of the starved individual regaining access to food. In contrast, the physiological response to trauma generates a hypermetabolic state in which reserves of fat and protein are rapidly mobilized in order to fend off infection and promote tissue repair (Table 1.1). Physiological stresses to the body, including infection, bone fracture, burns or other tissue injuries, elicit a common metabolic response regardless of their nature. Thus, a minor surgical procedure will produce the same pattern of metabolic response as a viral infection. It is the magnitude of the response that is variable, and this is largely determined by the severity of the trauma (Romijn, 2000).

The hypermetabolic response to trauma is driven by endocrine changes that promote the catabolism of protein and fat reserves. Following the initial physiological insult, there is an increase in circulating concentrations

<table>
<thead>
<tr>
<th>Acute phase response</th>
<th>Markers of the response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic change</td>
<td>Catabolism of protein, muscle wastage. Amino acids converted to glucose for energy or used to synthesize acute phase proteins. Catabolism of fat for energy</td>
</tr>
<tr>
<td>Fever</td>
<td>Body temperature rises to kill pathogens. Hypothalamic regulation of food intake disrupted, leading to loss of appetite</td>
</tr>
<tr>
<td>Hepatic protein synthesis</td>
<td>Acute phase proteins synthesized to combat infection (e.g. C-reactive protein, α1-proteinase inhibitor, caeruloplasmin). Liver reduces synthesis of other proteins, including transferrin and albumin</td>
</tr>
<tr>
<td>Sequestration of trace elements</td>
<td>Zinc and iron taken up by tissues to remove free elements that may be utilized by pathogens</td>
</tr>
<tr>
<td>Immune cell activation</td>
<td>B cells produce increased amounts of immunoglobulins. T cells release cytokines to orchestrate the inflammatory response</td>
</tr>
<tr>
<td>Cytokine production</td>
<td>Tumour necrosis factor-α and the interleukins 1, 2, 6, 8 and 10 work to produce a hypermetabolic state that favours production of substrates for immune function, but inhibits reproduction and spread of pathogens</td>
</tr>
</tbody>
</table>
of the catecholamines, cortisol and glucagon. Increased cortisol and glucagon serve to stimulate rates of gluconeogenesis and hepatic glucose output, thereby maintaining high concentrations of plasma glucose. The breakdown of protein to amino acids provides gluconeogenic substrates and also leads to greatly increased losses of nitrogen via the urine. Lipolysis is stimulated and circulating free fatty acid concentrations rise dramatically. These are used as energy substrates, along with glucose. The response to trauma is essentially an inflammatory process, and, as such, the same metabolic drives are noted in individuals suffering from long-term inflammatory diseases including cancer and inflammatory bowel disease (Richardson and Davidson, 2003). The inflammatory response serves two basic functions. Firstly, it activates the immune system, raises body temperature and repartitions micronutrients in order to create a hostile environment for invading pathogens (Table 1.1). Secondly, it allocates nutrients towards processes that will contribute to repair and healing.

The inflammatory response is orchestrated by the pro-inflammatory cytokines (e.g. TNF-α, IL-1 and IL-6) and the anti-inflammatory cytokines (e.g. IL-10). Whenever injury or infection occurs, the pro-inflammatory species are released by monocytes, macrophages and T helper cells. The level of cytokines produced is closely related to the severity of the trauma (Lenz et al., 2007). The impact of pro-inflammatory cytokines is complex. On the one hand, they activate the immune system and protect the body from greater trauma. On the other, at the local level of any injury, they increase damage by stimulating the immune system to release damaging oxidants and other agents that indiscriminately destroy invading pathogens and the body’s own cells. The production of pro-inflammatory cytokines therefore has to be counterbalanced as an excessive response can lead to death (Grimble, 2001). This is the role of the anti-inflammatory cytokines and some of the acute phase response proteins, several of which inhibit the proteinases released during inflammation and therefore limit the breakdown of host tissues.

In addition to stimulating proteolysis and lipolysis within muscle and adipose tissue, the cytokines have a number of actions that impact upon nutritional status. Firstly, they increase the basal metabolic rate. An element of creating a hostile environment for pathogens includes raising the core temperature of the body (fever). This greatly increases energy demands. The capacity to meet those demands through feeding is reduced as cytokines also act upon the gut and the centres of the hypothalamus that regulate appetite, effectively switching off the desire to eat. As can be seen in Table 1.2, the increased metabolic rate associated with the response to trauma greatly increases the demands of the body for both energy and protein. In severe cases, requirements can be doubled, even though the critically ill patient will be immobilized and not expending energy through physical activity. This can pose major challenges for clinicians managing such cases as the injured patient may be unable to feed normally, and due to the anorectic influences of pro-inflammatory cytokines, the patient may be unable to feed normally, and due to the anorectic influences of pro-inflammatory cytokines, the capacity to ingest sufficient energy, protein and other nutrients is greatly reduced. Enteral or parenteral feeding is therefore a mainstay of managing major injuries.

With more severe trauma, the mobilization of reserves can produce marked changes in body composition. Muscle wasting may occur as the calcium-dependent calpains and ubiquitin–proteasome break down proteins rapidly to make amino acids available for gluconeogenesis and the synthesis of important antioxidants such as glutathione (Grimble, 2001). Body composition changes are beneficial to the injured patient as they primarily generate glucose. This is the optimal energy substrate for these circumstances, not least because it can be metabolized anaerobically to produce ATP in tissues where blood flow may be compromised and oxygen delivery impaired.

In the short term, the hypermetabolic response and the accompanying anorexia of illness are unlikely to impact significantly upon the nutritional status of an individual, although nutritional status prior to onset of trauma would be an important consideration. For example, the nutritional consequences of a fractured femur in a young, fit adult male may be dramatically different to those in a frail elderly woman. Prolonged

### Table 1.2 The metabolic response to injury and infection increases requirements for energy and protein.

<table>
<thead>
<tr>
<th>Nature and severity of trauma</th>
<th>Increase in energy requirement (x basal)</th>
<th>Increase in protein requirement (x basal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor surgery or infection</td>
<td>1.1</td>
<td>1.0–1.5</td>
</tr>
<tr>
<td>Major surgery or moderate infection</td>
<td>1.3–1.4</td>
<td>1.5–2.3</td>
</tr>
<tr>
<td>Severe infection, multiple or head injuries</td>
<td>1.8</td>
<td>2.0–2.8</td>
</tr>
<tr>
<td>Burns (20% BSAB)</td>
<td>1.5</td>
<td>—</td>
</tr>
<tr>
<td>Burns (20–40% BSAB)</td>
<td>1.8</td>
<td>2.0–2.8*</td>
</tr>
</tbody>
</table>

*Dependent upon level of nitrogen losses in tissue exudates and age of patient. Children with burns have higher requirements. BSAB, body surface area burned.
periods of disease accompanied by inflammatory responses that drive hypermetabolism will promote states of protein–energy malnutrition, such as kwashiorkor, or can produce the emaciated state of cachexia. Cachexia is characterized by loss of weight, decline in appetite and muscle atrophy due to mobilization of muscle protein. It is generally associated with underlying chronic illnesses such as cancer, tuberculosis or untreated AIDS. Nutritional support (i.e. supplemental feeding) of chronically ill individuals or those who have suffered more acute trauma can limit the impact of the hypermetabolic response upon body composition and overall nutritional status. However, the catabolic metabolism cannot be reversed until the injury or illness is resolved, so the priority in these scenarios is limiting weight loss and loss of muscle mass, rather than achieving weight gain.

1.2.3.3 Compromised supply and deficiency

Clearly, there is a direct relationship between the supply of a nutrient to the body and the capacity of the body to carry out the physiological functions that depend upon the supply of that nutrient. As can be seen in Figure 1.2, the range of nutrient intakes over which optimal function is maintained is likely to be very broad, and there are a number of stages before functionality is lost. It is only when function can no longer be maintained that the term nutritional deficiency can be accurately used.

A nutrient deficiency arises when the supply of a nutrient through food intake is compromised to the extent that clinical or metabolic symptoms appear. The simplest example to think of here relates to iron deficiency anaemia in which low intakes of iron result in a failure to maintain effective concentrations of red blood cell haemoglobin, leading to compromised oxygen transport and hence the clinical symptoms of deficiency that include fatigue, irritability, dizziness, weakness and shortness of breath. Iron deficiency anaemia, like all deficiency disorders, reflects only the late stage of the process that begins with a failure of supply through intake to meet demands (Table 1.3). Once the body can no longer maintain function using nutrient supply directly from the diet, it will mobilize stores. In the case of iron, this will involve the release of iron bound to the protein, ferritin, to maintain haemoglobin concentrations. No change in function will occur at this stage, but the individual will now be in a state of greater vulnerability to deficiency. A further decline in supply through intake may not be matched through mobilization of stores, and so full deficiency becomes more likely. This situation in which intakes are sufficiently low that, although there are no signs of deficiency, biochemical indicators show that nutrition is subnormal is generally referred to as marginal nutrition, or subclinical malnutrition.

1.2.3.4 Malnutrition

Malnutrition describes the state where the level of nutrient supply has declined to the point of deficiency and normal physiological functions can no longer be maintained.
The manifestations of malnutrition will vary depending on the type of nutrient deficiencies involved and the stage of life of the malnourished individual. In adults, malnutrition is often observed as unintentional weight loss or as clinical signs of specific deficiency. In children, it is more likely to manifest as growth faltering, with the affected child being either underweight for their age (termed wasted) or of short stature for their age (termed stunted; see Section 6.2.2.1). Specific patterns of growth are indicative of different forms of protein–energy malnutrition. Wasting is associated with marasmus where a weight less than 60% of standard for age is used as a cut-off. Oedema with a weight less than 80% of standard for age is indicative of kwashiorkor.

From a clinical perspective, protein–energy malnutrition is the most serious undernutrition-related syndrome. Marasmus and kwashiorkor are the classical definitions of this form of malnutrition. Historically, marasmus was considered to be a pure energy deficiency and kwashiorkor to be protein deficiency, but it is now clear that the two are different manifestations of the same nutritional problems. Marasmic wasting is a sign of an effective physiological adaptation to long-term undernutrition. It is characterized by a depletion of fat reserves and muscle protein, along with adaptations to reduce energy expenditure. Children who become wasted in this way, if untreated, will generally die from infection as their immune functions cannot be maintained during the period of starvation. Kwashiorkor is a more rapid process, often triggered by infection alongside malnutrition. The metabolic changes with kwashiorkor are strikingly different to marasmus as the adaptation to starvation is ineffective. Fat accumulates in the liver, and expansion of extracellular fluid volume, driven by low serum albumin concentrations, leads to oedema. Micronutrient deficiencies often occur alongside protein–energy malnutrition and may partly explain why individuals with kwashiorkor, unlike those with marasmus, are unable to adapt successfully to malnutrition.

