Food composition data

PRODUCTION, MANAGEMENT AND USE

H. Greenfield and
D.A.T. Southgate

Second edition

INFOODS

FAO
FIAT PANIS

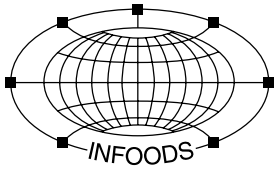
Food composition data

Food composition data

PRODUCTION, MANAGEMENT AND USE

by
H. Greenfield
University of New South Wales,
Sydney, Australia
and
D.A.T. Southgate
Formerly of the Agricultural and Food
Research Council Institute of Food Research,
Norwich, United Kingdom

Technical editors:
B.A. Burlingame and U.R. Charrondiere



Food and Agriculture
Organization of
the United Nations
Rome 2003

Editing, design and production
by the FAO Publishing Management Service

ISBN 92 5 104949 1

First edition published in 1992
by Elsevier Science Publishers

Contents

Foreword to the first edition

About 20 years ago, there was a recognition in Europe that real benefits would flow from coordinating the manner in which food composition tables were produced in the various countries of Europe. Subsequent development of computerized nutritional databases has further highlighted the potential advantages of working together. Such cooperation could lead to improved quality and compatibility of the various European nutrient databases and the values within them. This realization was one of the driving forces behind the development of the EUROFOODS initiative in the 1980s when those people in Europe interested in data on food composition began working together. This initiative received further impetus with the establishment of the EUROFOODS-Enfant Concerted Action Project within the framework of the FLAIR (Food-Linked Agro-Industrial Research) Programme of the Commission of the European Communities.

It was quickly recognized that the draft guidelines for the production, management and use of food composition data, which had been prepared under the aegis of INFOODS (International Network of Food Data Systems, a project of the United Nations University), would be especially applicable to the objectives of the Concerted Action. The guidelines have been written by two recognized experts. Many people associated with FLAIR EUROFOODS-Enfant have added constructive criticism and advice to that offered previously by those associated with INFOODS. Thus the guidelines are backed by a consensus in the community of those responsible for the production and use of food composition tables and nutrient databases.

I am sure that the book will be regarded by those concerned with the production and use of nutritional compositional data as a lighthouse on an ocean with poor visibility, many hazards and shipwrecks. It will provide invaluable light not only for people in Europe but also for those on other continents across the oceans.

Clive E. West
Project Manager
FLAIR EUROFOODS-Enfant Project
Wageningen, February 1992

Preface to the second edition

The first edition of this book was used extensively in training food composition analysts and compilers around the world, commencing with the first Food Composition training course held in Wageningen, the Netherlands, in October 1992. Five courses have been held subsequently in Wageningen and the course has also been transferred to developing regions including one each in Chile for LATINFOODS countries, Jamaica for CARICOMFOODS countries, Thailand for ASEANFOODS and SAARCFOODS countries, and three in South Africa for the countries of ECSAFOODS.

The use of the book in the United Nations University/INFOODS training courses revealed that changes were required to update the text and figures, in particular to make the book more user-friendly on an international basis. Increasingly, as time went by, the huge explosion in methods of analysis meant that the book was becoming rapidly out-of-date. Further, the establishment of food composition programmes around the world increased the pool of experience available. However, revision was not feasible as a commercial project. Although some tertiary courses, largely in the industrialized countries, were able to draw on the book in teaching, the prohibitive cost of the first edition meant that purchases of the book were mainly for libraries rather than by individuals or for local food composition programmes. When the first edition became out of print, the copyright reverted to the original authors.

In 2001, Dr Barbara Burlingame, the Director of INFOODS (Food and Agriculture Organization of the United Nations [FAO]), proposed a rescue package – which was taken up eagerly by the authors. The proposal was for the authors to revise and update the first edition in the light of the comments of trainees in the course over the previous decade, and to incorporate improved methods of analysis (while not excluding those older methods, which were still being used satisfactorily in those parts of the world where access to sophisticated and costly instrumentation was limited). It was also proposed that FAO make the print edition of the book available at an affordable price, oversee its translation into the main languages of the United Nations Organizations and, further, place the book on the FAO Web site for worldwide access. The authors were pleased to accept this proposal since the original concept of the book had always been wide availability at a price that placed it within the reach of students and workers, particularly those in developing countries.

The second edition was largely prepared by means of electronic communication interspersed with occasional face-to-face meetings to establish the roles of the authors and FAO and identify the new or revised material to include. David Southgate worked from a very large literature database compiled by Heather Greenfield for the period 1990 to the present, together with his unparalleled experience in the compilation of the United Kingdom tables

and discussions with trainees in courses held in the Netherlands and other parts of the world, to collate the first comprehensive draft of the revised edition, which included particular sections drafted by Heather Greenfield and inputs from members of the INFOODS mailing list.

A meeting of the authors with Barbara Burlingame in Norwich, United Kingdom, made possible an extensive review of the text, particularly to incorporate elements required by FAO. The draft chapters were reviewed by experts and the final version for publication was prepared through a long process of careful checking and revision conducted by Heather Greenfield, Barbara Burlingame and Ruth Charrondiere (FAO), working in collaboration by e-mail correspondence and, where possible, consultation with all the original sources of information. Barbara Burlingame oversaw the preparation of the final text for publication in various formats at FAO.

As in the first edition, the personal perspectives and prejudices of the authors doubtless show through. We believe that there is no *a priori* method of obtaining compositional data without analysis. The book recognizes that analytical facilities and resources are limited in virtually all countries and that, at the same time, there is a large amount of compositional data in the literature, in both published and unpublished sources and in other databases. It is essential to make proper use of this material. The book therefore devotes a considerable amount of attention to the evaluation of this published material to ensure that it is of the appropriate quality to use in combination with directly analysed values. We trust that this book, used in combination with other INFOODS texts, will be a key to the improved quality of food composition data worldwide.

Preface to the first edition

In 1972 a working party of the Group of European Nutritionists met in Zurich (Switzerland) to consider the principles that should be used in preparing national tables of food composition. A small book based on a working paper for this conference and describing guidelines for the preparation of such tables was subsequently published (Southgate, 1974).

During those discussions it became clear that in the future more tables providing international coverage (e.g. for all of Europe) would be needed. Since then, widespread advances in computer techniques have made the creation of such international databases technically feasible; their development is impeded, however, by the variable analytical quality, the incompatibilities, and even the unknown provenance of existing compositional data. Furthermore, large areas of the world remain where little information on food composition is available.

In 1983 a conference was held at Bellagio (Italy), under the auspices of the United Nations University, to identify the tasks that needed to be carried out in order for internationally valid, consistent and usable food composition data to become available. During the discussions the creation of an International Network of Food Data Systems (INFOODS) was proposed (Rand and Young, 1983).

One of the first tasks for INFOODS was to revise and extend the earlier Southgate (1974) guidelines, which addressed issues relevant to the central problem of data quality and compatibility. Accordingly, one of us (HG) spent four months as an INFOODS Fellow working with the original guidelines' author (DATS) at the Food Research Institute in Norwich (United Kingdom). This initial work, continued and completed by correspondence, drew information from production and management of food composition data in the United Kingdom and United States and from Australian experience of producing data. In January 1985, a partially completed version was reviewed by a working group in Washington, DC (United States). A revised version, prepared on the basis of this review, was reviewed again by a number of international authorities; their comments were used in the version prepared in 1986.

After reviews by experts in the computer field, and considerable inputs from participants in the FLAIR Concerted Action No. 12 EUROFOODS-Enfant Project, the final revised version was prepared by correspondence and meetings between the authors while one of us (HG) was a Visiting Scientist at the International Agency for Research on Cancer (IARC), Lyon (France) in connection with the Nutrition and Cancer Programme.

In preparing a document of this kind, personal feelings and prejudices inevitably emerge; they are the responsibilities of the authors alone, who nonetheless beg their readers to remember that these idiosyncrasies developed during lengthy consideration of nutritional compositional data, their production and use.

Acknowledgements

For the first edition
We are grateful to INFOODS (Dr N.S. Scrimshaw and Dr V.R. Young) for providing the initial impetus forthe project and for financial support which enabled its commencement. Thanks are also dueto Prof R.F. Curtis, AFRC Food Research Institute, Norwich (United Kingdom) for administrative support of the first phase of the project. In addition, thanks are due to the many people who contributed ideas, skills or information for the initial draft. They include: the INFOODS review committee members, N-G. Asp, R. Bressani, M. Deutsch, H. Herstel, J.C. Klensin, J. Pennington, W.M. Rand, R. Sawyer, W. Wolf, V.R. Young. In the United Kingdom: A. Broadhurst, D.H. Buss, J.R. Cooke, K.C. Day, R.M. Faulks, A.A. Paul, L. Stockley, G. Mason, E.M. Widdowson. In the United States: G. Beecher, F. Hepburn, J. Holden, B. Perloff, K.K. Stewart. In Italy: F. Fidanza, J. Perissé, W. Polacchi. In the Netherlands: R. Breedveld, A.E. Cramwinckel, M.B. Katan, M. van Stigt Thans, C.E. West. In Indonesia: D. Karyadi. In Thailand: A. Valyasevi, K. Tontisirin. In India: K. Pant, K. Doesthale, B.S. Narasinga Rao. In Australia: K. Cashel, R. English, G. Hutchison, A.R. Johnson, J.H. Makinson, A.S. Truswell, R.B.H. Wills, M. Wootton. In Sweden: Å. Bruce, L. Bergström.

We are particularly grateful to Dr C.E. West and the FLAIR Concerted Action No. 12 EUROFOODS-Enfant Project for enabling the completion and publication of this book and to Dr L. Tomatis (Director) and Dr E. Riboli (Head, Nutrition and Cancer Programme) of the International Agency for Research on Cancer for administrative support for completion of the book for publication. Our thanks are due to the participants of the FLAIR EUROFOODS-Enfant Concerted Action for reviewing the final draft: A. Amorin Cruz (Portugal), W. Becker (Sweden), H.K. Hendrickx (Belgium), P. Hollman (Netherlands), M.T. Fernández Muñoz (Spain), I. Martins (Portugal), D.L. Massart (Belgium), M.L. Ovaskainen (Finland), A.H. Rimestad (Norway), I. Torelm (Sweden) and C.E. West (Netherlands). We are very grateful for their comments, which have been extremely valuable in preparing the final version. Thanks are due also to W. Horwitz for comments on Chapter 5. We also acknowledge the advice of J. Cheney, B. Hémon and M. Friesen (IARC).

For the second edition
The authors would like to express their deep gratitude to B. Burlingame, Director of INFOODS (Food and Agriculture Organization of the United Nations [FAO]/United Nations University) for initiating and resourcing the second edition under the aegis of FAO. They also acknowledge the work of B. Burlingame and R. Charrondiere (FAO) for revisions and updates to the manuscript.

For this edition, the authors and editors are grateful to the following people for their reviews: W. Schüep (Switzerland), H. Schonfeldt and L. Smit (South Africa), S. Gilani (Canada), P.J.M. Hulshof (Netherlands), A. Sinclair (Australia), P. Finglas (United Kingdom), H. Boon (Australia) and the members of the INFOODS Food Composition mailing list for their responses to surveys. We also acknowledge the work of G. di Felice (FAO) and S. Debreczevi (UNSW) for secretarial assistance.

Introduction

A knowledge of the chemical composition of foods is the first essential in dietary treatment of disease or in any quantitative study of human nutrition.

(McCance and Widdowson, 1940)

This statement is as true now as it was in 1940, when it formed the first sentence in the introduction to the book that has now evolved into the United Kingdom National Nutritional Database (Food Standards Agency, 2002a).

The source of information on the composition of foods was, traditionally, printed food composition tables; these are now being replaced by computerized compositional databases from which the printed versions are usually produced. The information is widely used in the health, agriculture and trade sectors.

The data are used in research studies of the effects of diets on health, reproduction, growth and development. Data are also used for devising diets with specific nutrient composition in clinical practice, in the formulation of ration scales and in the devising of emergency food supplies. Nationally and internationally, compositional data are used in the assessment of the nutritional value of the food consumed by individuals and populations.

The recognition of the involvement of diet in the development of many diseases (McGovern, 1977) has led to an expansion in the number and range of studies of the relationship between diet and health and disease, which has led to a greater focus on nutrient data. Willett (1998) has drawn attention to this and to the need for databases to be reviewed regularly: "Diets of human populations are extremely complex … Maximal insight into the relation between diet and disease will usually be obtained by examining diets both as constituents and as foods. Calculations of intakes of nutrients and other constituents require a food composition database that is complete and current."

The evidence that has emerged from these epidemiological studies has led to a growth in the production of national and international guidance on choosing a healthy diet. Composition data provide the foundations for the development of education programmes on choosing healthy diets. As part of this guidance to consumers, many governments have implemented

the nutrition labelling of foods. Some countries require the producers of food products to provide their own analytical data on the composition of their products.

However, in appropriate cases, most regulations allow the use of compositional data taken from an authoritative compilation, such as a national food composition database, as a substitute for direct analysis. This development has added a quasi-regulatory role to food composition databases and strengthens the need for maintenance of data quality in terms of both the representativeness of the samples and the quality of the analytical data.

Establishing the composition of foods often has advantages for the trade in foods because importing countries with nutrition-labelling regulations prefer (and may require) that imported foods conform to the standards expected of locally produced foods.

Computerized databases have substantial advantages over printed food composition tables: they can contain a greater volume of information and the data can be used in calculations much more easily. The information can also be reformulated in different ways relatively easily to accommodate the needs of different users.

These advantages of calculation from computerized databases are especially important for nutritional epidemiologists, who frequently have to work with very large numbers of subjects and a large number and variety of food consumption records.

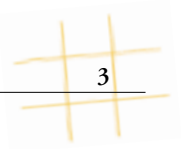
The power of epidemiological studies can be greatly enhanced when they are implemented at the international scale. For this to be effective requires, first, compatible records of food consumption and, second, national databases that are compatible. Compatible in this context implies "capable of being used together".

Achieving a worldwide system of compatible food composition databases lies at the heart of the INFOODS programme. INFOODS – the International Network of Food Data Systems – was established in 1984 on the basis of the recommendations of an international group, and it operates under the auspices of the Food and Agriculture Organization of the United Nations (FAO) and the United Nations University (UNU) (Scrimshaw, 1994). Its goal is to stimulate and coordinate efforts to improve the quality and availability of food analysis data worldwide and to ensure that anyone, anywhere, would be able to obtain adequate and reliable food composition data. It has established a framework for the development of standards and guidelines for the collection, compilation and reporting of food component data.

This book is a continuation of the INFOODS effort, building upon earlier books (Klensin *et al.*, 1989; Rand *et al.*, 1991; Klensin, 1992; Greenfield and Southgate, 1992). The principles and guidelines contained in this book are intended to aid individuals and organizations concerned with the construction of food composition databases. The primary objective is to show how to obtain information that will meet the requirements of a database system that is compatible with systems that have already been, or are being, developed worldwide.

The book focuses on the areas of information-gathering that are critical in determining data quality and must therefore be closely controlled.

It is important to recognize that the term "guidelines" is not used in a prescriptive sense but in the sense of the "principles" of preparing databases. These principles draw on and are



a result of experience gained in the preparation of databases over many years and in different countries. The guidelines do not set out detailed sampling or analytical protocols but provide examples of approaches that have been used successfully. In many countries, the protocols that should be followed are set out within a legal framework that must, of course, be followed. However, by discussing and setting out the available options the guidelines may suggest where established programmes might be revised.

The nutritional and analytical sciences are developing continuously and these developments may indicate better approaches than those set out in these guidelines. It is expected that these principles will serve as a framework for the future development of food composition data programmes.

The structure of the book follows the stages in an idealized programme of work in preparing a food composition database. Chapter 1 describes the variety of uses of a food composition database that the compilers (those with executive responsibility for collecting and assessing the data to be used in the database and their presentation) have to meet. Chapter 2 describes the overall design of programmes for creating, or revising, a food composition database. Subsequent chapters deal with the selection of foods for inclusion (Chapter 3) and the selection of nutrients (Chapter 4). Chapter 5 describes the principles of sampling foods and Chapter 6 deals with the selection of analytical methods and their evaluation. Chapter 7 presents a review of the methods available for the nutrients, focusing on methods that have been shown to be compatible internationally. Chapter 8 describes the principles of assessing the quality of analytical data. Chapter 9 describes the presentation of data and the modes of expression that are central to producing compatible data. Chapter 10 discusses the compilation of data for inclusion in the computerized database. The processes and design of computerized systems for compositional databases lie beyond the scope of this book. Chapter 11 deals with the intrinsic limitations of nutrition databases that constrain their use. The chapter also provides guidance on the proper use of the food data. Finally, Chapter 12 discusses the future needs in the area of food composition.

Chapter 1

Food composition data and food composition databases

Early food composition studies were carried out to identify and determine the chemical nature of the principles in foods that affect human health. These studies were also concerned with the mechanisms whereby chemical constituents exert their influence and provided the basis for the early development of the science of nutrition (McCollum, 1957), and they continue to be central to the development of the nutritional sciences. Current knowledge of nutrition is still incomplete, and studies are still required, often at an ever-increasing level of sophistication, into the composition of foods and the role of these components and their interactions in health and disease.

Somogyi (1974) reproduced a page of the earliest known food composition table, dated 1818. Ever since, it has been customary to record food composition data in printed tables for use by both specialists and non-specialists. While printed tables will continue to be produced, computerized data systems have replaced them in some settings because of the ease with which data can be stored, and the facility with which the large amounts of data can be accessed and processed.

These systems are increasingly used to generate printed and computerized food composition tables and data files. Computerized and printed tables generally contain a subset of nutrients and foods and often no further documentation. A single computerized data system can generate a variety of tables and files, each containing specific subsets of numeric, descriptive and graphical information. Examples are the different user databases released by New Zealand (Burlingame, 1996).

Studies of the relationship between diet and health have led to increased interest in the range of biologically active constituents present in foods that accompany the nutrients, and data for these constituents are often required, as are data for additives and contaminants. A well-designed data system can accommodate non-nutrient data, although this should not detract from the primary objective of the database programme – the provision of data on the nutrient content of foods.

Methods of compiling food composition databases

Early food composition tables were based on analyses carried out in the laboratories of researchers such as Von Voit in Germany, Atwater in the United States of America and Plimmer in the United Kingdom (UK) (Somogyi, 1974; Atwater and Woods, 1896; Widdowson, 1974). Later, the United States moved towards compiling tables from scrutinized data produced by a number of laboratories. An element of this procedure was introduced into the UK tables, where the third edition of McCance and Widdowson (1940) included vitamin and amino acid values drawn from the literature. Southgate (1974) distinguished these two methods as the direct and indirect method of compiling tables. These methods, and other procedures for compiling food composition data, were described by INFOODS (Rand *et al.*, 1991).

Direct method

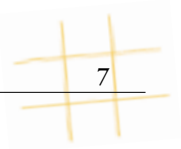
The advantage of the direct method, in which all of the values are the results of analyses carried out specifically for the database being compiled, is that close control of the sampling, analysis and quality control procedures yields highly reliable data. Early UK food composition workers analysed different purchases of the same food separately, but without duplicate determinations, with the intention of gaining some limited information on nutrient variation in each food (McCance and Shipp, 1933). In subsequent versions of the UK tables, however, the various purchases of the food were combined, reducing costs and increasing the number of foods that could be analysed in a given period of time (McCance, Widdowson and Shackleton, 1936). Even with this procedure, the direct method remains costly and time-consuming, and imposes pressure on the analytical resources available in many parts of the world.

Indirect method

The indirect method uses data taken from published literature or unpublished laboratory reports. There is consequently less control over the quality of the data, which may be uneven. Great care must therefore be taken in their appraisal for inclusion in the database. In some cases, values are imputed, calculated (see below), or taken from other tables or databases, and it may be impossible to refer back to the original source; these values carry a lower degree of confidence. The indirect method is most commonly employed when analytical resources are limited, or the food supply is largely drawn from food imported from other countries where compositional data are available. Although the indirect method is clearly less demanding of analytical resources than the direct method, the level of scrutiny required often makes it time-consuming and costly.

Combination method

Most food composition databases nowadays are prepared by a combination of the direct and indirect methods, containing original analytical values together with values taken from the literature and from other databases as well as imputed and calculated values. This combination method is the most cost-effective and is particularly successful when staple



foods are analysed directly, and data for less important foods are taken from the literature (including that from other countries, if necessary). However, minimization of the amount of imputed and calculated values in principle increases the reliability and representativeness of the database.

Types of food composition data

Food composition databases currently available contain compositional values of differing quality, reflecting the different ways in which they were obtained. If data are to be used internationally they must be of consistent and compatible quality so that they can be used in combination for collaboration between individuals and countries in nutritional research, nutrition education, food regulation, and food production and processing. Data types and sources can be identified in food composition databases by codes (USDA, 2003a; Burlingame *et al.*, 1995a), as is done in many countries, and by reference (Wu Leung, Butrum and Cheng, 1972). In general order of preference, the sources of data are:

Original analytical values

These are values taken from the published literature or unpublished laboratory reports, whether or not they were from analyses carried out explicitly for the purpose of compiling the database. They may be assimilated into the database unmodified, or as a selection or average of analytical values, or as combinations weighted to ensure that the final values are representative. Original calculated values are included in this category (e.g. protein values calculated by multiplying the nitrogen content by the appropriate factor, or fatty acids per 100 g food calculated from fatty acid values per 100 g total fatty acids).

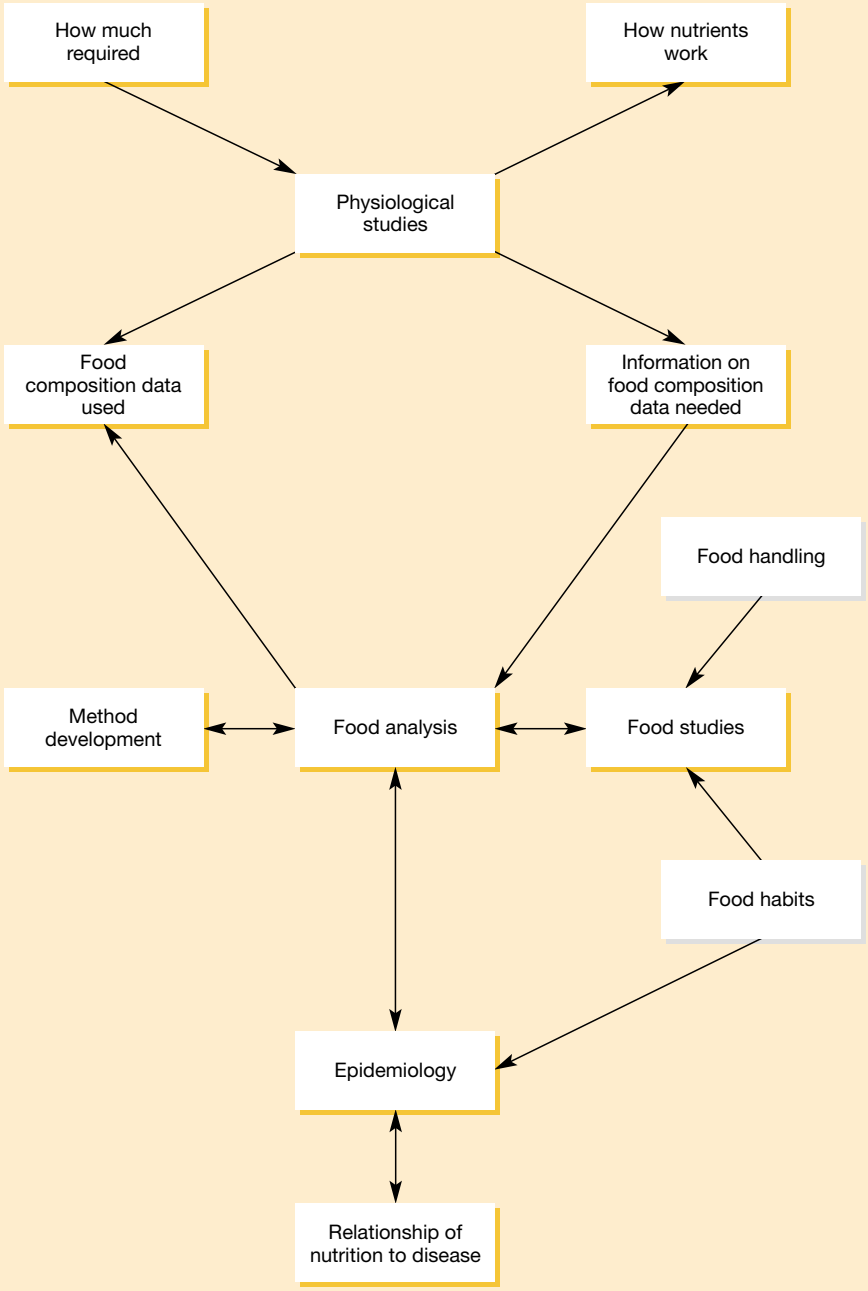
Imputed values

These data are estimates derived from analytical values obtained for a similar food (e.g. values for peas used for green beans) or for another form of the same food (e.g. values for "boiled" used for "steamed"). They may also be derived by calculation from incomplete or partial analyses of a food (e.g. carbohydrate or moisture by difference, sodium derived from chloride values or, more commonly, chloride calculated from the value for sodium). Similar calculations can be made by comparing data for different forms of the same food (e.g. "dried" versus "fresh" or "defatted" versus "fresh").

Calculated values

These are values derived from recipes, calculated from the nutrient contents of the ingredients and corrected for preparation factors: loss or gain in weight, usually referred to as yields, and micronutrient changes, usually referred to as retention factors. Such values are only rough estimates, because the preparation conditions for recipes vary dramatically, such as cooking temperature and duration, which will significantly affect yield and retention. Another

Figure 1.1 The integration of nutritional analyses of foods into food and nutrition research



calculation method is the calculation of the nutrient values of cooked foods based on those of raw foods or foods cooked in a different way, using specific algorithms, retention and yield factors.

Borrowed values

These are values taken from other tables and databases where reference back to the original source may or may not be possible. Adequate reference to original sources is necessary to justify a borrowed value. In some cases, the borrowed values should be adapted to the different water and/or fat contents.

Presumed values

These are values presumed as being at a certain level or as zero, according to regulations.

Sources of food composition data

Foods are chemically analysed for a variety of purposes. Food composition databases rely on nutritional and toxicological analyses conducted by government, academia and industry to determine the potential contributions of foods to the diet, and to determine compliance with regulations concerning composition, quality, safety and labelling. Foods may also be analysed for the purpose of ongoing monitoring of the food supply (e.g. Bilde and Leth, 1990). All of these compositional studies produce data that can be considered for entry into a food composition database.

Nutritional evaluation of foods

In human nutrition studies, the composition of foods is investigated, ideally, in a research setting interacting with one or more other areas of nutrition research (Figure 1.1). The data are most useful when they represent foods in the forms generally consumed (see Chapter 5, Sampling).

In agriculture, factors such as disease resistance and yield, rather than nutritional value, have tended to dominate decision-making regarding policies and programmes. Similarly, in food technology economic considerations such as consumer appeal and profitability have been the major influences on product development. Attitudes are changing, however, and nutritional quality is now one of the factors considered in cultivar selection and the development of processed foods.

The production, handling, processing and preparation of foods profoundly affect their nutritional quality. Extensive literature covers agricultural practices (climate, geochemistry, husbandry, post-harvest treatments); processing methods (freezing, canning, drying, extrusion); and stages in food preparation (holding, cutting, cooking). Most nutritional studies in these areas, however, cover a limited range of nutrients (notably labile vitamins); very little information is provided on the broad range of nutrients (Henry and Chapman, 2002; Harris and Karmas, 1988; Bender, 1978; Rechigl, 1982). Nevertheless, data from these types of studies can often

be useful in food composition databases, either as data *per se*, or in establishing relevant yield and retention factors for calculations (see Chapter 9).

Food regulations

Levels of certain nutrients, additives and contaminants in foods are monitored for several reasons. Some nutrients, for example, may react adversely under particular processing conditions, producing poor sensory quality or affecting the safety of the food (e.g. *trans* fatty acids). Labelling regulations also require prescribed levels of nutrients in specific foods (e.g. vitamins and minerals in fortified foods, polyunsaturated fat levels in margarine). Certain toxic substances are limited to prescribed levels and are monitored by government, industry and other laboratories. The nutrient content of manufactured foods is rarely made available in electronic format to compilers, and care must be exercised when compiling databases using information provided on food labels.

Management of food composition data

Food composition tables were, in the early development of nutrition, the major resource of food composition data; they are, however, constrained physically by the growing volume of compositional data, and their attendant documentation, or metadata. They are also expensive to update and thus older data can remain in use for longer than is desirable. The most significant disadvantage of food composition tables is that calculations made using the data they contain can only be made with considerable additional work. Computerized compositional databases do not suffer from these disadvantages and are often used instead of the printed tables as the primary sources of compositional data for foods. A comprehensive food composition database should be the repository of all numeric, descriptive and graphical information on the food samples.

This book is concerned with the production and assessment of food composition data intended for entry into a computerized database, but it is equally applicable to data intended for printed food composition tables, because the principles involved are virtually identical.

Food composition data can be managed at four different levels, which together provide an effective way of handling the data (Table 1.1). This approach has advantages for assessing the quality of the data. They form a sequence of stages.

Level 1: data sources

These are the published research papers and unpublished laboratory and other reports containing analytical data, together with their bibliographic references. Normally, the data sources are part of the reference database.

Level 2: archival data

These records (written or computerized) hold all data in the units in which they were originally published or recorded, and are scrutinized only for consistency as would be normal in the refereeing of scientific papers prior to publication. Foods should be coded or annotated to

Table 1.1 Stages in food composition data management

Stages	*Description*	*Format*
Data source	Public and private technical literature containing analytical data, including published and unpublished papers or laboratory reports	As presented by original authors
Archival record	Original data transposed to data record without amalgamation or modification; scrutinized for consistency	One data set per original source to include details of origin and number of food samples, food and analytical sample handling, edible portion, waste, analytical methods and quality-control methods
Reference database	Data from all records for one food brought together to form the total pool of available data	Common format
User database	Data selected or combined to give base mean values with estimates of variance for each food item	Common format

assist in identification, and values should be annotated to indicate unit, calculation, mode of sampling, numbers of food samples analysed, the analytical methods used and any quality assurance procedures in place. Any bibliographic references relevant to the data source are noted. At this stage it is possible to make a preliminary assessment of the data quality (see Chapter 8).

Such records should make it unnecessary to refer back to the original data sources whenever a query arises. Normally, the archival data are used in the preparation of the reference database.

Level 3: reference database

The reference database is the complete pool of rigorously scrutinized data in which all values have been converted into standard units and nutrients are expressed uniformly, but in which data for individual analyses are held separately. This database should include all foods and nutrients for which data are available, and provides links to sampling procedures and analytical methods, laboratory of origin, date of insertion and other relevant information, including bibliographic references to the data sources. The data will usually be expressed according to the conventions, units and bases adopted for the user databases (see Chapter 9).

The reference database will usually be part of a computer database management system, with computer programs or written protocols developed to calculate, edit, query, combine, average and weight values for any given food. It is from this database and its programs that the user databases can be prepared.

The database will be linked to records on analytical methods and records for other constituents, for example non-nutrient constituents such as biologically active constituents, additives and contaminants. Records of physical characteristics such as pH, density, non-edible portion or viscosity that are often collected in food technology papers should also be linked to the reference database. Conversion factors, calculations and recipes should also be stored.

Level 4: user database, printed and computerized tables

In general, the user database is a subset of the reference database, and the printed form often contains less information than the computerized form. Many professional users of food composition data would require the information recorded in the reference database, but most require only a database containing evaluated food composition data that, in some cases, have been weighted or averaged to ensure that the values are representative of the foods in terms of the use intended. Moreover, values for nutrients in each food may, if appropriate, be amalgamated (e.g. total sugars, ratios of the different classes of fatty acids) rather than shown as individual constituents. These databases may contain indications of data quality based on assessment of the sampling and analytical procedures.

These databases should include as many foods and nutrients as possible, with preference being given to complete data sets. Methods, sampling procedures and literature sources should be coded at nutrient level so the user can perform an independent evaluation or comparison with other databases. The data, of course, must be expressed in uniform, standard units (see Chapter 9). The defining feature of a user database may be considered as a database that gives one series of data per food item.

Simplified food composition database or tables

Simplified databases or tables can be produced from the main user database. In these, fewer nutrients are covered, and some reductions of food categories may be possible (e.g. for meat cuts data may appear only for "medium cooked," omitting "rare" and "well cooked"). Values can appear as units per 100 g of food or per average serving, expressed in household units or portion sizes. Modified versions of the database can also be produced to assist manufacturers in food labelling. Various types of database or printed table can be produced from the same comprehensive database, ranging from a fairly extensive version for the professional user to a smaller version for consumers or for users involved in large-scale food preparation.

Special-purpose food composition tables and databases

Tables and databases restricted to selected nutrients can be produced for people with special dietary needs or interests (e.g. for diabetics, or for people with kidney disorders for whom a diet controlling protein, sodium and potassium is required, or for nutrition educators, or for people wishing to lose weight). Data may be presented per 100 g of food, or per portion size or common household measures. Such tables and databases might be produced showing foods with ranges of nutrients – high, medium and low levels, for example. Data could also be given in other useful units (e.g. sodium and potassium in millimoles for renal patients).

Types of food composition database programme

National

Ideally, each country should have an established programme to manage its own food composition data, the data being considered an important national resource, as important as any other national collection of data.

While the level of certain nutrients in some foodstuffs will vary little between countries (e.g. the amino acid composition of lean meats), other nutrients, even in foods that are available worldwide, will vary greatly because of differing cultivars, soils, climates and agricultural practices. Recipes for composite dishes with the same name vary between countries. Different technological practices are also used; flour, for example, is produced and used at different extraction rates and may be fortified to different levels with different nutrients (Greenfield and Wills, 1979). Some countries have unique foods, food products or processing procedures (Somogyi, 1974). For these and other reasons, it is essential to develop a national food composition database programme, and to ensure that such a programme draws on data from other countries only when those values are considered applicable to nationally consumed foods.

Although attempts are being made to develop common food standards (e.g. the Joint FAO–World Health Organization [FAO/WHO] Food Standards Programme, Codex Alimentarius (FAO/WHO, 2003a,b), differences in food descriptions will continue to occur between countries.

Regional

The preparation of regional food composition databases is of great importance. Many countries, particularly in the developing world, lack the resources needed for a full-scale national food composition programme, but share a similar food supply to that of neighbouring countries. Cooperation between United States government departments, the Institute of Nutrition of Central America and Panama (INCAP) and FAO has produced some early regional food composition tables for Latin America (Wu Leung and Flores, 1961), Africa (Wu Leung, Busson and Jardin, 1968), East Asia (Wu Leung, Butrum and Cheng, 1972) and the Near East (FAO, 1982). More recently, this cooperation with FAO/UNU/INFOODS has led to the publication of regional tables for Pacific island countries (Dignan *et al.*, 1994), Latin America (LATINFOODS, 2000) and Southeast Asia (Puwastien *et al.*, 2000).