The causes of malnutrition are complex and are not simply related to a limited food intake. Where intake is reduced, this is often due to food insecurity associated with famine, poverty, war or natural disasters. Reduced food intake can also arise due to chronic illness leading to loss of appetite or feeding difficulties. Malnutrition will also arise from malabsorption of nutrients from the digestive tract. This, again, could be a consequence of chronic disease or be driven by infection of the tract. Losses of nutrients are an important consequence of repeated diarrhoeal infections in areas where there is no access to clean water and adequate sanitation. Malnutrition may also be driven by situations that increase the demand for nutrients including trauma (as described earlier), pregnancy and lactation, if those increased demands cannot be matched by intake.

Malnutrition is most common and most deadly in the developing countries, where it is the major cause of death in children. Stunting and wasting among malnourished children have long-term consequences too, as often the reduction in stature is not recovered, leading to reduced physical strength and capacity to work in adult life. As poverty is the most frequent cause of malnutrition, a self-perpetuating cycle can be established, as the stunted child becomes the adult with reduced earning capacity, whose children will live in poverty. Stunted, underweight women will also have children who are at risk due to lower weight at birth. Pregnancy is a time of high risk for malnutrition in women living in developing countries. Stunting is commonplace among women in South and Southeast Asia and is often accompanied by underweight. For example, in India and Bangladesh, up to 40% of women of childbearing age have a body mass index (BMI) of less than 18.5 kg/m² (Black et al., 2008). Iron deficiency anaemia is endemic among pregnant women in developing countries, with prevalence of between 60 and 87% in the countries of southern Asia (Seshadri, 2001). Maternal and childhood malnutrition are believed to cause 3.1 million deaths among the under-fives every year (Black et al., 2013).

Developed countries also have a burden of malnutrition among vulnerable groups. At greatest risk are the elderly, who may develop protein–energy malnutrition or micronutrient deficiencies due to specific medical

<table>
<thead>
<tr>
<th>Stage</th>
<th>Biochemical indicators and reference ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal iron status</td>
<td>Haemoglobin 14–18 g/dl (men) and 12–16 g/dl (women), serum ferritin 40–280 µg/l, transferrin saturation 31–60%</td>
</tr>
<tr>
<td>Depleted iron stores</td>
<td>Falling serum ferritin. Normal ranges for haemoglobin and transferrin saturation. Ferritin 13–20 µg/l</td>
</tr>
<tr>
<td>Iron deficiency</td>
<td>Transferrin saturation falls as transport of iron declines. Haemoglobin normal. Serum ferritin &lt;12 µg/l. Transferrin saturation &lt;16%</td>
</tr>
<tr>
<td>Iron deficiency anaemia</td>
<td>Haemoglobin synthesis cannot be maintained and declines to &lt;13.5 g/dl (men) and &lt;12 g/dl (women). Serum ferritin &lt;10 µg/l. Transferrin saturation &lt;15%</td>
</tr>
</tbody>
</table>
conditions or through low intakes associated with frailty or loneliness (see Section 9.5.1). Surgical patients are at risk of protein–energy malnutrition as a result of the inflammatory response to trauma. As in the developing countries, poverty increases the risk of malnutrition among children and immigrant groups. There are many ways of targeting these at-risk groups, for example, monitoring the growth of infants or including regular weighing and nutritional assessments of hospital patients. Malnutrition is easily treated through appropriate nutritional support.

The prevention of malnutrition is a major public health priority on a global scale. While a lack of food security and the risk of protein–energy malnutrition remain a major issue for many populations, there have been a number of success stories in the battle to prevent clinically significant malnutrition. The basic approaches that can be used to prevent nutrient deficiency are diet diversification, supplementation of at-risk individuals and fortification. The basis for these approaches and their use in the attempt to eradicate vitamin A deficiency is described in Research Highlight 1.1. Similar strategies have been used to reduce the occurrence of iodine and iron deficiency diseases.

Iodine deficiency is an important issue for populations in all continents except Australasia. Availability of iodine is essentially limited by the iodine content of the soil and hence uptake by plants and animals. Iodine deficiency disorders, including cretinism and goitre, are a major manifestation of malnutrition, with approximately 740 million affected individuals worldwide. Fortification has been the cornerstone of the fight against iodine deficiency, with the Universal Salt Iodization (USI) programme providing iodized salt (20–40 mg iodine/kg salt) to 70% of households in affected areas. Where the iodized salt is consumed, marked improvements in iodine status of the population are rapidly noted (Sebotsa et al., 2005). Although there are still significant numbers of individuals at risk of iodine deficiency disorders, due to lack of coverage of the USI programme, this fortification approach is widely considered to be a public health nutrition success for the World Health Organization (WHO).

### 1.2.4 Classical balance studies

Nutritional status with respect to a specific nutrient can be measured using balance studies. These have classically been used to determine requirements for some

---

**Research Highlight 1.1 Strategies for combating vitamin A deficiency.**

Vitamin A deficiency (VAD) is one of the most common forms of malnutrition on a global scale (West, 2003), with greatest prevalence in Africa, Central and South America and South and Southeast Asia. Subclinical VAD blights the lives of up to 200 million children every year and is a causal factor in up to half a million cases of childhood blindness and up to a million deaths of children under the age of 5 years. VAD is also responsible for stunted growth in children and may cause blindness in women with increased demands for vitamin A, due to pregnancy or lactation. In 1990, the World Health Organization pledged itself to the virtual elimination of VAD by the year 2000. The strategies used to achieve this goal provide useful examples of how all common nutrient deficiencies might be prevented at a population level. Three main approaches have been used to tackle VAD:

1. **Diet diversification.** For many populations in areas where VAD is common, the range of staple foods consumed is very limited. For example, rice is the basis of most meals for many in southeast Asia. Rice is a poor vitamin A source. Diversification programmes include health education and promotion of consumption of a greater range of foodstuffs and the development of home gardening to provide vitamin A sources. Faber et al. (2002) showed that a home gardening programme in South Africa increased knowledge and awareness of VAD, improved availability of vitamin A sources and increased serum retinol concentrations in young children.

2. **Supplementation.** In most countries where VAD is common, children are now supplemented with vitamin A, using an oil capsule, two or three times a year, often coupling supplement doses with other public health activities such as immunizations. Berger et al. (2008) highlighted the major disadvantage of supplementation, which is that it fails to reach all those in need of supplements. For VAD, those most at risk are preschool children who have less access to school-based supplementation programmes. Often, the poor and those most in need of supplements are least likely to receive them. Supplementation is expensive, which may reduce efficacy of the approach in impoverished countries (Neidecker-Gonzales et al., 2007).

3. **Fortification.** Fortification involves the addition of nutrients to staple foods at the point of their production, thereby increasing the amount of nutrient delivered to all consumers of that foodstuff. VAD in several countries has been tackled using this strategy. Red palm oil is widely available in many VAD-affected areas and is a rich source of β-carotene. In India and parts of Africa, the addition of this oil to other oils traditionally used in cooking, and to snacks, has been shown to effectively increase vitamin A intake by the general population (Sarojini et al., 1999). Zagré et al. (2003) showed that introducing red palm oil to a population in Burkina Faso was highly effective in reducing occurrence of VAD. A similar approach involves increasing the vitamin A content of crops such as rice, either through genetic modification (e.g. ‘golden rice’) or traditional plant breeding (Mayer, 2007).
nutrients in humans. Essentially, the balance method involves the accurate measurement of nutrient intake for comparison with accurate measures of all possible outputs of that nutrient via the urine, faeces and other potential routes of loss (Figure 1.3). If there is a state of balance, that is, intake and output are at equilibrium, it can be assumed that the body is saturated with respect to that nutrient and has no need for either uptake or storage. This technique can be applied to almost any nutrient, and by repeating balance measures at different levels of intake, it is possible to determine estimates of requirements for specific nutrients. The balance model works on the assumption that in healthy individuals of stable weight, the body pool of a nutrient will remain constant. Day-to-day variation in intake can be compensated by equivalent variation in excretion. The highest level of intake at which balance can no longer be maintained will indicate the actual requirement of an individual for that nutrient.

Nitrogen balance studies were used to determine human requirements for protein (Millward et al., 1997). Such studies involved experiments in which healthy subjects were recruited and allocated to consume dietary protein at specified levels of intake. After 4–6 days of habituation to these diets, urine and faeces were collected for determination of nitrogen losses over periods of 2–3 days. On this basis, it was possible to state dietary protein requirements for different stages of life as being the lowest level of protein intake that maintained nitrogen balance in healthy individuals, maintaining body weight and engaging in modest levels of physical activity. Nitrogen balance studies are problematic in several respects, including the fact that 24-h urine collections used in such studies are often incomplete, because studies may fail to allow sufficient time for subjects to habituate to their experimental diet and because factors such as unobserved infection, stress or exercise may increase demand for protein. It has also been impossible to use balance studies to examine protein requirements for all age groups and in all health situations, so requirements for pregnant and lactating women and for children are based on balance studies in young adults and make estimates of allowances for tissue deposition, growth and milk synthesis and secretion.

1.2.5 Overall nutritional status

The diet delivers a multitude of components rather than single nutrients, and it is unlikely that any individual will have a diet that perfectly achieves balance for all of them. For example, an individual can be in balance for protein while consuming more energy than is required and insufficient iron to meet demand. Hence, it is often not appropriate to discuss overall nutritional status of an individual without consideration of nutritional status with respect to specific nutrients.

Whether considering the overall nutritional status of an individual, examining nutritional status with respect to a specific nutrient or investigating the nutritional status of a population, it is important to take into account a broad range of factors. It should be clear from the previous discussions that intake is just one component of the supply side of the balance model. Nutritional status is only partly determined by the food that is being consumed. Nutritional status also depends upon the activities and health status of the individuals concerned. Trauma and high levels of physical activity will increase demand, while a sedentary lifestyle will decrease demand. Most important though is the stage of life of the individuals under consideration. Physiological demands for nutrients vary to a wide degree, depending on age,

---

**Figure 1.3** Determining nutrient requirements using the balance method. Precise measurements of nutrient intake and of output by all possible routes enable determination of nutrient requirements. The highest level of intake at which balance can no longer be maintained will indicate the actual requirement of an individual for that nutrient.
body size and gender. The impact of variation within the diet upon health and well-being is largely, therefore, governed by age and sex.

1.3 The individual response to nutrition

The ingestion of food and nutrients is the beginning of a chain of events involving digestion and absorption, metabolic processing and physiological responses. For example, consumption of carbohydrate brings about an increase in blood glucose concentrations followed by insulin secretion and the uptake of glucose into cells and tissues for use in energy metabolism that will drive muscle contraction, thermogenesis and a host of other processes. Ingestion of sodium will result in haemodynamic changes that impact upon blood pressure and kidney function. The nature of the metabolic and physiological changes that occur following ingestion means that the response to nutrients plays a fundamental role in determining health and disease. All individuals have unique characteristics that shape the nature of the response to nutrition. This poses a problem in establishing guidelines for health nutrition as the desire is to provide general guidance on a ‘one-size-fits-all’ basis, but the reality is that some guidance will be inappropriate for a significant proportion of any population. Many factors will contribute to the individual response to nutrition (Figure 1.4), but genetic background and lifespan-related factors are of major significance.

1.3.1 Stage of the lifespan

Nutritional status is determined by the balance between the supply of nutrients and the demand for those nutrients in physiological and metabolic processes. So far in this chapter, we have seen that both sides of the supply–demand balance equation can be perturbed by a variety of different factors. Intake, for example, can be reduced in circumstances of poverty, while demand is elevated by physiological trauma. The main determinants of demand are, however, shaped by other factors such as the level of habitual physical activity (which will increase energy requirements), by gender, by body size and by age. It is this latter factor that provides the focus of this book.

The demand for nutrients to sustain function begins from the moment of conception. The embryonic and fetal stages of life are the least understood in terms of the precise requirements for nutrition, but it is clear that they are the life stages that are most vulnerable in the face of any imbalance. Demands for nutrients are high in order to sustain the rapid growth and the process of development from a single-celled zygote to a fully formed human infant. An optimal balance of nutrients is essential, but the nature of what is truly optimal is difficult to dissect out from the competing demands of the maternal system and the capacity of the maternal system to deliver nutrients to the fetus. The embryo and fetus represent a unique life stage from a nutritional perspective, as there are no nutrient reserves and there is a total dependence upon delivery of nutrients, initially by the yolk sac and later by the placenta. The consequences of undernutrition at this stage can be catastrophic, leading
to miscarriage, failure of growth, premature birth, low weight at birth or birth defects (MRC Vitamin Study Group, 1991; Godfrey et al., 1996; El-Bastawissi et al., 2007). All of these are immediate threats to survival, but it is also becoming clear that less than optimal nutrition at this stage of life may also increase risk of disease later on in life (Langley-Evans, 2015).

After birth, the newborn infant has incredibly high nutrient demands that, in proportion to body weight, may be two to three times greater than those of an adult. These demands are again related to growth and the maturation of organ systems as in fetal life. Growth rates in the first year of life are more rapid than at any other time, and the maturation of organs such as the brain and lung continues for the first 3–8 years of life. Initially, the demands for nutrients are met by a single food source, milk, with reserves accrued from the mother towards the end of fetal life compensating for any shortfall in supply of micronutrients. In later infancy, there is the challenge of the transition to a mixed diet of solids (weaning), which is a key stage of physiological and metabolic development. The consequences of imbalances in nutrition can be severe. Infants are very vulnerable to protein–energy malnutrition and to micronutrient deficiencies, which will contribute to stunting of growth and other disorders. Iodine deficiency disorders and iron deficiency anaemia can both impact upon brain development, producing irreversible impairment of the capacity to learn. Obesity is now recognized as a major threat to the health of children in the developed countries. In this age group, it is not simply a product of excessive energy intake and low energy expenditure. Increasingly, we are seeing that the type or form of foods consumed at this time can influence long-term weight gain, with breastfed infants showing a lower propensity for obesity than those who are fed artificial formula milks (Arenz et al., 2004; Bayol et al., 2007).

Beyond infancy, nutrient demands begin to fall relative to body weight but still remain higher than seen in adulthood through the requirement for growth and maturation. These demands are at their greatest at the time of puberty when the adolescent growth spurt produces a dramatic increase in height and weight that is accompanied by a realignment of body composition. Proportions of body fat decline and patterns of fat deposition are altered in response to the metabolic influences of the sex hormones. Proportions of muscle increase and the skeleton increases in size and degree of mineralization. Nutrient supply must be of high quality to drive these processes, and in absolute terms (i.e. not considered in proportion to body size), the nutrient requirements of adolescence are the greatest of any life stage. However, adolescents normally have extensive nutrient stores and are therefore more tolerant of periods of undernutrition than preschool children (1–5 years).

The adult years have the lowest nutrient demands of any stage of life. As growth is complete, nutrients are required solely for the maintenance of physiological functions. The supply is well buffered through stores that protect those functions against adverse effects of undernutrition in the short term to medium term. In developed countries, and increasingly so in developing countries, the main nutritional threat is overweight and obesity, as it is difficult for adults to adjust energy intakes against declining physiological requirements and the usual fall in levels of physical activity that accompany ageing. Reducing energy intake, while maintaining adequate intakes (AI) of micronutrients, is a major challenge in elderly individuals. Chronic illnesses associated with ageing can promote undernutrition through increased nutrient demands, while limiting appetite and nutrient bioavailability.

For women, pregnancy and lactation represent special circumstances that may punctuate the adult years and increase demands for energy and nutrients. Nutrition is in itself an important determinant of fertility and the ability to reproduce (Hassan and Killick, 2004). In pregnancy, provision of nutrients must be increased for the growth and development of the fetus and to drive the deposition of maternal tissues. For example, there are requirements for an increase in size of the uterus, for the preparation of the breasts for lactation and for the formation of the placenta. To some extent, the mobilization of stores and adaptations that increase absorption of nutrients from the gut serve to meet these increased demands, but as described earlier, imbalances in nutrition may adversely impact upon the outcome of pregnancy. Lactation is incredibly demanding in terms of the energy, protein and micronutrient provision to the infant via the milk. As with pregnancy, not all of the increase in supply for this process depends upon increased maternal intakes, and in fact, women can successfully maintain lactation even with subclinical malnutrition. Adaptations that support and maintain breastfeeding may impact upon maternal health. For example, calcium requirements for lactation may be met by mobilization of bone mineral and, if not replaced once lactation has ceased, could influence later bone health. However, although nutritionally challenging, most evidence suggests that lactation is of benefit for maternal health and actually contributes to reduced risk of certain cancers and osteoporosis (Ritchie et al., 1998; Danforth et al., 2007).

Lifespan factors clearly impact upon nutritional status as they are a key determinant of both nutrient requirements and the processes that determine nutrient supply.
In studying relationships between diet, health and disease, one of the major challenges is to assess the quality of nutrition in individuals and at the population level. Tools used for these nutritional assessments will be described in the next section.

1.3.2 Genetics

Long-term disease states such as obesity, cancer or coronary heart disease are the products of a number of risk factors, working together, against a battery of protective factors. Disease is promoted by a poor diet, smoking, sedentary lifestyle and accumulated experiences across the lifespan. These factors all overlay the genetic background of the individual to determine how the body responds to nutrition and other environmental factors. The genotype of each individual comprises a complex set of traits that might be disease-promoting (susceptibility genes) or disease-suppressing (protective genes). For most disease states, more than one gene will be driving the components of risk.

Due to the complexity of the genetic determinants of physiological function, individuals will respond to nutritional signals in different ways. For example, some individuals will have a genetic make-up that promotes high energy expenditure. This enables them to maintain a healthy weight at a level of energy intake that is sufficient to promote obesity in other individuals, who may instead carry obesity-promoting genes.

Some of the risk of chronic disease is determined by single nucleotide polymorphisms (SNPs) (pronounced ‘snips’), which are variants in the sequences of genes which control specific aspects of physiological and metabolic function (Joost et al., 2007). SNPs are inherited sequences which differ from the most common sequence of a gene by just one base (e.g. a C replaced by a T; Figure 1.5). Such a change may generate a protein product of dramatically altered tertiary structure and impact significantly upon physiological function. SNPs are now well characterized as having interactions with components of the diet, and some detailed examples will be discussed in later chapters.

The C677T SNP in methylenetetrahydrofolate reductase (MTHFR) is one of the best studied examples of a genetic variant that impacts upon the variability of the individual response to diet. MTHFR is an enzyme that plays an important role in the metabolism of folates and effectively controls the availability of one-carbon donors in intermediary metabolism. Within the population, there will be three distinct populations based upon variants of C677T, namely, individuals carrying CC, CT or TT genotypes. For those carrying TT, circulating concentrations of the amino acid homocysteine will tend to be higher as activity of the enzyme is lower than with the CC variant (Figure 1.6). Hyperhomocystaemia is a known risk factor for cardiovascular disease, unless the diet delivers sufficient folate to offset this risk.

The contribution of a SNP to risk of disease should not be overestimated. Often, the influence of SNPs is miniscule compared to the impact of lifestyle factors. For example, a variant of the calpain-10 gene is associated

---

**Figure 1.5** Single nucleotide polymorphisms (SNPs) arise when there are single base changes in the DNA sequence of a gene. As all individuals carry two copies of a gene, the polymorphism can result in individuals being homozygous or heterozygous for specific gene variants.
with a 20% increase in risk of type 2 diabetes, which is dwarfed by the 4- to 30-fold risk that is associated with obesity (Joost et al., 2007). It should also be borne in mind that other genetically determined factors may modulate the influence of the SNP. In the case of the C677T polymorphism of MTHFR, the influence on disease risk varies between ethnic groups (Klerk et al., 2002). Some SNPs may increase risk of one chronic disease yet protect against another. Cardiovascular disease-prone TT MTHFR variant-carrying individuals appear to have some protection against cancers of the large intestine (Joost et al., 2007).

Understanding all of these processes is important. Throughout this book, you will encounter examples of uncertainty and variation in the effectiveness of nutritional interventions to prevent or manage disease. This is because the impact of diet upon physiological and metabolic processes will vary due to the genetic variation in the population and because even straightforward nutrient–gene interactions may generate opposing responses to nutrient signals under different circumstances. Some polymorphisms related to disease risk sometimes appear to have influences that are specific to gender, menopausal status or race. It is also the case that the expressed phenotype may be life stage specific. The Bsm1 SNP in the vitamin D receptor is strongly predictive of bone health and the response to calcium in children, but has no association with bone health in adults (Ferrari et al., 1998). It is therefore clear that the genotype can be an important determinant of disease risk. However, the instances of where it is the sole or major driver of risk are rare, and in most cases, genetic inheritance is just one of the many components that determine the overall risk profile for chronic disease. Research Highlight 1.2 describes one well-known gene polymorphism and its relationship to disease.

1.4 Assessment of nutritional status

The assessment of nutritional status is necessary in a variety of different settings. Working with individuals in a clinical setting, it may be necessary to assess dietary adequacy in order to plan the management of disease states, or to make clinical diagnoses. Public health nutritionists require data on dietary adequacy at a group level in order to make assessments of the contribution of nutritional factors to disease risk in the population and to develop public health policies or intervention strategies. Nutritional assessment is also a critical research tool used in determining the relationships between diet and disease. These situations, which rely on considerations of the likelihood of nutritional deficit or excess at the individual or population level, use tools that aim either to measure intakes of nutrients or the physiological manifestations of nutrient deficit or excess within the body.
Tools for nutritional assessment include anthropometric measures, dietary assessments, determination of biomarkers and clinical examination.

### 1.4.1 Anthropometric measures

Anthropometric methods make indirect measurements of the nutritional status of individuals and groups of individuals, as they are designed to estimate the composition of the body. Table 1.4 provides a summary of the commonly used anthropometric techniques. Many of these have been designed to estimate the lean or fat mass which are present within the body. Information about relative fatness or leanness can be a useful indicator of nutritional status since excess fat will highlight storage of energy consumed in excess, while declining fat stores and loss of muscle mass are indicative of malnutrition. Extremes within anthropometric measures, for example, the emaciation of cachexia or morbid obesity, are useful indicators of disease risk or progression in a clinical setting. In children, serial measures of height and weight can provide sensitive measures of growth and development that can be used to highlight and monitor nutritional problems. The most robust anthropometric measures are challenging as they require specialist equipment. As a result, most surveys and research projects that examine large groups of people use simply determined measures such as BMI (weight in kg divided by height in metres²). As shown in Figure 1.7, BMI is widely used as a measure of body fatness and to classify overweight and obesity, but it is a non-specific measure that can be misinterpreted (see Sections 6.4 and 8.4.1).