Some countries are collaborating on food composition analyses among themselves – for example, those in the North European region and those in the South Pacific region (Becker, 2002; South Pacific Commission, 1982). Other regional programmes may be those serving participating countries in multicountry epidemiological studies (Slimani *et al.*, 2000). Simplified national programmes can be derived from such international or regional programmes.

Criteria for a comprehensive food composition database

The current high level of interest in nutrition requires that food composition databases meet the following criteria:

1. **Data should be representative**
   Values should represent the best available estimate of the usual composition of foods in the forms most commonly obtained or consumed. Ideally, some measure of variability in the composition of the food should be given.
2. **Data should be of sound analytical quality**
   Original analytical data from rigorously scrutinized sources are the ideal. Values from other databases, and imputed or calculated data should be included only when original analytical data are not available or are known not to be of sufficient quality.
   High-quality analytical data are those produced by methods that have been shown to be reliable and appropriate to the food matrix and nutrient in question. These methods must be applied proficiently, and evidence of this proficiency is required to assure data quality. It is also desirable that the analyst and the laboratory satisfy criteria of good laboratory practice. Further, evidence is required to show that the food sample was representative and was collected and handled properly. However, for existing data, documentation on sampling, source or analytical method is often not available, at least electronically.
   Chapters 5, 6, 7 and 8 contain more specific guidelines for sampling procedures, methods of analysis and assurance of data quality; these three areas should always be considered in determining the quality of analytical food composition data.
3. **Coverage of foods should be comprehensive**
   The database should include all foods that form a major part of the food supply and as many as possible of the less frequently consumed foods. The selection of foods for inclusion in a database is discussed in Chapter 3.
4. **Coverage of nutrients should be comprehensive**
   Values should be included for all of the nutrients and other components known or believed to be important to human health. National priorities regarding health will have a major role in deciding which nutrients should be included. The criteria for selecting nutrients to be covered are discussed in Chapter 4.
5. **Food descriptions should be clear**
   To be easily identified, foods must be unambiguously named and described. (Food nomenclature is discussed by McCann *et al.* [1988]; Truswell *et al.* [1991]; Møller and Ireland [2000a,b]; and Unwin and Møller [2003].)
6. **Data should be consistently and unambiguously expressed**
   The data should be unambiguous in mode of expression and consistent in the use of units, factors used in calculation, and procedures used in rounding values.
7. **Origins of data should be provided at nutrient value level**
   Information should be given on the sources of the data, noting whether data are analytical,

calculated or imputed, and, as appropriate, on the procedures of any calculation and imputation, and the methods of sampling and analysis. Confidence or quality codes for the values should also be supplied.

8. **Tables and databases should be easy to use**
   In addition to having clear terminology and systematic expression, databases and computerized tables must be easily accessible and readily understood. Printed tables should be of clear legibility and manageable size and weight.
9. **The content of different databases should be compatible**
   The descriptions of foods, modes of expression and derivations of values should conform as closely as possible to existing international standards (e.g. the INFOODS tagnames) and to other major comprehensive food composition databases. Scientific needs require computerized databases and tables to be constructed with a view to using them in combination with other such systems.
10. **Database should have few missing data**
    It follows from the above that any food composition database or table should aim to have as few gaps as possible because missing data can significantly distort the resultant nutrient intake estimations. It may be better to include imputed or borrowed data, always clearly identified as such, than no data at all. On the other hand, practical considerations often dictate that an incomplete database or table be produced to meet immediate needs. Information besides nutrient data (e.g. data on toxic substances or additives), though useful, is not essential at this stage.

Uses of food composition data

Food composition data are used primarily for the assessment and the planning of human energy and nutrient intakes. In both cases, the approach is most useful when applied to groups rather than individuals. Assessment and planning can be divided into several subcategories for which the precise requirements of the database differ and for which additional information is required.

Assessment of nutrient intakes (nutritional analysis)

When the weights of consumed foods are known, food composition data permit the intake of each nutrient to be calculated by multiplying the weight of each food by the concentration of the nutrient in that food and then adding the results, according to the equation:

$$I = \sum(W_1C_1 + W_2C_2 + W_3C_3 + ...W_nC_n)$$

where: I = intake of the nutrient, W_1 = weight consumed of food 1, C_1 = concentration of the nutrient in food 1, etc.

Knowledge of nutrient intakes is required at several levels, as outlined below.

Individual level

A person's nutrient intake can be calculated by the use of food composition data and food intake data (estimated from a dietary history or dietary recall or measured in a weighed intake study) (Cameron and van Staveren, 1988; Nelson, 2000). This information can show gross dietary adequacy or inadequacy, or dietary imbalance, and is important in the determination of dietary advice or in prescription of a therapeutic diet. The user must be aware, however, that due to the natural variability of foodstuffs, food composition data may not predict the composition of a single portion of any particular food with accuracy.

Group level

Foods consumed by populations can be measured by various techniques (Marr, 1971) and translated, by means of food composition data, into nutrients consumed. The results give one indication of the nutritional status of the group (Jelliffe and Jelliffe, 1989; Gibson, 1990) and may be used to explore the relationship of a diet to a variety of health indices – sickness and death patterns, growth rate, birth weight, measures of clinical nutritional status, physical performance, etc. Examples of groups usually studied in this way are:

a) physiological groups, such as growing children, pregnant and lactating women, elderly people;
b) socio-economic groups (e.g. racial, caste, income or occupational);
c) clinical groups, such as patients and healthy controls;
d) intervention groups, usually drawn from the preceding categories, which receive a dietary supplement or other programmes;
e) cohorts in epidemiological studies of diet and health (Riboli and Kaaks, 1997).

Data drawn from studies of groups are used not only for identification of nutritional problems and planning of nutrition interventions to counteract them; they can also be employed in research that seeks to identify nutrient intakes desirable for good health. The results of such studies may feed back into food and nutrition policy in the form of food supplement programmes for children, food stamps for low-income groups, dietary advice to pregnant women, preventive diets for reducing heart disease rates, etc.

National and international levels

National statistics for agricultural production, adjusted for exports, imports, non-food use and gross wastage, are multiplied by nutrient composition data and divided by the total population to produce estimates of gross nutrient availability per capita. These data permit an assessment of the gross adequacy or inadequacy of the national food supply and indicate shortfalls or excesses. Food monitoring systems (e.g. Bilde and Leth, 1990) can follow the consumption of desirable and undesirable substances over a period of years.

Data from individual nations can be assembled to give cross-national or worldwide pictures of food and nutrient availability; such data are used in formulating food and nutrition policy, in setting goals for agricultural production, in formulating guidelines for consumption and particular policies such as food fortification or food supplementation (Buss, 1981).

Internationally, this information has implications for trade and for the development of assistance policies. In research, comparisons of nutrient intakes of different countries, together with other epidemiological data, enable further elucidation of the role of dietary constituents in health and disease. At present, long-term changes in the food supply can only be monitored adequately by the use of up-to-date food composition tables and databases. For example, the fat and iron content of meat have been altered in Western countries by changes in methods of husbandry and butchering. Comparison of today's cuts with those of ten years ago can be made by reference to past food composition tables (Vanderveen and Pennington, 1983).

Subnational and community levels

Similar calculations can be made to provide estimates of the distribution of nutrients within a country. These findings can indicate actual or potential nutritional problems. Such studies are often critically important for developing countries that have diverse geographical regions. Periodic surveys, as part of a full system of nutritional surveillance, can monitor nutritional change and the effectiveness of food and nutrition policies.

Planning, advising or prescribing food and nutrient intakes (nutritional synthesis)

The physiological requirements or recommended intakes of most nutrients have been estimated (e.g. FAO/WHO/UNU, 1985), and it is the job of the nutritionist to translate these requirements or recommendations into desirable food intakes, at varying levels of cost. Again, this task can be performed at several levels, as outlined below.

Prescription of therapeutic diets

A therapeutic diet must be nutritionally balanced and adequate while at the same time controlling the intake of one or more specified nutrients. The prescription of therapeutic diets, therefore, requires professional training and a detailed understanding of the composition of foods. Table 1.2 lists types of disorder that require therapeutic diets, together with the dietary components that must be controlled. Unfortunately, most available food composition tables and databases do not hold information on all of the components listed in Table 1.2, and primary data sources may have to be consulted to obtain the required information.

Planning of institutional diets

Food composition data are used to translate recommended nutrient intakes into cost-limited foods and menus. Large sectors of the population (e.g. military establishments, workplace cafeterias, hospitals, prisons, schools, day-care centres and hotels) are provided with meals in this way.

Table 1.2 Examples of clinical conditions that require food composition information for the planning of therapeutic diets

Clinical condition	*Composition information required*
Requiring general dietary control	
Diabetes mellitus	Energy value, available carbohydrate, fat, protein, dietary fibre
Obesity	Energy value, fat
Hypertension	Energy value, sodium, potassium, protein
Renal disease	Protein, sodium, potassium
Deficiency states	
Anaemia	Iron, folate, vitamin B_{12}
Vitamin deficiencies	Specific vitamin contents
Metabolic disorders	
Haemochromatosis	Iron
Hyperlipidaemias	Fat, fatty acids, cholesterol
Inborn errors of amino acid metabolism	Amino acids
Gout, xanthinuria	Purines
Gall bladder disease	Fat, calcium, cholesterol, dietary fibre
Wilson's disease	Copper
Intolerances	
Disaccharides, monosaccharides	Individual sugars, especially sucrose, lactose, fructose, galactose
Gluten (and other specific proteins)	Gluten, specific proteins
Migraine	Monoamines
Allergies	Specific proteins

Note: This list is not intended to be inclusive.

National food and nutrition policy

A national food and nutrition policy will often define goals for the intake of certain nutrients. These goals must be translated into food production targets for the agriculture sector or into food consumption targets for the market or the public health sector (e.g. through increased subsidy or promotion of certain foods).

Nutritional regulation of the food supply

Food regulators use nutritional data on primary foods or "traditional" food products as a reference point for desirable nutrient levels for processed and newly introduced foods. For example, consumers should be able to rely on a traditional dairy product having certain levels of calcium and riboflavin; new processing techniques should not significantly alter the essential

nutritional quality of the well-recognized product. Similarly, a manufactured or fabricated substitute should provide the same nutritional value as the food it is intended to replace (Vanderveen and Pennington, 1983).

A food composition database can also provide a preliminary check on label information or claims. For example, a food may be advertised as high in nutrient X, and information on the composition of its listed ingredients will indicate whether that food product could be high in nutrient X without fortification (for which special regulations may exist). Further, data on "new" cultivars being evaluated for widespread commercial introduction can be compared with data for traditional cultivars.

Some countries permit the nutrition data used in labelling certain composite foods to be calculated from nutrient data for ingredients taken from food composition tables and databases. In such cases, it must be ensured that nutrient values from the food composition tables and databases are comparable with those of the food regulations concerning food labelling.

Planning of nutrition intervention programmes

Nutrition interventions, such as food aid programmes, supplementation schemes and disease prevention programmes, require the use of food composition data in order to translate specific nutrient needs into food requirements. Note that such programmes may require confirmation by direct analysis, particularly at the research level.

Limitations of food composition databases

The limitations of food composition tables or databases are often not sufficiently understood by many users. Foods, being biological materials, exhibit variations in composition; therefore a database cannot accurately predict the composition of any given single sample of a food. Hence, although food composition tables and databases can be used to devise a diet, meal or supplement, the levels of nutrients are essentially estimates. For metabolic studies a direct analysis is usually necessary to obtain the required accuracy in the measured intake of the nutrients being studied.

Further, food composition databases and tables are limited in their usefulness for regulatory as well as scientific purposes. They cannot predict accurately the nutrient levels in any food; this is especially true for labile nutrients (e.g. vitamin C and folates) or constituents added or removed during food preparation (fat, moisture). Furthermore, the composition of a given food may change with time (e.g. a manufacturer's formulation may change) invalidating the use of the values in the database. Predictive accuracy is also constrained by the ways in which data are maintained in a database (as averages, for example).

Food composition databases frequently cannot be used as literature sources for comparison with values obtained for the food elsewhere. Values from one country should be compared with values obtained in other countries by reference to the original literature. Food composition

databases can be used more confidently when the values are known to be based on original analytical values. Any imputations, calculations, weightings or averaging must be clearly documented and, most important, food items must be adequately described to enable comparisons to be made.

It seems that, despite major efforts during the past 20 years on harmonizing food descriptions, nutrient terminology, analytical methods, calculation and compilation methods, values from existing food composition tables and databases are not readily comparable across countries. In addition, users may not always be aware of the difference in nutrient values between raw and cooked foods and might erroneously use the values for raw foods in place of those for cooked ones. This is often the case in countries using food composition tables that contain mainly raw foods.

Finally, there has been an increase in the consumption of manufactured foods and mineral and vitamin supplements, accounting for up to 60 percent of the total food intake, but these are rarely listed in food composition tables and databases (Charrondiere *et al.*, 2002). As a result, it can be assumed that nutrient intake estimations are increasingly unrepresentative of the actual nutrient intake.

Users

The users of food composition tables and databases vary greatly: economists, agricultural planners, nutritionists, dietitians, food service managers, food and agricultural scientists, manufacturers, food technologists, home economists, teachers, epidemiologists, physicians, dentists, public health scientists, non-specialist consumers and journalists. Access to different types of computerized tables and databases is required to suit these differing needs; this is now achievable due to the availability of computers.

Chapter 2

Initiation and organization of a food composition programme

Over the last decade food composition activities have increasingly been undertaken by a variety of agencies, programmes, projects and people, for an ever-growing number of reasons. Many national, regional and international agencies acknowledge the importance of food composition data and the need to interchange information that is unambiguous and useful to all those who need it (Rand and Young, 1983; Rand *et al.*, 1987; West, 1985; Lupien, 1994).

The creation of a food composition database calls for an integrated approach to the generation, acquisition, processing, dissemination and use of food composition data.

International level

INFOODS, the International Network of Food Data Systems, was established in 1983 by the United Nations University (UNU), with an organizational framework and international management structure that included a global secretariat and regional data centres. Its mandate is "to improve data on the nutrient composition of foods from all parts of the world, with the goal of ensuring that eventually adequate and reliable data can be obtained and interpreted properly worldwide" (INFOODS, 2003). In the mid-1990s, FAO joined UNU in the INFOODS effort. The main activities of INFOODS at the international level include development of technical food composition standards, assistance to regional data centres and individual countries in developing their food composition activities, and publication of the *Journal of Food Composition and Analysis* (Elsevier, 2003).

Most countries in the world participate in international fora and are signatories to international agreements that directly and indirectly relate to food composition. The World Declaration and Plan of Action for Nutrition adopted at the International Conference on Nutrition (FAO/WHO, 1992), the Rome Declaration on World Food Security and the World Food Summit Plan of Action (FAO, 1996), and the World Trade Organization's Agreements on Sanitary and Phytosanitary Measures and Technical Barriers to Trade (WTO, 1998a,b) are examples of such agreements.

Regional level

Currently, there are 17 regional data centres in operation (see Appendix 1). Regional food composition tables have been prepared, both electronically and in printed form (Dignan *et al.*, 1994; de Pablo, 1999; Puwastien *et al.*, 2000), and many regions undertake regular food composition coordination activities and have established technical task forces that involve individual countries in the region.

National level

Most countries now undertake activities relating to the production of food composition data. A national food composition programme is usually the result of the combination and coordination of activities, within a defined administrative framework, related to food composition data generation, compilation, dissemination and use. Many countries have established a steering committee to facilitate such a framework. A steering, or advisory, committee is ideally composed of individuals directly involved in food composition work, that is, the data users, generators, compilers and disseminators. The involvement of data users – agriculturalists, analysts, health professionals, dietitians, nutritionists, food industry personnel and consumer groups – is crucial to the effectiveness of a steering committee.

Often a single organization has overall responsibility for the management of a national food composition programme, yet it is rare that a single organization accomplishes all the activities itself. Regardless of their affiliations, laboratory-based data generators must interact closely with the data compilers, and compilers must interact closely with data users. Data compilers therefore serve the central function and usually also act as data disseminators (i.e. they publish the data, electronically and/or as printed tables). In most countries there also exist other agencies whose activities are directly or indirectly related to food composition data, and who operate in concert with the national programme. National food composition programmes also operate in conjunction with their regional data centres and with ongoing international activities.

The organizational framework of a national programme will depend on the policies and procedures already being followed in the country or region where the programme is being established. Indeed, the national food and nutrition policy of the country concerned may already favour the establishment or updating of a food composition database (e.g. Langsford, 1979); any new programme should generally aim to fit into the framework of the existing national policy.

Many countries will already have experience in the production of food composition data and their use in tables. In developing a database programme, the aim should be to build on this experience. Existing data on foods with known, relatively stable composition can be used in the new database, provided that these data are evaluated and meet the criteria for inclusion.

Programme initiation

A decision to embark on the production or revision of a food composition database may be made by government, or within a research institute or department, by professional groups of users (e.g. dietitians, epidemiologists) or, occasionally, an individual researcher.

The advocacy for newly establishing or revitalizing a database programme can effectively be presented in different ways:

a) a carefully researched document, submitted to a government department or committee by professional or scientific societies or by influential individual scientists;
b) published articles in local scientific or medical journals;
c) a conference or session at a conference, culminating in official resolutions addressed to a government committee, department or other authority;
d) production by users or analysts of an unofficial set of food composition tables or a computerized database;
e) establishment of a formal or informal committee, with representatives from all interested parties, to start up a programme.

Any submitted document should emphasize the potential benefits of such a programme, especially in terms of community health and welfare, national esteem and economic benefits accruing through reduced health costs and advantage to the food industry, agriculture and trade. The availability and usefulness of any existing data and resources should be stressed. In addition, cost estimates that take into account the costs of administration, analyses, data management and data dissemination will be required.

Objectives of a food composition database programme

Any group or individual with responsibility for a database programme should pursue the following objectives:

1. produce a system that meets the multiple needs of users in different sectors;
2. work in the most cost-effective manner possible, within a specified time;
3. maintain full and regular consultation with all interested parties to ensure acceptability of the final product;
4. provide for continuing revision or updating of the data system and for periodic revision of any derived database or tables, according to a specified timetable;
5. publicize the programme widely to ensure that the database and its outputs and updates are widely disseminated and adopted into use;
6. provide for continuous access of all users to the database and related products.

Definition of users' requirements

A food composition database should be defined by the uses for which it is intended. Because such a database is essentially a tool for nutritional work in the widest sense, it must be designed with all immediate and proposed uses clearly defined, and potential users must play a major role in its design.

Three aspects are fundamentally important:

a) the selection of foods to be included (see Chapter 3);
b) the selection of nutrients for which values are required (Chapter 4);
c) the modes of expression to be used (Chapter 9).

When a governmental committee decided to revise the database presented in *The composition of foods* (Paul and Southgate, 1978), a steering panel was set up to define the requirements of users. The panel consisted of users (government departments, dietitians and research nutritionists) and compilers, as well as the person in charge of the analytical work and those responsible for the design of the computerized database. The steering panel consulted major users of the existing tables (dietitians, researchers, food industry) by questionnaire (Paul and Southgate, 1970) and in personal discussions, and invited comments by advertisements in the scientific and food press. The compilers collated this information and used it to plan the revision.

A user questionnaire was also used in the early stages of the Pacific Island Food Composition Programme (Bailey, 1991). Other methods for obtaining suggestions from users are to hold a public meeting (Greenfield and Wills, 1981) or national conference (Food and Nutrition Research Institute/National Research Council of the Philippines, 1985), or to solicit submissions from scientific societies (Bernstein and Woodhill, 1981).

Users' contributions to the programme should be continuous, to ensure that the database is both relevant and practical. It may therefore be useful for professional associations of users (or a consortium of them) to form a committee that would continue to supply information and monitor the programme. Including a session or workshop on the subject at an annual national or regional nutrition conference (e.g. the Sociedad Latinoamericano de Nutrición conference), or holding food composition conferences of the type held annually in the United States (USDA, 2003b), may be useful as a forum for this purpose.

This overall strategy in the design of a database programme and definition of users' requirements is illustrated in Figure 2.1.

Stages of the programme

The stages of an ideal food composition database programme are set out in Figure 2.2. Funding must be obtained and procedures established for communication between all relevant parties. All existing food database programmes and facilities in the country should ideally be coordinated, because much of the analytical work can be done cooperatively by government, research

Figure 2.1 Initiation of database programme: definition of users' requirements

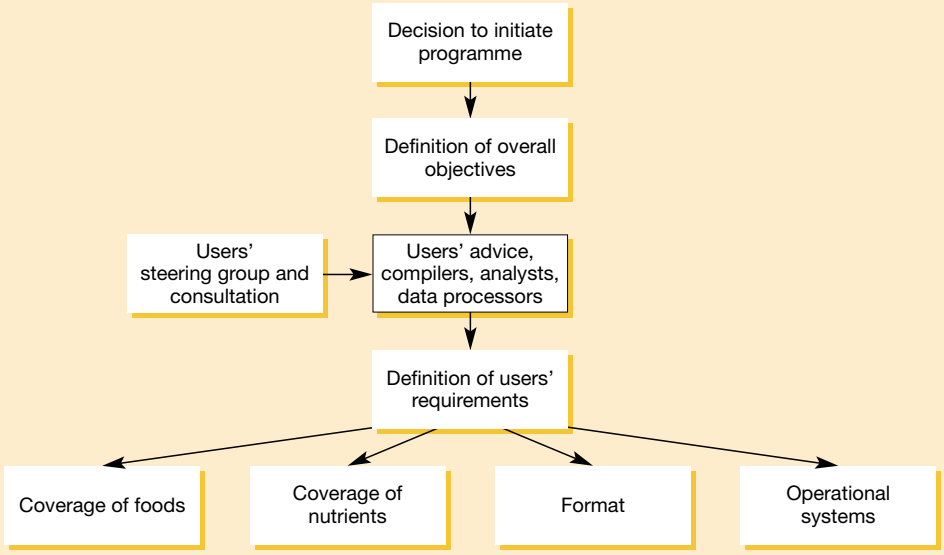
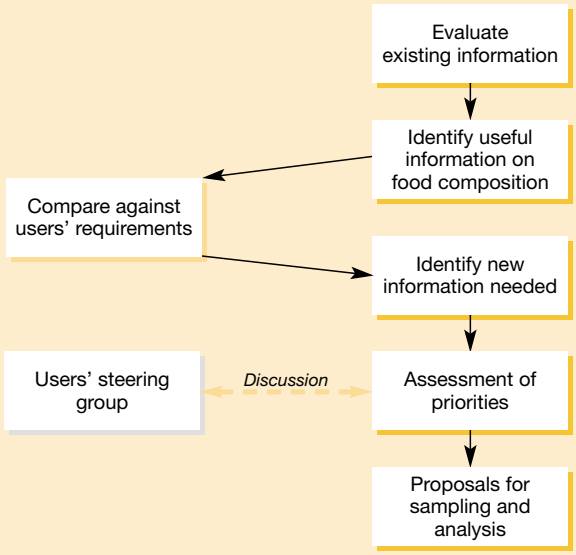


Figure 2.2 Establishing priorities for sampling and analysis



Box 2.1 Major elements in the budget of a food composition database programme

- Meetings (of compilers, analysts, committees)
- Compilers (salaries, support staff, other overheads)
- Food sample purchase and transport
- Analytical programme (salaries, equipment, consumables)
- Expert consultants
- Submissions from users (including attendance at committee meetings)
- Data management and processing costs (including outside contractors)
- Publication costs (print, computer, and online formats)
- Publicizing, dissemination, marketing

institutes, or industry laboratories working in food research or related fields. Facilitation of this collaboration should be an early, important priority.

Obviously, a budget will have to be drawn up; Box 2.1 lists the various items that need to be provided for.

Reviewing, collecting and compiling existing information

Usually, information on the composition of locally available foods already exists, even in countries that have no formal national tables of food composition. The first stage is therefore to evaluate this information, both published and unpublished, for its suitability as data sources (see Chapter 10 for the principles guiding this evaluation). Consideration of user requirements reveals what new information is required, and proposals for new sampling and analytical programmes are made. In most countries it is necessary at this stage to define priorities; this will require further input from the users of the data system.

Sampling and analytical programmes

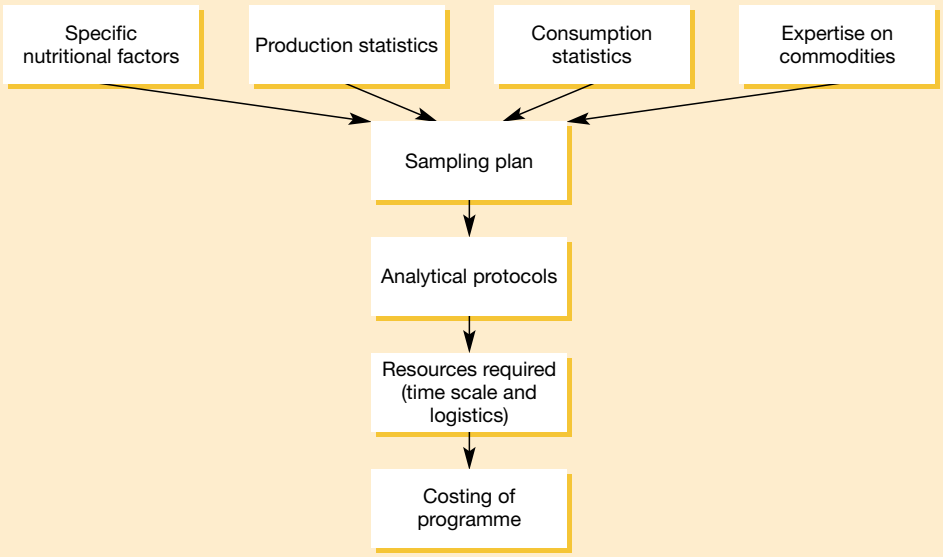
Sampling and analysis should be considered together, both to ensure data quality (Chapters 5, 6, 7 and 8) and also because the resources required for sampling and analysis need to be estimated together.

In developing the sampling plan and protocols (Chapter 5), a considerable range of inputs is essential, and the compilers need to consult widely. If, as happens in many countries, part of the programme is assigned to a contractor, the compiler must ensure that the contractor is aware of user requirements and the quality standards that have been set for data entering the system.

Sampling and analytical programmes are most conveniently focused on specific foods or groups of foods. This focus on specific foods is also useful in defining the experience required of groups invited to tender contracts. This stage is shown schematically in Figure 2.3. The proposed time scale for the work will determine resource requirements, and logistical factors need to be considered carefully. Once these factors have been assessed it is possible to estimate the costs of the different sections of the programme and submit a budget for approval.

Analysts must plan carefully to ensure that a balance is kept between personnel, laboratory space, equipment, running costs, and so on. Analysts preparing budgets or submitting

Figure 2.3 Development of sampling and analytical programmes



contract proposals should highlight funds needed for meeting any specific requirements for their laboratories, as it is unlikely that any laboratory will already be ideally suited to carry out the work. Budgetary considerations will vary from country to country. Where labour is expensive, investment in automated equipment may be most advisable. Where labour is inexpensive, more staff can be employed. Wet chemical methods may be more appropriate if it is difficult to service and obtain parts for instruments.

Tasks in addition to chemical analyses include the regional collection of foods, determination and preparation of edible portions of foods, estimation of serving sizes and consideration of cooking methods (see Chapter 3). Groups with the appropriate technical facilities can carry out this work separately from the analytical programme, if necessary.

Supervision of the analytical programme

In principle, the concept of data quality is built into the analytical procedures (Chapters 7 and 8), and the users' steering group will ensure that the analysts are aware of the detailed requirements of users. Nevertheless, it is useful to review analytical programmes regularly to reinforce the overall objective of the analyses – the construction of a food composition database for many different types of user.

Conversely, analysts should keep the users' steering group informed of both the limitations of, and improvements in, analytical methodology, in order to ensure that the group works with realistic expectations.

Arrangements must be made for regular reports from the analytical laboratories. Requirements for reports must be carefully specified so that all analytical data are provided. For example, a protein value alone should not be accepted if the method used was nitrogen (N) determination. In this case, the N value and the factor used or suggested by the laboratory should be provided along with the calculated protein value. Units and rounding criteria must also be specified for reports. Policies must be established regarding the publication of laboratory results before their release in the food composition database. It is generally desirable for the work to be published independently so that the scrutiny of referees will strengthen its scientific validity.

Evaluation of analytical reports

Data provided by the analytical laboratories are subjected to initial evaluation (Chapter 9), ideally in discussion between compilers and analysts, to ensure consistency. Difficulties that may have arisen during the execution of the work can also be discussed at this time. Inevitably, problems will have required those involved in sampling or analysis to depart from the formal protocols. It is vital that the compilers be fully aware of such changes.

Compilation of the reference database

Once sufficient information has been accumulated, it is desirable to initiate reviews by the users' steering group and by external specialists in the relevant commodity or food. The users' review provides an assessment of whether the objectives defined by the users are being met; furthermore, it provides a means of managing the progress of the programme.

The external review serves as a conventional peer review and ensures that the data being acquired are compatible with specialized knowledge (which may not be nutritionally oriented) regarding the commodity or foods. Where proprietary products are involved it is desirable to submit the data to the manufacturer for comment. This step will identify inconsistencies with the manufacturers' quality-control data and will indicate whether the food samples analysed were representative of normal production.

Compilation of a user database

The compilers should work closely with the users' steering group. A review by users of sections of the database as they are prepared is highly desirable. These reviews enable users to alert the compilers to problems regarding format, user-friendliness and adequacy of data, and enable the compilers to alert users to problems of inadequate data or to indications that further analytical work is needed. As the database nears completion, pilot trials of its operation become desirable. These trials can be organized through the users' steering group.

Operation of the database

Maintenance

Once the database starts to be used, a series of operational studies is desirable. Although studies designed specifically to test the database are valuable (see Chapter 10), the real tests come with regular use, and provision should be made to collect and collate information on difficulties or inconsistencies encountered by users. Errors must be centrally recorded so that the database can be corrected. It is especially important that the database maintenance be seen as a continuous operation.

Updating

It is also desirable to establish a permanent users' group, familiar with the programme's original criteria, which will periodically consider extension and revision of the database.

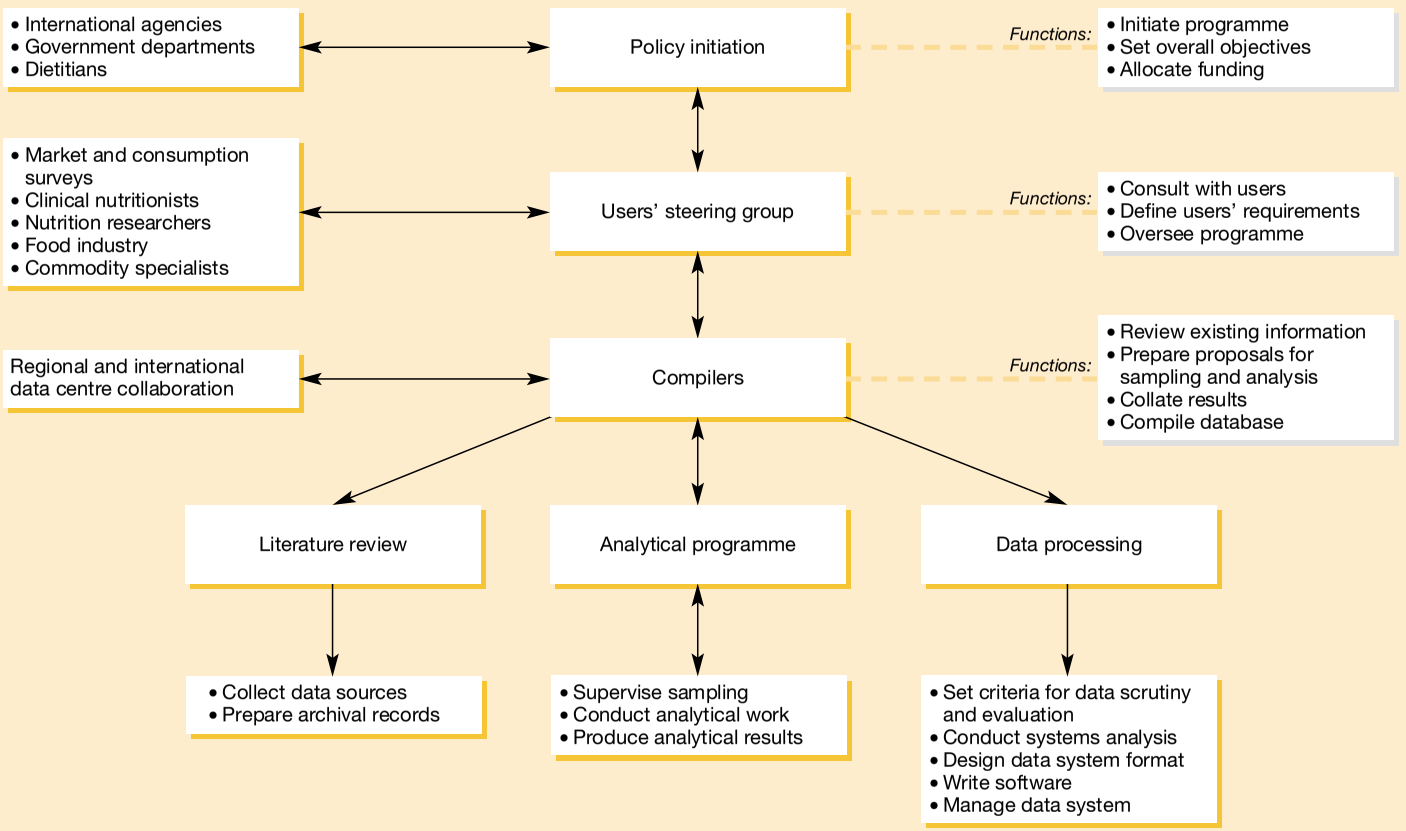
Continuous or periodic revision is essential for several reasons. The level of a food's consumption can change, particularly with the appearance of "new" foods (e.g. instant noodles). The nutritional quality of a traditional food may also change (e.g. changes in animal husbandry and butchering affect the fat content and micronutrient quality of meats). New methods for preparing convenience foods may have striking effects on a food's nutrient composition (e.g. extruded potato-based snacks, depleted of vitamin C) or on its nutritional consequences for sensitive individuals (e.g. the swing towards fructose syrups and sweeteners). Moreover, in addition to changes in foods themselves, advances in analytical methodology may indicate a need to re-analyse foods for a particular nutrient. These trends necessitate continuous nutritional monitoring of the food supply (Paul, 1977) and indicate that a database should be revised from time to time or on a continuous basis. The advent of computer database systems simplifies, in principle, the continuous updating of a database and periodic production of derived databases or tables.

Copyright and other conventions

In view of the fact that copyright and intellectual property legislation varies from country to country (Ricketson, 1995), database compilers will need to familiarize themselves with the national and interna-tional provisions and abide by them. Such provisions may include the need to seek permission to use the data, the format of acknowledgement required and the payment of a royalty. Further, normal scientific conventions should be followed regarding the acknowledgement of all data sources so that users can refer directly to the original source.

The organization responsible for the food composition programme, with the endorsement of the national steering committee, will generally publish the food composition data in various printed and electronic forms, and may charge users for the material cost of the publications. The USDA National Nutrient Database for Standard Reference (USDA, 2003a) is an example of a database that is freely available in the public domain. At the same time, provision should be made for licensing the data for commercial users (Greenfield, 1991b), such as diet analysis software developers, who may then on-sell their product with the data.

Figure 2.4 Schematic overview of the organization of a database programme



Overview of programme structure and organizational requirements

The schematic outline of the programme in Figure 2.4 shows the organizational elements of a food composition database programme and some of the responsibilities of each component. The whole programme requires communication back to the higher level and, indeed, constant interaction as proposals are made, priorities established, work designed and executed, and the final product reviewed. The compilers form the executive members of the programme, ensuring that objectives defined by the users' steering group are met and that quality is maintained.

In practice, the compilers may be several individuals, each responsible for a single area (e.g. literature review, supervision of analytical programmes or data on certain nutrients, commodities or foods). If resources permit this division of labour, which enables specialized knowledge to develop, it is essential to have a good line management so that the senior compiler has a clear overview of the work as a whole.

Continued interaction with the relevant regional data centre is usually helpful in ensuring that standards are maintained and that data are compatible.

Chapter 3

Selection of foods

Most users of food composition databases would like them to be "comprehensive". The objective of the food composition programme is to ensure that the database includes a range of food items that covers as completely as possible the foods eaten by the population for which the database is being prepared. However, the ideal of a truly "comprehensive database" is, in fact, an impossible objective, primarily because of the very large number of foods forming the human diet, especially if one includes all possible variations in the range of cooked mixed dishes. The continuous development of new food products by the food industry and new plant varieties and animal husbandry techniques by the agricultural industry means that analysts and compilers are aiming at a constantly moving target. The volume of analytical work required for comprehensive coverage and the resource implications of this work also make it impracticable. Therefore, those involved in the food composition programme – through a national steering committee or other consultative means – have to develop a strategy for establishing priorities in selecting food items for inclusion.

The approach described below is suitable for use in preparing a database *de novo*. In practice, however, this is very rare because most countries or regions have some existing information available in the form of food composition tables or a computerized database. However, the strategy suggested is equally valid for use in the revision or extension of existing information.

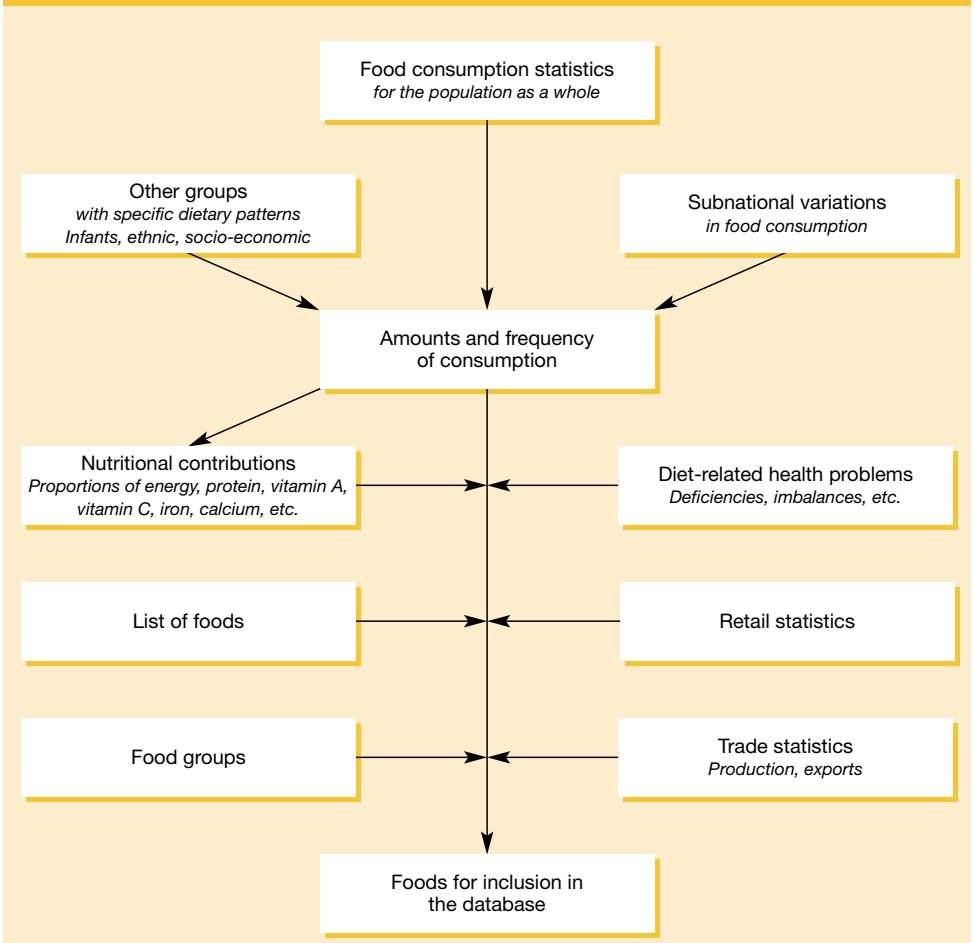
Establishing priorities

A range of different sources of information needs to be considered when establishing priorities. These are summarized in Figure 3.1 on page 34.

Food consumption statistics

The ideal is, first, food consumption statistics. Foods that are most commonly consumed in terms of both frequency and amounts consumed, provide a list of "core foods". In identifying these foods it is necessary to look beyond the statistics for the total population to the

Figure 3.1 Stages involved in the selection of foods for inclusion in a food composition database



consumption patterns of specific subgroups, particularly infants and those with specific dietary requirements. Within the population, ethnic groups with distinctive dietary patterns also need to be considered, as do different socio-economic and regional groups. Data at the commodity level are available from the FAO Statistical Databases (FAO, 2003), and data from household or individual surveys are often available from government ministries of statistics, health or agriculture.

Nutrient contributions

The food consumption statistics should then be used to estimate the nutrient contributions of the different foods (Chug-Ahuja *et al.*, 1993; Schubert, Holden and Wolf, 1987).

The USDA has developed a procedure using food consumption data and nutrient values for developing the *Key Foods* list (Haytowitz *et al.*, 1996). Key foods have been defined as those foods that contribute up to 80 percent of any one nutrient. When total nutrient contributions from the key foods are aggregated, they should account for approximately 90 percent of the nutrient content of the diet for the nutrients examined. This method utilizes existing nutrient profiles and nationally representative data collected from food consumption surveys. More samples are collected and prepared for foods that provide important amounts of nutrients of public health significance to the diet, and not every sample is analysed for all the nutrients currently in the database (Haytowitz, Pehrsson and Holden, 2000). This key-foods approach forms the core of the current USDA nutrient analyses contracts (Haytowitz, Pehrsson and Holden, 2002), and many other countries are adopting this method (Galeazzi *et al.*, 2002).

Nutrients of public health significance in the country

The contribution to energy intakes should be the first to be examined; this establishes the foods that may be considered as the staples in the diet. Other nutrients should be examined in a sequence related to their public health significance. In some countries, protein would be considered next; in other countries, the preferred focus would be the nutrients that are not evenly distributed in foods, for example vitamin A (retinol), vitamin C, iron and calcium. Where iodine deficiency is a public health issue, sources of iodine will need to be included. Vitamin A deficiencies would indicate the need to consider foods that are rich in provitamin carotenoids in addition to sources of retinol. The numbers of additional foods will progressively be reduced using this sequential key-foods type of approach.

Trade and economic factors

The importance of food trade needs to be considered when preparing a list of foods. In food-exporting countries, the list may also need to include the foods most important to the export economy, particularly processed foods in view of the fact that nutrition labelling is required for these by many importing countries.

Preparing a list of foods

Food consumption statistics may be very limited for many populations and in establishing priorities alternative strategies may be needed. One useful approach is to prepare a list of foods consumed and make subjective estimates of their importance. The list needs to be compiled using a number of sources, e.g. government departments, university researchers. As food consumption patterns are largely determined by socio-economic factors, it is important to involve those sectors of the community in preparing the list.

Food production and retailing statistics may also be useful sources of information to assist in constructing the list. The Food Balance Sheets and Food Supply databases published by FAO, which are available for most countries, also provide data on national domestic availability of foods and their per capita contributions to the energy, protein and fat supplies (FAO, 2003).

Table 3.1 Examples of major food groups used in food composition databases and tables

FAO food tables for the Near East[1]	*Pacific Islands food composition tables*[2]	*United Kingdom food tables*[3]
Cereals and grain products	Cereals and cereal products	Cereals and cereal products
Starchy roots and tubers	Starchy vegetables	(included in vegetables)
Dry grain legumes and legume products	Legumes	(included in vegetables)
Nuts and seeds	Nuts and seeds	Nuts
Vegetables	Other vegetables	Vegetables
	Green leaves	
Fruits	Fruits	Fruit
Sugars, syrups and sweets	Confectionery	Sugars, preserves and snacks
Meat and poultry	Meat and poultry	Meat and meat products
Eggs	Eggs	Eggs and egg dishes
Fish and shellfish	Fish	Fish and fish products
	Seafood	
Milk and milk products	Milk and milk products	Milk and milk products
Oils and fats	Fats and oils	Fats and oils
Beverages	Beverages	Beverages
		Alcoholic beverages
	Herbs, spices, sauces	Herbs and spices
Miscellaneous		Soups, sauces and miscellaneous foods
	Processed foods	
	Mixed cooked dishes	
	Coconut products	
	Wild animal foods	

Sources:
[1] FAO, 1982.
[2] Dignan *et al.*, 1994.
[3] FSA, 2002.

Use of food groups

It is often convenient to structure a food composition database using food groups. This ensures that the diet as a whole is considered and that the focus is not distorted by emphasizing one food group at the expense of the diet as a whole.

There is no internationally standardized approach to food groupings. At the 16th International Congress of Nutrition, the INFOODS presentation reported on the issue of food groupings (Burlingame, 1998).

Most food composition databases have between 10 and 25 food groups. Even though the concept of food grouping seems to be internationally agreed upon, the actual classification of food has been shown to be highly culturally dependent and most national databases have unique examples.

The Pacific Islands food composition tables (Dignan *et al.*, 1994), for example, have coconut products as a group because of the economic and cultural importance of this food and the diversity of products. Other countries divide coconut products between several different food categories such as fats and oils for coconut oil; nuts and seeds for coconut flesh; beverages for coconut water. The Central America and Panama (INCAP) database has three groups that are unique: bananas, maize and cornbreads (FAO/LATINFOODS, 2002). The ASEAN food composition database has edible insects as a group (Puwastien *et al.*, 2000).

Researchers and nutritionists in international organizations often report population nutrient intakes by food group rather than by individual foods, suggesting the importance of standardization for international data comparison. The food groups used in the past by FAO (1982), and currently in the UK (Food Standards Agency, 2002), and Pacific Islands food composition tables (Dignan *et al.*, 1994) are shown in Table 3.1.

Identifying priorities for revision of an existing database

The procedure when revising an existing database is very similar to that of compiling a new one, but it will be necessary to consider also which foods may need updated values.

Changes in food consumption patterns should be taken into account, and the values for food items for which there is evidence, even presumptive evidence, that the food has changed in composition since the last database was prepared should be reviewed. Changes in food production – both primary in agriculture, and secondary in food processing, marketing and storage – will also need to be considered. Consultation with the food industry and, where possible, with research groups specializing in the study of specific commodities, often provides useful information on changes that have taken place.

Selection of foods within food groups

Figure 3.1 (page 34) illustrates the stages in the establishment of priorities and the selection of foods for inclusion in the database. At the level of specific foods in each group the strategy requires knowledge of the marketing and consumption of foods. This information will also be required in drawing up the sampling protocols, which is discussed in Chapter 5.

Information will be required from departments of agriculture, commodity boards, trade associations and research groups involved in the study of specific foods. Retail trade journals and consultations with food manufacturers can also provide information on the relative market shares of different brands of the same product. The inclusion of proprietary or brand-

Table 3.2 Examples of possible groups and subgroups for food composition databases and tables

Food group	*Possible subgroups*	*Comments*
Cereals and cereal products	Grains and flours	Including cereal-based prepared foods
	Cereal products (breads, pasta, tortillas, sweet biscuits, savoury biscuits, cakes, doughs, crispbread)	
	Breakfast cereals	
Vegetables and vegetable products	Roots, tubers, stems, corms, plantains	Including textured vegetable protein, leaf protein, soy products, fungi, vegetable juices, algae
	Leafy vegetables	
	Legumes and their seeds	
Fruits and fruit products	Fresh fruits (berries, citrus fruit, etc.)	
	Processed fruits, including juices	
Nuts and seeds		Including oilseeds
Oils and fats	Seed oils, marine oils, margarines	Including ghee, butter, oilseeds
Fish and fish products	Fish and their eggs	Including echinoderms and other marine animals
	Molluscs and their eggs	
	Crustacea and their eggs	
	Processed fish (dried, salted, smoked, canned)	
Meat and meat products	Subgroups for various meat species	Including amphibians, reptiles, marsupials
	Poultry and game	
	Offal	
	Processed meat products	
Eggs	Subgroups for various species	Including egg-based dishes
Milk and milk products	Subgrouped by species; creams, yoghurts, cheeses, milk-based cream desserts	Including ice creams
Sugars and syrups	Sugars, syrups, confectionery, desserts, jams, jellies, preserves	
Beverages	Teas, coffees, cordials, soft drinks, fruit-flavoured drinks	Including carbonated drinks but excluding milk and fruit and vegetable juices
Alcoholic beverages	Beers, wines, fortified wines, spirits, liqueurs	
Miscellaneous	Herbs, spices, condiments, leavening agents	

(Continued)

Table 3.2 (Continued)

Food group	*Possible subgroups*	*Comments*
Subgroups based on types of use		
Fast foods	Kebabs, tacos, hamburgers, fried chicken, pizza	
Infant foods	Infant formulas, prepared infant foods	
Special dietary foods	Reduced energy foods, diabetic foods, low-sodium foods	Including parenteral and enteral feeds, therapeutic meal replacements
Manufactured foods	Processed meals, snack foods, packet mixes, soups, sauces, gravies	
Prepared foods	Institutional meals (restaurant meals), domestic meals, recipe-based meals	
Non-cultivated foods	Wild plants and animals	

name foods should be restricted to stable, well-established lines if frequent revision or updating of the database is not possible. It may be possible to include brand-name foods where these products are unique, or combine foods such as cheeses (e.g. hard cheeses, blue-vein cheeses) or biscuits (e.g. sweet, savoury, filled) into generic compositional types.

Once a clear idea of the relative importance of various foods has been reached and a provisional list of candidate foods for inclusion drawn up, existing compositional data on these foods should be examined following the principles set out in Chapter 10. This process will review the quality of the data and their applicability to the food currently consumed and will establish whether or not sampling protocols need to be developed to provide the necessary data for their inclusion.

It is often useful to group the foods at this point into subgroups as outlined in Table 3.2. These may be arranged according to the type or use of the foods. Subgroupings of foods with similar matrix and nutrient characteristics often provide a convenient basis for developing common sampling and analytical approaches.

Presentation of foods in the database

The different levels of use of compositional databases require compositional data to be given for foods in the raw state, in the processed state, and as prepared for consumption. Where resources are limited, priority should be given to providing data for the most important foods in their raw state and the most common forms in which they are consumed.

Where foods are commonly consumed in more than one form (e.g. peeled and unpeeled; boiled, fried or roasted), values should ideally be given for all these forms where resources permit. A pragmatic approach may need to be adopted to conserve resources by preparing one form of the food in one way and another type in another way and then extrapolating the composition for the different methods of preparation. For example, different cuts of bacon may be analysed in their raw state and one cut analysed after frying and another after grilling, with the observed changes being extrapolated to all cuts.

The human diet typically includes a wide range of prepared foods with often complex recipes, and it is rarely possible to analyse all the different types of dish. In such cases, it may be decided to calculate the composition of the dishes from the recipes, taking into account the changes in weight on cooking and nutrient retention factors.

The most common cooking methods and the major nutritional changes associated with each are listed in Table 3.3. The table indicates the information required to calculate the composition of the cooked foods from the raw food or ingredients. In some instances calculation is not really suitable and complete analysis should be undertaken if the food is sufficiently important in the diet.

The food preparation may be carried out in a laboratory but it is essential that local cooking methods be reproduced as closely as possible, if examples of the cooked food cannot be collected (e.g. Greenfield and Kosulwat, 1991). Some traditional methods are difficult to replicate in a laboratory, e.g. the Pacific Island earth oven (Kumar *et al.*, 2001) and great care is needed in obtaining values using these methods. In such cases, local knowledge of food cultures, and possibly the advice of anthropologists, is essential to guide the process.

Preparation of edible material

Most databases use analytical values obtained by analysis of the edible material. During the selection of food for inclusion in a database, it is therefore necessary to identify the edible matter to be analysed. This will often be influenced profoundly by the cultural norms of the population for whom the database is being prepared. The inedible portion, or refuse, should also be measured and recorded in the database, since many users, particularly those in food service management, will be calculating nutrient content on the basis of foods as purchased. Table 3.4 provides examples of edible and inedible portions of some foods.

Food nomenclature

Accurate use of any database requires that the food items are correctly identified; thus compilers need to consider carefully how foods are named in the database. Several authors have discussed the issue of food nomenclature (Arab, Wittler and Schettler, 1987; McCann *et al.*, 1988; Truswell *et al.*, 1991).

Table 3.3 Principal cooking methods and estimation of cooking factors

Method	*Description*	*Expected yield*	*Expected retention*	*Experimental measurements*
Boiling, simmering in excess water	Cooked by immersion in boiling water and separated by draining	Loss or gain of water, loss of solids	Loss of water-soluble and heat-labile micronutrients	Measure water content before and after cooking
Water absorption	Cooked by immersion in boiling water, which is absorbed completely	Gain of water	Loss of heat-labile micronutrients	Measure water content before and after cooking
Baking	Cooked by dry heat in enclosed oven	Loss of water	Loss of heat-labile micronutrients. Concentration of components	Measure water and fat contents before and after cooking
Earth oven	Food buried in hot solids	Loss of water	Loss of heat-labile micronutrients. Concentration of components	Measure water and fat contents before and after cooking
Deep frying	Immersed in hot fat	Loss of water, gain/loss of fat	Loss of heat-labile and other micronutrients. Concentration of components	Measure water and fat contents of cooked food. Complete analysis. Weigh remaining fat/oil after cooking if possible
Shallow frying	Cooked in shallow fat on hot surface	Loss of water, gain/loss of fat	Loss of heat-labile and other micronutrients. Concentration of components	Measure water and fat contents of cooked food. Complete analysis. Weigh remaining fat/oil after cooking if possible
Steaming	Wrapped or unwrapped, cooked in moist heat, above boiling water or hot quenched stones	Loss or gain of water	Loss of heat-labile micronutrients.	Measure water content before and after cooking
Roasting	Cooked by dry heat with or without addition of fat	Loss of water, loss or gain of fat	Loss of heat-labile and other micronutrients. Concentration of components	Measure water and fat contents of foods before and after cooking. Complete analysis

(Continued)

Table 3.3 (Continued)

Method	*Description*	*Expected yield*	*Expected retention*	*Experimental measurements*
Grilling	Cooked on rack under/over direct heat	Loss of water and fat	Loss of heat-labile and other micronutrients. Concentration of components	Complete analysis
Microwave	Cooked in enclosed oven by electromagnetic radiation at 915 or 245 MHz	Loss of water	Loss of heat-labile micronutrients. Concentration of components	Measure water content before and after cooking
Braising	Cooked in closed vessel with added liquid and/or fat; may be pre-cooked in fat	Loss or gain of water and fat, loss of solids	Loss of heat-labile and other micronutrients.	Measure water and fat contents before and after cooking
Stewing	Simmered in water in closed vessel on heat source for some time	Loss or gain of water	Loss of water-soluble and heat-labile micronutrients	Measure water content before and after cooking
Open-fire roasting	Cooked on rack or spit over open fire	Loss of water and solids, especially fat	Loss of heat-labile micronutrients. Concentration of components	Complete analysis
Griddle or dry-frying	Cooked on heated metal surface, without added fat	Loss of water, fat and solids	Loss of heat-labile micronutrients. Concentration of components	Measure water and fat contents before and after cooking or complete analysis
Cooking in fire	Cooked in fire	Loss of water and fat, gain of ash	Loss of heat-labile and other micronutrients. Concentration of components	Measure water, fat and ash contents before and after cooking. Complete analysis
Tandoori	Dry-cooked in sealed or covered clay vessel	Loss of water; loss of solids	Loss of heat-labile and other micronutrients. Concentration of components	Measure water and fat contents before and after cooking
Pressure cooking	Cooking in sealed vessel; moist at elevated pressure	Loss or gain of water and fat	Loss of heat-labile and other micronutrients	Measure water and fat contents before and after cooking

Note: All foods and/or ingredients need to be weighed before and after cooking.

Table 3.4 Examples of edible and inedible portion of foods

Food	*Inedible portion*	*Edible portion*
Banana	Peel	Flesh
Cabbage	External yellow or wilted leaves, thick stalks	Remaining leaves and stalk
Canned vegetables in brine	Brine	Drained vegetables
Cheese	(Rind)	(Rind), inner part
Chicken	Bones, (skin from back), some fat pads, (tail), connective tissue	Muscle, skin from breast and leg, subcutaneous fat
Fish		
fresh	Bone, viscera, (head), fins, (skin)	Muscle, roe, (head), (skin)
canned in brine or oil	Bones, brine, (oil), (nil)	Flesh/bones, (oil)
dried, small	Nil	All
Fruit, canned in syrup	Nil	All (solids and liquid may be analysed separately)
Insects	Legs, wings, (head)	Flesh, carapace, (head)
Liver	Blood vessels, connective tissue	Remaining tissue
Meat	Bone, gristle, (fat)	Muscle, (fat), connective tissue
Orange	Peel, albedo, central pith	Segments, residual albedo
Passion fruit	Peel, (seeds)	Flesh, (seeds)
Pineapple	Peel, tuft, base, core	Flesh
Potato, sweet potato	(Peel)	Flesh, (peel)
Pumpkin	Peel, (seeds)	Flesh, (seeds)
Sugar cane	Woody layers, pith	Juice

Note: The inedible portions usually include damaged material. The decision whether a part is edible or not depends on cultural norms and individual preference. The portions in parentheses may or may not be discarded.

Consumers in different parts of a country often give foods different names and the same names are occasionally used for different foods. Provision for a thesaurus of alternative names should therefore be made early in the database compilation process. The names of foods should, as far as possible, be those used by the intended users. Foods covered by legislation with regard to labelling and/or composition should be named in the legally approved way.

Use of faceted descriptor system

The name of a food is frequently insufficient for its unequivocal identification, especially

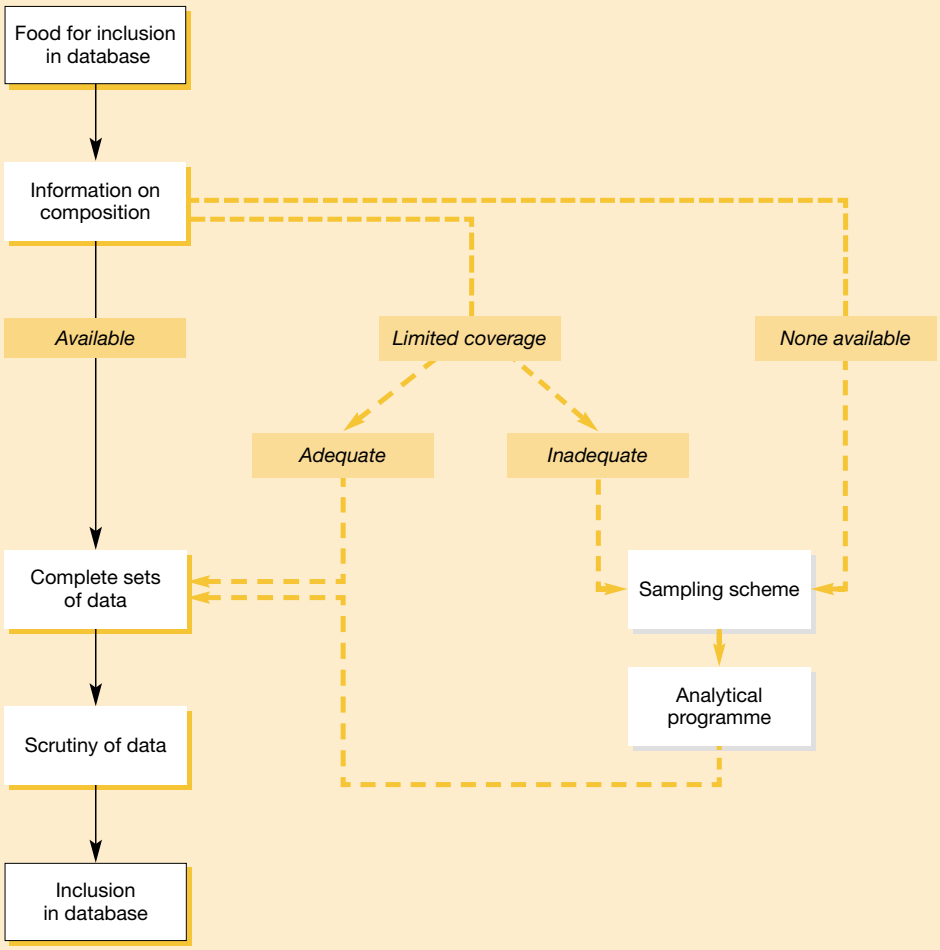
Table 3.5 Facets for use in food nomenclature for identifying foods

Essential facets	*Desirable facets*
	Group, subgroup
Common name (e.g. can be a fixed name, or a string of facets)	Other names, name in local language(s), brand names
Scientific name: genus, species, variety	
Kind/type (e.g. animal source for processed meat)	
Part (e.g. seed, stem, leaf, leg, shoulder, wing)	Maturity
Name of portion analysed (e.g. with or without peel/skin, tissue fat/lean)	Grade
Nature of edible and edible portion	
Origin (country, region)	Husbandry (e.g. pasture-fed, hydroponic)
Processing technique	Added ingredients
Preparation technique	Details of techniques
Special descriptors (low-fat, unsweetened)	
Physical state, shape, size, form, temperature	Extent of preparation (e.g. frozen, thawed, reheated)
Type of fat used in recipe	
Type of liquid used in recipe	
Packaging medium (e.g. brine, syrup)	Pack date, container residence time (from pack date to analysis), shelf life, type of surface in contact with food (important for contaminants)
Short name (fixed character length for outputs such as concise tables)	

Note: This list is not exclusive; all facets that aid identification should be included.

when a national database is used internationally. Food descriptors are therefore needed to identify the foods more clearly and identify the type of preparation used. The use of a systematic series of facets (i.e. properties or attributes) is recommended. A faceted descriptor system permits better searching of large databases, where the same word can represent very different things (e.g. "green" can be a kind of pepper, or a state of maturity), and, when standardized, also facilitates data interchange. Various attempts have been made internationally to standardize systems for naming and describing foods (Truswell *et al.*, 1991; Ireland and Møller, 2000), but international agreement has not yet been reached. The most usual facets are listed in Table 3.5, although any facet that aids identification may be used. Information relating to these facets must be compiled during the collection of samples and their analysis; this has important implications for record-keeping during sampling, which will be discussed in Chapter 5.

Figure 3.2 Decision tree for selection of foods for analysis



Resource implications

The priorities for inclusion of foods in a database need to be considered alongside the priorities for inclusion of nutrients and other constituents because the combined requirements will have implications for the total sampling and analytical resources needed. If a large number of nutrients are to be included this may limit the number of foods that can be analysed using the usually finite resources available, and vice versa. Figure 3.2 illustrates the selection of foods for analysis.

The first essential step is to evaluate any existing information. This may show that complete information, which is still valid for the current food supply, is already available. It may also indicate that where a food is imported it may be possible to use data from the country of origin.

However, the information may be limited, or deemed inadequate, and may need to be supplemented by additional analyses – for example, when a constituent has not been measured before, or where the method of analysis used previously is no longer considered reliable. In such cases sampling and analytical protocols will need to be devised.

Where no information is available and the food is judged important, sampling and analytical protocols will clearly need to be devised.

Finally, all the available data will be scrutinized to ensure that they are of compatible quality. This step also has resource implications, as highly trained personnel will be needed to undertake this important last step.

Chapter 4

Selection of nutrients and other components

The aim of food composition databases should be to include all nutrients or other bioactive food components that are known or believed to be important in human nutrition. This ideal can rarely be achieved, especially where resources are scarce, and therefore decisions must be made on priorities. Some measure of selectivity is both desirable and practicable, particularly in respect of analytical work, which constitutes the major demand on resources.

The following considerations, in addition to the availability of resources, will govern the selection of nutrients and other food components:

a) the basic need for information;
b) health problems in the country concerned;
c) the state of current thinking in the nutritional and toxicological sciences;
d) the availability of existing data;
e) the existence of adequate analytical methods;
f) the feasibility of analytical work;
g) national and international nutrition-labelling regulations.

The stages in this process are outlined schematically in Figure 4.1.

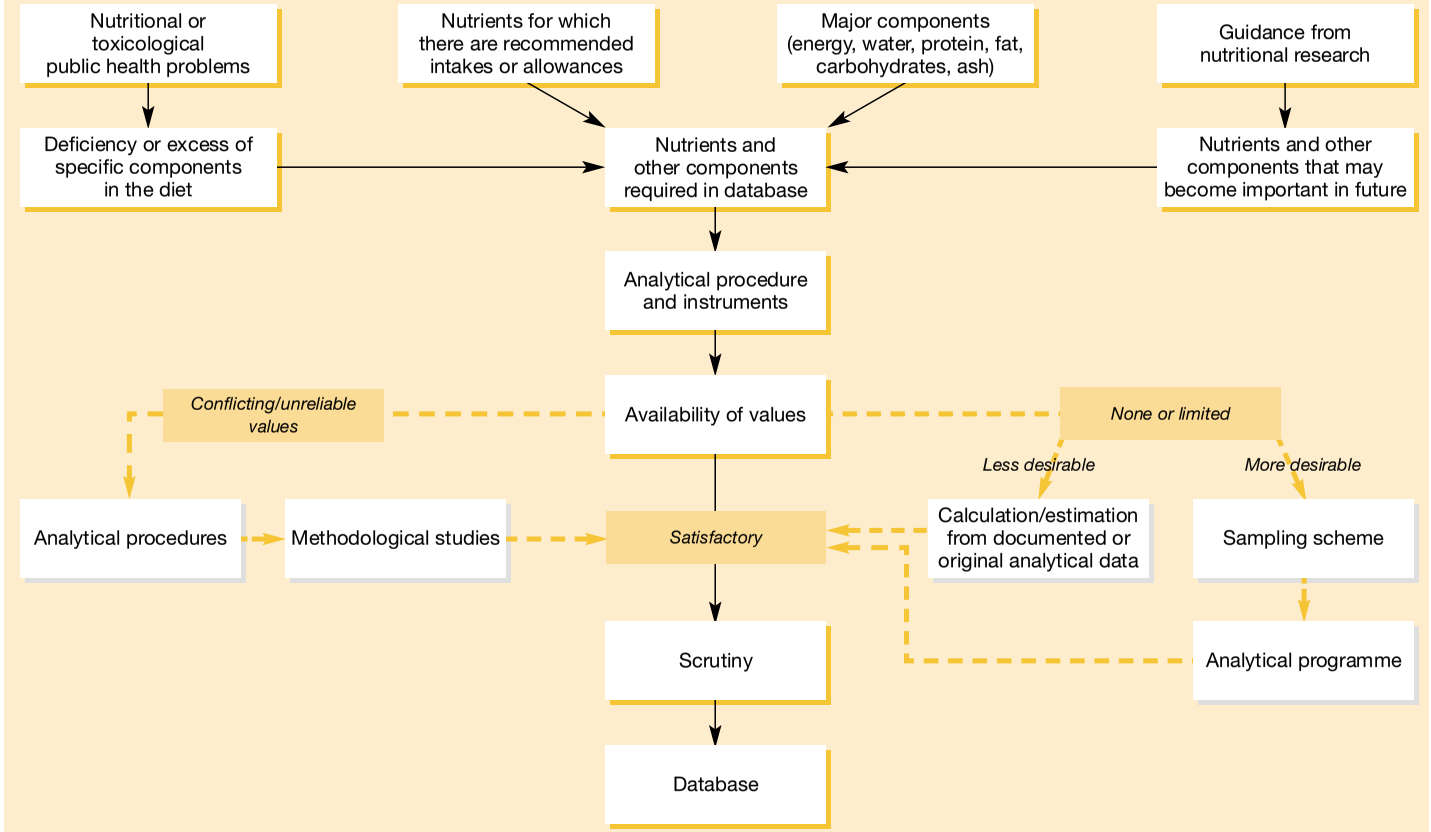
The basic need for information

In all countries, information on water, protein, fats, carbohydrates and energy will be required as a minimum base.

Health problems in the country concerned

In countries where deficiency diseases are a pressing problem, information on key vitamins (e.g. vitamin A) and minerals (e.g. iron) will be required. In industrialized countries, however, where problems such as cardiovascular disease, diabetes mellitus, hypertension and cancer

Figure 4.1 Schematic outline of the process of selecting nutrient values



are predominant, data on energy, fat, fatty acids, cholesterol, individual carbohydrates and sodium may be seen as top priority. All countries with long dark winters, or where sunlight is prevented from reaching the skin for cultural or other reasons (e.g. *purdah*, institutionalization), food levels of vitamin D will be required. This range of constituents will be required worldwide if a complete epidemiological evaluation of degenerative diseases is to be made and guidelines for preventive dietary practices are to be established (Rand and Young, 1983). In a country where toxicological problems have been identified, relevant data on food toxins (e.g. goitrogens) or contaminants (e.g. mycotoxins [Van Egmond, 1984; Van Egmond and Speijers, 1999], heavy metals), may need to receive high priority.

The state of nutritional and toxicological sciences

Food components to be included should also reflect the general state of nutritional and toxicological thinking. A comprehensive database should include all nutrients for which recommended intakes have been established nationally and, where appropriate, internationally.

In addition, those involved with the preparation of databases should try to anticipate needs for data. Interest in "new" or "rediscovered" components of food can escalate rapidly (Southgate, 1985); thus those who are responsible for database programmes must be aware of current developments and the interests of nutritional and clinical scientists. There is now, for example, significant interest in values for the glycemic indices of foods (Brand-Miller *et al.*, 1999). These give a measure of the rate at which carbohydrates are digested (see Chapters 6 and 7) and some tables have been produced (Foster-Powell and Miller, 1995). Caution may be necessary in interpreting responses to questionnaires, however. For example, when Paul and Southgate (1970) reviewed the requests of some users of the United Kingdom food composition tables, they discounted advice to exclude nutritionally unavailable carbohydrates, because they were aware of the growing interest in dietary fibre.