### 1.4.2 Estimating dietary intakes

Estimation of dietary intakes, either to determine intakes of specific macro- or micronutrients or to assess intakes of particular foods, is a mainstay of human nutrition research. A range of different methods are applied, depending on the level of detail required. All approaches are highly prone to measurement error.

#### 1.4.2.1 Indirect measures

The least accurate measures of intake are those that make indirect estimates of the quantities of foodstuffs consumed by populations. These techniques are used to follow trends in consumption between national populations or within a national population over a period of time.

Food balance sheets are widely used by the United Nations Food and Agriculture Organization (FAO) to monitor the availability of foods, and hence nutrients.
within most nations of the world and are published on an annual basis. They allow temporal trends to be monitored easily and apply a standardized methodology on a global scale. A food balance sheet is essentially compiled from government records of the total production, imports and exports of specific foodstuffs. This allows the quantity of that foodstuff available to the population to be calculated (available food = production + imports – exports). Dividing that figure by the total number of people in the population allows the daily availability per capita to be estimated. Figure 1.8 shows data abstracted from the 2004 FAO food balance sheets, indicating how daily availability of protein from plant and animal sources varies with different regions of the world. Food balance sheets are subject to considerable error due to assumptions that are made in their compilation. It will be assumed that the nutrient composition of a food will be the same regardless of where it is produced,

Table 1.4 Anthropometric measures used to estimate body composition and nutritional status.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Component of body composition estimated</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body mass index (weight/height²)</td>
<td>Weight relative to height</td>
<td>Does not distinguish between lean and fat mass. Does not measure the composition of the body</td>
</tr>
<tr>
<td>Skinfold thicknesses</td>
<td>Fat mass</td>
<td>Requires skill in measurement. Makes assumptions about the even distribution of fat in the subcutaneous layer</td>
</tr>
<tr>
<td>Waist circumference or waist/hip ratio</td>
<td>Fat distribution</td>
<td>A good indicator of abdominal fat deposition. Requires set protocols for measurement</td>
</tr>
<tr>
<td>Mid-upper arm circumference</td>
<td>Muscle mass</td>
<td>Prone to measurement error. Unsuitable for some groups (e.g. adolescents) with rapidly changing fat and muscle patterns. Good indicator of acute malnutrition</td>
</tr>
<tr>
<td>Bioimpedance</td>
<td>Fat mass</td>
<td>Influenced by hydration status of subjects</td>
</tr>
<tr>
<td>Underwater weighing</td>
<td>Body density, fat and lean mass</td>
<td>Requires subjects to undergo training for an unpleasant procedure. Underestimates fat mass in muscular individuals</td>
</tr>
<tr>
<td>Isotope dilution</td>
<td>Body water</td>
<td>Influenced by fluid intake of subject. Analytically difficult and expensive</td>
</tr>
<tr>
<td>Scanning techniques (NMR, computed tomography, DXA)</td>
<td>Proportions and distribution of lean and fat mass</td>
<td>Expensive, restricted access to scanners. Use ionizing radiation so unsuitable for children and pregnant women</td>
</tr>
</tbody>
</table>

Figure 1.7 Body mass index (BMI) is commonly used to define and classify overweight and obesity.

<table>
<thead>
<tr>
<th>Healthy weight</th>
<th>Over weight</th>
<th>Class I obesity</th>
<th>Class II obesity</th>
<th>Class III obesity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI 20–24.9</td>
<td>BMI 25–29.9</td>
<td>BMI 30–34.9</td>
<td>BMI 35–39.9</td>
<td>BMI 40+</td>
</tr>
</tbody>
</table>
which is clearly incorrect. For example, the selenium content of cereals from North America is considerably greater than in the same cereals from Europe as European soils are relatively impoverished in this mineral. The balance sheets also assume that all available food will be completely consumed by humans and do not allow for wastage or feeding to animals. It is also fallacious to assume that available food will be equally distributed to all people in a population and the sheets make no distinction between food available to men and women, to adults and children or to rich and poor.

Food accounts are a similar approach to estimating food availability, but instead of collecting data on a national scale, they are used to measure the food available to a household or an institution (e.g. a nursing home). By compiling an inventory of food stored at the start of a survey, monitoring food entering the setting (often measured by looking at invoices and receipts from food shopping) and taking into account any food grown in the setting, it is possible to calculate the food available per person over the period of the survey. As with the food balance sheet, this method does not allow accurate estimation of individual food intakes and does not allow for food wastage, but the food account can provide data on dietary patterns of families or similar groups at low cost and over an extended period of time.

### 1.4.2.2 Direct measures

Direct measures of nutrient intake collect data from individuals or groups of individuals and, in addition to their obvious application to clinical circumstances, are well suited to research in human nutrition and epidemiology. Although more robust than the indirect estimates described earlier, all direct measures of intake are prone to bias and error and results must always be interpreted with caution. These methods may be particularly useful for studying individuals or populations in different settings and study types (Table 1.5).

#### 1.4.2.2.1 Dietary recall methods

The dietary recall method is not only one of the best methods for examining nutrient intakes in a clinical setting, but it may also be used in research. One of the major disadvantages of the method is the need for a trained interviewer to spend a period of time with the patient or research subject to elicit detailed information on all food and drink consumed over a recent period of time (Table 1.5). An interview will explore all food and beverages consumed during that period and ask for descriptions of cooking methods, portion sizes, use of condiments and eating between meals. Most dietary recalls will be based upon intakes over the preceding

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary recall</td>
<td>Inexpensive; can be detailed; useful in clinical settings; can be repeated in the same individual; does not influence food intake</td>
<td>One recall seldom representative; relies on memory; intensive data entry; prone to under- and over-reporting; requires trained interviewer</td>
</tr>
<tr>
<td>Food frequency questionnaire</td>
<td>Can be self-administered; can use automated data entry; inexpensive for large population studies; represents usual intake over long periods of time</td>
<td>May not capture portion sizes; not useful for estimating absolute nutrient intake; subject must be literate; multiple foods may be covered by a single listed item; depends on memory</td>
</tr>
<tr>
<td>Food record</td>
<td>Does not depend on memory; data can be detailed and precise; captures intake over several days; can estimate nutrient intakes with good precision</td>
<td>Act of recording may alter diet; subject must be literate; intensive data entry; high burden on subject can lower response rates; prone to under- and over-reporting</td>
</tr>
</tbody>
</table>
24 h but in some cases may look at 48- or 72-h periods. Information obtained in this way can then be coded for detailed analysis of energy and nutrient intakes using appropriate nutritional analysis software or food tables. Dietary recall methods can generate detailed information on the types of food consumed and portion sizes. The use of photographic food atlases showing portion sizes for commonly consumed foods can enhance the quality of this quantitative information. Spending time interviewing a subject also makes it relatively easy to obtain recipes used in cooking and information about cooking techniques (e.g. use of oils in frying). Like all methods of estimating nutrient intake, the dietary recall is prone to inaccuracy due to under-reporting and over-reporting of food intake by certain groups of people. It is also dependent upon the memory of the subject and so loses accuracy when attempting to estimate habitual intake.

1.4.2.2.2 Food record methods

Food records, or diaries, administered to subjects for completion in their own time are widely regarded as the most powerful tool for estimation of nutrient intakes. Subjects keep records for extended periods of time (usually 3–7 days) and note down all foods and beverages consumed at the time they are consumed. Portion sizes can be recorded in a number of different ways, with the subject most frequently either noting an estimated intake in simple household measures (e.g. 2 tablespoons of rice, 1 cup of sugar) or an intake estimated through comparison to a pictorial atlas of portion sizes. To improve the quality of the data, intake can be accurately determined by weighing the food on standardized scales, taking into account any wastage (a weighed food record). Froebisher and Maxwell (2003) found that in studying intakes of children aged 6–16, a food record with a photographic atlas of portion sizes gave a good level of agreement with weighed records. In some settings, it is possible for a researcher to do the weighing, thereby reducing influences upon the subject consuming the food. Inaccuracies in estimates of portion sizes are a major problem associated with food record methods, particularly with some subgroups in the population, and methods should be chosen that best serve the purpose of the dietary survey. Surveys of small groups of well-motivated people in a metabolic unit lend themselves well to weighed record methods, while in large surveys of free-living individuals, these are rarely practical.

Food records have a number of strengths compared to other methods of estimating intake. Complex data on meal patterns and eating habits can be obtained through study of food diaries, and this information can supplement estimates of nutrient intake. By obtaining records for periods of 5–7 days, the intakes of most micronutrients can be estimated with some degree of confidence (Table 1.5), in addition to energy and macronutrients. For some nutrients, it is suggested that records of 14 or more days may be required (Block, 1989). The major disadvantage of the food record approach is the reliance upon the subject to complete the record fully and accurately. Maintaining a food record is burdensome, and it is often noted that the degree of detail and hence accuracy will be greater in the first 2–3 days of a 7-day record compared to later days. The act of recording intake, especially if a weighed record is used, can change the eating behaviour of subjects and hence lead to an underestimate of habitual intakes.

Like other direct methods, the food record is prone to under-reporting and over-reporting of energy and nutrient intakes among certain subgroups in the population due to the tendency of individuals to report intakes that will reflect them in the best possible light to the researcher. Bazelans et al. (2007) studied a group of elderly individuals, comparing self-reported intakes on a 24-h food record to estimates of likely energy intake based upon the subjects’ basal metabolic rates calculated using the Schofield equation. It was found that approximately 20% of men and 25% of women significantly under-reported or over-reported their energy intakes. Subjects with a BMI under 25 kg/m² (i.e. in the ideal weight range) were most likely to under-report, while 13% of those with BMI in the overweight range and 27% of those with a BMI in the obese range were found to have under-reported their energy intake. Obese and overweight women are frequently found to under-report intakes in dietary surveys.

1.4.2.2.3 Food frequency questionnaire methods

Food frequency questionnaire methods involve the administration of food checklists to individuals, or groups of individuals, as a means of estimating their habitual intake of foods, or groups of foods. Subjects work through the checklist and, for each foodstuff, indicate their level of consumption (i.e. number of portions) on a daily, weekly or monthly basis (Figure 1.9). Semi-quantitative food frequency questionnaires also collect information on typical portion size.

Food frequency questionnaires can vary in their complexity and length. Often, a questionnaire will consist of 100–150 food items and will therefore allow for a comprehensive coverage of the dietary patterns of a subject. Some questionnaires are much shorter and may be focused upon a particular food group or the main sources of a specific type of nutrient. For example, Block and colleagues (1989) developed a questionnaire with
just 13 items in order to identify individuals who had high intakes of fat. This was used as a preliminary screening tool to select subjects for a more detailed investigation.