Although these guidelines are primarily concerned with the provision of nutritional information, there is growing recognition that a wider range of constituents play an important role in the relationship between diet and health (Ames, 1983). These include naturally occurring biologically active constituents such as a range of phytochemicals including phytates, oxalates, flavonoids, glucosinolates and phytosterols. Some of these components, such as goitrogens (Gaitan, 1990; Speijers and Van Egmond, 1999) alter the nutritional values of foods, through interactions in the food or gut, or during metabolism. There is also interest in including information on food additives and contaminants in databases (Louekari, 1990; Burlingame, 2001). The amounts of additives in foods are highly brand-sensitive and often subject to variation with time, so it is particularly important for these data to be date-marked. The distribution of contaminants is often more complex than the distribution of the naturally occurring constituents within foods and representative values may be difficult to establish. Furthermore, sampling procedures for contaminants are often designed to identify maximum likely exposure in a population, and it may be misleading to list contaminant values in the

same record as nutrients. For these reasons, these guidelines make only limited reference to contaminants, although their importance is recognized (Young, 1984).

Availability of existing data

A great deal of information is available for certain nutrients or non-nutrient components that have been the focus of research or have been measured for regulatory purposes. These data should be employed, provided that they meet the programme's quality criteria. Where resources are limited and preclude inclusion of all components in the user database, it would still be useful to store all available data at the archival levels of the data system.

Existence of adequate analytical methods

The availability of reliable analytical methods is an essential determinant of components for inclusion (Stewart, 1980) (see Chapters 6 and 7). It will not be cost-effective to analyse foods for a particular nutrient, however high its priority, if methods are untried or yield conflicting values. When methods are in doubt, it may be appropriate for methodological studies to be implemented as part of the database programme.

The emergence of a reliable new or improved method for measuring a nutrient may create the need for analysis (or reanalysis) of foods that are important in the food supply or that are known or suspected to be good sources of the nutrient concerned.

Feasibility of analytical work

The commissioning of analyses for each nutrient must be governed by practical factors: the cost and time required, and the availability of equipment, trained personnel, chemicals, etc. These are major considerations, especially in some developing countries. Costs must always be weighed against the nutritional or clinical requirements for particular nutrients. Where resources are limited it may be useful to search out other laboratories, such as governmental regulatory laboratories or those working on soil chemistry, for collaboration. Borrowing or calculating values would be the final option.

National and international nutrition-labelling regulations

Nutrition labelling has emerged in recent years as one of the more important and demanding areas involving food composition. The key international body concerned is the Codex Alimentarius Commission (FAO/WHO, 2003), operated jointly by FAO and WHO. Complete

food labelling text, with a section on nutrition labelling, is available in print and electronic form (FAO/WHO, 2001). Compliance with Codex Alimentarius is voluntary, and many countries have their own unique nutrition-labelling regulations (FDA, 2001; EC, 1990; FSANZ, 2001). It is useful for food composition programmes to include all the nutrients required in their national nutrition labelling as well as those required in the labelling regulations in countries within their region. For food exporting countries, the nutrients required in the regulations of major trading partners are also important for inclusion in the food composition database.

Coverage at different stages of data management

As noted earlier, ideally, a food composition database system should include values for as many nutrients and other components as possible, with technical provision for adding more information as it becomes available. However, because a comprehensive database system is a national reference resource, it is useful to list the values for individual forms of nutrients separately, where separate analytical values are available or can be obtained, particularly in a reference database. The factors used for converting the different forms of a nutrient to a single value to give an indication of its biological value may change as the state of nutritional science advances. If only the calculated (derived) value is recorded in the database management system, it will not be possible to recalculate the putative total biological activity; thus, it is desirable that the measured values appear in addition to calculated values. In any event, all conversion factors used should be listed in numeric data fields as equivalent to components, or in the documentation sections of the database.

Component data can be expressed on many different bases. For example, amino acids can be expressed as mg per g nitrogen (N) (or as g per 16 gN) and fatty acids as percentages of the total fatty acids, and this is the preferred format for entering such data, if this is the way in which they were obtained from the analytical laboratory. However, at the user level, it is often more useful to present all the data for a particular food as g per 100 g edible portion (or per 100 ml for some beverages, along with density values). User databases (or, more usually, printed tables) will vary in complexity and coverage; hence specific decisions must be made on each component for the different data outputs. Thus, data may be presented as "total" or "available" values for nutrients, for which several forms exist, calculated using appropriate factors and a documented algorithm.

Analogously, in simplified printed tables it may be desirable to regroup some components, such as fatty acids and cholesterol, into separate sections. This will almost certainly be the case when printing costs are a constraint.

In the case of special-purpose tables, many formats are possible. In tables for non-specialists, values may be grouped (e.g. fat <1 g, 1–5 g, 5–10 g, etc.), or foods may be listed according to their ranking as sources of nutrients (excellent, good, fair, poor) depending on the proportion of the recommended daily allowance present in an average serving.

Suggested coverage of nutrients for different levels of data management is given in Table 4.1, and Table 4.2 provides examples of data dissemination formats. Comments on some of these components follow, and further details can be found in Chapters 6 and 7.

Water

It is essential to give values for water content in published tables and papers on food composition and at all levels of data management, including the comprehensive user database. Variations in water content are important determinants of the levels of other components, and data on water content make it possible to compare nutrient values (e.g. for different foods or different analyses of the same food) on a similar moisture basis. This information is also essential when data from different sources are being compared or combined. Analyses for some nutrients are conveniently performed on the dry matter (DM) sample. Therefore laboratory data may be reported per 100 g DM, and recorded in the reference database in this way. However, each DM value must be related to the analysed water content of the same sample, so that nutrient values can be recalculated to their appropriate fresh-weight basis. In simplified printed tables it may be unnecessary to list water content, but it should only be omitted when space is a critical constraint.

Protein

Values for protein are required at all levels of the data system. Conventionally, they are based

Table 4.1 Constituents required at different levels in a database system*

Concise user database	*Comprehensive user database*	*Reference database*[a]
Major components		
Water	Water	
Protein	Nitrogen, total	Protein (protein N x factor) Non-protein N Components of non-protein N
	Protein (total N x factor, sum of amino acids)	
	Nitrogen conversion factor	
	Amino acids	
Fat, total (or fat as triacylglycerols equivalent)	Fat, total (or fat as triacylglycerols equivalent) Fatty acid conversion factors	Phospholipids, sterols, stanols, other lipid classes
Total saturated fatty acids, total monounsaturated fatty acids, total polyunsaturated fatty acids	*Trans* fatty acids, individual fatty acids, total saturated fatty acids, total monounsaturated fatty acids, total polyunsaturated fatty acids	Isomers of unsaturated fatty acids

(Continued)

* Constituents listed for the comprehensive user database are also common to the reference database

Table 4.1 (Continued)

Concise user database	*Comprehensive user database*	*Reference database*[a]
Major components (continued)		
Carbohydrate, available and/or total	Carbohydrate, available and/or total	
Sugars, total	Sugars, total Individual mono-, di- and oligosaccharides Polyols, total and individual Glycemic index	
Polysaccharides	Starches, including glycogen Polysaccarides	Rapidly digestible starch Resistant starch
Dietary fibre[b]	Dietary fibres[b] and their fractions	Non-cellulosic polysaccharides Cellulose Lignin Monosaccharide components of non-starch polysaccharides
	Organic acids, total	Individual organic acids
Alcohol	Alcohol	
Metabolizable energy	Metabolizable energy with energy conversion factors	Individual energy conversion factors Determined heat of combustion
Ash, total	Ash, total	
Inorganic constituents		
Sodium	Sodium	
Potassium	Potassium	
Calcium	Calcium	
Magnesium	Magnesium	
Iron	Iron, haem Fe, non-haem Fe	
Zinc	Zinc	
	Phosphorus	
	Chloride, fluorine, nitrate, nitrite, sulphate	
Iodine (if public health concern)	Iodine	
Selenium (if public health concern)	Essential trace elements (Cr, Mn, B, Co, Se)	
	Inorganic contaminants (Pb, Cd, As, Hg, Ni, Al)	
Vitamins		
Vitamin A (RE) Retinol Beta-carotene equivalents	Vitamin A (RE), retinol, beta-carotene equivalents, beta-carotene, other provitamin A carotenoids,[c] all activity factors	Other retinoids with activity factors

(Continued)

Table 4.1 (Continued)

Concise user database	*Comprehensive user database*	*Reference database*[a]
Vitamins (continued)		
Vitamin A (RE) Retinol Beta-carotene equivalents (continued)	Individual carotenoids, including non-provitamin A carotenoids	Isomeric forms
Vitamin D	Cholecalciferol (vitamin D_3), 25-hydroxy-vitamin D_3, ergocalciferol (vitamin D_2), 25-hydroxy-vitamin D_2, activity factors.	
Vitamin E	Vitamin E (and activity factors), tocopherols and tocotrienols	
Vitamin K[d]	Vitamin K[d]	
Vitamin C	Vitamin C, individual vitamers (e.g. ascorbic and de-hydroascorbic acids)	
Thiamin	Thiamin	
Riboflavin	Riboflavin	
Niacin, total	Niacin, total; preformed niacin; potential niacin from tryptophan	Tryptophan value, conversion factor
Folates, total[e]	Folates, total; individual vitamers; activity factors[e]	
Vitamin B_6	Vitamin B_6 total; individual vitamers	
Vitamin B_{12}	Vitamin B_{12}, individual isomers	
	Pantothenic acid	
	Biotin	
Other components		
	Bioactive substances (e.g. flavonoids, phytoestrogens)	Bioactive substances (e.g. flavonoids, phytoestrogens)
	Organic contaminants, pesticides and other residues	Organic contaminants, pesticides and other residues
	Additives	Additives

Notes:

[a] This might include contaminants and additives and all constituents that exhibit biological activity, particularly dietary phytochemicals. In most cases the data sets will cover a limited number of foods.

[b] These values need to be defined by the analytical method used.

[c] Some users require estimates of total vitamin A activity; because the calculations of activity are uncertain it is better to give measured retinol and carotene values separately.

[d] Values for all vitamin K forms are not available, at present K_1 are adequate.

[e] These values need to be defined by the mode of calculation and/or analytical method used.

Table 4.2 Examples of data dissemination formats

Output form and user	*Foods*	*Components*	*Basis*	*Numeric data*	*Source/quality/ confidence codes*
Tables[a], concise Consumers and professionals	Limited subset, including aggregates (e.g. hard cheese, soft cheese)	Small subset: core nutrients	Per 100 g and up to two other measures	Mean	Desirable at food level
Tables, abridged Consumers and professionals	Large subset, disaggregated foods (e.g. individual cheeses)	Large subset: nutrients, factors, non-nutrients	Per 100 g and one or more other measures	Essential: mean Desirable: standard deviation and/or standard error, number of samples	Desirable at value level
Tables, unabridged Professionals	All	All	Per 100 g and one or more other measures, per g N[b], per g TFA[c]	Mean, standard deviation and/or standard error, number of samples	Essential at value level
Electronic files, customized Professionals/ specialists (e.g. clinicians)	All, or according to user requirements	Large subset, according to user requirements	Per 100 g and other measures as user selection, per g N, per g TFA	Essential: mean Desirable: standard deviation and/or standard error, number of samples; according to user requirements	Desirable at value level
Electronic files, comprehensive Professionals (e.g. researchers)	All	All	Per 100 g and other measures as user selection, per g N, per 100 g TFA	Mean, standard deviation and/or standard error, number of samples	Essential at value level

Notes:
[a] In all cases, "Tables" implies fixed format for visual presentation, printed or Web-based.
[b] N = nitrogen, for amino acids expressed in units mg/g N.
[c] TFA = total fatty acids, for individual fatty acids expressed in units mg/g TFA.
Source: INFOODS Web site, adaptation of Burlingame (1996).

on total nitrogen values using a nitrogen conversion factor (FAO/WHO, 1973), with all factors being recorded at the food level in the database. Values can also be based on the total nitrogen minus the non-protein nitrogen multiplied by a specific factor related to the amino acid composition of the food, or as the sum of amino acids (see Chapters 6 and 7). New amino acid data used in conjunction with the ratio of total amino acid residues to amino acid nitrogen seem to suggest that the nitrogen conversion factor should be lowered. Sosulski and Imafidon (1990) suggest a global conversion factor of 5.7 and Salo-Väänänen and Koivistoinen (1996) of 5.33, both with individual factors for different foods and food groups. At this time no new international agreement on conversion factors had yet been reached.

Total fat

Values for total lipids vary considerably with analytical method (see Chapters 6 and 7) and may be of limited nutritional significance; nevertheless, they are widely used and should be included at all levels of the database.

Fat (-acylglycerols). Inclusion of this item is desirable in the reference database, primarily for use in the calculation of food energy value, and also because of the interest in triacylglycerols from animal and vegetable sources. The widespread and increasing use of mono- and acylglycerols in manufactured foods is an additional reason for its inclusion.

Phospholipids. Values for the different classes of these substances should be included at the reference database level because of their wide use as emulsifying agents, and because of their physiological properties.

Sterols. Although cholesterol was once considered the most important sterol from a nutritional viewpoint, the significance of the other sterols (e.g. sitosterol) is now recognized; they should be included at the user database level.

Fatty acids. Data for individual fatty acid stereoisomers should be included in the reference database. At this level, the most convenient mode for expressing fatty acid values is as g fatty acid per 100 g total fatty acids. In user databases, however, expression as g fatty acid per 100 g of food is more useful. In simplified user databases the fatty acids may be grouped into total saturated, total mono-unsaturated and total polyunsaturated acids, or the ratio between the groups may be cited together with the total fat value. Another grouping of major interest is as n-9, n-6 and n-3 families of unsaturated fatty acids (Gurr, Harwood and Frayn, 2002).

Carbohydrates

Values for available (glycemic) and unavailable (non-glycemic) carbohydrates derived by analysis are desirable throughout the database system. The earlier practice of including carbohydrate calculated "by difference" has proven to be scientifically unsound and should be phased out as soon as possible (FAO/WHO, 1998).

Available carbohydrates (glycemic). These include all the sugars (glucose, fructose, sucrose, lactose and maltose) known to be glucogenic in humans and the polysaccharides (starch and partially hydrolysed starches, and glycogen) hydrolysed by the endogenous secretions of the human digestive tract (Table 4.3).

Unavailable carbohydrates (non-glycemic). These include all the polysaccharides that are not hydrolysed by the endogenous secretions of the human digestive tract: components of the plant cell wall (cellulose, non-cellulosic polysaccharides, pectic substances and hemicelluloses) and a range of polysaccharides used as food ingredients or food additives. These together are the non-starch polysaccharides (NSPs), which are often used as a definition of dietary fibre. There are several other definitions of dietary fibre, each identified by a different methodology, and each measuring different amounts of the non-glycemic carbohydrates, and other non-carbohydrate material (e.g. lignin).

Oligosaccharides. There is growing recognition of the potential nutritional importance of this group and an associated need to start assembling values for these components. Oligosaccharides include tri-, tetra- and pentasaccharides of the raffinose series, analogous maltoderivatives and a range of fructose polymers, including those at the lower end of the polysaccharides. Individual oligosaccharides need to be recorded separately because they are metabolized differently.

Polyols (sugar alcohols). These comprise a group of polyhydric alcohols structurally related to the sugars where the reducing group has been reduced to a hydroxyl compound. Very small amounts of them occur naturally in foods, but they are widely used as food additives for their humectant properties or as a replacement for sugars in reduced-energy products, low cariogenic sweets and foods for diabetics. Under the labelling regulations of some countries, polyols are included in the carbohydrate declaration, but in a nutritional database it is preferable to list them separately under their specific trivial names. Table 4.3 indicates the more important polyols used in foods.

Organic acids

These are important in relatively few foodstuffs, and their inclusion in a user database should be selective. Values should be given for fruits, fruit products (including juices), a few vegetables (particularly those preserved in acetic acid), and other manufactured products, such as vinegar, salad dressings that have organic acids listed as major ingredients, soft drinks and yoghurt. In these cases, organic acids should be included in energy calculations.

Alcohol

Alcohol (ethyl alcohol) may be a significant energy contributor; levels must be determined and used in energy calculations for alcoholic beverages, and for confectionery and desserts containing alcohol.

Table 4.3 Carbohydrates in foods

Chemical grouping	*Classes*	*Types present in the diet*	*Relative importance*	*Nutritional classification*	*INFOODS tagnames*
Sugars					
Free sugars	Monosaccharides	Monosaccharides	Major	Glycemic and non-glycemic	MNSAC
	Pentoses (monosaccharides)	Arabinose	Rare	Non-glycemic	ARAS
		Xylose	Rare	Non-glycemic	XYLS
	Hexoses (monosaccharides)	Glucose	Major	Glycemic	GLUS
		Fructose	Major	Glycemic	FRUS
		Galactose	Minor	Glycemic	GALS
	Disaccharides	Disaccharides	Major	Glycemic	DISAC/DISACM
		Sucrose	Major	Glycemic	SUCS/SUCSM
		Lactose	Minor[1]	Glycemic	LACS/LACSM
		Maltose	Minor[2]	Glycemic	MALS/MALSM
Oligosaccharides	Contain between 3 and 9 monosaccharide residues	Oligosaccharides, total available	Minor	Glycemic and non-glycemic	OLSAC/OLSACM
		Maltotriose and higher	Minor[2]	Glycemic	MALTRS/MALTRSM
		Raffinose	Minor[3]	Non-glycemic	RAFS/RAFSM
		Verbascose	Minor[3]	Non-glycemic	VERS/VERSM
		Stachyose	Minor[3]	Non-glycemic	STAS/STASM
Polyols	Polyols (formerly called sugar alcohols)			Non-glycemic	POLYL
	Trihydric	Glycerol	Minor	Non-glycemic	GLYRL
	Pentahydric	Xylitol	Minor[4]	Non-glycemic	XYLTL
		Galactitol (dulcitol)	Minor	Non-glycemic	GALTL

(Continued)

Table 4.3 (Continued)

Chemical grouping	*Classes*	*Types present in the diet*	*Relative importance*	*Nutritional classification*	*INFOODS tagnames*
Polyols (continued)					
	Hexahydric	Mannitol	Minor	Non-glycemic	MANTL
		Sorbitol (glucitol)	Minor[5]	Non-glycemic	SORTL
	Disaccharide alcohols	Lactitol	Minor[6]	Weakly glycemic	LACTL
		Maltitol	Minor[6]	Weakly glycemic	MALTL
Polysaccharides					
Reserve polysaccharides	Starches	Starches	Major	Glycemic	STARCH/STARCHM
		Amylose (linear)	Major	Glycemic	AMYS/AMYSM
		Amylopectin (branched)	Major	Glycemic	AMYP/AMYPM
		Partially hydrolysed starches	Major in processed foods	Glycemic	STAHY/STAHYM
		Glycogen	Minor from meats, etc.	Glycemic	GLYC/GLYCM
		Resistant starch	Major	Glycemic	STARES
	Fructans	Fructan	Minor	Non-glycemic	FRUTN
		Inulin and higher fructo-oligosaccharides	Minor	Non-glycemic	INULN
	Mannans	Mannan	Minor	Non-glycemic	MANN
		Gluco mannan	Minor	Non-glycemic	GLUMN
		Galacto mannan[7]	Minor	Non-glycemic	GALMN
Structural polysaccharides (plant cell wall constituents)	Non-cellulosic polysaccharides	Pectic substances[8]	Water soluble, uronic acid rich	Non-glycemic	PSACNCP

(Continued)

Table 4.3 (Continued)

Chemical grouping	*Classes*	*Types present in the diet*	*Relative importance*	*Nutritional classification*	*INFOODS tagnames*
Polysaccharides (continued)					
Structural polysaccharides (continued)	Non-cellulosic polysaccharides	Hemicelluloses[9]	Water insoluble, mainly xylans and glucans, uronic acid poor	Non-glycemic	HEMCEL
	Cellulose	Various degrees of polymerization		Non-glycemic	CELLU
Modified starches[10]	Cross-linked esters, ethers and phosphates			Some may be glycemic or partially glycemic	STAMO/ STAMOM
Gums and mucilages	Gums Mucilages	Wide range of water-soluble substances[9]		Non-glycemic	GUMS MUCIL
Algal polysaccharides	Sulphated	Carrageenan[10]		Non-glycemic	CARGN
		Agar[10]		Non-glycemic	AGAR
	Unsulphated	Alginates[10]		Non-glycemic	ALGNT

Notes:

1 This sugar is derived from milk and milk products and the consumption of these foods will determine its importance.
2 These sugars are derived from foods containing glucose syrups and may be more important when consumption of these foods is high.
3 These oligosaccharides are present in many vegetables.
4 This polyol is widely used in low cariogenic confectionery and the consumption of these products will increase its importance.
5 This polyol is used in some foods designed for diabetic patients.
6 These are widely used as bulking agents and are weakly glycemic.
7 Linear mannans with single side chains widely used as thickeners in processed foods.
8 Wide range of polysaccharides, galacturonans, galacturonorhamnans, arabinans, galacatoarabinans.
9 Wide range of polysaccharides, linear and branched heteroglycans, especially xylans and glucans, widely used as bulking agents in processed foods.
10 Used as ingredients to control the physical properties of many processed foods.

Source: Modified from Southgate, 1991.

Inorganic constituents

Total ash. Values for ash are frequently given in data sources and the values should be entered into the database system primarily because they can be used in internal checks on the sum of all the proximate components, the calculation of total or available carbohydrate by difference and the mineral content. Because the values are not of nutritional significance, they need not appear in simplified tables.

Individual inorganic constituents. All the essential inorganic elements should be included. Current instrumental techniques provide information on a wide range of minor trace constituents with little extra cost, and it is desirable to include a comprehensive list. The forms in which some trace elements occur are important in relation to their bioavailability and should therefore be recorded when this information is available.

Vitamins

Many vitamins occur in several active forms called vitamers; if it is technically possible, the vitamers should be analysed separately and the values held separately in the database system, in some cases at the user database level. In simplified tables, it will usually be enough to provide a value for the total activity of the vitamin in question. It is, however, essential to document the algorithms used to calculate these estimates of total activity.

Non-nutrient constituents

Contaminants. Contaminants include mycotoxins, heavy metals and residues of pesticides, herbicides and animal growth promoters. The distribution of contaminants in foods is such that the concept of representative values for contaminants differs from that for nutrients. It may be misleading to list contaminant values in the same record as nutrients. Listing in archival and/or reference auxiliary data records is preferred.

Bioactive substances. There has been a growing interest in the range of dietary phytochemicals in recent years, particularly in view of their possible protective action against cardiovascular diseases and certain cancers. These include isothiocyanates, polyphenols, flavonoids, isoflavones, lignans, saponins and coumestrol (AICR, 1996; Pennington, 2002). Consequently, there is a parallel interest in the inclusion of phytochemicals in food composition databases (Ziegler, 2001). The collection of data from data sources is useful, although it may not be possible to find complete data sets.

Antinutrients and toxicants. Some constituents have undesirable physiological effects, for example, goitrogens, haemagglutinins, antivitamin factors, trypsin inhibitors, oxalic acid and phytic acid. Data for these components should be included for the relevant foods. Other important natural toxicants include solanine, cyanides, glucosinolates, lathyrogens, mimosine

and nitrosamines. Ideally, data for these natural components should be incorporated in the reference database.

Additives. Many additives are measured, in whole or in part, during the course of nutrient analyses. Salts, for example, are included in analyses for various cations and anions; protein additives are determined in nitrogen analysis; and some emulsifiers and thickeners are included in analyses for nitrogen, starch and unavailable carbohydrates. Clearly, specific analyses are preferable. However, the need for data on additives and other non-nutrient components of foods may relate to priorities regarding food safety and not necessarily to nutritional priorities.

Miscellaneous. Where data exist for other compounds of interest, such as caffeine, theophylline, theobromine, tannins and other bioactive compounds (carnosine, carnitine, creatinine), they should be listed in the database at least up to the reference level.

Chapter 5

Sampling

The quality of sampling and analytical data is a major determinant of database quality. Sampling foods for inclusion in a compositional database is one of the more demanding and difficult aspects of database preparation and often requires the compilers to make intuitive judgements and compromises. This chapter reviews the objectives of sampling and discusses the various aspects for consideration in making these judgements.

Where the necessary information on the composition of a food is not available (as is often the case in developing countries) or is inadequate (e.g. it is no longer applicable to the current food supply or the analytical values need to be measured using more recent methods), then sampling and analytical protocols need to be devised.

Ideally these should be developed in conjunction with each other because the requirements of the analysts will determine the amounts of foods necessary for the analyses and how the foods should be stored and, if necessary, preserved.

Objectives in sampling

Users of compositional databases require representative values for the composition of the foods consumed by the population for whom the database is being prepared.

The primary objectives in sampling, therefore, are to collect food samples that are representative and then to ensure that changes in composition do not take place between collection and analysis.

All foods are biological materials and exhibit natural variations in composition. A secondary objective may be to document this variability as it relates to factors such as season, geography, cultivar and husbandry. Such variations are to be expected and should not be confused with variations associated with the analytical conditions. The combined protocols – that is, for sampling and analysis – should also ensure that the representative attributes are maintained in the portions taken for analysis.

Some basic terms

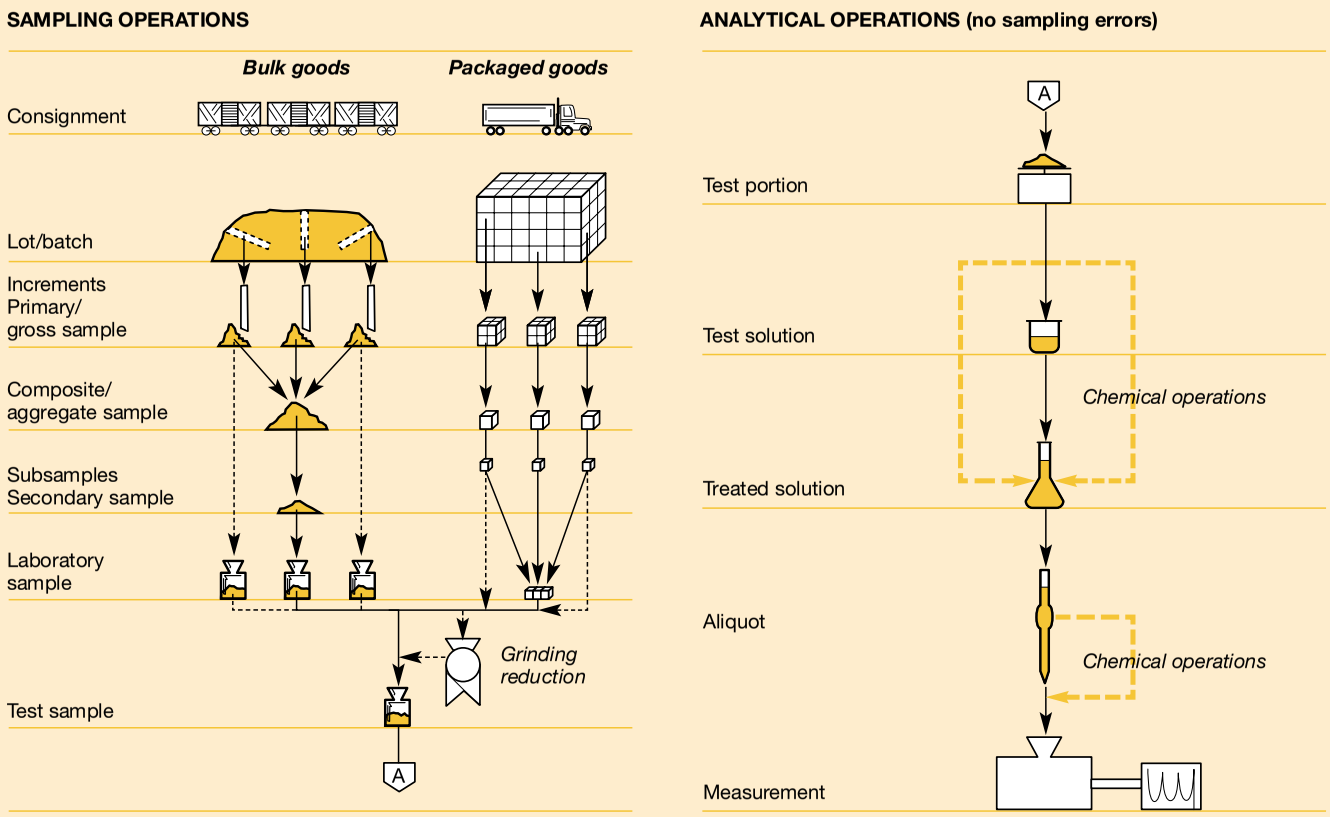
In the context of the following account, the term *sampling* is used to describe the activities involved in the selection and collection of items of food defined in terms of number, weight and nature of the material to be analysed. Much of the formal terminology developed for use in sampling was designed for use in the commercial sector for the purposes of surveillance and determination of contamination (Horwitz, 1990). Some of these terms have little relevance for nutrient database work and therefore are not discussed further. Table 5.1 outlines the steps involved in the sampling process and provides definitions of the terms that will be used later in this book. Figure 5.1 illustrates the different stages in sampling and analysis, indicating where sampling errors may arise as distinct from analytical errors.

Because of the variability and heterogeneity of foods, all sampling is associated with some degree of error when the results are extrapolated back to the composition of the whole

Table 5.1 Definition of terms used in sampling of food for a nutritional database

Term	*Definition*	*Comments on application in food composition studies*
Sample	A portion selected from a larger quantity of material	A general term for a unit taken from the total amount (the population) of a food
Sampling protocol	A predetermined procedure for the selection, withdrawal, preservation and preparation of the sample	Sometimes called a sampling plan
Characteristic	The property or constituent that is to be measured or noted	Description of the food, nutrient and other analyses
Homogeneity	The extent to which a property or constituent is uniformly distributed	Foods are usually heterogeneous or must be assumed to be so
Sampling error	The part of the total error associated with using only a fraction of the total population of food and extrapolating it to the whole population. This arises from the heterogeneity of the population	Because of the heterogeneous nature of foods, replicate samples must always be taken when estimating the composition of the population of a food
Batch	A quantity of food that is known, or assumed, to be produced under uniform conditions	Batch numbers should always be noted when sampling foods
Unit	Each of the discrete, identifiable units of food that are suitable for removal from the population as samples and that can be individually described, analysed or combined	These units form the basis of most food analysis work (e.g. an apple, a bunch of bananas, a can of beans, a prepared dish)

Figure 5.1 The relationships of the operations involved in sampling and analysis
The lower A of the sampling operations continues with the upper A of the analytical operations



Source: IUPAC recommendations from Horwitz, 1990.

population of a food. Sampling can merely provide data that define the probability that the values will apply to any one isolated unit of the food.

The approach to sampling

The selection of a representative sample and the combined protocols for sampling and analysis must be based on a clear understanding of the nature of the foods and the population of food being studied (i.e. all the individual units of the food). A database will be used for a considerable period of time and the values derived from the combined protocols will be used as if they were representative, in both space and time, over the lifetime of the database (and often for much longer). The design of the protocols therefore represents a monumental task and one in which it may be necessary to accept compromises. It is essential that such compromises are based on knowledge of the food in question.

Sources of food

The principal sources of food samples are summarized in Table 5.2. These groupings correspond to the levels at which databases are used.

Bulk commodities

Compositional data obtained from analyses of bulk commodities have wide-ranging uses. They are commonly used in commerce or for surveillance of imports for contamination with agrochemicals or the misuse of growth stimulants. These data also provide the basis for calculating the nutrient values in food disappearance statistics and sometimes in household and industrial recipes. Standard sampling procedures have been defined for many commodities and these should be followed: International Organization for Standardization (ISO, 2003); Official Methods of the Association of Analytical Communities (AOAC International, 2002, 2003); Codex Alimentarius (FAO, 1994; FAO/WHO, 2003). Care should be taken to ensure that samples are truly representative of the bulk commodity. Several samples may need to be taken from separate sacks, cases, packages or carcasses, and at several points in a silo or container. Random sampling is preferable to the collection of readily accessible units. Collectors should take packages from several randomly identified cases or packages, for example. This level of sampling presents logistical problems that are best overcome by taking samples during the loading or unloading of a consignment. Special probes or triers are required (Horwitz *et al.*, 1978) for sampling finely particulate foods (e.g. sugar, grain), fluids (e.g. milk) or solids (e.g. cheese).

Nutrient analyses at this level are often limited to major components, but generally involve many analysed samples (sometimes in the hundreds), and therefore result in very high-quality values.

Table 5.2 Major sources of food samples for analysis for a food composition database

Source level	*Examples*	*Level of use of compositional data*
Bulk commodities	Meat carcasses, bulk consignments of grain, fruit, vegetables, wine, edible fats	Used mainly to assess nutritional value of food supplies and for food disappearance statistics. Also useful for intake assessment
Wholesale commodities and foods	Meat carcasses, prime cuts, bulk packs of foods, often for institutional use	
Retail foods	Foods as sold to the consumer, e.g. meat cuts, vegetables, fruits, wine, processed foods	Used mainly to assess household and individual food and nutrient intake. Also useful for food supply statistics
Field, garden or wild foods	Foods grown or gathered, hunted animals	
Foods as consumed	Foods at the level of consumption, e.g. cooked dishes (single or multiple ingredients), street foods	Used to assess individual food and nutrient intake

Wholesale foods

Sampling of wholesale foods generally follows the principal approaches used with bulk commodities. Randomization of sampling is essential.

Retail foods

These foods constitute the majority of foods included in food composition databases in industrialized countries. For primary products such as meats, fruits or vegetables, the major concern of the sampling protocol is to ensure that the complete range of sales outlets is represented. The primary sample should be made up proportionately of the volumes of food passing through the different outlets. The potential for regional variation also needs to be covered in the design of the sampling protocols.

In non-industrialized countries where food distribution systems may be less developed, regional considerations assume greater importance and variations in composition from one rural market to another may be substantial. Regional stratification (see below) of the sampling may be considered a more useful approach in view of the regional variation in the composition of produce. In many cases presenting data that are representative of a very diverse population may not be acceptable.

Proprietary foods constitute an important range of foods in many countries and their composition should be included in the database. Where a database is prepared by government personnel there is often reluctance to include brand names. In practice, for many proprietary foods, the brand name is essential for identification. In some countries, the range of branded items of a food is very numerous, and covering all the different brands increases the analytical workload. Compositional data supplied by the manufacturer may be acceptable provided

that they meet the criteria set for analytical quality, and that the manufacturers can assure the compilers that the samples analysed were representative of products as sold retail. Problems can arise using this approach because many proprietary foods are reformulated at frequent intervals and database values rapidly become out of date. Many compilers prefer to restrict this type of database entry to foods that are stable and well-established. In some cases, pooling the different brands according to market share is considered appropriate.

When collecting samples, care must be taken to ensure that the full range of retail outlets is properly represented. When available, retail sales statistics can be useful. In many cases proprietary products are produced under such strict quality control that limited sampling is satisfactory.

Field or garden produce

These sources of food are often ignored in industrialized countries, but in many countries food produced by the family constitutes an important component of the diet and should therefore be considered by database compilers. These foods tend to be much more variable – the composition of plant foods is especially dependent on the soils and fertilizer treatments. Such factors therefore need to be taken into account in the design of sampling protocols. Most field or garden produce is eaten seasonally as fresh and then preserved according to traditional methods that can differ substantially from commercial practice.