Food frequency questionnaires have many desirable attributes for researchers wishing to estimate intakes in large populations (Table 1.5). They are self-administered by the subject, are generally not time consuming and are unlikely to influence eating behaviours. Data entry can sometimes be automated, reducing the analytical burden for the researcher. Moreover, the food frequency questionnaire provides an estimate of habitual intake over a period of months or even years, as opposed to the snapshot obtained by looking at a food record representing just a few days. However, the food frequency questionnaire can be a weak tool when considering portion sizes and is therefore less effective for estimating micronutrient intakes than a food record. Food frequency questionnaires must also be valid for the population to be studied as the range of foods consumed will vary with age and various other social and demographic factors. For example, if attempting to survey nutrient intakes in a population with a wide ethnic diversity, the foods and food groups included on the questionnaire need to reflect that level of diversity. A questionnaire that fails to include staple foods consumed by particular ethnic groups will inevitably underestimate their intake. For this reason, new food frequency questionnaires undergo extensive validation that includes comparing food frequency data with parallel analysis of dietary recalls and/or weighed food records in the same individuals.

### 1.4.3 Biomarkers of nutritional status

Biomarkers of nutritional status are measures of either the biological function of a nutrient or the nutrient itself in an individual or in samples taken from individuals. These measures can often provide the earliest indicator of a nutrient deficit as they register subnormal values ahead of any clinical symptoms. Biomarkers are therefore useful in monitoring the prevalence of nutrient deficiency, measuring the effectiveness of the treatment of deficiency and assessing preventive strategies. Given the huge difficulties of making accurate assessments of dietary intakes, as described earlier, biomarkers provide a useful means of validating dietary data and are often measured as adjuncts to dietary surveys. For example, in the UK National Diet and Nutrition Survey of preschool children (Gregory et al., 1995), measurements of circulating iron status were used to back up food record data collected on iron intakes. The doubly labelled water method (Koebnick et al., 2005) can be used to validate energy intakes estimated using dietary records or other means (Figure 1.10).

---

**Figure 1.9** A food frequency questionnaire is used to estimate the habitual consumption of foodstuffs within the diet of an individual.
Biomarkers of nutritional status are often regarded as being more objective than other indices. They include functional tests and measurements of nutrient concentration in easily obtained body fluids or other materials. The latter type of measurement can be a static test, which is performed on one occasion, or may be repeated at intervals to monitor change over time. The relative merits of these approaches will be discussed later in this section.

Functional tests measure biological processes that are dependent upon a specific nutrient. If that nutrient is present at suboptimal concentrations in the body, then it would be expected that the specific function would decline. The dark adaptation test is classic example of a functional test, which determines vitamin A status. The dark adaptation test measures visual acuity in dim light after exposure to a bright light that desensitizes the eye. The reformation of rhodopsin within the retina is dependent upon the generation of cis-retinol, and thus, the visual adaptation in the dark will be related to vitamin A status. Measurement of the excretion of xanthurenic acid is a functional test for vitamin B6 (pyridoxine) status. Xanthurenic acid is a breakdown product of tryptophan and kynurenine and is formed via pyridoxine-dependent reactions.

Non-functional measures of biomarkers typically involve direct measures of specific nutrients in simply obtained samples from individuals. These are most commonly samples of blood (plasma, serum or red cells) or urine but could include faeces, hair or, more rarely, biopsy material from the adipose tissue or muscle. Static tests provide a snapshot of the nutrient concentration in the sample at a given point in time and could be misleading as they often provide an indicator of immediate intake rather than habitual intake. For example, plasma zinc concentrations will vary hugely from day to day, reflecting ongoing metabolic fluxes, and fall by up to 20% following a meal (King, 1990). Wherever possible, repeated tests should be taken to increase confidence in the measured biomarker, or tests should be performed in a sample that provides a stronger indicator of habitual intake. In the case of zinc, plasma measurements are of limited value as most zinc is held in functional forms within tissues and less than 1% of the total pool is in circulation. Red or white blood cell zinc concentrations could be used as a more robust biomarker, as could white cell metallothionein concentrations (metallothionein is a key zinc-binding protein). Hair zinc concentrations give a better intake of long-term status. Zinc is deposited in hair follicles slowly over time, and so using this sample source removes the influence of shorter-term fluctuations in status. Lakshmi Priya and Geetha (2011) measured zinc, selenium, magnesium, copper, lead and mercury in hair samples from children as part of a study to assess their putative role in autism. Similarly, the EURAMIC study (Kardinaal et al., 1993) used measures of α-tocopherol and β-carotene in biopsies of adipose tissue to assess intakes of these vitamins. As fat-soluble vitamins are stored in this tissue, this gave an indicator of habitual intake over several weeks.

Figure 1.10 The doubly labelled water method is a technique used to assess energy expenditure. Subjects consume water containing stable isotopes of hydrogen and oxygen. This water reaches equilibrium with the body water. Measures of the doubly labelled water in saliva and urine enable estimation of the loss of 18O₂ from body water. That loss can only occur through production of labelled carbon dioxide. Carbon dioxide production is a measure of metabolism.

Biomarkers of nutritional status are often regarded as being more objective than other indices. They include functional tests and measurements of nutrient concentration in easily obtained body fluids or other materials. The latter type of measurement can be a static test, which is performed on one occasion, or may be repeated at intervals to monitor change over time. The relative merits of these approaches will be discussed later in this section.

Functional tests measure biological processes that are dependent upon a specific nutrient. If that nutrient is present at suboptimal concentrations in the body, then it would be expected that the specific function would decline. The dark adaptation test is classic example of a functional test, which determines vitamin A status. The dark adaptation test measures visual acuity in dim light after exposure to a bright light that desensitizes the eye. The reformation of rhodopsin within the retina is dependent upon the generation of cis-retinol, and thus, the visual adaptation in the dark will be related to vitamin A status. Measurement of the excretion of xanthurenic acid is a functional test for vitamin B6 (pyridoxine) status. Xanthurenic acid is a breakdown product of tryptophan and kynurenine and is formed via pyridoxine-dependent reactions.

Non-functional measures of biomarkers typically involve direct measures of specific nutrients in simply obtained samples from individuals. These are most commonly samples of blood (plasma, serum or red cells) or urine but could include faeces, hair or, more rarely, biopsy material from the adipose tissue or muscle. Static tests provide a snapshot of the nutrient concentration in the sample at a given point in time and could be misleading as they often provide an indicator of immediate intake rather than habitual intake. For example, plasma zinc concentrations will vary hugely from day to day, reflecting ongoing metabolic fluxes, and fall by up to 20% following a meal (King, 1990). Wherever possible, repeated tests should be taken to increase confidence in the measured biomarker, or tests should be performed in a sample that provides a stronger indicator of habitual intake. In the case of zinc, plasma measurements are of limited value as most zinc is held in functional forms within tissues and less than 1% of the total pool is in circulation. Red or white blood cell zinc concentrations could be used as a more robust biomarker, as could white cell metallothionein concentrations (metallothionein is a key zinc-binding protein). Hair zinc concentrations give a better intake of long-term status. Zinc is deposited in hair follicles slowly over time, and so using this sample source removes the influence of shorter-term fluctuations in status. Lakshmi Priya and Geetha (2011) measured zinc, selenium, magnesium, copper, lead and mercury in hair samples from children as part of a study to assess their putative role in autism. Similarly, the EURAMIC study (Kardinaal et al., 1993) used measures of α-tocopherol and β-carotene in biopsies of adipose tissue to assess intakes of these vitamins. As fat-soluble vitamins are stored in this tissue, this gave an indicator of habitual intake over several weeks.
1.5 Nutritional epidemiology: Understanding diet-disease relationships

Epidemiology is the branch of medical science that studies the causes of health and disease in populations rather than in individuals. At the simplest level, epidemiology can be used to examine geographical and temporal trends in disease to determine initial clues to the causes of disease and then use more sophisticated techniques to examine those possible causes. Nutritional epidemiology focuses on nutrition and nutrition-related factors as both causal and preventative factors in disease. Understanding nutritional epidemiology is important as the findings from such work are used as the main evidence base to develop public health strategies, health education advice and government policy on nutrition. For example, international policies to combat iodine deficiency through USI would not be possible without robust epidemiology to show that the strategy would be both effective and safe.

Nutritional epidemiology studies focus on two key measurements: the exposure and the outcome. Exposure refers to a nutrition-related factor that may be related to disease. This could be a marker of body composition (e.g. BMI), a specific nutrient (e.g. folic acid), a dietary pattern (e.g. vegetarianism) or a food-related behaviour (e.g. alcohol consumption). The outcome is the disease of interest, which can be measured as confirmed diagnosis of ill health (morbidity), as risk factors for disease (e.g. raised blood pressure as a risk factor for heart disease) or as death from a disease (mortality).

1.5.1 Cause and effect

Establishing which foods and nutrients may be causally related to disease processes is a critical aim for nutritional epidemiology. Without knowledge of how nutrients and nutrition-related factors influence disease processes, it is impossible to design effective interventions that prevent disease or nutrition-based treatments for established disease. It can be relatively simple to identify associations or correlations between factors but less easy to determine which are biologically or clinically meaningful. It is always important to appreciate that correlation does not indicate causal relationship. For example, as shown in Figure 1.11, we might look for evidence that the amount of meat in the diet is related to risk of developing oesophageal cancer and in a simple sense could correlate the meat intakes of different populations with the occurrence of oesophageal cancer in those populations. One approach might be to find populations with low rates of cancer and populations with high rates of cancer and compare their meat
The correlation we observe may be genuine or may be spurious, and much more information needs to be considered in designing a robust study to evaluate the relationship (e.g. the size of the sample needed, the method used to collect data on meat intake and the duration of the study) and interpreting the data obtained (considering alternative reasons for the correlation such as obesity or smoking habits).

1.5.2 Bias and confounding

Figure 1.11 shows the complexity of establishing a robust study for investigating a diet–disease relationship. Components of that figure such as sample size and who to sample, duration of study and consideration of the accuracy of measurement are all representatives of a phenomenon called bias, which can limit the usefulness of epidemiological studies. Studies which fail to control different types of bias, either at the point where the study is designed and initiated or during the analysis of the data, may draw spurious conclusions and be of no value in identifying diet–disease relationships.

Different types of bias may be present within a study, and bias is generally classified as ‘selection bias’, ‘measurement bias’ and ‘confounding bias’. Selection bias refers to factors that relate to how the people involved in the study were recruited. It is rarely possible to assess diet and disease in a whole population, and so inevitably a smaller sample has to be assessed. We may, for example, be interested in the links between cancer and meat intake across the whole of the United Kingdom but cannot possibly examine all 60 million people in the population. Instead, a study is likely to consider a few thousand individuals to represent that whole population. To avoid such bias, a wide age range should be sampled. Measurement bias (also called information bias) occurs when there are errors in the measurement of exposure or outcome that lead to misclassification of individuals. For example, if the method for measuring meat intake is inaccurate, individuals could be classed as high consumers when really they are...
Measurement bias is a major problem for nutrition–disease studies as the methods for considering food intake are prone to misreporting by study participants (either deliberate or due to poor memory (specifically termed recall bias)) or differences in the measurement between interviewers. In the case of the meat–oesophageal cancer relationship (Figure 1.11), for example, considering oesophageal cancer as a single outcome as opposed to two different diseases would also lead to measurement bias. Adenocarcinoma may have a different relationship to meat intake than squamous cell carcinoma.

Confounding bias is another form of bias and describes the situation where a third factor explains the relationship between exposure and outcome. To be truly considered as a confounding factor, the additional factor must be related to both exposure and outcome but not lie on the causal pathway between the two. Confounding bias is another form of bias and describes the situation where a third factor explains the relationship between exposure and outcome. To be truly considered as a confounding factor, the additional factor must be related to both exposure and outcome but not lie on the causal pathway between the two. The classic example of confounding is shown in Figure 1.12 where the relationship between alcohol and cancer may be explained by the fact that tobacco smoking causes cancer and people who consume high amounts of alcohol are also more likely to be heavy smokers. Some epidemiological studies are able to limit confounding bias at the design stage (see the following text), but usually confounding is adjusted for when analysing data.

### 1.5.3 Quantifying the relationship between diet and disease

As will be outlined in the following text, there are a variety of approaches used in nutritional epidemiology to explore the relationships between diet and disease. Each of these approaches yields information which quantifies the impact of specific nutrition-related factors (dietary patterns, specific nutrient intakes, obesity) upon disease outcomes (disease diagnosis, death) or risk factors for disease (blood pressure, elevated circulating biomarkers). Understanding the nature of these measured outcomes is critical for interpretation of the findings of epidemiological studies.

Some studies will yield relatively simple measures of outcome. For example, a study which measures blood pressure or elevated cholesterol concentrations in a group of people exposed to a factor and a group who are not will make a straightforward measure of whether the markers differ in the two groups of people. Considering differences in diagnosis or death, however, involves measures of ‘risk’ which are less familiar to those new to the subject.

One approach to quantifying risk is to provide data on prevalence or incidence rates (Table 1.6). These are essentially measures of the likelihood that individuals in a population will develop a disease (morbidity) or die from a disease (mortality), and an epidemiological study might, for example, compare the prevalence of cancer in meat eaters compared to non-meat eaters. Prevalence and incidence have different meanings. Prevalence is the measure of how likely an individual is to have a particular disease and is calculated as the number of cases in a population divided by the number of people in the population. So if in the United Kingdom there are 9600 cases of oesophageal cancer, then the prevalence is 9600 divided by 60000000, or 0.00016. This would be simplified to 16 per 100000 population. Incidence describes the probability of a person being diagnosed with a disease during a given period of time, or in effect the number of newly diagnosed cases. For example, if over the course of 1 year there are 2500 new cases of oesophageal cancer diagnosed in the United Kingdom, the incidence would be 2500 divided by 60000000, or 0.000042. This simplifies to 4.2 new cases per 100000 people per year.

A more commonly used approach to express the degree of risk associated with a nutrition-related factor is represented by odds ratios, relative risk and hazard ratios (Table 1.6). These are all variations on the same theme whereby risk is expressed as the likelihood in one group (e.g. meat eaters) relative to a reference group (e.g. non-meat eaters). This is generally easy to understand as an odds ratio of 10, indicating that people exposed to a particular factor are 10 times more likely to experience a disease event. An odds ratio of 0.2 means that people exposed to a particular factors are 80% ($1 - 0.2 = 0.8 \times 100 = 80\%$) less likely to experience a disease event. However, all calculated ratios are estimates and as shown in Figure 1.13 must be interpreted with reference to the confidence intervals that are calculated. These give a measure of how reliable those estimates

---

**Figure 1.12** A confounding factor is an additional factor that may explain the relationship between an exposure and an outcome. The confounding factor is related to both outcome and exposure, but does not lie on the causal pathway between the two.
Table 1.6 Definitions of key terms in epidemiology.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative risk</td>
<td>An indicator of the risk of an event (e.g. disease) occurring in one group compared to another. It is the ratio of the probability* of an event occurring in an exposed group and the probability of an event occurring in a control group.</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>An indicator of the risk of an event (e.g. disease) occurring in one group compared to another. It is the ratio of the odds† of an event occurring in an exposed group and the odds of an event occurring in a control group.</td>
</tr>
<tr>
<td>Hazard ratio</td>
<td>An indicator of the risk of an event (e.g. death) occurring in one group compared to another. It is calculated from the rates at which events occur over a period of time in the two groups.</td>
</tr>
<tr>
<td>Incidence rate</td>
<td>Incidence measures the risk of developing a disease within a given measure of time (e.g. the number of new cases of cancer per year).</td>
</tr>
<tr>
<td>Prevalence rate</td>
<td>Prevalence measures the proportion of people within a population who have a particular disease. It is usually expressed as the number of affected people for a given population size (e.g. the number of cancer cases per 100,000 population).</td>
</tr>
</tbody>
</table>

*Probability is a measure of how likely an event is and is calculated as number of adverse outcome/total number of outcomes. It is usually expressed as a value from 0 to 1.
†Odds are also measures of how likely an event is but are calculated as probability of an event/1 minus the probability.

Figure 1.13 Understanding odds ratios. a) Odds ratio (OR) is a descriptor of the risk of event compared to a reference group. For the reference, the OR is set at 1.0. If OR is less than 1.0, that indicates decreased risk. If more than 1.0, it indicates increased risk. b) OR is an estimate of risk and the quality of that estimate will depend upon methodological factors and biological variation. The range of possible values for OR is represented by 95% confidence intervals (CI). If the range of confidence intervals includes the value of 1.0 then we cannot state that risk is significantly different to the reference group.
are. There are important differences between odds ratios, hazard ratios and relative risk ratios, and they are all calculated in different ways (Table 1.6). As a result, they are not interchangeable terms and they are often used with specific study types. Odds ratios are calculated in most case–control studies, while relative risk is more often used in much larger cohort studies or with randomized controlled trials (RCTs).

### 1.5.4 Study designs in nutritional epidemiology

There are a number of different approaches that can be taken to explore diet–disease relationships in human populations, and these vary in their capacity to determine causality of relationships (Table 1.7). Figure 1.14 shows a hierarchy of research designs that has the study designs with the lowest methodological quality (animal studies and in vitro studies, ecological studies) at the bottom and the highest methodological quality (RCTs and meta-analyses) at the top. Data from all studies needs to be interpreted with this hierarchy in mind. Work performed in animals must always be viewed through the lens of species differences between animals and humans. Studies of large populations of free-living individuals (cohort studies) will inevitably be subject to unaccounted-for confounding factors and other bias, which can only be eliminated though an RCT.

<table>
<thead>
<tr>
<th>Study design</th>
<th>Approach taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecological study</td>
<td>Nutritional exposures and disease outcomes are considered in populations that are grouped by geographical area or time period. Only population averages and not individual data are analysed</td>
</tr>
<tr>
<td>Cross-sectional study</td>
<td>A descriptive study which measures nutritional exposure and disease outcome in a single population at a specific point in time</td>
</tr>
<tr>
<td>Case–control study</td>
<td>A study which compares nutritional exposures in a population with a specific disease to a similar reference population without disease</td>
</tr>
<tr>
<td>Cohort study</td>
<td>An observational study which follows a population over a period of time. This allows nutritional exposure measured at the beginning (baseline) to be related to disease that develops over the course of the study. Follow-up from baseline may be over many years</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>An experimental study in which a randomly selected group of people are administered nutrients, foodstuffs or other interventions focused on nutrition and are compared to a matched control group over a period of follow-up</td>
</tr>
</tbody>
</table>

**Table 1.7** Study designs in nutritional epidemiology.

**Figure 1.14** Hierarchy of evidence in nutrition–disease studies. Experiments in animals or in vitro have the lowest methodological quality, while randomized controlled trials and meta-analyses are of highest quality.
1.5.4.1 Ecological studies
An ecological study is an observational study which seeks to compare the general characteristics of whole populations in order to determine the factors which explain variation in disease risk between those populations (Table 1.7). For example, there may be gross differences in coronary heart disease death among the countries of Europe, and we may wish to determine whether those differences arise due to variation in diet. Alternatively, we may wish to determine why rates of coronary heart disease death in one location in 2012 are much lower than in 1992. To approach these questions, data about outcome will generally be extracted from government or international databases, and data about exposure will be collected from national diet and nutrition surveys or international sources such as the FAO food balance sheets. Importantly, this is summary data about large groups of people rather than data that has been collected on individuals (Figure 1.15). This data is then used to examine simple correlations between exposures and outcomes.

Ecological studies are ideal for examining new ideas and developing hypotheses that can be explored using approaches with greater methodological quality and capacity for determining causality. They have many weaknesses, however, not least the fact that the findings generated from group data are not necessarily applicable to individuals in a population (this is termed the ecological fallacy). The way in which data is collected leads to numerous problems, for example, the data on dietary exposure may have been collected in different ways in different countries (heterogeneity of exposure) and may not be truly comparable. The exposure and outcome may not have even been determined in exactly the same populations (e.g. data on diet in 1992 was collected in England, Wales and Scotland, but data on heart disease was collected in the whole United Kingdom – England, Wales, Northern Ireland and Scotland). Ecological studies are also highly prone to uncontrollable confounding factors.

1.5.4.2 Cross-sectional studies
A cross-sectional study is an observational study in which the exposure and outcome are measured simultaneously in a group of individuals (Figure 1.15). This data is then used to examine simple correlations between exposures and outcomes.

Figure 1.15 Research designs in epidemiology.
diet–disease relationships. For example, it could very quickly show a relationship between BMI and energy intake. However, it is not possible to state with any certainty whether a relationship is causal. Is high energy intake a cause of high BMI, or do people of high BMI consume more energy to meet a greater demand? Another problem of the cross-sectional study is that the individuals who have a disease may change their behaviour because of the disease and hence mask the relationship. Subjects with high blood pressure, for example, may be advised by their doctors to reduce salt intake, thereby hiding a simple relationship between salt and blood pressure.

1.5.4.3 Case–control studies

Case–control studies are observational studies which deliberately sample a group of people who are confirmed to have a disease of interest (cases) for comparison with a group of people without the disease (controls; Table 1.7). The exposure of the two groups to dietary factors in the past will then be compared to see if risk of disease can be related to those factors (Figure 1.15). For example, a group of people with high blood pressure can be compared to a group with normal blood pressure by considering salt intakes over the preceding 10 years. This approach could address the question of whether long-term, habitual salt intake is a causal factor in development of high blood pressure.

The case–control study is a quick and inexpensive approach to epidemiology as a relatively small number of subjects are required. This is because the selection of people with the disease ensures a good representation of the diseased population which may not be possible with a cohort study (see the following text). When the disease of interest is rare, the case–control design becomes especially powerful as it draws together a population that is unlikely to be sampled in a cohort study. The case–control study can also consider more than one exposure variable in the diet. There are significant problems with case–control designs however, of which the most important is the fact that while outcome is usually well defined, the exposure must be established retrospectively. As described earlier, assessment of diet in individuals is problematic, and this becomes even more prone to recall bias when looking back over many years. The other major problem is the recruitment of a suitable control group. Matching closely to the cases (e.g. similar ages, same sex) is important for reducing confounding bias. The control group should also come from the same geographical area as the cases so that it represents the same population as the cases were recruited from. Often, they are patients with other conditions who are recruited from the same clinics and hospitals as the cases. Good studies will carefully examine the health of the controls to ensure that they do not have the disease of interest (undiagnosed cases). Poor selection of controls can greatly undermine the quality of a case–control study.