Uncultivated and wild foods

Many communities, especially those living a "hunter-gatherer" or semi-nomadic style of life, consume substantial quantities of wild plant and animal foods. Such foods account for a significant proportion of daily consumption, and their inclusion in a database can be very useful for those studying the nutrition of such groups. Collecting samples of these foods can pose particular problems. They may be difficult to identify properly and also tend to be variable in composition and maturity (Brand-Miller, James and Maggiore, 1993). Frequently random sampling is virtually impossible and "convenience" sampling, as the opportunity arises, is the only option. Provided that this approach is documented in the database, it is acceptable. Documentation will alert users to the limitations of the data and minimize the possibility of them being used inappropriately.

Foods as consumed

Many dietary intake studies, especially epidemiological investigations, require the measurement of food and nutrient consumption at the individual level, i.e. foods as directly consumed. These foods – "on the plate", as they are often called – comprise cooked foods of all kinds, including complex mixed dishes. The latter are often prepared using a variety of recipes and cooking methods, which poses difficulties in selecting representative samples. Simulation of the cooking procedures in the laboratory or dedicated kitchens is often used to prepare samples for analysis. This approach is generally satisfactory, although in the domestic context being simulated, food preparation is not always carried out in a controlled fashion and decisions

on when cooking is complete are a matter of individual preference and judgement. Nevertheless, laboratory-based sample preparation allows for detailed documentation of all the relevant conditions (cooking temperature, duration, end-point internal temperature, etc.). Collection of cooked dishes from a randomly selected range of households would provide more representativeness, and is sometimes, therefore, the preferred approach (Greenfield, 1990b). However, this approach also presents its own logistical problems.

Samples of institutionally prepared foods from, for example, hospitals, industrial and public canteens and educational establishments, are more easily obtained. Samples from fast food establishments and of "take-away" foods are also easier to collect. The difficulties in sampling, the enormous range of possible variation among cooked foods and financial constraints have frequently led compilers to use calculations from recipes to estimate the composition of cooked dishes.

Major sources of variability in nutrient composition

Foods are inherently variable in composition, and the approach to sampling and the design of the sampling and analytical protocols need to take account of this factor.

Geographical samples

In a single country there may be a wide diversity of soil and climatic conditions, resulting in significant variance in food composition. Variations in food marketing and food preparation within different parts of a country – or among countries in the case of a multicountry database – may also produce notable variance. For these reasons, geographically-specific data may be presented in the database as a supplement to nationwide and/or regionwide averages. In other countries, the variations may be of similar magnitude to those due to other causes, in which case the national sample could be weighted according to the proportions of the population living in the regions or the proportions of the total consumption of the foods.

Seasonal samples

Seasonal variations in nutrient composition need to be accommodated in the combined protocols. Plant foods are especially prone to variation, particularly in their water, carbohydrate and vitamin content. Fish also show seasonal variations, especially in fat content, and milk and milk products exhibit variations in vitamin content primarily due to seasonal differences in feeding patterns. The collection of samples needs to be organized, in terms of timing and frequency, to reflect these variations. In some cases, seasonal data need to be given separately in the database. The analytical measurements of the seasonal samples can often be restricted to those nutrients showing variation.

Physiological state and maturity

The states of maturity of plants and animal foods cause variation in composition: in the

concentrations of sugars, organic acids and vitamins in many plants, and of fats and some minerals in animal foods. Some of these variations are a consequence of seasonal effects.

The storage of plant foods also often affects water and vitamin contents and levels of some organic nutrients due to residual plant metabolism in storage.

Cultivar and breed

These may be a significant source of variation for some nutrients and the combined protocols will need to provide for this variation. It is desirable to document this cultivar or breed variation within the database. Some research organizations sample specifically to capture cultivar and breed differences. The significance of the differences attributable to cultivar or breed can only be ascertained by controlling for other factors that can influence variation, and by sampling and analysing individually, not in composite, a large number of samples.

Methods of sampling

The main sampling methods used for nutrient composition databases are summarized in Table 5.3.

Random sampling

Random samples are collected in such a way as to ensure that every item in the population of the food being sampled has an equal chance of being collected and incorporated into the sample to be analysed. This is difficult to achieve in practice because it is difficult to visualize the entire population of, say, all the cabbages in a country let alone ensure that each one has an equal chance of being selected. It is more usual to set up a stratification (see below) of the food population.

Stratified sampling

In this method the population of food is classified into strata, taking into account the most important causes of variation.

Stratification by geographical area may be useful even where there are no known significant regional variations (Smits *et al.*, 1998). Stratification according to the distribution of the consuming population, among rural and urban sources, or by type of retail outlet, are other useful examples (Torelm, 1997). The sampling of branded foods can be stratified according to manufacturing plant. Where different brands of the same food are not expected to show significant variation, the sample can be weighted according to market share.

Where this information is not available, extrapolating from similar foods or an intuitive assessment will be required.

Selective sampling

Selective sampling is widely used in experimental studies of plant and animal husbandry and

Table 5.3 Main sampling methods used in nutrient composition studies

Method	*Definition and characteristics*	*Notes on application*
Random sampling	Samples are taken in a way that ensures that any one unit has an equal chance of being included	The theoretical ideal but rarely practicable when sampling foods for nutritional databases
Stratified sampling	Units of sampling are taken from defined strata (subparts) of the parent population. Within each stratum the samples are taken randomly	Often the most suitable method for use in database work. Strata may be regional, seasonal, retail sale point, etc., as defined by knowledge of the food being studied
Selective sampling	Samples are taken according to a sampling plan that excludes material with certain characteristics or selects only those with defined characteristics	Most commonly used in the analysis of contaminants. Can be used, with caution, for database work
Convenience sampling	Samples are taken on the basis of accessibility, expediency, cost or other reason not directly concerned with sampling parameters	Rarely suitable for database work but may be the only practicable way to sample wild or uncultivated foods or composite dishes from households

in home economics. The resultant data are valuable guides for the design of sampling protocols but since they are not generally representative of the foods available, they require careful documentation when included in the database.

Where, however, it is clear that the methods of husbandry and the storage of the foods are comparable with current practice for the production of food the data may be useful.

This method is often legitimately used in the analysis of contamination, where the objective may be to identify maximal exposure to contaminants. The distribution of contaminants in foods is frequently highly skewed. Random sampling will therefore often include samples in which the concentration of the contaminant is below the level of detection. This is the primary reason why data on the levels of contaminants are often held separately from representative nutrient data in the database.

Samples of foods prepared in a laboratory can be regarded as selective samples. Laboratory preparation may be the only practicable way to obtain data on the composition of certain foods and therefore the derived data may be useful in databases. Generally, however, samples collected from cooks working in domestic or industrial kitchens are to be preferred as they can be regarded as more representative of foods generally available for consumption.

Convenience sampling

The collection of samples from conveniently accessible points is a very common, and possibly misleading, practice in compositional studies. This method may be acceptable as a preliminary

exercise to obtain estimates of variation in composition, but in general data obtained using this method should be regarded as low quality.

Convenience sampling may be the only option in the case of wild or uncultivated foods; provided the sources of the samples are fully documented the values can be used in a database.

Limits of all sampling methods

In all methods the compositional data obtained can only be an estimate of the composition of the food and are subject to limitations imposed by the variation in the composition of foods.

Designing combined sampling and analytical protocols

The objective is to prepare well-documented protocols that provide the basis for those involved in collecting and handling the samples, from their collection in the field through to the laboratory. This process serves to ensure that the data generated meet the objectives of the compilers and the requirements of the database users.

Responsibility for preparing the combined protocols

In some countries the database compilers control the sampling and analytical work and are responsible, in collaboration with the analysts, for preparing the written combined protocols. In most countries, however, the sampling and analytical work will be carried out under contract(s); here the compilers' input may be restricted to establishing the broad outlines of the work required. These initial specifications should set out the principles of the database requirements with regard to representativeness and the analytical data quality standards that the reports from the contractors must meet.

Detailed combined protocols are then prepared by the contractors in consultation with the compilers. The sampling may be contracted to local sampling groups (e.g. where the database covers a large country or region); again, it is essential that the subcontractors are fully conversant with the sampling objectives.

Where the analytical work is subcontracted, either for all or selected nutrients, the subcontractors must be aware of the preferred analytical methods and have in place the proper data quality assurance schemes. Where the subcontractors wish to use other methods with which they may be more familiar or experienced, they should provide evidence that these are compatible with the preferred methods.

It is of paramount importance that units and modes of presentation of the results are predefined and written into the contracts. For example, laboratories may use ppm (parts per million, mg/kg) or ppb (parts per billion, microgram/kg) to express the results of trace metal analysis, and others use IU (International Units) for some vitamins. Fatty acids should always be reported as units of mass (mg/100 g) and may additionally be reported as a percentage of total fatty acids. It should also be predetermined whether results should be reported on a dry weight basis or wet weight basis. In either case, water content values must be reported.

Choice of sampling method

Some form of stratified sampling will generally be the method of choice. Even where there is no evidence of regional differences in composition, a stratification based on collecting samples on a regional basis of the population of the food consumed will be included in the sampling. For pragmatic reasons it may be necessary to restrict the extent of sampling and most compilations devote the most extensive sampling to the most important "core foods" or "key foods" and those foods that are major sources of particular nutrients, (Chug-Ahuja *et al.*, 1993; Schubert, *et al.*, 1987; Haytowitz, Pehrsson and Holden, 2002; Pennington and Hernandez, 2002; Perry *et al.*, 2000) where, for example, there are public health concerns. Foods that are relatively minor components of the diet are usually less emphasized in the protocols. Many proprietary or branded foods, which are produced in a few factories, can clearly be sampled more simply than, say, meat products which are often "core foods" and which can show great variability, necessitating much more detailed and extensive protocols. Vegetables and fruits, which show seasonal variations in composition, will need to have a seasonal stratification. Each group of foods must be considered on a case by case basis. The logistics of the analytical work often make it desirable to sample foods on a food group basis because sample handling and the actual methods used will be common across the group.

Table 5.4 Summary of stages in sampling and preparation of samples in food composition studies

Terms	*Description*	*Main use in food composition studies*
Primary sample	The collection of one or more units initially taken from the total population of the food	The usual starting point in compositional studies. The ideal is the collection of several replicates that are treated separately. Primary samples are often mixed to form composites
Reduced sample	A representative part of the primary sample obtained by a division or reduction process	Frequently used to reduce the primary sample to a more manageable weight
Composite sample	Mixtures formed by combining primary samples	Frequently used in food composition studies. Composites may be samples of the same food or combinations of different brands or cultivars
Laboratory sample	The sample sent to or received by the laboratory	The primary sample (or a reduced sample) often requires further handling in the laboratory (e.g. thawing, cooking, separation of inedible matter). The edible portion may need further reduction or mixing
Analytical sample	The portion prepared from the laboratory sample from which the portions for analysis are taken	This is usually the form in which the food samples are prepared for analysis
Analytical portion	The quantity of food of the proper weight for each analytical measurement	The analysis of duplicate analytical portions is the minimum acceptable; several replications are preferable

During the course of describing the sampling process a number of stages are met, each of which uses the terms "sample". Table 5.4 sets out a summary of the stages and some suggested definitions which may be used to make it clear which type of sample is meant at the different points in sampling and analysis.

Size and number of samples

Size. The total amount of food required for the different analyses forms the basis for deciding the size of individual samples. In practice, because foods are heterogeneous, taking small portions at the primary sampling stage can lead to error. For many foods the individual items for collection are readily identifiable; in other cases they will need to be defined. In practice, 100–500 g represents a convenient guide to the size of a primary sample, with preference being given to the upper end of this range. Some food items, for example certain cuts of meat, are much larger than this and cannot easily be reduced to a smaller but still representative unit; for the purpose of the primary sample these should be used in their entirety.

Number. In order to calculate the number of samples needed, information is first required on the variability of the composition of the food (Proctor and Muellenet, 1998). This also assumes that the concentration of the nutrient is uniformly distributed in the food, which is a reasonable assumption for many nutrients but often not true for trace elements.

In practice, the required information is often incomplete and one has to proceed intuitively. Furthermore, many nutrients, especially vitamins, show greater variability than, say, protein, so the number of samples required formally will be greater.

An example of how the calculations are performed is provided in Appendix 2.

Most sampling schemes adopt a standard of at least ten units and the United States requires data for nutrition labelling to be based on 12 units. However, strictly speaking the number depends on the variability of the nutrients being measured and thus different numbers of food samples are required for certain nutrients.

Preparing the protocols

The protocols are written documents that describe the sampling process: the identity of the food, the size and weight of units to be collected, the stratification to be used and the distribution of sampling sites. Tables 5.5a–5.5d give the information that is required for preparation of the sampling protocol, commencing with the description of the primary food sample (Greenfield, 1989; McCann *et al.*, 1988).

Table 5.5a deals with the identification of the food. The record of the collection is recorded in Table 5.5b, a detailed description of the food collected in Table 5.5c, and the handling in the laboratory in Table 5.5d.

Table 5.5a Suggested food sample record for food composition studies: identification

Common name of food	
Sample code number	
Date of receipt in laboratory	

Food identification	*Examples of record*
Alternative names	Other common names (in language of country of origin) and English equivalent where possible
Scientific name	Genus, species, variety
Plant food	Entire plant, or part of plant (root, stem, leaves, flower, fruit, seeds)
Animal food	Entire animal, or part (leg, head, internal organ)
State of maturity	Immature, ripe, etc.
Grade	Where appropriate
Other details	Any details that the collector thinks may be relevant

The volume of information recommended in this documentation may seem excessive, but experience suggests that information from different stages is very critical when assessing the quality of sampling and subsequent analyses. Moreover, if the details are not recorded at the appropriate time they cannot be recovered retrospectively.

Identification

Table 5.5a sets out the information required. The first section constitutes a label that should be securely and permanently attached to the sample. The laboratory may subsequently add an acquisition number. Most of the information required is self-evident.

Record of collection

Table 5.5b sets out the information to be recorded during sample collection. The items recorded correspond to the sampling plan as set out in the combined protocols. This indicates the designed stratification and the method for achieving randomization within the strata. The use of random number tables is one useful approach. The protocol must also specify the procedure to be followed if the defined sample item is not available for collection. This may be the nomination of a replacement item or the need to choose an alternative sampling point.

Most of the items are self-evident. A record of the purchase price can be useful for auditing purposes and for household budget studies. A photographic record, with a measurement scale and colour standard (e.g. Pantone sheet), if available, is recommended to facilitate the identification of the sample (Burlingame *et al.*, 1995b). If photographic records are not practicable, a simple line drawing may suffice (McCrae and Paul, 1996).

Table 5.5b Suggested food sample record for food composition studies: record of collection

Common name of food	
Sample code number	
Date of receipt in laboratory	
Collection details	***Examples of record***
Date and time of collection	
Name of collector	
Place of origin	If known, (village, district, province, map reference)
Sampling point	Type (field, garden, roadside stall, farm market, shop, warehouse, supermarket, take-away food bar, restaurant, household, deep sea, shoreline)
Address(es) of sampling point(s)	
Conditions of cultivation	Where known (altitude, rainfall, fertilizer treatment, irrigation, feed regime)
Season	Time of year, dry or rainy season
Purchase price	If relevant
Graphical record	Visual record with scale; line drawing may be sufficient
Transport conditions	Details, including mode and conditions of transport and storage
Other details	Any details that the collector considers relevant

The combined protocol identifies the arrangements for transporting primary samples from the collection sites to the laboratory. The logistical aspects of handling what may be large amounts of food require careful consideration; the storage procedures, including choice of containers and modes of transport, should be specified in consultation with the analysts. These and all other aspects of the combined protocols need to be rehearsed or at least taken through a "paper exercise" with the participation of all those involved. Secure storage in inert containers, which can be heat-sealed using simple equipment, is preferable. Ideally, the samples should be cooled with crushed ice or solid CO_2. If this is not possible, they should be transported to the laboratory with minimum delay. In some cases, the limitations of the sampling and transport arrangements may preclude the analysis of nutrients that are likely to be changed by metabolism (see Table 5.6 on page 80).

Where the distance to the laboratory is short, road or rail transport may be suitable but, where longer distances are involved, air transport may be the only alternative. (This will involve liaison with the airlines to ensure that the storage conditions are compatible with airline safety regulations.) In other cases considerable ingenuity may be required to suit local conditions.

Table 5.5c Suggested food sample record for food composition studies: description of samples collected

Common name of food	
Sample code number	
Date of receipt in laboratory	
Description	***Examples of record***
Food type	Food grouping (legume, fruit juice, milk products, etc.)
Local use of food	In festivals, famine, etc.
Physical dimensions	
Physical state	Shape, form (e.g. liquid, solid, whole, divided, particle size)
Process and preservation method	Canned, smoked, sun-dried, etc.
Preparation method for consumption	Cooking method
Extent of preparation	Raw, uncooked, partially cooked, fully cooked, thawed, reheated
Packing medium	Brine, oil, syrup, water
Container or wrapping	Can, glass, paper, foil, leaves
Contact surface	Glass, type of plastic, foil
Label or list of ingredients	Retain label, estimated by inspection
Batch number	For branded foods
For branded or pre-packed food	
Weight of food collected	
Number of items	
Weight of individual items	
Weight of common measure or portion	
Other details	Any details that the recorder considers relevant (e.g. after fresh samples were collected they were vacuum sealed)

The personal security of the samplers should also be considered, as they often carry relatively large amounts of money to pay for the samples that they are collecting; indeed, the large amounts of food they carry may also be a target for theft. Payment for samples can often be arranged by credit, thereby eliminating one of these concerns.

Description of samples collected

Most of the information suggested in Table 5.5c may be added once the samples have arrived

Table 5.5d Suggested food sample record for food composition studies: record of handling in laboratory

Common name of food	
Sample code number	
Date of receipt in laboratory	
Handling stage	***Examples of record***
Weight and nature of inedible matter	Prior to further preparation (e.g. head and feet of poultry, outer wilted leaves)
Weight and nature of edible matter	Prior to further preparation (e.g. remainder of poultry carcass)
Method of preparation	Preparation of raw sample or cooking method, type, time, temperature and end-point temperature of foodstuff
Weight before cooking	
Ingredients added, if any	
Weight after cooking	
Weight and nature of edible portion of prepared food	
Weight and nature of inedible material	Bone, gristle, etc.
Method of mixing and reduction	Grinding, homogenizing in blender (type of blades)
Details of preparation of composite sample, if applicable	Simple mixing of equal weights or weighting of primary samples from the designated strata
Type of storage	Addition of preservatives, temperature of storage, etc.
Method used to take analytical samples	
Storage of analytical samples or further processing	
Name and signature of person completing record	
Date of record	
Other details	Any details that the collector thinks may be relevant

at the laboratory, but the details concerning local use and preparation method may need to be added during sampling.

Labels and lists of ingredients should be retained because they provide key information that may prove useful in explaining analytical discrepancies (e.g. foods where supplementary ingredients have not been added and the labelling is incorrect, differences in formulation of branded foods given the same names).

Record of handling in laboratory

Table 5.5d provides a record of the early preparation of samples in the laboratory leading up

to the preparation of the analytical samples. The laboratory may wish to add its own laboratory acquisition number. Laboratory record-keeping constitutes the first stage of a laboratory quality assurance programme, which will be discussed in detail in Chapters 6, 7 and 8. For this reason it is essential to preserve the linkage between the sample ID number and any laboratory acquisition number.

The primary samples will need to be unpacked and the sample compared with the information recorded in Tables 5.5a, 5.5b and 5.5c.

The protocol will specify whether the primary samples are to be analysed individually or combined in some way. Individual analysis of primary samples provides valuable information on the extent of variations in nutrient content, thus helping to define the confidence limits that can be ascribed to the mean values recorded in most databases. Individual analyses require substantial resources, however, and for many databases composite samples are analysed instead. The composite samples may comprise a simple combination of equal weights of all primary samples, or weighted amounts of primary samples from different strata or sampling points according to information on food consumption or production.

Throughout this handling stage, the principal objectives of the sampling process must remain foremost in the minds of everyone involved, namely to ensure the representativeness of the sample and to protect it from changes in composition and contamination. Table 5.6 summarizes the major effects of sample storage and preparation, the nutrients affected and the precautions to be observed.

The samples should be thawed carefully and handled as quickly as possible. Once again rehearsal of these procedures should always be carried out.

In separating the edible and inedible matter the cultural norms of the population consuming the food need to be considered. Complete documentation is essential for later use in the database.

When cutting, mincing or grinding food samples, protective measures must be taken to exclude the possibility of contamination. The procedures should be tested in advance (Wills, Balmer and Greenfield, 1980). The use of plastic or Teflon® coated tools may be necessary. Metal implements should not be used where iron and trace elements are to be analysed; some trace elements may be introduced by the use of stainless steel.

The physical characteristics of the sample are among the important factors to consider in preparing the samples. Lichon and James (1990) have reviewed and evaluated a range of 12 homogenization methods. One should also carry out pilot studies to check on the homogeneity produced by the chosen procedure and that fractionation of the samples has not occurred. Each food will need to be considered case by case.

Storage of the analytical samples

The logistics of sampling preparation usually mean that it is more convenient to store the analytical samples prior to analysis. At least three sample replicates should be stored. Storage in a frozen state is usually the minimum acceptable with preference given to –40 or even –70 °C, which is current common practice. Storage at –20 or –30 °C is acceptable for fat

Table 5.6 Effects of sample storage and preparation on nutrient content and precautions required to minimize them

Effects	*Potential changes*	*Nutrients affected*	*Precaution*
Drying out	Loss of water	All nutrients	Design of protocol. Keep samples in sealed containers or covered. Weigh food at start and during preparation
Absorption	Gain of water	All nutrients, especially in low-moisture and hygroscopic foods	Design of protocol. Keep samples in sealed containers
Microbial activity	Degradation/autolysis Synthesis	Losses of carbohydrates, proteins. Gains in thiamin, vitamin B_6, niacin and vitamin B_{12}	Storage at low temperature. Pasteurization or addition of inhibitors may be necessary
Oxidation	Destruction of unsaturated fatty acids	Alterations in profile of fats	Store at –30 °C in sealed containers under nitrogen. Addition of antioxidants or bacteriostatic agents
	Loss of vitamins	Losses of vitamin C, riboflavin and folates	
Acid	Hydrolysis	Losses of sucrose and higher oligosaccharides	Store at low temperature. Neutralize acid
Alkaline	Destruction	Loss of thiamin	Avoid alkaline conditions and SO_2
Light	Photodegradation	Loss of riboflavin	Protect from light
Contamination during sampling	From cooking vessels, soil, dust, etc.	Increases in inorganic nutrients	Design protocol to minimize contamination, gently rinse with distilled water
Contamination (from metallic blades, milling equipment, glassware, etc.)	Increase in inorganic nutrients	Increase in major trace elements	Select apparatus with care. Clean all utensils thoroughly before use and store in plastic bags
Separation	Separation of fats. Fractionation of particles	Changes in composition overall, alteration in fibre content	Avoid overvigorous mixing and thaw/freeze cycles
Enzymatic and metabolic activity	Changes in organic nutrients	Losses of sugars, vitamin C, folate deconjugation	Store at low temperatures. Protect folates with ascorbate

analyses. The container must be closely sealed with the minimum of headspace. When the samples are taken from storage any sublimed water above the sample must be carefully reincorporated in the mass.

Where freeze-drying is possible, storage of the freeze-dried samples in frozen or chilled conditions is satisfactory. Air-dried samples should be stored in such a way as to prevent uptake of water or contamination with insects or mites.

Preparation of analytical portions

In producing values for a compositional database a range of analytical procedures will be performed, requiring a number of analytical portions – often over a considerable time period (unless a large number of analytical staff are available). The procedures for taking the portions and their size will usually be defined by the nature of the analytical method to be used. It is imperative that all portions taken are representative and the methods used follow procedures defined by an established quality control programme.

Where analytical portions are repeatedly taken from the stored analytical samples the risks of contamination or taking an unrepresentative portion increase. It is therefore desirable to store a number of identical analytical samples and to minimize the number of staff involved in taking portions from them.

It is impossible to specify the sampling procedures for all methods and nutrients, but some typical procedures are given as examples in Appendixes 3 and 4.

Resource implications

The combined protocols provide a detailed basis for estimating the resources required for the sampling and analytical work. It may be necessary to revise the protocol, either by reducing the number of samples or being selective about the range of analyses to be carried out. This will require a re-examination of the processes used to establish the priorities described in Chapters 3 and 4. Combinations of analyses or extrapolation from related samples may be necessary.

Many compilers adopt the strategy of using a simplified sampling protocol for foods that are minor components of the diet and restricting the complete sampling protocols for core foods, foods that are major sources of nutrients and foods that are of greater importance in terms of public health.

Training

It is essential that all those involved in the sampling process are familiar with the objectives of the work and are clear about their roles. This can be done by rehearsal of the procedures if only as a "paper exercise". This process will identify aspects that are unclear or impracticable and require modification.

Table 5.7 Major sources of error in sampling

Source	*Examples*	*Precautions*
Food sample identification	Poor labelling of samples	Maintenance of documentation throughout sampling and analytical process
Nature of sample	Samples do not conform to the defined sampling protocol	Explicit instructions in sampling protocol, training of sampling staff
Transport and handling	Samples contaminated, degraded or depleted during transport or storage. Loss of samples	Protocol specifies conditions to be maintained, supervision
Analytical sample preparation	Incorrect mixing or homogenization	Proper supervision in laboratory. Laboratory quality assurance systems
Analytical sample storage	Incorrect storage of samples	Proper laboratory techniques and supervision

Table 5.7 summarizes the major sources of error in sampling. These highlight the central importance of documentation, staff training and supervision of the various stages. The sampling stages form the first and critical phases of a fully developed quality assurance programme (see Chapters 6, 7 and 8). Unless the samples are collected and handled correctly the analytical work – however well-executed – will be wasted because the values obtained will not relate to representative samples. It is however a truism that "one cannot inspect-in quality [by supervision], it must be built in". This depends on adequate staff training so that individuals fully understand their roles in the overall process.

Chapter 6

Choice of analytical methods and their evaluation

Reliable data on the nutrient composition of foods can only be obtained by the careful performance of appropriate, accurate analytical methods in the hands of trained analysts. The choice of the appropriate methods carried out under quality assurance schemes is the second crucial element in ensuring the quality of the values in a food composition database.

For many nutrients, several alternative analytical methods are available that, it is often assumed, give comparable results. In fact, methods vary in their suitability for a given analysis and different food matrices. Before the relative merits of particular methods are discussed in Chapter 7 it is necessary to consider the principles involved in method selection. In doing so it is recognized that the analysts' choices may be limited by the resources available; this makes it all the more important to understand the principles involved in method evaluation, particularly the need to define the limitations of any given method.

The evaluation of methods is not the purview of the analysts alone. The technical and scientific advisers to the database programme should be thoroughly conversant with the underlying principles of analytical methodology and the various methods themselves, sharing the responsibility with the analyst for choosing a method.

Compilers should also endeavour to be knowledgeable about the analytical methods used. They are responsible for scrutinizing methods when assessing non-commissioned data or published analyses to assess their suitability for inclusion in the database and to devise the specification for contracts for the preparation of sampling and analytical protocols.

It is also desirable that the professional users of a database should have some understanding of the analytical methods used, and that specialist users should be conversant with the methods used for the nutrient(s) relating to their special interests.

At present there are a number of methodological limitations in the production of data for certain nutrients. Based on a review of methods, Stewart prepared a table summarizing the position in 1980 and 1981, which was later extended by Beecher and Vanderslice (1984). In the table the nutrients were grouped according to the availability of valid methods to measure them. The expanded interest in nutrient composition in legislation and for use in epidemiological research has resulted in further work on method evaluation and development. In the United States, the Association of Analytical Communities (AOAC International) carried

Table 6.1 Availability of methods for nutrient analysis (adequacy of methods)

Nutrient	*Good*	*Adequate*	*Not adequate for certain foods*	*Lacking*
Moisture	Moisture			
Nitrogenous constituents	Total nitrogen, amino acids		Protein, non-protein nitrogen	
Lipid constituents	Fatty acids	Cholesterol, phospholipids, *trans* fatty acids, individual triacylglycerols	Some isomeric fatty acids	
Carbohydrates and dietary fibre	Individual sugars, starch, non-starch polysaccharides	Total dietary fibre, individual non-starch polysaccharides, resistant starch		Lignin
Inorganic nutrients	Sodium, potassium, calcium, magnesium, phosphorus, iron, copper, zinc, boron, chloride	Selenium, manganese, fluorine	Chromium, haem iron, cobalt, molybdenum	
Vitamins	Thiamin, riboflavin, niacin	Vitamin C, retinol, carotenoids, vitamin E, vitamin D, vitamin B_6, total folates, folic acid, biotin, pantothenic acid, vitamin B_{12}	Some carotenoid isomers, vitamin K	Some folate isomers

out a review of methods for use in nutrition legislation (Sullivan and Carpenter, 1993), and major reviews of micronutrient methods were undertaken by the Food Standards Agency in the United Kingdom (2002).

Studies for the development of standard reference materials (SRMs) undertaken in the United States by the National Institute of Standards and Technology (NIST) and in Europe by the Community Bureau of Reference (BCR) have also contributed to method development.

Stewart's original assessments have been updated in Table 6.1, which presents a revised version based on a review undertaken to assess the compatibility of methods (Deharveng *et al.*, 1999). In the table, "good" methods have been extensively evaluated in collaborative trials, "adequate" methods have been subjected to more limited study, and methods categorized as "not adequate for certain foods" have not been studied on a wide range of food matrices. It is important to note that these assessments hold true only when the analyses are carried out by trained analysts and that they do not include any consideration of speed or costs.

The table does not include the wide range of biologically active constituents that are now considered as candidates for inclusion in food composition databases. The methodologies for most of these constituents have not yet been widely studied in collaborative trials.

Choice of methods for nutrients

The primary objective of food composition databases is to provide their users with compositional information on nutrients; therefore the primary factor in the choice of methods is the appropriateness of the analysis in terms of providing the information required by the users. The measurements must provide values that can be used to assess the nutritional value of foods. This means that the database users' requirements may differ from those concerned with the regulation of food composition or the quality control of food in production. Thus, while the measurement of crude protein (total nitrogen multiplied by a factor) is adequate for many purposes, amino acid data would provide a better assessment of the nutritional value of a food. A value for total lipids may be adequate in relation to food quality control, whereas a nutritionist would require assessments of triacylglycerols, sterols and phospholipids separately and detailed fatty acid data. Similarly, while total carbohydrate values may be adequate for food quality control, a nutritionist would require specific values for the different carbohydrates (FAO/WHO, 1998). As a consequence, more biochemically orientated methods are often required when obtaining values for food composition databases.

In some countries, the choice of method may be prescribed by national legislation. In other countries, the regulations often permit the use of methods that give comparable, i.e. similar, values to those obtained by the official methods.

Other considerations will also influence the choice of method. The use of some of the most advanced methods may require substantial capital investment to provide the necessary instrumentation. Considerable resources are also required in the form of trained staff to operate and maintain the instrumentation. The development of such instrumental methods

represents a preference for investing in capital rather than in recurrent staff costs and for reducing the cost per analysis by speeding up analysis.

It is incorrect to give the impression that nutrient analyses cannot be performed without such sophisticated instrumentation; for many nutrients classical manual methods are available that give equally sound values. These methods are labour-intensive rather than capital-intensive.

It is true that analyses of certain nutrients, fatty acids for example, do require instrumentation; where this is lacking a laboratory would need to seek collaborative arrangements to acquire the data.

Laboratories in developing countries may lack funds for capital outlay (especially as foreign currency) and lack the resources for the specialized maintenance and supplies necessary for high-technology instrumentation. On the other hand, local funds may be available for technical staff with the necessary background for carrying out non-instrumental methods that provide valid data. A comprehensive range of compatible methods has therefore been covered in Chapter 7.

Laboratories should focus their attention on evaluating and improving the quality and performance of the methods currently employed rather than attempting to institute a wide range of methods using new, untried, methods or losing confidence because of their lack of sophisticated equipment. In many cases, implementing a data quality assurance system and training staff are often better ways to produce good-quality compositional data.

The formal training of food analysts, where it is carried out, usually focuses on the highly accurate detection of compounds appropriate for food regulations. These compounds are often contaminants, which are present at low levels, and the choice of method generally emphasizes levels of detection, sensitivity and precision. In nutrient analysis for a food composition database, the requirements for accuracy and precision may be orientated more towards the recommended intake of a nutrient and the relative importance of the food being analysed in the diet (Stewart, 1980). Analysts may, for example, spend considerable effort measuring vitamins in foods at levels that are nutritionally insignificant.

This difference in emphasis underlines the need for all individuals involved in producing data to be familiar with the objectives of the work, from sampling through to analysis. Sampling protocols should specify the levels of accuracy that are expected. It is also important to maintain a regular dialogue between compilers and the sampling and analytical teams throughout the duration of the work.

While the appropriateness of the method may be a primary factor in method selection, it is also necessary to take into account the analytical attributes of the method.

Criteria for choice of methods

It is useful to consider a number of points suggested by Egan (1974):

1. Preference should be given to methods for which reliability (see below) has been established by collaborative studies involving several laboratories.

2. Preference should be given to methods that have been recommended or adopted by international organizations.
3. Preference should be given to methods of analysis that are applicable to a wide range of food types and matrices rather than those that can only be used for specific foods.

The analytical method selected also needs to have adequate performance characteristics. Büttner *et al.* (1975) summarize these as reliability criteria (specificity, accuracy, precision and sensitivity) and practicability criteria (speed, costs, technical skill requirements, dependability and laboratory safety).

Thus "reliability" represents a summation of the more conventional measures of method performance. Many analysts would also consider another attribute as falling within this summation: "robustness" or "ruggedness". This attribute is described below.

Attributes of methods

(adapted from Horwitz *et al.* [1978], with permission)

Reliability

This is a qualitative term expressing a degree of satisfaction with the performance of a method in relation to applicability, specificity, accuracy, precision, detectability and sensitivity, as defined below, and is a composite concept (Egan, 1977). It represents a summation of the measurable attributes of performance. The analyte and the purposes for which the analyses are being made determine the relative importance of the different attributes. Clearly, the analysis of a major constituent such as protein, fat or carbohydrate in foods does not require the same low limit of detection as that needed for the measurement of a carcinogenic contaminant. Conversely, the measurement of a constituent at low levels in foods (e.g. most trace elements, selenium, chromium or vitamins such as vitamin D, vitamin B_{12} and folates) cannot be expected to deliver the same high accuracy or precision as found with the major constituents.

Horwitz, Kamps and Boyer (1980) found from a study of the results of a large number of collaborative studies undertaken under the auspices of the AOAC that there was a strong empirical relationship between the concentration of an analyte and the observed precision obtained by experienced analysts. The relationship they found was:

$$CV = 2^{(1 - 0.5 \log C)}$$

where CV is the coefficient of variation and C the concentration g/g.

Many workers use this relationship when assessing the performance of methods for nutrients present at low concentrations.

Applicability

This is also a qualitative term. A method is applicable within the context in which it will be used, for example the analysis of a specific food matrix. Applicability relates to the freedom

from interference from other constituents in the food or from the physical attributes of the food matrix that would make extraction of the analyte incomplete. Applicability is also determined by the usable range of the method. Methods that are applicable at high concentrations may not be applicable at low concentrations. Equally, a method may be applicable to one matrix (e.g. meat) but be inappropriate for another (e.g. a cereal product).

All unfamiliar methods or methods described for a specific food must be checked carefully when used for a matrix that is different from those for which it has been used previously.

Specificity

Specificity is the ability of a method to respond exclusively to the substance for which the method is being used. Many methods are "semi-specific", relying on the absence of interfering substances in the food being examined. Sometimes a method with poor specificity is acceptable when the purpose of the analysis is to measure all similar substances within a group (e.g. total fat, ash).

Accuracy

Accuracy is defined as the closeness of the value obtained by the method in relation to the "true value" for the concentration of the constituent. It is often expressed as percentage accuracy. Inaccuracy is, as a corollary, the difference between the measured value and the "true value".

The concept of a "true value" is, of course, hypothetical because the "true value" for a nutrient in a food is not known. All analytical values are therefore estimates of that value.

Büttner *et al.* (1975) take the view that there exists a true value for all constituents in a sample of food. This is fundamental to the analysts' art; it is not true that the value for a defined analytical sample of a food is the "true value" for all samples of that food. The sampling error and the analytical errors for any specific method determine the confidence limits for all determined values.

The accuracy of a method is usually determined by reference to standard amounts of the analyte and preferably by the analysis of standard reference materials (SRMs) or certified reference materials (CRMs) that have been analysed, often using several compatible methods, by a group of skilled analysts to provide certified values together with the confidence limits of that value.

Precision

Precision is a measure of the closeness of replicated analyses of a nutrient in a sample of food. It is a quantitative measurement of "scatter" or analytical variability. Strictly speaking, it is imprecision that is measured by carrying out replicate analyses on the same sample (which must be homogeneous and stable). The measurements may be made by one analyst within one laboratory when the assessment is designated "repeatability" (that is, within-laboratory precision) or by several analysts in different laboratories when it is designated "reproducibility" (that is, between-laboratory precision). Comparisons can also be made among different analysts in one laboratory (called "concordance"), and by one analyst on different occasions.

In each case the standard deviation (SD) of the analytical values is calculated (which means that there must be a sufficient number of replications). The SD is customarily divided by the mean value to give a relative standard deviation (RSD), or multiplied by 100 to give the coefficient of variation (CV). In analytical literature, the RSD is used for reproducibility and rsd for repeatability.

It is important to recognize the distinction between accuracy (see the definition above) and precision. One can have very high precision (a low RSD) and poor accuracy and, conversely, have high accuracy with poor precision where the confidence limits of the value obtained will be wide. The ideal is to combine high precision (low RSD) with high accuracy (as judged by the value obtained with an SRM).

Detectability

Detectability is defined as the minimum concentration of analyte that can be detected. This is rarely an issue in nutritional studies, as very low concentrations of nutrients, even some trace elements or vitamins, are not usually nutritionally significant. These are customarily recorded as "trace" in many printed food composition tables. However, it is useful to know whether or not a nutrient is present, and at what level one can confidently record zero in a database. The detectability limit of a method is the concentration at which the measurement is significantly different from the blank. Since blank values also show some variability, the limit can be defined as greater than +2SD (of the blank measurements) above the blank level. The detection limit is below the concentration at which measured values can be made; that is, it is outside the usable range of the method.

Sensitivity

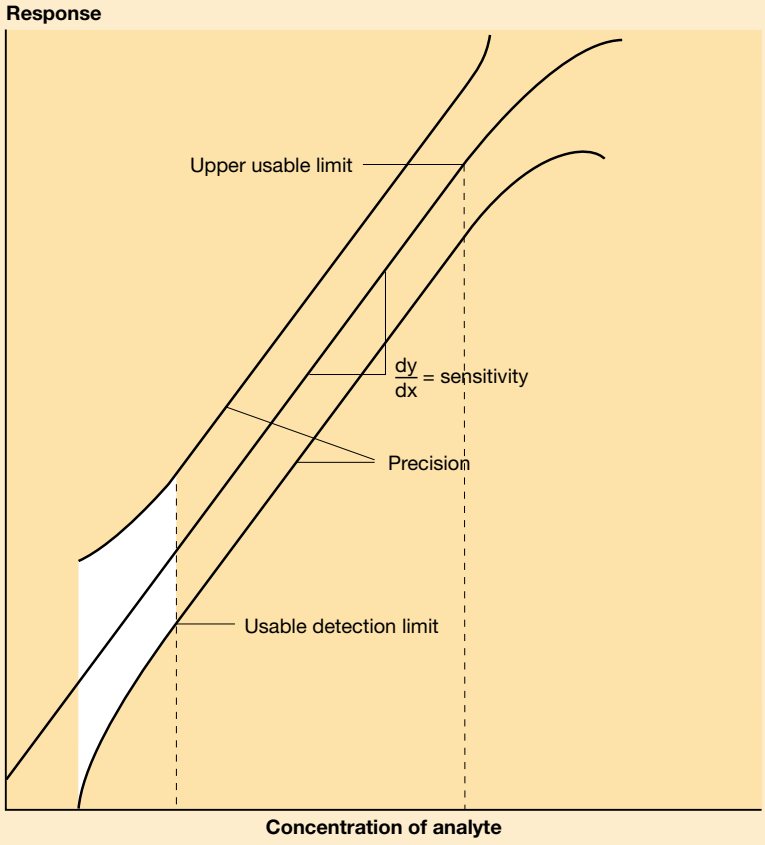
Sensitivity in analytical terms is the slope of the response–concentration curve or line (Figure 6.1). If the slope is steep the method has a high sensitivity; conversely, if the slope is shallow the method has a low sensitivity. When a narrow range of concentration is of interest, a high sensitivity is often desirable; for a wide range of concentrations, a low sensitivity may be preferable. In most nutritional composition studies, trace element analysis requires high sensitivity. In practice, this can often be achieved by increasing the response signal strength by electronic amplification or through chemical concentration of the element.

High sensitivity is usually required for the analysis of contaminants. While contaminants are not usually included in food composition databases, they may become more important in the future, especially those with antinutritional or toxicological properties.

Robustness (ruggedness)

This is a qualitative attribute and refers to the capacity of a method to perform adequately in the face of fluctuations in the analytical protocol. Such fluctuations could include the timing of stages, changes in temperature, or the precise concentrations of reagents. It also includes variations in the skill, training and experience of the analysts carrying out the method. Ideally, during the initial development of a method its authors should have explored and

Figure 6.1 Response as a function of concentration, illustrating the attributes of methods



Source: Modified and reproduced with permission from Stanley L. Inhorn, ed., *Quality assurance practices for health laboratories*. Copyright 1978 by the American Public Health Association.

documented the capacity of the method to withstand these types of fluctuation and to perform under a variety of conditions. Methods are available for examining such variations (Youden and Steiner, 1975).

Authors of analytical methods should identify the stages in their methods that require strict attention and control, and document these in the published description of the method.

Summary of attributes

Figure 6.1 provides a diagrammatic summary of the attributes. In the figure the response

(height, area, weight, volume, time, optical density or another type of measurement) is shown as primarily a linear function up to a certain level that defines the usable range of the method. Where only a single analyte elicits the response, the method is specific; this specificity may be inherent in the method or may be achieved by chemical separation from interfering substances. This, therefore, is a property of the chemistry of the analyte and of potential interfering substances. The sensitivity of the method is indicated by the slope of the response line. The confidence envelope indicates the precision of the method and the difference between the response line and the hypothetical true line represents the measure of accuracy. The confidence envelope can be calculated at any level, but 95 and 99 percent are commonly used. In the former case, only 1 in 20 measurements can be expected to fall outside the envelope and in the latter only 1 in 100. The white area represents the region of uncertainty where the relative standard deviation is so large that no certainty can be assigned to a value.

Validating analytical methods

Even well-established methods need to be evaluated by the analysts themselves, using their own staff, reagents and equipment (Wills, Balmer and Greenfield, 1980). An evaluation of the attributes of the method should be established under the conditions prevailing in the laboratory and the performance characteristics that are relevant to the purpose of the analyses should be quantified.

Reviewing the method as a whole

In the first stage of the evaluation, the analysts should familiarize themselves with the method as described in the formal protocol for the method concerned. This begins with a "paper exercise" to ensure that the principle of the method is understood and that the various stages are clear in the analysts' minds. The list of reagents required should be checked against the procedures. Occasionally, a common reagent will be omitted from the reagent list because the authors assume that all laboratories will have it to hand. Standardization of some reagents may be needed before the method is started. At the same time, the analysts should check the equipment required and any specifications listed for the equipment.

Finally, the analysts should go through each stage, familiarizing themselves fully with its purpose. At this point it is suggested that an assessment of the criticality of each stage is made, as recommended in the ANALOP approach (Southgate, 1995); this exercise will determine the possibility for error or uncertainty that might occur if the conditions described are not followed precisely.

Timing may or may not be critical. For example "leaving overnight" may imply a specific time period, say from 18.00 to 09.00 the following day (i.e. 15 hours), or merely that when

this point is reached the method can be left until the following day – an indeterminate time period. Timing may represent a minimum time period; alternatively, "heat for 10 minutes in a boiling water-bath", for example, may mean "exactly 10 minutes" or "while the analyst takes coffee". Understanding the critical timed stages is especially important when a method is carried out for the first time and until it becomes "routine".

Analogously, the concentrations of certain reagents are also critical, especially when the reagent must be used in excess for a reaction to be fully completed.

Using the published description of a method as one would follow a recipe in cooking can be fraught with disaster. The analyst must understand the logic of a method. Running through a method as a trial and discounting the results is useful for checking the stages, especially with regard to timing. Less-experienced staff may take time to adjust themselves to a procedure for which the published account of the method suggests that there are many critical operations (e.g. as in the non-starch polysaccharide method [Englyst, Quigley and Hudson, 1994], where the mixing stages are critical). Once this assessment is completed, the analyst will be in a better position to evaluate the various performance attributes.

Applicability

The application of an unfamiliar method to a food matrix other than that for which it was developed or used previously requires careful consideration. It will be necessary to decide, often intuitively, how the matrix will behave in an extraction phase and whether there is any likelihood of interfering substances being present. The chemistry of the analyte and the expected range of the nutrient in the "new" food will therefore need to be considered.

Such matters cannot always be decided intuitively, however, and the method must be tested on the food material. The use of different analytical portions will provide evidence of interference or indicate possible problems with extraction or inadequate concentrations of reagents.

The recovery of standard amounts of the analyte added to the sample can establish whether extraction is complete. Recovery tests are not completely adequate because the added analyte may be more easily extractable than the intrinsic nutrient. Poor recoveries indicate problems; good recoveries may be regarded as encouraging but not conclusive.

Comparisons with values reported in the literature for the matrix may be helpful, as may collaborative studies with another laboratory.

Specificity

Assessing this attribute requires knowledge of the chemistry of the analyte and the food matrix. A value may be required for a group of substances, such as total fat (lipid solvent soluble) or sugars, in which case a semi-specific method may be adequate. Values for triacylglycerols or individual sugars, however, require a much more specific method. Certain vitamin values must include all the active forms; for example, vitamin A (retinol) values should include other active retinoids. Here again, specificity is critical.

Accuracy

This is a difficult attribute to measure because its true value is unknown. The first stage is to analyse standard amounts of the pure analyte. Recovery studies of standards added to the foods are useful, especially if a series of different amounts is used and then a comparison made of the sensitivity of the method for pure standards and the added standards. Recovery studies, as mentioned above, do not provide unequivocal proof of the accuracy of a method because they assume that the added nutrient may be extracted with the same efficiency as the intrinsic nutrient (Wolf, 1982).

Analysis of authentic samples

Analysis of authentic samples that have already been analysed by another laboratory is a useful guide for analysts using a method for the first time. This procedure forms what might be regarded as a simple type of collaborative study.

Analysis of standard reference materials

Standard reference materials are unique materials with a range of food matrices (limited at present but increasing in numbers) that have been produced by a national or regional organization such as the National Institute of Standards and Technology (NIST, 2003a) in the United States or the Community Bureau of Reference (BCR) for the European Union (BCR, 1990; Wagstaffe, 1985, 1990). The samples have been very carefully homogenized and rigorously tested for homogeneity and stability under different storage conditions for different lengths of time (Wolf, 1993). They are then analysed using well-defined analytical methods. Where possible, a number of different compatible methods based on different principles are used. The values generated are then certified with defined confidence limits for the values. The range of nutrients for which SRMs or CRMs are available is limited (but increasing). Coverage is good for many constituents, including some trace elements, some fats, fatty acids, total nitrogen and cholesterol.

SRMs (or CRMs) are expensive to produce and therefore too costly to use routinely (say, with every batch of analyses – which would be the ideal). Each laboratory (or a group of local laboratories) should therefore consider preparing in-house reference materials using similar approaches to that used to produce SRMs (Southgate, 1995).

The homogenized material is stored in a large number of individual containers and used routinely in the application of the method and occasionally alongside the SRM. Recording the values obtained over time on a control chart will help identify any trends towards high or low values. A control chart usually has a central line indicating the control limits for a statistical measure (SD for example) for a series of analyses (American Society for Quality Control, 1973). The laboratory results are plotted on the vertical axis against time (days,

hours, etc.) on the horizontal axis. The horizontal scale should provide for at least three months of data and the chart should be checked regularly for evidence of runs above or below the central line or any evidence of lack of randomness (Mandel and Nanni, 1978; Taylor, 1987). Theoretically, the values should be randomly distributed about the central line. When they fall consistently above (or below) the line, they represent possible indicators of systematic bias in the method, which should be investigated.

The preferred materials for in-house reference materials are non-segregating powders such as non-fat milk powders, gelatine, flours, powder mixes for parenteral feeds (Ekstrom *et al.*, 1984) and food matrices common to the local food supply, e.g. soybean meal and fishmeal for ASEANFOODS (Puwastien, 2000). Torelm *et al.* (1990) describe the production of a fresh reference material based on a canned meat.

One alternative is to carry out analyses using standard samples on a routine basis using a control chart to alert laboratory personnel to problems requiring remedial action.

Precision

The original published description of a method usually gives some indication of the level of precision achieved in collaborative studies, thus providing a "standard of achievement". Each laboratory, once its personnel are familiar with the method, should evaluate its own levels of precision.

The first step is for each analyst to assess their repeatability by analysing several replicates (preferably at least ten) of the same material and calculating the relative standard deviation. Second, all the analysts within the laboratory should analyse several replicates (preferably ten) of the same material to assess concordance within the laboratory. When setting up a method for the first time, it is useful to test repeatability and concordance using standards. Using blind concentrations of standards prepared by colleagues gives further confidence when using an unfamiliar method.

Finally, participation in a collaborative trial to assess the reproducibility of the method and to evaluate the laboratory repeatability with other analysts is a valuable approach that can be useful as part of the development of analytical skills.

Formal schemes exist for the collaborative analysis of some nutrients; samples for analysis are provided on a regular basis by NIST (2003a) in the United States and by the National Accreditation of Measurement and Sampling (NAMAS) in the United Kingdom (UKAS, 2003). In addition, Wageningen University in the Netherlands is the base for the International Plant-analytical Exchange (IPE, 2003), which provides a basis for developing analytical proficiency, especially for trace elements.

Difficulties may be encountered with regard to the entry of food materials into certain countries and most schemes are quite expensive, which may be a prohibiting factor where resources are limited. In such cases, the organization of local collaborative studies should be considered.

Collaborative studies

There are three major types of collaborative study. The first type, sometimes known as a "round robin", or "ring test", provides comparative assessments of laboratory performance. Homogeneous samples of food, often with their identities concealed, are distributed centrally, together with guidance on the preparation of standards and the calculation of results. The results are then collected centrally and analysed statistically. The results are usually provided to the participating laboratories in the form of charts showing the performance of each laboratory against the analyses as a whole. Each laboratory is given a code number and can assess its own performance. Outliers where the values obtained are significantly different from the mean and reproducibility found in the trial are also indicated. This type of collaborative study is of most benefit to laboratories involved in compositional analysis that wish to test and improve their performance.

A second type is that used by the Association of Analytical Communities (Thompson and Wood, 1993; AOAC International, 2003) to establish the performance of a method. In this case the collaborating analysts analyse a series of food samples supplied centrally, using a common analytical protocol. Standards and some reagents, where the specifications are critical (such as enzymes), are also supplied centrally, as are forms for calculating, expressing and recording the results. At least eight, but preferably more, analysts and laboratories are involved in such a study. The results are collected and analysed statistically, usually by an associate referee. The performance characteristics are used in the assessment of the method before it is accepted into the Official Methods manual.

A third type of study is used by the BCR in the European Union, primarily in the development of standard certified materials. Here, a group of laboratories analyses samples provided centrally, initially using their routine methods. Standards may be distributed together with forms describing how the results should be expressed. The results are collected centrally and analysed statistically. The findings are distributed and the analysts subsequently called to a meeting. The object of the meeting is to assess the different methods and identify where laboratories using the same methods found different values. Agreement is then reached on protocols that should be followed in a second round.

The results from the second round of the study will often identify methods that give satisfactory reproducibility and those methods that give similar results, although a third round may be required. These methods are then used in a carefully controlled certification study of food materials intended for potential reference materials. The ideal is to have a number of methods, based on different principles, that are compatible. In some instances the certification can only be given for values obtained by only one method.

It is important that the analysts involved in collaborative studies of this nature see the primary objectives of the studies as raising standards of analytical performance and furthering the development of analytical skills and not as a management tool for checking the performance of analysts.

Checking calculations and analyses

When anomalous results appear in collaborative studies or in routine analyses, for example on the control charts, the first step is to go through the logic and application of the calculations, as these are the most frequent causes of anomalous results. Most collaborative studies define the calculations explicitly to avoid such problems, but they still occur. For this reason the calculation procedures should be set out in a logical fashion within the analytical protocols.

The second stage is to repeat the analyses with a series of freshly prepared standards. Improper dilutions or weighing are frequent causes of error.

In the third stage the analyses are repeated by another, more experienced, analyst. Repeating the analyses using a portion from an earlier stage of the analyses does not constitute a rigorous check; ideally, fresh analytical portions should be used. Neither does simple repetition provide an adequate check because any bias related to the standard or the food matrix may be replicated.

If the results still appear anomalous the analyst should analyse the sample blindly using only its sample code number and, if possible, a colleague should be asked to introduce a "blind" replicate. Southgate (1987) has identified a range of laboratory practices that may lead analysts to believe, erroneously, that they have achieved good repeatability and how these practices can be changed (Table 6.2).

All these operations form part of a data quality assurance scheme and their documentation is vital for database compilers when they come to assess the quality of the analytical data, which is discussed in Chapter 8.

Table 6.2 Operational practices that may lead to systematic errors

Operation	*Common practices*	*Remedy*
Size of analytical portion	Identical or closely similar analytical portions	Work with replicates of different sizes
Reagents used	Always from same batch	Vary sources of reagents
Standard solutions	Prepared from same stock or same series of dilutions	Prepare fresh standards regularly
Replication of analyses	Analysed in same batch or at the same time	Analyse replicates in different batches or different days. Participate in collaborative studies
Analyst	Only one analyst	Carry out analysis with different analysts regularly. Collaborate with other analysts Exchange samples
Choice of procedure	Only one procedure	Where possible, use methods based on different principles. Collaborate with other laboratories

Source: Modified from Southgate, 1987.

Chapter 7

Review of methods of analysis

This review of analytical methods presents assessments of their applications, limitations and the resources required. The objective of the review is to provide guidance on the selection of compatible methods for the nutrients and some other constituents. The continuous developments in analytical chemistry make it almost impossible to ensure that the review is comprehensive and takes into account all recent developments. The review does not provide detailed analytical protocols; for these the reader needs to consult the relevant specialist texts.

In this review, for each nutrient (or group of nutrients), tables summarize the available methods. Estimates of capital costs have been given in three categories: low, where the method requires basic equipment that would usually be found in a laboratory; medium, where specialized instrumentation is required but normally costing less that US$5 000; high, indicating the need for specialized equipment usually costing more than US$10 000.

The proximate system of analysis

The proximate system for routine analysis of animal feedstuffs was devised in the mid-nineteenth century at the Weende Experiment Station in Germany (Henneberg and Stohmann, 1860, 1864). It was developed to provide a top level, very broad, classification of food components. The system consists of the analytical determinations of water (moisture), ash, crude fat (ether extract), crude protein and crude fibre. Nitrogen-free extract (NFE), more or less representing sugars and starches, is calculated by difference rather than measured by analysis.

Although some of the methods used historically in the proximate system of analysis are not recommended for the preparation of food composition databases (e.g. crude fibre), it is useful to consider the concepts involved as they have dominated views on the composition of foods and food analysis. This system was developed at a time when the chemistry of most food constituents was only partially understood, and the growth of nutritional sciences has shown that for nutritional studies a more detailed and biochemically oriented approach to

Table 7.1 Methods of analysis for water

Procedure	*Applicability*	*Limitations*	*Capital costs*	*Selected references*
Physical removal of water				
Air oven at 100–105 °C	Most foods, except those rich in sugars and fats	Caramelization of sugars, degradation of unsaturated fats, loss of other volatiles	Low	AOAC International, 2002; Anklam, Burke and Isengard, 2001; Nielsen, 1998
Vacuum oven at 60 °C	Most foods	Loss of volatiles	Low	As above
Freeze-drying	Most foods	Slow, residual water in samples	Medium	As above
Microwave oven	Medium or high moisture	Charring	Low	As above
Dean & Stark distillation	Foods high in volatiles	Safety of solvents used	Low	As above
Chemical reactivity				
Karl Fischer	Low moisture, hygroscopic foods		Low	As above
Physical methods				
NMR	Most foods	Need for calibration with specific food	High	Bradley, 1998; Hester and Quine, 1976
NIR	Established for cereals and some other foods	Need for extensive calibration with specific food. Particle size dependence	High	Williams, 1975
Chromatography				
GLC	Meat and meat products		High	Reineccius and Addis, 1973
GSC	Some meat products		High	Khayat, 1974

Notes:
References selected provide detailed procedures, evaluations or reviews.
NMR = nuclear magnetic resonance; NIR near infrared reflectance; GLC = gas–liquid chromatography; GSC = gas–solid chromatography.
Low, Medium, High capital costs are described in the text.

food analysis is needed. Nevertheless, proximate analysis, including the original methods, still forms the basis for feed analysis, and the analysis of foods for legislative purposes in many countries.

Many people find the concept and term "proximates" useful to represent the gross components that make up foods; the actual analytical methods then become independent. Others believe that the definition of proximates is based on the original methods prescribed by Henneberg and Stohmann, and that method substitution, e.g. dietary fibre instead of crude fibre, negates the use of the term.

Water

Values for water remain an essential constituent in food composition databases because water content is one of the most variable components, especially in plant foods. This variability affects the composition of the food as a whole. The range of methods for water analysis is summarized in Table 7.1.

The methods are based on the direct or indirect measurement of water removed from the food, changes in physical properties that change systematically with water content, or the measurement of the chemical reactivity of water (Egan, Kirk and Sawyer, 1987; AOAC International, 2002; Sullivan and Carpenter, 1993; Southgate, 1999; Bradley, 1998).

For the majority of foods in food composition databases, drying methods are adequate; although slight methodological differences can be observed, these differences are rarely significant. The AOAC Official Methods recommend a lower drying temperature (70 °C) for plant foods to minimize the destruction of carbohydrates. Where this occurs it is usually better to use vacuum drying or freeze-drying.

Vacuum drying is most efficient if a slow leak of dry air is passed through the oven. This approach has the advantage that the analytical portions can be left unattended for long periods. Vacuum drying at 60–70 °C is preferable to drying in an air oven, particularly for foods that are rich in sugars. However, for most foods drying in an air oven is satisfactory for food composition database purposes.

Freeze-drying requires more capital investment but has the advantage that it dries the foods under mild conditions. Freeze-dried material is light, easily transported and can also be ground very easily. The process does, however, usually leave some residual moisture in the freeze-dried material, which must be removed to give values that are comparable with other drying methods.

Drying in a microwave oven is very quick but requires continuous surveillance to avoid charring. Drying with infrared lamps has been very successfully automated (Bradley, 1998). Both of these methods, however, are more suitable for routine quality control.

All the methods mentioned so far are unsuitable for foods with a high content of volatile components because these are driven off with the water. The Dean and Stark method can be used for such foods where a value for the moisture content is required. In this method the

Table 7.2 Methods of analysis for nitrogen and protein

Procedure	*Applicability*	*Limitations*	*Capital costs*	*Selected references*
Total nitrogen				
Kjeldahl	Manual, all foods	Minor interference from inorganic nitrogen	Low	AOAC International, 2002; Sullivan and Carpenter, 1993
	Automated, at several levels of complexity	Minor interference from inorganic nitrogen	Medium	
Dumas	Automated, all foods	Includes inorganic nitrogen. Analytical portion size	High	AOAC International, 2002
Radiochemical methods	Most foods	Instrumentation required	Very high	Pomerantz and Moore, 1975
Protein				
Total N × factor	All foods	Variations in NPN	Low	FAO/WHO, 1973
Protein N × factor	Preferable for vegetables, some fish, yeast foods, insect foods, breastmilk	Choice of procedure for measurement of NPN. Better to use amino acid N	Low	Koivistoinen *et al.*, 1996; Bell, 1963
Methods applicable to specific foods				
Formol titration	Dairy products	Specificity	Low	Taylor, 1957; AOAC International, 2002; Chang, 1998
Biuret	As above	Specificity	Low	Noll, Simmonds and Bushuk, 1974; as formol
Folin's reagent	As above	Specificity	Low	Lowry *et al.*, 1951; Huang *et al.*, 1976; as formol
Alkaline distillation	Cereals	Specificity	Low	Chang, 1998
Dye-binding	Specific foods, some cereals, some legumes	Specificity	Low	As formol
NIR	Established for some foods	Number of calibration samples	High	Hunt *et al.*, 1977a

Notes: References selected provide detailed procedures, evaluations or reviews. NPN = non-protein nitrogen; NIR = near infrared reflectance.

water is distilled off as an azeotropic mixture with an immiscible solvent such as toluene, xylene or tetrachloroethylene. The method is an AOAC-approved method for spices and cheese, and has achieved good levels of precision (AOAC International, 2002).

The Karl Fischer method is especially useful for foods with very low moisture content and for hygroscopic foods that are difficult to dry using conventional methods. The levels of accuracy achieved are rarely required for food composition databases.

The physical methods for measuring water content require expensive, highly specialized instrumentation and are most suitable where there is a very high throughput of similar samples.

Near infrared reflectance (NIR) methods, for example, have been widely applied for the analysis of cereal grains. The method requires calibrating with a large number of samples with moisture values measured by conventional methods to develop the analytical equations. Nuclear magnetic resonance (NMR), gas–liquid chromatography (GLC) and gas–solid chromatography (GSC) methods also require detailed calibration and are of greatest value in measuring the distribution of water in foods and identifying the forms of water in meats.

Nitrogen and nitrogenous constituents

Lakin's (1978) review still provides a comprehensive account of the analysis of nitrogen and nitrogenous constituents, and the methods are discussed briefly by Sullivan (1993) when reviewing the AOAC Official Methods, by Chang (1998) and by Southgate (1999). The range of methods is summarized in Table 7.2.

Total nitrogen

The proximate system, where "protein" is measured as total nitrogen multiplied by a specific factor, continues to dominate food composition studies. Most cited values for "protein" in food composition databases are in fact derived from total nitrogen or total organic nitrogen values. In the majority of cases, total nitrogen is measured using some version of the Kjeldahl (1883) method (which measures total organic nitrogen). In this method the organic matter is digested with hot concentrated sulphuric acid. A "catalyst mixture" is added to the acid to raise its boiling point, usually containing a true catalytic agent (mercury, copper or selenium) together with potassium sulphate. All organic nitrogen is converted to ammonia, which is usually measured by titration or, more rarely, colorimetrically. In the original method, a relatively large analytical portion (1–2 g) was used, but this requires large amounts of acid. Micro-Kjeldahl methods are much more commonly used as they produce a reduced amount of acid fumes and also require less acid and catalyst mixture. Environmental considerations exert considerable pressure to ensure the safe disposal of mercury and, especially, to minimize acid usage.

The micro methods can be automated at several levels (Egan, Kirk and Sawyer, 1987; Chang, 1998). Automation of the distillation and titration stages works well but automation of the digestion has proved quite difficult.

The Dumas method measures the total nitrogen as nitrogen gas after complete combustion of the food. Comparison of the results obtained with those obtained using the Kjeldahl method shows good agreement (King-Brink and Sebranek, 1993). The method has been successfully automated and, although the instrumentation is expensive, a high throughput of samples is possible, with good precision. The equipment uses very small analytical portions, and a finely divided analytical portion is essential.

NIR can also be used to measure nitrogen in some foods, although a large number of calibration samples is required.

Protein

Since the development of the proximate system of analysis, "crude protein" values have been calculated by multiplying the total nitrogen (N) by a certain factor. This factor was originally 6.25, based on the assumption that proteins contained 16 percent of N. It has been known for a considerable time that proteins of plant origin (and gelatin) contain more N and therefore require a lower factor. Jones, Munsey and Walker (1942) measured the nitrogen content of a wide range of isolated proteins and proposed a series of specific factors for different categories of food. These factors have been widely adopted and were used in the FAO/WHO (1973) review of protein requirements. These are listed in Table 7.3. Several authors have criticized the use of these traditional factors for individual foods (e.g. Tkachuk, 1969). Heidelbaugh *et al.* (1975) evaluated three different methods of calculation (use of the 6.25 factor, use of traditional factors and summation of amino acid data) and found variations of up to 40 percent. Sosulski and Imafidon (1990) produced a mean factor of 5.68 based on the study of the amino acid data and recommended the use of 5.70 as a factor for mixed foods.

In principle, it would be more appropriate to base estimates of protein on amino acid data (Southgate, 1974; Greenfield and Southgate, 1992; Salo-Väänänen and Koivistoinen, 1996) and these were incorporated in the consensus document from the Second International Food Data Base Conference held in Lahti, Finland, in 1995, on the definition of nutrients in food composition databases (Koivistoinen *et al.*, 1996).

If these recommendations are to be adopted, the amino acid data should include values for free amino acids in addition to those for protein amino acids because they are nutritionally equivalent. The calculations require very sound amino acid values (measured on the food) as discussed below, and involve certain assumptions concerning the proportions of aspartic and glutamic acids present as the amides and correction for the water gained during hydrolysis. Clearly, this approach would not be very cost-effective when compared with the current approach.

At the present time it is probably reasonable to retain the current calculation method, recognizing that this gives conventional values for protein and that the values are not for true protein in the biochemical sense. However, it is important to recognize also that this method is not suitable for some foods that are rich in non-amino non-protein nitrogen, for example cartilaginous fish, many shellfish and crustaceans and, most notably, human breastmilk, which contains a substantial concentration of urea.

Table 7.3 Factors for the conversion of nitrogen values to protein (per g N)*

Foodstuff	*Factor*
Animal products	
Meat and fish	6.25
Gelatin	5.55
Milk and milk products	6.38
Casein	6.40
Human milk	6.37
Eggs	
whole	6.25
albumin	6.32
vitellin	6.12

Foodstuff	*Factor*
Plant products	
Wheat	
whole	5.83
bran	6.31
embryo	5.80
endosperm	5.70
Rice and rice flour	5.95
Rye and rye flour	5.83
Barley and barley flour	5.83
Oats	5.83
Millet	6.31
Maize	6.25
Beans	6.25
Soya	5.71
Nuts	
almond	5.18
Brazil	5.46
groundnut	5.46
others	5.30

* (Where a specific factor is not listed, 6.25 should be used until a more appropriate factor has been determined.)

Source: FAO/WHO, 1973.

A number of direct methods for protein analysis have been developed for specific foods based on reactions involving specific functional groups of the amino acids present; these are thus not applicable to the measurement of proteins in general. Such methods include formol titration (Taylor, 1957) and the biuret reaction (Noll, Simmonds and Bushuk, 1974). A widely used group of colorimetric methods is based on reaction with Folin's reagent, one of the most widely used biochemically in the dairy industry (Lowry *et al.*, 1951; Huang *et al.*, 1976). These methods are most commonly calibrated with bovine serum albumin, which is available at high purity.

Dye-binding methods have been widely applied in the dairy industry (Udy, 1971); dye-binding can be made more sensitive by extracting the dye (McKnight, 1977), and the methods have been included in the AOAC Official Methods. Most of these methods depend on calibration against the Kjeldahl method. Pomeranz, Moore and Lai (1977) have published a comparison of biuret, NIR, dye-binding and alkaline distillation in the measurement of protein in barley and malt. Ribadeau-Dumas and Grappin (1989) have published a review of protein measurements in milk. In general, dye-binding methods have their widest application

Table 7.4 Methods of analysis for amino acids

Procedure	*Applicability*	*Limitations*	*Capital costs*	*Selected references*
Ion-exchange chromatography after acid hydrolysis	All foods	Hydrolytic losses of more labile amino acids and slow release of branched chain amino acids	High	AOAC International, 2002; De Geeter and Huyghebaert,1992.
High-performance liquid chromatography after acid hydrolysis	All foods	As above	High	As above
Gas chromatography after acid hydrolysis and derivatization	Most foods	Choice of derivatives is critical	Medium to high	As above
(Sulphur amino acids) Acid hydrolysis after oxidation of sulphur amino acids.	Most foods	Hydrolytic losses	High	As above
(Tryptophan) Alkaline hydrolysis and ion-exchange chromatography	Most foods	Hydrolytic losses of other amino acids	High	Moore and Stein, 1948; Landry and Delhave, 1993
(Tryptophan, S amino acids) Colorimetry	Most foods		Low	Blackburn, 1968; Christie & Wiggins, 1978
(Available lysine) Colorimetry	Most foods		Low	Carpenter, 1960; Booth,1971

Notes: References selected provide detailed procedures, evaluations or reviews.

in the routine quality control of analysis of large numbers of similar types of sample (Van Camp and Huyghebaert, 1996).

Amino acids

Before the development of ion-exchange chromatography (IEC) individual amino acids were measured by colorimetric methods or by microbiological assay. Although these methods yielded acceptable results they have been almost completely superseded by chromatography procedures (Moore and Stein, 1948). These use automated systems that give complete analyses rapidly and with reasonable levels of precision.

The amino acids in the protein must first be released by hydrolysis and this constitutes the most critical stage of the analysis. Acid hydrolysis, usually with 6M HCl in an oxygen-free solution, gives complete release of most amino acids. Tryptophan is completely degraded in acid conditions and threonine, serine and the sulphur amino acids are partially degraded. Alternative hydrolysis conditions must therefore be used to measure tryptophan. Cystine and methionine are usually protected by specific oxidation before hydrolysis. Losses of threonine and serine are time-dependent and it is necessary to carry out serial hydrolyses to estimate the rate of degradation and correct the values accordingly. Conversely, the branched-chain amino acids are slowly released on hydrolysis, and serial hydrolyses are necessary to estimate complete release (Neitz, A., personal communication). Williams (1982) reviewed the development of IEC techniques and discusses the use of high-performance liquid chromatography (HPLC) as an alternative.