1.5.4.4 Cohort studies

Cohort studies are also referred to as longitudinal studies as they seek to study a large group of people over a period of time (Figure 1.15). They identify exposures in a population and then by following the cohort over time identify outcomes as they occur and compare the incidence of disease in exposed individuals with unexposed individuals (Table 1.7). The Framingham Heart Study, for example, recruited 5209 men and women aged 30–62 in the town of Framingham, Massachusetts, in 1948. These subjects have been examined every 2 years since the inception of the study, allowing determination of the major cardiovascular risk factors that predict heart disease morbidity and mortality (blood pressure, blood cholesterol, obesity, diabetes, physical inactivity).

Cohort studies may report findings from prospective cohorts or retrospective cohorts. A prospective cohort is generally recruited for a specific purpose and collects baseline data and conducts follow-up measurements at intervals over several years before reporting final outcomes. A retrospective cohort will generally involve collecting disease outcome data in a large group of people and then tracing back events by collecting historical records that provide the exposure data. For example, a prospective cohort to examine the relationship between diet and cancer would recruit a population, characterize their diet and follow for a period to see if occurrence of cancer was related to diet many years before. A retrospective cohort would look at a population of people, including people with and without cancer, and look back to see if the diets of subjects from 10 to 20 years earlier differed among cancer sufferers and those without cancer. Retrospective cohorts are generally smaller and less expensive to study than prospective cohorts but are more prone to confounding bias due to the passing of time and incomplete historical record-keeping.

1.5.4.5 RCTs

RCTs are generally regarded as the gold-standard method for an epidemiological study (Table 1.7). RCTs are essentially clinical experiments in which individuals are randomly assigned to be a control (no treatment) or to receive some form of intervention as a controlled exposure (Figure 1.15). In nutritional epidemiology, that intervention will generally be the administration of
a supplemental nutrient (e.g. antioxidant, essential fatty acid) or a behavioural intervention related to food intake or weight control (e.g. increased physical activity, adoption of a meat-free diet).

The nature of the RCT design means that it can provide strong evidence of cause and effect. If control and intervention groups are carefully matched for key characteristics, there will generally be no issue of confounding, and as the experimenters have control over the exposure, there are few issues of bias in the measurement of that exposure. Problems can arise due to non-compliance with the protocol. Participants may, for example, not take supplements as directed. RCTs in nutrition are, however, often less effective than similar studies where the treatments are drugs (clinical trials). This is because often the diet–disease relationships can be quite weak and take years to develop. It is generally not feasible to carry out a supplementation trial over many years due to expense, because the subjects may become disaffected and drop out (e.g. the consumption of fish oil capsules causes bad breath and so cannot be tolerated for more than a few months) and because the nature of the intervention may become apparent to the control group prompting them to change their diet and behaviour in a way that detracts from the analysis (e.g. although not allocated to receive a folate supplement, the controls may consume more natural sources in the belief it will be healthier). While drug trials can be easily blinded so that the participants and researchers are not aware of allocation to control or test groups, this is sometimes not possible within a nutrition trial (e.g. exercise intervention cannot be disguised as something else). Thus, the intervention can introduce behavioural changes among participants, which can distort study findings.

1.5.4.6 Systematic review and meta-analysis

Systematic review is an approach to research that makes use of existing evidence to address a specific research question. In the same way as a laboratory experiment, clinical trial or single epidemiological study, the systematic review will use rigorous methodology to ensure that the quality of the finished review is reliable and robust. Systematic reviews synthesize the findings of all available research on a particular topic in an unbiased manner and produce an impartial summary of the findings, which fully considers flaws and gaps in the evidence base. This is an extremely powerful tool for evaluating the balance of evidence, particularly in areas where there is apparent controversy and uncertainty. The use of explicit, systematic methods in reviews limits bias (systematic errors) and reduces chance effects. This provides more reliable results upon which to draw conclusions, as these can be based on the totality of the available evidence rather than the elements that suit the bias of a narrative review. Systematic reviews use modern electronic database searches to access all published works related to a research question. These works and any material obtained from other sources (e.g. direct communication with experts in the field to obtain unpublished material) comprise the data that is subsequently analysed.

Systematic reviews are often combined with the analytical technique called meta-analysis. Meta-analysis enables researchers to exploit greater statistical power by effectively combining the results of several studies to generate a larger and robust sample. This enables inconsistency between studies to be both quantified and overcome and more precise measures of risk to be calculated. An example of the power of meta-analysis to resolve contentious issues is shown in Research Highlight 1.3.

Other chapters of this book will make frequent reference to meta-analysis.

The 2007 World Cancer Research Fund Expert Report detailed a systematic review of the evidence on salt and stomach cancer, which included data from 3 cohort studies, 21 case–control studies and 12 ecological studies. This systematic review was able to generate a meta-analysis that included data from two of the cohorts and
nine of the case–control studies (WCRF, 2007). Using the combined data sets, the expert group found that there were a small but significant relationship apparent in the cohort studies (relative risk 1.08, 95% CI 1.00–1.17) and a close to significant relationship apparent in the case–control studies (relative risk 1.01, 95% CI 0.99–1.04). On this basis, the reviewers were able to conclude that there is a probable causal relationship between high salt intake and stomach cancer. Without a robust, unbiased review process and well-conducted meta-analysis, this conclusion would have been impossible to reach.

1.6 Dietary reference values

Dietary reference values (DRVs) are standards that are set by the health departments of governments in a number of countries around the world. DRVs are guidelines that can be used to define the composition of diets that will maintain good health. There are many complex systems of DRVs used in different countries. These vary according to national health priorities and policies; predominant health status, socioeconomic status, body mass and rates of growth; and local factors, for example, the composition of foods or other lifestyle influences, that determine the absorption and hence bioavailability of nutrients (Pavlovic et al., 2007).

DRVs are used in a variety of different ways. While some systems, such as those developed for the United Kingdom, are generally intended to be used only with populations or subgroups within populations, others (e.g. the US Dietary Reference Intakes) are widely used in providing dietary guidance for individuals. On a population level, the DRVs are useful yardsticks with which to assess the adequacy of the diet of a population and hence protect individuals within that population against the adverse consequences of either deficiency or excess. By using DRVs as standard measures against which dietary survey data can be compared, it is possible to estimate the prevalence of risk of deficiency for specific nutrients within a population.

In some countries, there are regular surveys of national dietary patterns among age- and gender-specific groups, for example, the UK National Diet and Nutrition Surveys (Bates et al., 2014) or the US National Health and Nutrition Examination Surveys (Centers for Disease Control, 2014). Findings from such surveys are compared to the DRVs in order to highlight potential nutrient deficiencies. In other countries, food supply data at the national level, such as the food balance sheets collected by the FAO, can be used to crudely estimate the average per capita availability of energy and the macronutrients and compared to international standards. Although such data are prone to error, as described earlier, they can be used for tracking trends in the food supply and determining availability of micronutrient-rich foods. By comparison of such data with DRVs, it is possible to uncover evidence of gross inadequacies in the quality of the diet across whole populations (but not subgroups such as children or the elderly). Standards for nutrient provision based upon DRVs can also be used in the planning of food supplies to regions (e.g. in humanitarian aid) or in menu planning for caterers in hospitals, schools or other institutional settings. Many of the food labelling schemes used in supermarkets are based upon published DRVs for specific nutrients.

1.6.1 The UK DRV system

In 1979, the United Kingdom set a series of DRVs termed the recommended daily amounts (RDAs). In 1991, a new series of DRVs were published to replace these RDA values, as they were considered to be prone to misunderstanding and misuse. The term ‘recommended’ wrongly suggests a level of intake that an individual must consume on a daily basis in order to avoid adverse consequences. The new system of DRVs produced by the Committee on Medical Aspects of Food Policy (COMA) (DoH, 1991) therefore dropped the word recommended and was developed to indicate different levels of intake that would be suitable for healthy populations, broken down by age and gender.

In setting the DRVs, the COMA reviewed research for each macro- and micronutrient in order to determine the levels of intake that are necessary to maintain normal health and physiological function. In considering the available evidence, the key issues to be explored for each nutrient were as follows: (1) What level of intake is necessary to maintain circulating or tissue concentrations within normal ranges? (2) What level of intake is necessary to avoid clinical deficiency in individuals or in populations? (3) What level of intake has been established as being effective in treating clinical deficiency? (4) What level of intake has been shown to maintain normality in a biomarker of adequacy?

As shown in Figure 1.16, the relationship between nutrient intake and disease risk is not linear. At low levels of intake, the probability of adverse consequences (deficiency disease, loss of physiological function) is elevated. With rising intakes, the probability of such consequences declines to zero as intakes provide the requirements of most of the individuals in a population. At higher intakes, the probability of adverse consequences associated with overnutrition begins to rise. In
developing a set of DRVs appropriate for a population like the United Kingdom, in which the economic wealth of the population makes overnutrition more likely than undernutrition, this continuum between risk and intake must be recognized.

In common with the United States and other countries (see the following text), the UK DRVs were developed to map onto the expected distribution of nutrient requirements in a population. As shown in Figure 1.17, this would usually be expected to follow a normal distribution, which actually relates to the left-hand side of the distribution of risk plotted against intake (Figure 1.17). In this context, the mean value (midpoint) in a normal distribution would represent a level of nutrient intake at which the requirements of 50% of the people in a population would be met. Within the UK DRV system, this point is termed the estimated average requirement (EAR). When a population is consuming a nutrient at a level close to the EAR, it can be assumed that for 50% of people, this will be sufficient, but that for up to 50%, nutritional status would be compromised.

The other DRVs are set at points that are two standard deviations either side of the mean. The reference nutrient intake (RNI) is the upper value and within the normal distribution would represent a level of intake that would meet the requirements of 97.5% of the population. When a population is consuming a nutrient at a level close to the LRNI, it could be assumed that for most individuals, this will be insufficient and that deficiency disease would be rife.

For some nutrients (e.g. pantothenic acid, biotin and molybdenum), the COMA had insufficient data to be able to derive estimates of requirements but recognized the biological importance of these compounds in the diet. In the absence of extensive information, the safe intake was set. This is an upper level (UL) of intake set at a point likely to prevent deficiency and avoid toxicity. Safe intakes are of greatest importance to vulnerable groups in the population such as infants and children (DoH, 1991).