The conditions for acid hydrolysis require pure acid and a high ratio of acid to analytical portions of the food. Even so, high-carbohydrate foods often react with the amino acids during hydrolysis, leading to losses that are difficult to quantify (Silvestre, 1997). Vapour phase hydrolysis has been suggested as an approach that minimizes the degradative losses. In this method the dried food (or protein) sample is hydrolysed by condensing acid. 6M HCl corresponds to the constant boiling mixture for the acid (De Geeter and Huyghebaert, 1992).

Sulphur amino acids are usually oxidized with performic acid before hydrolysis. Some chlorination of tyrosine can occur and the addition of phenol to the acid is often used to reduce this. The hydrolysis should be carried out under nitrogen or, preferably, in sealed tubes.

Hydrolysis must be carried out for three different time periods – 24, 38 and 48 hours – to allow correction for slow release and degradation losses. If pure bovine serum albumin is hydrolysed as a standard, this should also be hydrolysed for the same time periods.

Tryptophan is measured after alkaline hydrolysis (KOH, $Ba(OH)_2$ or LiOH) (Landry and Delhave, 1993). It is usual to measure the leucine in the hydrolysate to adjust the values to be consistent with the acid hydrolysis. A number of alternative reagents and pre- and post-column derivatives have been used, but ninhydrin, despite its instability, is probably the most widely employed. Most other reagents vary in their sensitivity. Capillary gas chromatography has also been used, but most of the reagents vary in their rates of reaction with different amino acids.

In calculating the results of amino acid analyses it is important to express the amino acid values as mg amino acid per g nitrogen applied to the column. As a check on the analyses it is also important to calculate the recovery of nitrogen as amino acids and ammonia from the measured amino acids. There will usually be some losses during hydrolysis and the chromatography. If the losses are found to exceed 10 percent, repeating the hydrolysis should be considered.

Since 1990, HPLC methods of derivatized amino acids have replaced IEC for the analysis of protein hydrolysates in most laboratories as they offer reduced analysis time and improved limits of detection of about 1 picomole (pmol) (Cohen and Strydom 1988; Davey and Ersser 1990; Sarwar and Botting, 1993).

HPLC may be used to separate amino acids on ion-exchange columns with postcolumn derivatization with ninhydrin or OPA (o-phthaldialdehyde) (Ashworth, 1987) or by precolumn derivatization followed by separation on reversed-phase octyl- or octadecyl silica (Cohen and Strydom, 1988). For the analysis of amino acids in protein hydrolysates, reversed-phase HPLC with precolumn derivatization with PITC (phenylisothiocyanate) is becoming established as a cheaper alternative to commercial amino acid analyses using IEC. The PITC derivatization method enables the accurate determination of all nutritionally important amino acids except tryptophan in 12 minutes, while a liquid chromatographic method requiring no derivatization enables the determination of tryptophan in about eight minutes (Sarwar and Botting, 1993).

The range of methods is summarized in Table 7.4.

Available lysine

Lysine can become nutritionally unavailable under certain conditions that lead to the ε-amino group reacting with carbohydrate. This reaction reduces the biological value of the protein. Using the Carpenter method (1960) available lysine can be measured by its reaction with 2,4-fluorodinitrobenzene. This method has been the subject of many modifications (Williams, 1982). HPLC separation of ε-DNP lysine is described by Peterson and Warthesen (1979).

Other nitrogenous substances

Several groups of foods, fish and other marine foods, meats, fungi and vegetables contain a range of nitrogenous materials, amines (Steadman, 1999) and nucleic acids. Many of these react with ninhydrin and can be separated by IEC. Methods for nucleic acids were reviewed by Munro and Fleck (1966). They may also be separated by HPLC and detected by their strong ultraviolet (UV) absorption.

Lipid constituents

FAO/WHO (1994) recommended that adequate food composition data on fats should be widely accessible and that standard methods and reference materials should be used for the analysis of

fatty acids and preparation of nutrient databases. The report provides good coverage of the compounds and nutritional issues of interest. Christie (2003) is a key reference for lipid analysis.

In the proximate system of analysis, 'fat' is measured as the fraction of the food that is soluble in lipid solvents. The extracted material contains a range of different classes of substances. For nutritional purposes the measurement of 'total fat' has limited value; nevertheless, it still is widely reported and is retained in many requirements for food labelling and the regulation of food composition.

The range of methods is summarized in Table 7.5.

Total fat

The values obtained for total fat or total material soluble in lipid solvents are very method-dependent. Carpenter, Ngeh-Ngwainbi and Lee (1993), in their review for the AOAC of methods for nutritional labelling, set out the nature of the problems encountered. Gurr (1992) and Gurr, Harwood and Frayn (2002) discuss in detail the methods available for separating the different classes of lipids.

The classical method is based on continuous extraction performed on dried samples of food in a Soxhlet extractor, sometimes preceded by acid hydrolysis. This technique is time-consuming and subjects the extracted lipids to long periods at high temperatures. Its main drawback, however, is that it yields incomplete lipid extractions from many foods, especially baked products or those containing a considerable amount of structural fat. The extractant used is often petroleum spirit (which is less flammable that diethyl ether and less likely to form peroxides), which requires completely dry analytical portions and the removal of mono- and disaccharides. Values obtained using this method require close scrutiny before their inclusion in a database and their continued use is not recommended.

Other solvents, for example, trichloroethylene, are used in a number of automated systems of the 'Foss-Let' type; these appear to give more complete extractions (Pettinati and Swift,1977).

The use of mixed polar and non-polar solvents has been shown to extract virtually all the lipids from most foods. In the case of baked (cereal) products, however, incomplete extraction of fat may occur. Chloroform–methanol extraction is well known (Folch, Lees and Stanley, 1957; Bligh and Dyer, 1959); this combines the tissue-penetrating capacity of alcohol with the fat-dissolving power of chloroform. The resultant extracts are complete but may also contain non-lipid materials and require re-extraction to eliminate these. This extraction method is preferred when the extract is to be subsequently measured for fatty acids and sterols (Shepherd, Hubbard and Prosser, 1974). The method is effective for composite foods and is included in the AOAC Official Methods. It has been shown to be useful for foods such as brain and egg that are rich in phospholipids (Hubbard *et al.*, 1977). The measurement of lipids after acid (Weibull and Schmid methods) or alkaline (Röse-Gottlieb method) treatment also provides good extraction from many foods. These techniques are recognized as regulated methods by the AOAC and the European Union. Alkaline methods are almost exclusively used for dairy foods and are the approved method for such foods. The extracts from acid and alkaline treatments are not suitable for fatty acid analysis because some oxidation and losses

Table 7.5 Methods of analysis for lipids

Procedure	*Application*	*Limitations*	*Capital costs*	*Selected references*
Total fat				
Continuous extraction (single solvent)	Low moisture foods (dry analytical samples)	Incomplete extraction from many foods. Time consuming. Extracts cannot be used for fatty acid studies	Low	Sullivan and Carpenter, 1993
Acid hydrolysis	All foods except dairy and high sugar products	Some hydrolysis of lipids. Extracts cannot be used for fatty acid studies	Low	AOAC International, 2002; Sullivan and Carpenter, 1993
Hydrolysis and capillary GLC	Most foods (NLEA compliance)		High	Ngeh-Ngwainbi, Lin and Chandler, 1997; House, 1997
Mixed solvent extraction	Rapid, efficient for many foods. Extract can be used for fatty acid measurements	Complete extraction from most foods. Extracts often need clean-up	Low	Bligh and Dyer, 1959; Hubbard *et al*., 1977
Alkaline hydrolysis	Dairy foods	Validated for dairy foods only	Low	AOAC International, 2002
NIR	Established for cereals	Requires extensive calibration against other methods	High	Hunt *et al*., 1977a
Triacylglycerols				
Range of chromatographic methods	All foods	Free fatty acids can interfere. TLC checks useful	Medium	Gurr, Harwood and Frayn, 2002
Fatty acids				
GLC	All foods after transmethylation	Validated for most foods	High	AOCS, 1998
HPLC	Under development	Not found to have advantages over GLC at present	High	Gurr, Harwood and Frayn, 2002
***Trans* fatty acids**				
GLC with infrared analyses	All foods	Availability of authentic standards for some isomers	Medium to High	As above
Infrared absorption	All food	Some interference	High	As above
GLC	All food	Capillary techniques are required	High/medium	As above

Notes: References selected provide detailed procedures, evaluations or reviews. GLC = gas-liquid chromatography; NLEA = United States Nutrition Labeling and Education Act; NIR = near infrared reflectance; TLC = thin-layer chromatography; HPLC = high-performance liquid chromatography.

due to (acid) hydrolysis of fats may occur. The AOAC has adopted methods for determining total fat (also saturated, unsaturated and monounsaturated fats) in foods using acid hydrolysis and capillary gas chromatography (Ngeh-Ngwainbi, Lin and Chandler, 1997; House, 1997) to comply with the Nutrition Labeling and Education Act (NLEA) definition of fat as the sum of fatty acids expressed as triacylglycerols.

Lipid classes show strong carbonyl absorption bands in the infrared region. NIR has been used for legumes (Hunt *et al.*, 1977a) and various other foodstuffs (Cronin and McKenzie, 1990). The effective use of the method depends on extensive calibration against comparable matrices using another approved method; for this reason the technique is most commonly applied in routine analyses of large numbers of very similar samples, for foods such as cereals and dairy products.

Triacylglycerols

Although it is probable that the composition of triacylglycerols (triglycerides) has nutritional significance, few databases contain compositional information. Methods for separating the individual components have not been extensively developed (Gurr, Harwood and Frayn, 2002). Thin-layer chromatography in combination with chromatography has been used. Total values can be found by separating the free fatty acids from the total lipid and can be used to give a "by difference" value. HPLC techniques have been proposed for the complete fraction of triacylglycerols (Patton, Fasulo and Robbins, 1990a,b; Gonzalez *et al.*, 2001).

Fatty acids

Separation by GLC of the methyl esters of the fatty acids prepared by transmethylation of the lipid extracts from foods is the method of choice. The development of column packing materials, capillary techniques and detector amplification systems has extended to application of the method for the separation of isotopic forms and longer-chain fatty acids. The technique published by the International Union of Pure and Applied Chemistry (IUPAC) (Paquot and Hautfenne, 1987) forms the basic procedure.

The exact method chosen will depend on the food to be analysed and the fatty acids of particular interest. Many users will be particularly interested in n-3 and n-6 fatty acids, *trans* acids and levels of long-chain fatty acids such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). Automation of sample injection and the computerization of the chromatographs have added to the costs of the analytical apparatus but greatly improve accuracy, precision and analytical throughput. The American Oil Chemists' Society (1998) methods are: Method No. Ce 1-62 (packed column method for methyl esters of C9–C24 acids, and animal fats), Method No. Ce 1b-89 (capillary method for marine oils and for ethyl or methyl esters of C14–C24 acids (percentage relative values and mg/g levels of EPA and DHA), Method No. Ce 1c-89 (capillary method for fatty acids, *trans* isomers and *cis*, *cis* methylene-interrupted isomers in vegetable oils), Method No. Ce 1e-91 (capillary method for C4–C24 fatty acids), and Method No. Ce 1f-96 (capillary method for *cis*- and *trans* fatty acids in hydrogenated and refined oils and fats).

Infrared detectors are useful in the measurement of *trans* fatty acids (AOAC International, 2002). The major difficulty is the assignment of unequivocal identity to isomers. This requires good standards or combining the GLC separation with mass spectrometry (Beare-Rogers and Dieffenbacher, 1990), which may make it impractical for some developing countries.

Infrared absorption is currently the preferred method for the measurement of *trans* fatty acids in hydrogenated fish oils. GLC measurement of *trans* fatty acids in partially hydrogenated vegetable oils using a flame ionization detector (FID) often underestimates the *trans* fatty acid content, even on very long, highly polar, capillary columns (Aro *et al.*, 1998).

Food composition laboratories lacking GLC instrumentation do not usually undertake fatty acid measurements but may seek cooperation with a laboratory with the necessary capital resources. The samples may be transferred to the laboratory as fat (which requires cold storage during transit and the addition of an antioxidant) or methyl esters (which also need to be protected against oxidation). It is important to verify these arrangements with the analysing laboratory to avoid interference by the antioxidants during chromatography.

The unsaturation of a fat can be estimated by iodine value determination (IUPAC, 1979; AOAC International, 2002); this remains a useful technique when full fatty acid analyses are not undertaken.

Sterols

In the past, nutritional analyses emphasized the measurement of cholesterol, but there is an increasing focus on the measurement of other sterols, especially phytosterols.

Cholesterol. The older techniques, using gravimetric and colorimetric methods, are now regarded as obsolete and are no longer used. The preferred methods are chromatographic, with widespread use of GLC of a range of derivatives separated on low-polarity columns (Punwar, 1975; Hubbard *et al.*, 1977). One problem with sterol analysis in general is that the greater proportion of other lipids in most foods limits the application of the methods to the lipid extract directly.

Saponification is required before the preparation of derivatives. The use of trimethysilyl (TMS) derivatives met the standards required by the AOAC (Carpenter, Ngeh-Ngwainbi and Lee, 1993) for use with mixed foods. The procedures are somewhat complex and simplified methods have been proposed that require shorter sample preparation times (Thompson and Merola, 1993).

Improvements in the development of capillary GLC have provided the basis for developing procedures that do not require derivatization and that meet the appropriate standards (Jekel, Vaessen and Schothorst, 1998).

Other sterols. The method described above can also be used for the separation and measurement of the range of phytosterols found in the diet (Jonker *et al.*, 1985), as can derivatization with TMS (Phillips, Tarrogo-Trani and Stewart, 1999).

Phospholipids

A comprehensive review of phospholipids published in 1973 (Ansell, Hawthorne and Dawkins.) summarized the analytical procedures available. Subsequently, HPLC techniques were developed (Hammond, 1982; Patton, Fasulo and Robbins, 1990a,b) and are now the methods of choice. Gunstone, Harwood and Padley (1994) provide an overview of methods for measuring the range of phospholipids.

Carbohydrates

The range of carbohydrates found in the human diet (see Table 4.3) illustrates the nature of the task facing the analyst who wishes to follow the recommendations published by FAO/WHO (1998) for measuring the carbohydrates in foods separately. Not all types of carbohydrates are, of course, present in all types of foods.

The distinctive metabolic and physiological properties of the different carbohydrates emphasize the fact that for nutritional purposes it is inadequate to consider the carbohydrates as a single component of foods.

The calculation of "carbohydrate by difference" using the Weende proximate system of analysis described at the beginning of the chapter was a reflection of the state of knowledge of carbohydrate chemistry at the time. Moreover, the system was designed for animal feedstuffs, especially for ruminants, and most of the carbohydrates (except lignin-cellulose of which crude fibre was an approximate measure) would therefore be digested in the rumen.

For nutritional purposes carbohydrates can be considered as falling into three groups based on the degree of polymerization:

- sugars (mono- and disaccharides);
- oligosaccharides (polymers of three to nine monosaccharide or uronic acid units);
- polysaccharides (polymers containing more that nine units), which fall into two broad categories: α-glucans (starches, starch hydrolysis products and glycogen) and a much more diverse group of non-α-glucans (non-starch polysaccharides [NSPs], which are the major constituents of dietary fibre).

These broad chemical groupings do not correspond precisely with physiological properties or with analytical fractions. The analyst faced with the analysis of carbohydrates, particularly NSPs, is "bound to make a compromise between the ideal of separating the many components and measuring them or a scheme that is entirely empirically based" (Southgate, 1969). In many cases, a food contains a limited range of carbohydrates and simpler procedures can be used for its analysis (Southgate, 1991).

The range of methods is summarized in Tables 7.6 to 7.8.

Sugars

A range of methods can be used for the analysis of the free sugars in foods; the choice depends primarily on the qualitative composition of the free sugars present in the food. Where a single

Table 7.6 Methods for the analysis of sugars

Procedure	*Application*	*Limitations*	*Capital costs*	*Selected references*
Specific gravity	Sugar solutions	Accurate for sucrose	Low	AOAC International, 2002; Southgate,1991
Refractive index	Sugar solutions	Empirical calibration required	Low	As above
Polarimetry	Single sugars, simple mixtures	Close attention to standardized methods is essential	Low	As above
Reductiometric	Reducing sugars, invert sugar mixtures	Non-reducing sugars	Low	AOAC International, 2002
Colorimetric	Single sugars, simple mixtures	Specificity	Low	Southgate, 1991; Hudson *et al*., 1976; Hudson and Bailey, 1980
Specific enzyme methods	Glucose, complex mixtures	Reagents can be expensive	Low	Bergmeyer, 1974
GLC	Complex mixtures	Need for derivatives	Medium	Englyst, Quigley and Hudson, 1994
HPLC	Complex mixtures	Choice of column, detectors	Medium to high	Southgate, 1991; Shaw, 1998; Englyst, Quigley and Hudson,1994

Notes: References selected provide detailed procedures, evaluations or reviews.
GLC = gas–liquid chromatography; HPLC = high-performance liquid chromatography.

carbohydrate species is present virtually any procedure can be used, but most foods contain a mixture of three or more components and separation of the components is required to produce accurate results. Specific enzymatic methods are available for the analysis of certain common mixtures without separation.

The methods for free sugars (and uronic acids) provide the end-analytical methods for most of the higher carbohydrate polymers after hydrolysis and separation of the components.

The evolution of the methods closely parallels the development of analytical techniques coupled with the pressures of the demands for analytical results. Thus the physical techniques were initially developed for the analysis of sucrose solutions in the sugar-refining industry. The reducing sugar methods were also developed for this industry and the methods were refined and their protocols codified under the auspices of the International Commission for Unified Methods of Sugar Analysis (ICUMSA, 1982). These methods still give satisfactory results providing the protocols are followed closely.

Colorimetric techniques were developed later, with the advent of improved methods for assessing optical density (although early measurement involved the visual matching of solutions). The range of chromogenic reagents for different monosaccharide classes and uronic acids mostly involve reactions in concentrated acids although colorimetric methods are based on reductiometric methods and a few on other reactions (Hudson *et al.*, 1976). The methods are not especially robust, but on simple sugar mixtures with proper quality control they give sound values. The methods are not truly specific and this limits their use for the analysis of mixtures (Hudson and Bailey, 1980).

Specific enzyme methods have been developed, the most notable being the glucose-oxidase method, which has a colorimetric end-point. A series of coupled reactions with NADPH–NADP using specific enzymes permits the analysis of mixtures of glucose/fructose and glucose/fructose/sucrose and maltose/galactose (Southgate, 1991).

Chromatography, initially on paper or silica plates, provided good separations and semi-quantitative methods, but ion-exchange techniques were difficult to develop.

Gas chromatographic analysis depended on the preparation of suitable volatile derivatives. Initially trimethyl-silylation provided suitable derivatives for the analysis of sugar mixtures, although the chromatograms were very complex. The most widely used and powerful method for the analysis of mixtures involves the reduction of the monosaccharides to the alditols and acetylation.

HPLC columns are now available that give good separation of sugar mixtures without the need for the preparation of derivatives. The first detectors used refractive indices to measure the eluted peaks, but these are relatively insensitive and have been superseded by the pulsed amperometric detector, which has improved sensitivity.

Polyols (sugar alcohols)

Polyols are not widely found in foods. Some can be measured by specific enzyme methods although HPLC methods are more commonly used.

Table 7.7 Methods for the analysis of polyols and oligosaccharides

Procedure	*Application*	*Limitations*	*Capital costs*	*Selected references*
Polyols				
Specific enzymatic methods	Limited to a few alcohols	Specificity of enzymes	Medium	
HPLC	Complex mixtures	Lack of standardized procedures; choice of column	Medium to high	Southgate, 1991
Oligosaccharides				
Specific enzymatic procedures	Selective hydrolysis and separation	Specificity of enzymes	Medium to high	Bergmeyer, 1974
GLC	Complex mixtures	Choice of column	Medium to high	Quigley, Hudson and Englyst, 1997

Notes: References selected provide detailed procedures, evaluations or reviews.
HPLC = high-performance liquid chromatography; GLC = gas–liquid chromatography.

Oligosaccharides

These are widely distributed, especially in vegetables, and the malto- series is found particularly in foods that have partial starch hydrolysates and glucose-syrup preparations as ingredients. The malto-oligosaccharides are hydrolysed by brush-border enzymes and are "glycemic carbohydrates" that need to be measured separately.

Fructo-oligosaccharides are increasingly used as ingredients and should be measured after hydrolysis with specific fructan hydrolases. The remaining galacto-oligosaccharides should also be measured after specific enzymatic hydrolysis. GLC and, particularly HPLC separation techniques also offer powerful methods for the analysis of these oligosaccharides (Quigley, Hudson an Englyst, 1997).

Polysaccharides

These are best considered, for nutritional purposes, under two headings – starch and non-starch polysaccharides (NSPs).

Starch. This category includes all the α-glucans, starches, partially hydrolysed starches and glycogen. The latter is a minor component of most animal products; it is found in significant concentrations in fresh liver and horse flesh and as traces in lean muscle.

Polarimetric methods are limited to some cereals, but with proper calibration and standardization can give satisfactory results (Fraser, Brendon-Bravo and Holmes, 1956; Southgate, 1991).

Dilute acid hydrolysis can be used for highly refined foods with low concentrations of NSPs, and virtually any monosaccharide method can be used to measure the glucose produced.

The use of a glucose-specific method such as glucose-oxidase extends the range of foods for which this method is useful (Dean, 1978; Southgate, 1991).

Enzymatic hydrolysis with specific amylolytic enzymes, followed by precipitation of the residual NSPs with ethanol, and measurement of the glucose produced, is the most satisfactory and widely applicable method. The choice of enzymes and the conditions for hydrolysis are critically important. If values for total starch are required, any enzymatically resistant starch must be treated with alkali or dimethyl sulphoxide (DMSO) before hydrolysis (Southgate, 1991).

Resistant starch. Although enzymatically resistant starch was first observed analytically, the current view is that it should be defined as resistant physiologically, that is, resistant to hydrolysis in the human gastrointestinal tract (Gudmand-Hoyer, 1991). Englyst, Kingman and Cummings (1992) have distinguished three types of resistance, due to physical enclosure of starch, starch granule structure, and retrogradation. The latter type is more common in processed foods. The most common approach is to measure starch before and after treatment with DMSO.

Rate of digestion. Englyst and his coworkers (1999) have proposed that the rate of digestion of starch is the major determinant of variations in the glycemic responses to food and proposed

Table 7.8 Methods for the analysis of polysaccharides

Procedure	*Application*	*Limitations*	*Capital costs*	*Selected references*
Starch				
Polarimetry	Some cereal foods	Needs very careful calibration	Low	Fraser, Brendon-Bravo and Holmes, 1956
Dilute acid hydrolysis using a general sugar method	Highly refined foods, low in NSP	Interference from any NSP present	Low	Southgate, 1991; Dean, 1978
Dilute acid hydrolysis and glucose-specific method	Foods low in ß-glucans	Presence of ß-glucans	Low	As above
Enzymatic hydrolysis and glucose-specific methods	All foods	Choice of enzymes and conditions	Medium	Wills, Balmer and Greenfield, 1980
Resistant starch				
Enzymatic hydrolysis of starch before and after treatment with alkali or DMSO		Choice of enzymes and conditions	Medium	Champ, 1992; Englyst, Kingman and Cummings, 1992
Rapidly digestible starch		Choice of conditions	Medium	Englyst, Kingman and Cummings, 1992
Slowly digestible starch		Choice of conditions	Medium	As above
Non-starch polysaccharides				
Enzymatic hydrolysis and removal of starch. Acid hydrolysis of NSP. GLC, HPLC separation of component monosaccharides. Colorimetric analysis of monosaccharides	Virtually all foods	Resistant starch must be treated before hydrolysis. GLC requires preparation of derivatives. Gives only total values	Medium to high	Englyst, Quigley and Hudson,1994; Southgate, 1995

Notes: References selected provide detailed procedures, evaluations or reviews.
DMSO = dimethyl sulphoxide; NSP = non-starch polysaccharides; HPLC = high-performance liquid chromatography; GLC = gas–liquid chromatography.

that the starch can be considered to fall within three classes: rapidly digestible starch, slowly digestible starch and resistant starch. While the rate can be distinguished *in vivo*, simulation analytically is quite difficult. Collaborative studies have shown that reasonable precision can be obtained (Champ, 1992).

Glycemic index. There has been great interest in including glycemic index (GI) values in food composition databases and a set of tables of GI values has been published (Foster-Powell and Miller, 1995). The GI values (strictly speaking, a ranking of the carbohydrates in foods) are based on their glycemic effect compared with that of a standard food. The GI is defined as "the incremental area under the blood glucose response curve expressed as a percentage of the response to the same amount of carbohydrates from a standard food taken by the same subject" (FAO/WHO, 1998). The standard food is usually white bread or glucose. FAO/WHO (1998) have published a protocol using six or more subjects and define carbohydrate as "glycemic (available) carbohydrate". A working definition used by the main Australian laboratory measuring GI defines carbohydrate as "total carbohydrate by difference minus the sum of dietary fibre plus resistant starch (if known) or the sum of starch plus sugars, including polyols and other slowly absorbable sugar derivatives" (Brand-Miller and Holt, personal communication).

In Australia, the use of a GI symbol on food labels is permitted and a Web site is available for consultation (http://www.glycemicindex.com). The GI of meals can be calculated but not of cooked recipe foods because the GI of a food is affected by cooking and processing.

Estimates of the different rates of digestion of starch in foods show some correlation with glycemic indices measured *in vivo*. These require a number of human subjects to have blood glucose levels measured at intervals for three hours after consumption of a fixed amount (50 g) of glycemic carbohydrates. The area under the curve is compared with the area under the curve for a 50 g glucose load or, better, 50 g of glycemic carbohydrates from white bread. White bread is preferred because glucose loads can be emptied slowly from the stomach because of osmotic effects. An interlaboratory study (Wolever *et al.*, 2003) showed that within-subject variation in glycemic response needs to be reduced to improve precision of the method.

An *in vitro* method for rapidly available glucose published by Englyst *et al.* (1999) showed a high correlation with glycemic response.

Non-starch polysaccharides. Methods for NSP analysis involve treatment of the sample to remove free sugars and starch by enzymatic hydrolysis. The unchanged NSPs are recovered by precipitation with ethanol (80 percent v/v), then washed and dried. The NSPs are hydrolysed using one of two methods: sequentially with dilute acid, which hydrolyses most of the non-cellulosic polysaccharides (NCPs), and with 12M H_2SO_4, which hydrolyses the cellulose; or the NSPs are hydrolysed completely using 12M acid (see "Measurement of NSPs" below for further details).

The monosaccharides are analysed by GLC after derivatization (as the alditol acetates [Englyst, Wiggins and Cummings, 1982]) or by HPLC, or as a total colorimetrically (Englyst, Quigley and Hudson, 1994). The methods are not very robust (Southgate, 1995), although

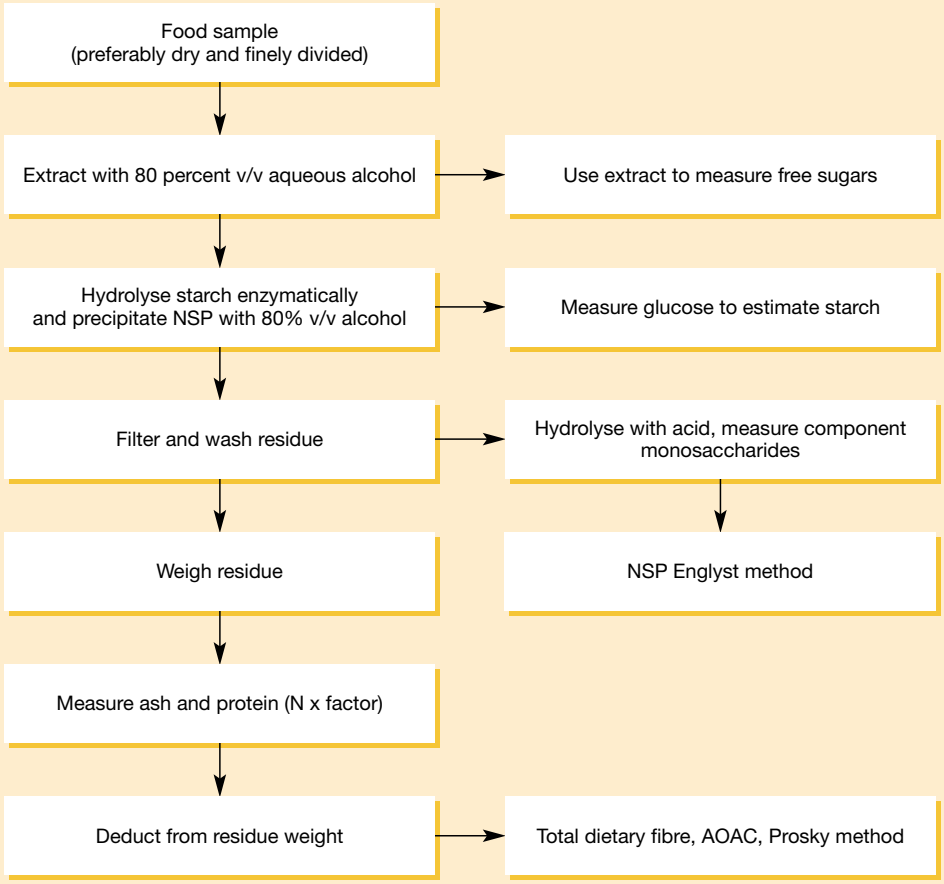
collaborative trials have shown that when careful attention is given to the protocol the methods give reasonable precision.

Choice of method for carbohydrates

There is no single method that meets the recommendations of the FAO/WHO (1998) review. Ideally, when planning to measure the carbohydrates in foods one should aim to measure the different carbohydrate species in a food sequentially using one analytical portion; this approach avoids the possibility of double measurement of an overlapping fraction.

The basic principles of such an approach are set out schematically in Figure 7.1.

Figure 7.1 Principles of measuring carbohydrates and dietary fibre



Notes: v/v = by volume; NSP = non-starch polysaccharides.

Extraction of free sugars, polyols and oligosaccharides. This could be done with an aqueous extraction, but this procedure will extract proteins with the result that subsequent analysis is more complex. The removal of fat is desirable for technical reasons, as this facilitates a more complete extraction of sugars. Extraction with aqueous alcohol is the most common approach: 80 percent v/v aqueous ethanol is most commonly used, but 85 percent v/v methanol is also useful, as is isopropanol. The extractions are usually made with a boiling solvent; care should therefore be taken to protect the analysts from solvent fumes. If the extract is likely to be acid it is important to neutralize the acid to avoid hydrolysis of di- and higher saccharides.

The aqueous alcohols will also extract some lower polysaccharides – short-chain polysaccharides as defined by Englyst and Hudson (1996). These should preferably be measured after selective enzymatic hydrolysis. Modern enzyme technologies have produced a wide range of very specific enzymes with high activity; many companies specialize in this area, for example, Boehringer Mannheim, Germany; Megazyme, Ireland; Nova, Denmark; and Sigma, the United States. Several of these companies prepare enzyme method "kits". The rate of development of enzyme technology is such that it is expected that selective enzymatic hydrolysis will become increasingly important analytically because of the specificity offered (McCleary and Prosky, 2001).

Starch hydrolysis. The next stage is to remove starch using selective enzymatic hydrolysis. A number of enzymes can be used for this purpose. A mixture of amylase and pullulanase has been used to give complete hydrolysis to glucose but many glucoamylases give virtually complete hydrolysis to glucose. The conditions for enzymatic hydrolysis are critical, both to ensure complete and rapid starch hydrolysis and to minimize hydrolysis of NSPs, especially β-glucans. Unhydrolysed NSPs are recovered by precipitation with ethanol to 80 percent v/v.

Measurement of NSPs. The precipitated NSPs are washed and dried gently and then hydrolysed. This may be done in boiling 1M H_2SO_4 followed by hydrolysis in 12M acid at ambient temperature. This produces, first, a hydrolysate containing the monosaccharides from the NCPs and, second, the monosaccharides from a cellulosic fraction. Alternatively, the NSPs may be hydrolysed in 12M acid followed by the dilute acid, which produces a hydrolysate containing the monosaccharides from the NSPs as a whole. Uronic acids are not hydrolysed completely by these methods, and colorimetric analysis is widely used (Englyst, Quigley and Hudson, 1994). Specific enzymatic hydrolysis of the uronic acid containing polymers is now possible (Quigley and Englyst, 1994).

Dietary fibre

Dietary fibre should be considered as part of the carbohydrates in foods. The major problem in the choice of method lies in the definition of dietary fibre and its interpretation in an analytical context. The term was first used in 1953, by Hipsley, to describe the sum of the

hemicelluloses, cellulose and lignin in food, in other words the components of plant cell walls in foods. Trowell, in 1972, took up the term for "the indigestible components of the plant cell wall in foods". Both these terms were too vague to use as a basis for an analytical method and in 1976 Trowell *et al.* (1976) proposed that it be defined as "the sum of the plant polysaccharides and lignin that are not digested by the enzymes of the gastrointestinal tract". This was closely analogous to the "unavailable carbohydrates" as defined by McCance and Lawrence (1929) and measurable by the procedures proposed by Southgate (1969).

In this method the aim was to measure the carbohydrates specifically using colorimetric techniques. Englyst developed this approach using the more specific GLC methods, which gave values for the non-starch polysaccharides and incorporated a stage to convert resistant starch to non-enzymatically resistant starch. The procedure was developed in a series of collaborative studies and the most recent protocols are described by Englyst, Quigley and Hudson (1994) and Southgate (1995). This method measures only the NSPs and does not include lignin.

In other parts of Europe, especially Sweden and Switzerland, and in the United States, the focus was directed at the "indigestibility of the polysaccharides and lignin". A gravimetric method was developed where the residue after starch removal is weighed to give a measure of total dietary fibre (TDF); this has evolved into the Official AOAC Method No. 982.29 (Prosky *et al.*, 1992). The method requires correction of the residue for undigested protein and for mineral contamination; total nitrogen and ash in the residue are measured and deducted to give the TDF values. These include lignin, resistant starch and all other indigestible carbohydrates (Guillon *et al.*, 1998). A modification has been introduced to include the measurement of indigestible oligosaccharides.

The Englyst NSP and the AOAC TDF procedures are not very robust, especially where low levels are present (Southgate, 1995). The NSP method uses analytical portions of 100–200 mg and the preparation and homogeneity of these portions is absolutely critical. The mixing procedures also require close attention during the execution of the method.

The AOAC gravimetric procedure requires great skill when measuring low levels but gives good precision with high-fibre foods such as bran and wholemeal products. The residue also includes heat-induced artefacts.