The DRVs are published as a comprehensive series of tables (DoH, 1999), which, for most nutrients, provide reference values for males and females separately and for different age groups (typically 0–12 months, 1–3 years, 4–6 years, 7–10 years, 11–14 years, 15–18 years, 19–50 years and 50+ years). To reflect increased demands for nutrients during pregnancy and lactation, some tables show additional increments of intake for pregnant and breastfeeding women. For micronutrients and trace elements, published values include all three terms (LRNI, EAR and RNI). With respect to protein, only EAR and RNI values were determined. Given that excess energy consumption is a driver of obesity and related disorders, it is undesirable to set reference values at an upper point such as the RNI, as a population that consumed energy at that level would be expected to have a high prevalence of related adverse effects such as obesity. DRV tables for energy therefore include only

![Figure 1.16 The association between risk of nutrition-related risk and level of nutrient intake. EAR, estimated average requirement; RNI, reference nutrient intake; UL, tolerable upper limit.](image-url)
Humans have a requirement for essential fatty acids, and children can develop clinical deficiency of linoleic acid. There are DRVs that indicate minimum intakes of essential fatty acids, but as low intakes of the majority of lipids are not associated with adverse health effects, the three main DRV terms are not applied to fats and carbohydrates. Instead, the COMA set population average guidelines for consumption of saturated, monounsaturated and polyunsaturated fats based on percentage of dietary energy provided by those sources. These guidelines represent maximum intakes in the light of the established risk of cardiovascular disease with high-fat intakes (see Chapter 8, Guidelines for healthy nutrition). This element of the UK DRV system differs from other components as it firmly indicates guidelines for individuals to follow rather than appropriate ranges for healthy populations (Whitehead, 1992). In the same way, the COMA set guidance values for sugars and complex carbohydrates based on percentages of dietary energy intake. Population averages are designed to encourage lower intakes of non-milk extrinsic (free) sugars and fats, while increasing intakes of starch and non-starch polysaccharides. Population averages for carbohydrate are likely to be revised following the 2014 report of the UK Scientific Advisory Committee on Nutrition (SACN, 2014) which proposed reductions in guidelines for free sugars (no more than 5% of dietary energy per day) and an increase for dietary fibre (25 g/day for adults).

In the United Kingdom, the DRVs are not intended to be guidelines for individuals. It is generally considered a fruitless activity to make estimates of nutrient intakes for individuals, given problems with obtaining accurate data on food intake and because it is impossible to estimate what the true requirements for any individual are likely to be. In making assessments of dietary intakes of groups within a population, the RNI is considered to be the most important benchmark for comparison. The nearer the average intake of a group within a survey is to the RNI, the less likely it is that any individual within that group will have an inadequate intake. However, the LRNI value provides a better indicator of the likely risk of widespread deficiency, whether clinical or subclinical. The nearer the average intake of the group is to the LRNI, the greater is the probability that some individuals within that group are not consuming that nutrient at a level adequate to meet their requirements.

Figure 1.17 The normal distribution as a basis for DRVs. UK DRVs are based upon an assumed normal distribution of individuals’ nutrient requirements and level of nutrient intake. The estimated average requirement (EAR) is set at the centre (mean) of the distribution. The lower reference nutrient intake (LRNI) and reference nutrient intake (RNI) values are placed two standard deviations below and above the mean, respectively. The nutrient requirements of all but 5% of the population should therefore be met by levels of intake between these two values.
An example of the DRVs in use is provided by the study of Cowin and colleagues (2000). This group assessed the nutrient intakes of 1026 18-month-old infants living in the southwest of England using a 3-day unweighed dietary record. By comparing recorded intakes with the RNI values for micronutrients, the survey concluded that intakes of most nutrients were adequate in this population group. However, for iron and vitamin D, it was noted that mean intakes were considerably below the RNI, suggesting that these nutrients could be a cause for concern in this population group. Indeed, for iron, where the LRNI is 3.7 mg/day for infants, it was noteworthy that the 2.5% of the population with the lowest intakes (i.e. the group who might be expected to be meeting their requirements despite low intake) consumed only 2.4 (girls) to 2.7 (boys) mg/day, figures well below the LRNI. Data of this kind can be the start point for further studies that identify the causes of deficiency and for formulating appropriate interventions and dietary recommendations (Cowin et al., 2001).

Although not intended for use with individuals, the DRVs could still be used in a clinical setting. When working with healthy individuals, assessments of dietary intakes that indicate intakes below or close to the LRNI could indicate a dietary problem and might be a stimulus for a more in-depth assessment of biochemical or clinical indicators of nutritional status. In planning a diet for an individual, the delivery of nutrients at the level of the RNI would be a basic priority to ensure optimal health.

1.6.2 DRVs in other countries

The UK system described earlier is just one example of DRVs defined with the purpose of guiding the provision of healthy nutrition on a population-wide scale. Many other countries use similar systems that have also been derived to map against the normal distribution of nutrient intakes against provision of nutrient demands. This approach is generally applicable for Westernized countries where the nutrition-related health concerns are usually focused on the consequences of nutrient excess rather than nutrient deficiency. Table 1.8 summarizes the dietary reference terms used in North America, Australia and New Zealand.

Among the countries of the European Union (EU), there is considerable variation in the terminology used to describe DRVs and in the precise nature of recommendations made for particular population groups, most particularly children. There are suggestions that the European countries should harmonize their DRV systems (Pavlovic et al., 2007) and that in the course of generating a common system, a further review of the evidence could be conducted to determine whether regional variation reflecting health status and other local issues is necessary or desirable. The EU Scientific Committee on Food defined three levels of DRVs: average requirement, population reference intake and lowest threshold intake. In general intent, these terms map against the UK EAR, RNI and LRNI values. In 2010, the European Food Safety Authority set some simple references for nutrient intakes, which included references for consumption of total carbohydrates, sugars, fibre, fats and water (EFSA, 2010).

As in the United Kingdom, the countries of North America reviewed their existing reference values, originally set in 1941, and replaced them with a new comprehensive format in the early 1990s (Kennedy and Meyers, 2005). In Canada and the United States, the EAR and RDA terms are exact equivalents of the UK EAR and RNI terms but are used in a different manner to that seen in the United Kingdom. EAR is a term that would be used to estimate the prevalence of inadequate intakes in a population, but RDA is a term specifically intended for use with individuals. A habitual intake below this level would be associated with increased risk of dietary inadequacy. In population surveys, however, comparing mean intakes to the RDA

<table>
<thead>
<tr>
<th>Region</th>
<th>Dietary reference terms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>LRNI</td>
<td>Lower reference nutrient intake</td>
</tr>
<tr>
<td></td>
<td>RNI</td>
<td>Reference nutrient intake</td>
</tr>
<tr>
<td></td>
<td>EAR</td>
<td>Estimated average requirement</td>
</tr>
<tr>
<td>United States/</td>
<td>Safe intake</td>
<td>Estimated average requirement</td>
</tr>
<tr>
<td>Canada</td>
<td>EAR</td>
<td>Estimated average requirement</td>
</tr>
<tr>
<td></td>
<td>RDA</td>
<td>Recommended daily allowance</td>
</tr>
<tr>
<td></td>
<td>AI</td>
<td>Adequate intake</td>
</tr>
<tr>
<td></td>
<td>UL</td>
<td>Tolerable upper limit</td>
</tr>
<tr>
<td>Australia/NZ</td>
<td>EAR</td>
<td>Estimated average requirement</td>
</tr>
<tr>
<td></td>
<td>RDI</td>
<td>Recommended daily intake</td>
</tr>
<tr>
<td></td>
<td>AI</td>
<td>Adequate intake</td>
</tr>
<tr>
<td></td>
<td>EER</td>
<td>Estimated energy requirement</td>
</tr>
<tr>
<td></td>
<td>UL</td>
<td>Upper level of intake</td>
</tr>
</tbody>
</table>
would tend to overestimate the likely prevalence of deficiency, as it is a figure set at a level where the requirements of 97.5% of the population are being met. This means that a significant proportion of the population is likely to be exceeding requirement (Kennedy and Meyers, 2005). For example, if the RDA for iron intake in children is 11.2 mg/day and the mean intake for a population is found to be 8.4 mg/day, it should not be assumed that deficiency will have a high prevalence. The majority of children in the population may be consuming well below the RDA value and still be achieving requirement. This could also be seen as a problem with the UK RNI. The tolerable UL term is defined as the highest average daily nutrient intake level that is unlikely to result in adverse health effects for almost all individuals in a population. Effectively, individuals could use this as a guide to limit their intake, and at the population level, it provides a benchmark against which estimates can be made of the likelihood of problems related to overnutrition. The AI term is similar to the UK safe intake in that it is used only where there is insufficient data to determine the EAR for a particular nutrient.

In Australia and New Zealand, the system of DRVs is broadly similar to that used in North America, except a fifth term (EER) is defined for energy. The EER comprises two separate terms. The estimated energy requirement for maintenance (EERM) is the energy intake that is estimated to maintain balance in healthy individuals or populations at a given level of physical activity and body size. The desirable estimated energy requirement (DEER) is the level of energy intake that should maintain energy balance in healthy individuals or populations of a defined gender, age, weight, height and level of physical activity, consistent with optimal health. Although complex, this is an important distinction as the EERM represents an actual energy requirement of an individual or group of individuals, while the DEER allows calculation of energy references that can be used to guide weight loss in a clinical situation (National Health and Medical Research Council of Australia, 2006).

In less affluent countries where there is a high burden of malnutrition-related disease, the priorities of governments are different, and DRVs are set at levels that are more appropriate for a setting where maintaining and monitoring food security are the main applications of the figures. Often, the values used in these situations are obtained from the FAO and focus heavily on setting levels of intake that will provide the basic requirements of most of the population and therefore avoid widespread clinical nutrient deficiency.

---

**SUMMARY**

- Nutritional balance depends upon the supply of nutrients being able to meet the physiological and metabolic demand for nutrients to be used as structural components or as substrates and cofactors for metabolism. Undernutrition or overnutrition arises through disturbance of this balance.
- Undernutrition can result from either a decrease in intake or an increase in the demand for nutrients. Increased demands are often a consequence of physiological insult or stressors, including trauma, pregnancy and lactation.
- Prolonged undernutrition can lead to micronutrient deficiency or malnutrition, which are common among infants and women in developing countries and among the elderly and poor in developed nations.
- The way in which the body responds to food and nutrients is strongly influenced by genetic factors. Single nucleotide polymorphisms result in common gene variants that can influence diet–disease relationships. This means that more individualized approaches to dietary guidelines may be more effective than general population guidelines.
- Stage of life is one of the most important determinants of nutritional status, as the nature of demands for nutrients and the way in which those demands are met undergo profound changes over the human lifespan.
- Nutritional status can be assessed by using anthropometric methods, by using different methods of measuring intake, through clinical examination or by measuring specific biomarkers. All methods are limited in their scope and are prone to inaccuracy.
- Exploring relationships between diet and disease relies on well-designed epidemiological studies. Simple ecological and cross-sectional studies can provide clues to diet–disease associations, but more robust cohort studies and randomized controlled trials are necessary to confirm causal relationships.
- DRVs are standards for nutrient intake, which are set by governments. They are widely used as the basis of nutrition-related advice and interventions. They can be used as research tools, as guidance for meal planners and caterers and for the monitoring of food security at a national level.

---

**References**


Additional reading

If you would like to find out more about the material discussed in this chapter, the following sources may be of interest.