In many countries, the choice of method for nutrition labelling will be defined by legislation. Nutritionally specific measurement of the different carbohydrate fractions is the preferred approach. The measurement of soluble and insoluble fractions is highly method-dependent and the FAO/WHO (1998) review concluded that there was no physiological justification for recording separate values based on solubility.

It is important to recognize that the hypothesis concerning the protective effects of dietary fibre was based on differences between diets (Burkitt and Trowell, 1975), i.e. it was a statement about the protective effects of diets that were rich in foods containing plant cell walls in a relatively unprocessed state. These diets are rich in many other components in addition to dietary fibre.

Alcohol

The classical method for measuring the alcoholic content of beverages is distillation of the de-gassed beverage and measurement of the specific gravity of the distillate. While this is still a valid and precise approach, measurement by GLC (which is simpler and quicker) or, alternatively, a specific enzyme procedure using alcohol dehydrogenase (Bergmeyer, 1974) are preferred methods as the distillation methods can be interfered with by other volatile constituents.

Organic acids

A variety of specific enzyme methods for different organic acids (Bergmeyer, 1974) remain valid, but these approaches have been superseded by HPLC methods (Wills *et al.*, 1983). In a food product that contains acetic acid, simple acid-base titration can be used (Sadler and Murphy, 1998).

Inorganic constituents

The majority of methods for inorganic constituents require the organic matrix of the foods to be removed, or extraction and concentration, before they can be applied. Destruction of the food matrix removes a large number of potential sources of interference and provides the inorganic material in a concentrated form. In classical food analysis the organic matrix was incinerated (usually in a muffle furnace at a controlled temperature) and the resultant inorganic residue was weighed to give a value for ash in the proximate system of analysis. The organic matrix can also be destroyed by being heated in concentrated acids. This procedure minimizes losses during the oxidation and avoids any reaction between the inorganic constituents and the vessel used for dry incinerations.

Once the organic matrix has been removed the inorganic constituents can be measured using a variety of techniques. These include classical gravimetric or volumetric methods, polarimetry, ion-selective electrodes, colorimetric procedures (which may or may not be highly specific) and instrumental methods (which offer an increase in speed of analysis, automation and good precision). Many of the instrumental methods can be used for analysis of a number of constituents. In using these methods it is important to ensure that interference from other constituents is eliminated and it is essential to use standard reference (or in-house reference) materials with a similar matrix and apply other quality control measures. This approach is of fundamental importance in the measurement of trace inorganic constituents.

Total ash

Nutritionally, there is little value in recording ash values other than to provide an approximate

Table 7.9 Methods of analysis for cations

Method	*Application*	*Limitations*	*Capital costs*	*Selected references*
Flame photometry	Na[a], K[a], Ca, Mg	Interferences	Medium	Dvorak, Rubeska and Rezac, 1971
AAS with electrothermal furnace	Na, K, Ca[a], Mg[a], Fe[a], Cu[a], Zn[a], Mn[a], Co[a], Cr[a]	Interferences from anions; special suppression techniques	Medium to high	Osborne and Voogt, 1978; AOAC, 1984
Hydride generation AAS	Se[a]		Medium to high	Foster and Sumar, 1996; Murphy and Cashman, 2001
Plasma emission spectrometry	Virtually all cations	Matrix effects need to be controlled	Very high	AOAC, 1984; McKinstry, Indry and Kim, 1999; Sullivan, 1993; Coni *et al.*, 1994; Suddendorf and Cook, 1984
Colorimetry	K[b], Mg, Fe, Cu, Zn[b]	Exacting techniques	Low to medium	Sandell, 1959; Paul and Southgate, 1978; Sullivan and Carpenter, 1993
Classical precipitation and titration	Ca, Mg	Size of analytical sample; skilled techniques	Low	Paul and Southgate, 1978

Notes:

References selected provide detailed procedures, evaluations or reviews.

AAS = atomic absorption spectrometry.

[a] Preferred method.

[b] Difficult and non-rugged methods.

estimate of the total inorganic material and to check for replication in the destruction of the matrix. A value for total ash is, of course, essential when it is necessary to calculate carbohydrate "by difference".

In dry ashing, the food is incinerated in a crucible, usually made of silica, although porcelain (can be used but less suitable) or platinum (very expensive but the least reactive) can be used. The food matrix must be destroyed by heating gently at first to char the sample and then at 500 °C in a muffle furnace (Wills, Balmer and Greenfield, 1980) to prevent foaming of lipids (and sugars) until a white (or light grey) residue is produced. Heating above 500 °C can result in the loss of alkali metals. The general procedure is described by Osborne and Voogt (1978) and in the AOAC Official Methods (see Sullivan and Carpenter, 1993).

In the case of "wet ashing" acid digestion, the food sample is heated with acid – usually a mixture of nitric and sulphuric acids. Perchloric acid is often included in the digesting acid mixture although this introduces the risk of explosion and the procedure must be carried out in a fume hood designed for the use of perchloric acid. Wet ashing offers the advantage that no reactions with the crucible can occur that can lead to the formation of insoluble silicates. Digestion can be carried out in a Kjeldahl flask but this requires a larger quantity of acid. Particularly for trace element analysis, digestion is best carried out in a sealed container. Tubes designed for this purpose are available from most laboratory suppliers. They are made from resistant glass and have a cap with a plastic insert to provide an inert gas-tight seal. The analytical portion and the acid are placed in the tube, which is then capped and may be heated in a conventional or microwave oven. The tube is then allowed to cool completely before the gases are released with care.

For trace element analyses, the acids used must be of the highest analytical quality; blanks should be run as a matter of course and digestion of the reference materials should be included.

The most widely used instruments are atomic absorption spectrophotometers, which are suitable for the analysis of most cations of nutritional interest. The more simple flame photometers can be used for the analysis of Na and K.

Plasma emission instruments such as inductively coupled plasma spectrometers are available that permit the analysis of a wide range of elements and have the capacity to handle a large number of samples and analytes (McKinstry, Indyl and Kim, 1999). They do, however, require high initial capital expenditure and routine maintenance. Ihnat (1982;1984) provides a detailed review of the application of these methods to foods. Sullivan (1993) discusses the use of these techniques in the AOAC's *Methods of analysis for nutrition labeling* (Sullivan and Carpenter, 1993).

Preparation of analytical portion. The residues from dry ashing are usually dissolved in dilute acid and made to volume before analysis. The solutions from wet ashing usually need dilution to a suitable volume before analysis.

Tables 7.9 and 7.10 show methods of analysis for cations and anions, respectively, in foods.

Table 7.10 Methods of analysis for anions

Application	*Method*	*Limitations*	*Capital costs*	*Selected references*
Phosphorus	Colorimetry		Low	Fiske and Subbarow, 1925
Chloride	Titrimetric		Medium	Cotlove, Trantham and Bowman, 1958
	Ion-specific electrode	Interferences	Medium	De Clercq, Mertens and Massart,1974
	Automated conductimetry		High	Silva *et al.*, 1999
Iodine	Microdistillation	Laboratory contamination	Medium	AOAC, 1984
	Ion-specific electrode		Medium	Hoover, Melton and Howard, 1971
	Alkaline dry-ashing		Medium	AOAC, 1984
	GLC		High	Mitsuhashi and Kaneda, 1990; Sullivan and Carpenter, 1993
Fluorine	Microdistillation	Time-consuming	Medium	AOAC, 1984
	Ion-specific electrode		Medium	Ferren and Shane, 1969; Kjellevold-Malde, Bjorvatn and Julshamn, 2001
	Polarography		Medium	Guanghan *et al.*, 1999
Sulphur	Gravimetric		Low	Paul and Southgate, 1978
	X-ray fluorescence		High	Isherwood and King, 1976
Nitrite	Colorimetry		Low	AOAC, 1980
	Ion-specific electrode		Medium	Pfeiffer and Smith, 1975; Choi and Fung, 1980
Nitrate	HPLC		High	Wootton, Kok and Buckle, 1985

Notes: References selected provide detailed procedures, evaluations or reviews.
GLC = gas–liquid chromatography; HPLC = high-performance liquid chromatography.

Cations

Sodium and potassium. Flame photometry and atomic absorption spectrophotometry (AAS) are the preferred techniques. Mutual interference can occur and interference from phosphorus has been observed. These can usually be overcome by the application of appropriate standards.

Calcium. Flame photometry and AAS techniques have similar sensitivities. Interference from phosphorus can occur but this can be suppressed by the addition of lanthanum salts or by the use of N_2O flames. Complexiметric titrimetric methods have been used and classical gravimetric methods can be used with foods rich in calcium.

Magnesium. AAS is the method of choice as this offers greater sensitivity than other procedures, with the exception of activation analysis.

Iron. This can be measured by AAS or inductively coupled plasma spectroscopy (ICP) instrumentally. There are, however, sound colorimetric methods available.

Zinc. While colorimetric methods are available, AAS or ICP are the better techniques to use.

Selenium. Hydride generation AAS has been widely used and is probably the method of choice at the present time (Foster and Sumar, 1996; Murphy and Cashman, 2001). Cathodic stripping voltammetry has also been proposed as a method (Inam and Somer, 2000).

Copper and other trace elements. These can be measured satisfactorily by AAS but may require the use of special conditions. ICP, when available, is a satisfactory technique (Coni *et al.*, 1994). Colorimetric methods for copper are quite sound (Sullivan and Carpenter, 1993).

Anions

Phosphorus. This can be measured by ICP but a well-established colorimetric method is the preferred method when applied to wet-digested samples (Fiske and Subbarow, 1925). If dry-ashed samples are used, the pyrophosphates formed during ashing must be hydrolysed.

Chloride. A range of methods can be used. Ion-specific electrode analysis represents the simplest approach, but the classical reaction by titration is also satisfactory (Cotlove, Trantham and Bowman, 1958). Procedures using automated conductimetry also seem to perform well (Silva *et al.*, 1999).

Iodine. This is regarded as one of the most difficult inorganic elements to measure. Dry ashing followed by titration or GLC has been used by the AOAC (Sullivan and Carpenter, 1993). Ion-specific electrodes offer some potential.

Table 7.11 Methods for the analysis of fat-soluble vitamins

Vitamin	*Method*	*Limitations*	*Capital costs*	*Selected references*
Vitamin A and carotenoids	Chromatography	Low recoveries of retinoids; overestimates of carotenoids	Low	AOAC, 1984; Carr and Price, 1926
	HPLC	Identification of carotenoids	Medium to high	Scott, 1992; Scott and Hart, 1993; Scott *et al.*, 1996; Wills and Rangga, 1996; Taungbodhitham *et al.*, 1998
Vitamin D	Bioassay	For low levels only; animal facilities required	Low to medium	Kodicek and Lawson, 1967; AOAC International, 1995
	Colorimetry	Lack of precision and sensitivity	Low	Nield, Russell and Zimmerli, 1940; Eisses and De Vries, 1969
	GC		Medium	Bell and Christie, 1974; Koshy, 1982
	HPLC	Lipid interference; two stages, preparative followed by analytical separation, needed for most foods	High	Mattila *et al.*, 1993, 1994, 1995; MAFF, 1997
	Radio-immunoassay		High	Bates, 2000
Vitamin E	Colorimetry	Interfering compounds	Low	Tsen, 1961; Christie and Wiggins, 1978
	GC		Medium to high	Christie, Dean and Millburn, 1973
	HPLC	Extraction techniques	High	Piironen *et al.*, 1984, 1987
Vitamin K	Colorimetry	Lack of specificity	Low	Irreverre and Sullivan, 1941; Hassan, Abd El Fattah and Zaki, 1975
	Column chromatography	Low		Matschiner and Taggart, 1967
	GC		Medium to high	Dialameh and Olson, 1969; Seifert, 1979
	HPLC	Lipid interference	High	Cook *et al.*, 1999; Indyk and Woollard, 1997; Piironen and Koivu, 2000; Koivu *et al.*, 1999

Notes: References selected provide detailed procedures, evaluations or reviews. GC = gas chromatography; HPLC = high-performance liquid chromatography.

Fluorine. Polarographic methods have been developed that produce a very good sensitivity (Guanghan *et al.*, 1999). Methods using ion-selective electrodes also seem to perform well (Kjellevold-Malde, Bjorvatn and Julshamn, 2001).

Sulphur. Sulphur may be measured via conversion to barium sulphate (Paul and Southgate, 1978) or by x-ray fluorescence (Isherwood and King, 1976).

Nitrate and nitrite. Methods include colorimetry (AOAC, 1980), HPLC (Wooton *et al.*, 1985) and capillary ion electrophoresis. Ion-specific electrodes can also be used (Marshall and Trenerry, 1996).

Vitamins

"Vitamin" is a physiological term rather than a chemical term, expressing a certain physiological activity that is related to the chemical substances responsible for this activity. Vitamin activity may be due to a group of chemical compounds, usually related structurally to one another (vitamers).

The analysis of vitamins presents a number of challenges to the analyst and considerable analytical activity has been, and still is, directed at achieving the ideal analytical method for providing chemical values that predict the physiological vitamin activity for human beings in the current context. The ideal method would measure the different vitamers separately so that a value could be calculated for the total vitamin activity (Brubacher, Müller-Mulot and Southgate, 1985). This ideal is rarely possible, in part because of the presence of interfering substances without vitamin activity.

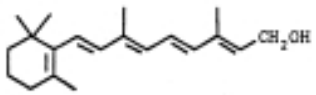
The discussion of methods for individual vitamins will emphasize the handling and preparation of samples for analysis; these are crucial factors because of the lability of some vitamins. Many vitamins are sensitive to light and some can be oxidized very rapidly. Heating can increase the rate of oxidation and may also lead to isomerization to inactive forms; unnecessary heating should therefore be avoided.

A number of detailed reviews on the analysis of vitamins in foods are available (Bates, 2000; Eitenmiller and Landen 1998; Machlin, 1984; Christie and Wiggins, 1978; Van Niekirk, 1982). Brubacher, Müller-Mulot and Southgate (1985) was the result of a collaborative European project which tried to establish a handbook of tested methods. A review of the AOAC Official Methods of vitamins is given by Sullivan and Carpenter (1993). Table 7.11 summarizes the methods for lipid(fat)-soluble vitamins and Table 7.12 summarizes those for the water-soluble vitamins.

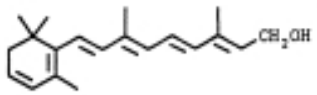
Lipid-soluble vitamins

These are the vitamins A, D, E and K, and the carotenoids with provitamin A activity. As nutrition interest is now also focused on the non-provitamin A carotenoids, it is also desirable to cover more of these carotenoids.

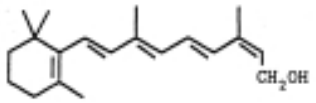
Figure 7.2 Structures of the main vitamin A-active retinoids



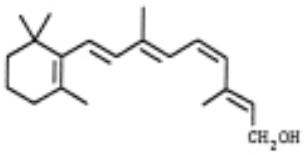
all-*trans*-retinol
(vitamin A_1 alcohol)



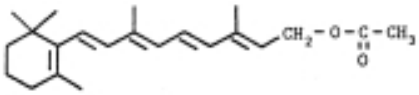
dehydroretinol
(vitamin A_2 alcohol)



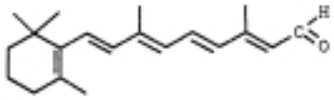
13-*cis*-retinol
(neo vitamin A_1 alcohol)



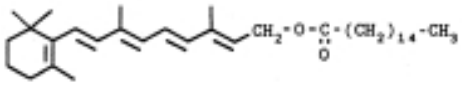
11-*cis*-retinol



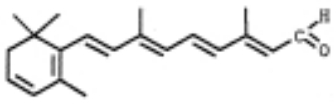
retinyl acetate



retinaldehyde



retinyl palmitate



dehydroretinaldehyde

Vitamin A. Vitamin A is a generic term that includes retinol, its esters and some isomers. The international standard for vitamin A is all-*trans*-retinol, for which the international reference IU was defined as 0.3 μg (= 0.344 μg retinol acetate) of this form of retinol. Other retinoids show some activity, including *cis*-isomers of retinol, retinaldehyde, retinyl ester, dehydroretinol and dehydroretinaldehyde. The structures for these substances are given in Figure 7.2. The activity of the vitamers is broadly similar and by convention they are given equal vitamin A activity as all-*trans*-retinol.

The older procedures relied on the colorimetric Carr–Price reaction of separation on ion-exchange columns. This reaction is highly prone to interference and the method of choice is now separation by HPLC with spectrophotometric measurement. Vitamin A is very sensitive

to light and all preparations of analytical portions must be carried out in subdued lighting, preferably gold lighting. The food samples are saponified in alcoholic potassium hydroxide with the addition of an antioxidant, ascorbic acid, butylated hydroxytoluene (BHT) or pyrogallol. The vitamins are extracted into a suitable organic solvent. The extract is evaporated with additional BHT at a controlled temperature. Both normal-phase and reversed-phase HPLC can be used for the separation. In normal-phase separations measurement is usually by fluorescence; in reversed-phase separations UV detection and measurement is preferred. Standards should be followed throughout the entire sample preparation and analysis and must be controlled regularly for purity (Brubacher, Müller-Mulot and Southgate, 1985).

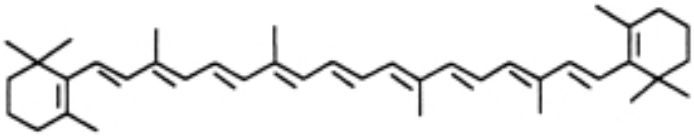
Nutritional interest originally focused on the carotenoids that demonstrated provitamin A activity, that is, were converted in the body to vitamin A. These are β-carotene, γ-carotene, α-carotene and β-cryptoxanthin (Figure 7.3). During the 1990s it was recognized that many other carotenes are biologically active as antioxidants and therefore this review is concerned with methods that permit the measurement of a wider range of carotenoids. There are some 600 carotenoid isomers (Bauemefeind, 1972), but many of these have restricted occurrence or are present in minor amounts in most common foods. Debate about how to present different carotenes and their relative activity in databases continues.

The classical method was to perform a simple chromatographic separation of the carotenes as a group, and measure spectrophotometrically against a common β-carotene standard (Brubacher, Müller-Mulot and Southgate, 1985). This has been replaced by more detailed separation using ion-exchange columns and HPLC. The conditions applied in saponification are critical and need to be carefully controlled using standard mixtures. If this is done, then comparable values can be obtained (Mangels *et al.*, 1993) with sufficient confidence to construct a database for the provitamin carotenoids (Chug-Ahuja *et al.*, 1993).

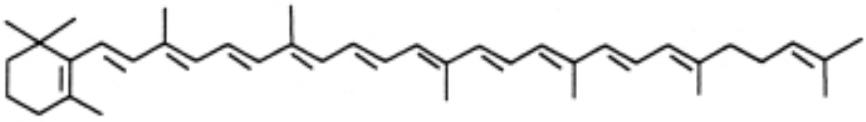
HPLC is now the most widely used and preferred method. Scott (1992) and his colleagues (Scott and Hart, 1993; Scott *et al.*, 1996), as part of an EU project to develop a SRM mixture of carotenoids, made an extensive series of studies on the various stages of the saponification extraction and HPLC analyses. Other analysts have also carried out detailed studies of the method (Wills and Rangga, 1996; Taungbodhitham *et al.*, 1998). These studies provide the basis for obtaining sound analytical values for the most important carotenoids. A revised system for evaluating published carotene values taking into account these studies has been proposed and the production of quality codes is now being evaluated.

Vitamin D. Two forms of vitamin D are found in foods, cholecalciferol (D_3) and ergocalciferol (D_2). One IU is equivalent to 0.025 □g of cholecalciferol or ergocalciferol. Vitamin D_3 is the more widely distributed (e.g. in fish oils, many fatty fish tissues, eggs, butter and cream cheese), and D_2 occurs naturally in low concentrations in fish oils and mushrooms, and is the form used in fortification. Some meats contain 25-hydroxy-cholecalciferol in concentrations that contribute to vitamin D activity and need to be considered. Figure 7.4 summarizes the structures of vitamin D. Estimates of the relative activities of cholecalciferol, ergocalciferol and their metabolites vary. The convention appears to be to attribute a factor of five times

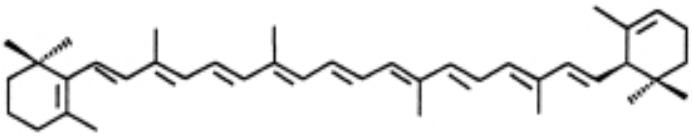
Figure 7.3 Structures of the main vitamin A-active carotenoids



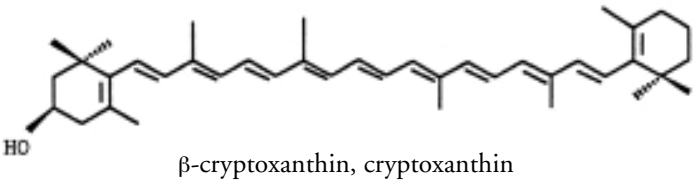
β-carotene



γ-carotene



α-carotene

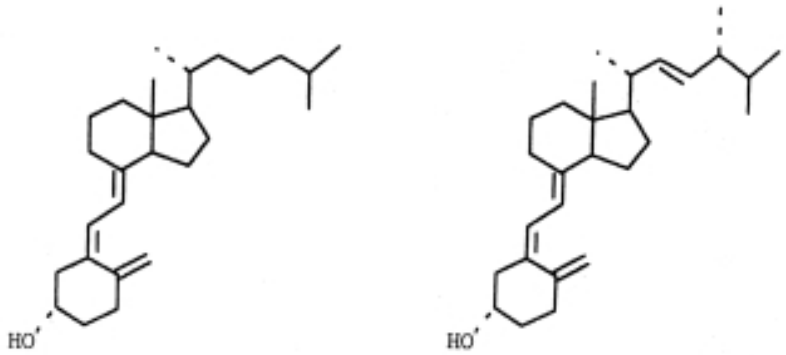


β-cryptoxanthin, cryptoxanthin

the activity of cholecalciferol to 25-hydroxycholecalciferol (Chan *et al.*, 1995, 1996). Therefore values for different forms should always be presented separately in analytical reports and reference databases.

Vitamin D in foods is found at a very low concentration, which makes its analysis difficult. The original methods were biological using chicks or young rats (e.g. Method No. 936.14 [AOAC International, 1995]). These methods are difficult to perform and had generally low precision. The major problem with vitamin D analysis is that most food sources contain other lipids that tend to interfere (Ball, 1998).

Figure 7.4 Structures of the main compounds in foods with vitamin D activity



cholecalciferol (vitamin D_3)

ergocalciferol (vitamin D_2)

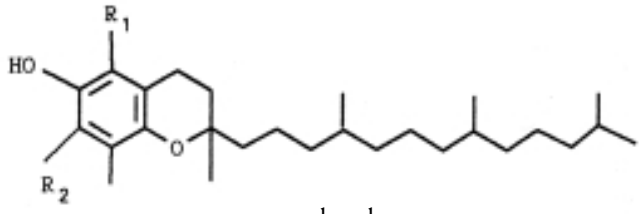
Gas chromatography is discussed by Koshy (1982), but HPLC is now the preferred technique and several methods have been published (cholecalciferol and 25-hydroxy-cholecalciferol in egg yolk [Mattila *et al.*, 1993], ergocalciferol and 25-hydroxyergocalciferol in edible mushrooms [Mattila *et al.*, 1994], and cholecalciferol, ergocalciferol and their 25 hydroxy metabolites in milk and meats [Mattila *et al.*, 1995]). Similar methods (unpublished) were used for meats in the United Kingdom food composition tables (Chan *et al.*, 1995, 1996) (V. Grace, UK Food Standards Agency, personal communication). The most useful method available involves a preliminary semi-preparative HPLC stage that eliminates much of the interference from other lipids. The food sample is saponified in alcoholic potassium hydroxide under nitrogen, with an antioxidant, ascorbic acid, hydroquinone, pyrogallol or BHT having been added before the saponification solution. The unsaponified lipids are extracted with a suitable organic solvent. An internal standard of the form of vitamin D not present in the sample is used. The unsaponified lipids are concentrated by rotary evaporation at low temperature. The extract is dissolved in the mobile phase of the semi-preparative HPLC. The conditions are carefully controlled to give a precise collection of the vitamin D.

The analytical separation may be carried out on normal or reversed-phase HPLC with UV detection. Reversed-phase is recommended for the analytical separation after normal-phase for the semi-preparation stage.

25-hydroxycholecalciferol can be measured by HPLC, as mentioned above (MAFF, 1997), but radio-immunoassay is probably the best choice at the present time where the necessary funds and equipment are available (Bates, 2000).

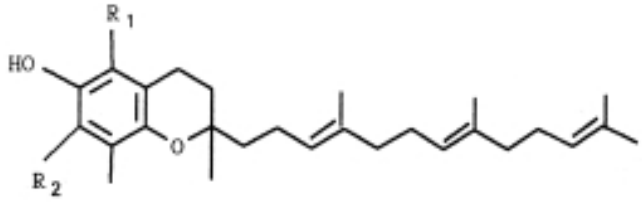
Vitamin E. Vitamin E activity is exhibited naturally by eight substances structurally based

Figure 7.5 Structures of the main compounds with vitamin E activity



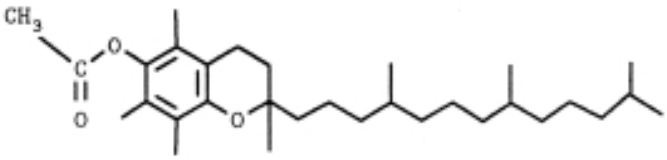
tocopherols

R_1	R_2	
CH_3	CH_3	α-tocopherol (α- T)
CH_3	H	β-tocopherol (β- T)
H	CH_3	γ-tocopherol (γ- T)
H	H	δ-tocopherol (δ- T)



tocotrienols

R_1	R_2	
CH_3	CH_3	α-tocotrienol (α- T_3)
CH_3	H	β-tocotrienol (β- T_3)
H	CH_3	γ-tocotrienol (γ- T_3)
H	H	δ-tocotrienol (δ- T_3)



α-tocopherol acetate

on tocopherols and tocotrienols (see Figure 7.5). Each vitamer has a different vitamin activity compared with α-tocopherol, which is seen as the primary structure. The preferred analytical method is therefore one that separates and measures all the different vitamers.

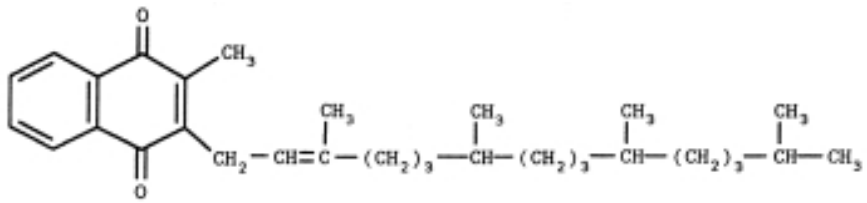
The food samples are saponified using alcoholic potassium hydroxide. The vitamin E vitamers are susceptible to oxidation at higher temperatures in alkaline conditions and should be protected by saponifying under nitrogen with the addition of antioxidants. The saponification conditions are similar to those used for vitamins A and D.

A colorimetric method, the Emmerie–Engel reaction with the reduction of ferric chloride and reaction with α, α'-dipyridine or 4,7-diphenanthroline, is also available. The complexes are rather unstable and give a total tocopherol value. The colorimetric method has been superseded by, first, GLC and, then, HPLC, which is now the preferred method.

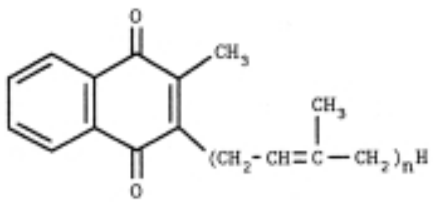
Both normal-phase and reversed-phase HPLC can be used, although the normal-phase represents the better approach and separates all the vitamers. Detection uses fluorescence (Piironen *et al.*, 1984, 1987). External standards are used and these need to be checked spectrophotometrically.

Vitamin K. Vitamin K activity is possessed by phylloquinone (K_1), the menaquinones (K_2 group) and menadione (synthetic K_3). The structures are shown in Figure 7.6.

Figure 7.6 Structures of the main natural compounds with vitamin K activity



phylloquinone (vitamin K_1)



menaquinone-n (MK-n, vitamin K_2)

Vitamin K is sensitive to alkali and UV radiation and the appropriate precautions need to be taken during analytical operations. Colorimetric procedures are available, but these lack specificity and have been replaced as the methods of choice. Most analytical attention has been given to the measurement of vitamin K_1. One major problem in the analysis is the presence of lipid, which must be removed by digestion with lipase before extraction with hexane (Indyk and Woollard, 1997). The solvent is evaporated under a stream of nitrogen and the residue dissolved in methanol, which is applied to a reversed-phase HPLC column. The eluate is reduced post-column with zinc and the fluorescence is then measured.

Semi-preparative separations have been used after digestions (Cook *et al.*, 1999) and dual electrode detection systems have also been proposed (Piironen and Koivu, 2000). Most authors comment on the great variability of the values obtained and emphasize the need for proper repeat sampling and replication of analyses (Piironen *et al.*, 1997; Jakob and Elmadfa, 1996).

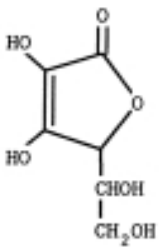
Water-soluble vitamins

These include vitamin C and a number of vitamins of the B-group. The study of vitamin C has a long history (Carpenter, 1986) and this vitamin is discussed first.

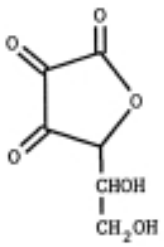
Vitamin C. Two substances show vitamin C activity, L-ascorbic acid and the first product of its oxidation – L-dehydroascorbic acid (Figure 7.7). The D-isomer (erythorbic acid), which is used as an antioxidant food additive, is not active. Ascorbic acid is a powerful reducing agent which is oxidized very quickly, especially at raised temperatures and in alkaline solutions. During the preparation of food samples for analysis it is especially important to minimize the losses due to oxidation (Brubacher, Müller-Mulot and Southgate, 1985).

In most fresh foods the amounts of dehydroascorbic acid are very low and for many purposes the measurement of ascorbic acid alone may be adequate. Thus, the reduction of 2,6-dichlorophenolindophenol is the simplest and most reliable method (AOAC Method Nos 967.21 and 985.33 [Sullivan and Carpenter, 1993]).

Figure 7.7 Structures of the common compounds with vitamin C activity



ascorbic acid



dehydroascorbic acid

The colorimetric method of Roe and Kuether (1943) involving the reaction with 2,4-dinitrophenyl hydrazine measures both ascorbic and dehydroascorbic acid.

The method of Deutsch and Weeks (1965) also measures both active forms fluorimetrically, after oxidation, and is recognized as an Official Method by the AOAC, both as originally described, and in a semi-automated version (Method Nos 984.26 and 967.22 [Sullivan and Carpenter, 1993]). Where the presence of erythorbic acid is not suspected, the fluorimetric method is probably the preferred method. HPLC techniques developed in the 1980s (Finley and Duang, 1981; Rose and Nahrwold, 1981; Keating and Haddad, 1982; Wimalasiri and Wills, 1983) for the separate measurement of ascorbic, dehydroascorbic and erythorbic acids are now widely used and give satisfactory performance (Schüep and Keck, 1990).

B-vitamins. This group includes a number of structurally distinct vitamins that were initially grouped together because they were water-soluble. The initial approach to the measurement of these vitamins, some of which are present at very low concentrations, was selective microbiological methods (Bell, 1974; Ball, 1994), and for some vitamins, total folates and vitamin B_{12}, microbiological assays remain the only practicable methods. For the remaining B-vitamins, more specific chemical procedures, especially HPLC, have been developed and collaboratively tested.

Thiamin. The structures of the substances showing thiamin activity (B_1) are shown in Figure 7.8. Thiamin is sensitive to heat and alkaline conditions and appropriate precautions must be undertaken during its analysis. Thiamin can be measured microbiologically using *Lactobacillus viridescens* or *L. fermentum*, but most analyses are based on its oxidation to thiochrome, which can be measured directly fluorimetrically. This is most conveniently carried out in conjunction with HPLC separation of interfering compounds. Thiamin, riboflavin and vitamin B_6 are present in foods as enzyme cofactors combined with phosphate and must therefore be hydrolysed and treated with phosphatase before analysis. In early descriptions of the methods for these vitamins different conditions were used, but a number of collaborative

Figure 7.8 Structures of thiamin (vitamin B_1)

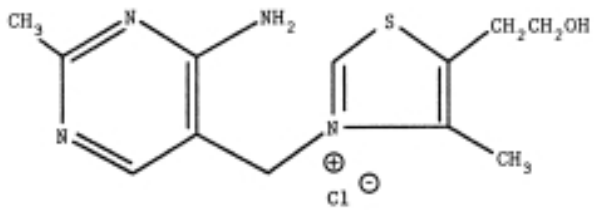


Table 7.12 Methods of analysis for water-soluble vitamins

Vitamin	*Method*	*Limitations*	*Capital costs*	*Selected references*
Vitamin C	Dye titration	Measures ascorbic acid only; pigments interfere	Low	AOAC, 1984
	Colorimetry	Measures inactive compounds also	Low	Roe and Kuether, 1943
	Fluorometry	Does not separate ascorbic and dehydroascorbic acids	Low	Deutsch and Weeks, 1965
	GLC		Medium	Schlack, 1974
	HPLC	Clean-up and separate detection of homologues add delays	High	Keating and Haddad, 1982; Wimalasiri and Wills, 1983; Speek, Schrijiver and Schreurs, 1984; Schüep and Keck, 1990
Thiamin	Microbiological	Time	Low	Bell, 1974
	Fluorometry		Low	AOAC, 1984
	HPLC		High	Fellman *et al.*, 1982; van den Berg *et al.*, 1996; Wimalasiri and Wills, 1985
Riboflavin	Microbiological	Time	Low	Osborne and Voogt, 1978; AOAC, 1984
	Fluorometry		Low	AOAC, 1984
	HPLC		High	Fellman *et al.*, 1982; Wimalasiri and Wills, 1985; Wills, Wimalasiri and Greenfield, 1985; Schüep and Steiner, 1988; van den Berg *et al.*, 1996
Niacin	Microbiological	Time	Low	Osborne and Voogt, 1978; AOAC, 1984; Sullivan and Carpenter, 1993
	Colorimetry	Hazardous reagent	Low	AOAC, 1984; Sullivan and Carpenter, 1993
	HPLC		High	Finglas and Faulks, 1987; Lahély, Bergaentzlé and Hasselmann, 1999; Rose-Sallin *et al.*, 2001

(Continued)

Table 7.12 (Continued)

Vitamin	*Method*	*Limitations*	*Capital costs*	*Selected references*
Vitamin B_6	Microbiological	Time; responses to different vitamers may not be equal; total values only	Low	Osborne and Voogt, 1978; Guilarte, McIntyre and Tsan, 1980; Sullivan and Carpenter, 1993
	HPLC		High	van den Berg *et al.*, 1996; Ndaw *et al.*, 2000
	Radiometric-microbiological		High	Guilarte, Shane and McIntyre, 1981
Vitamin B_{12}	Microbiological		Low	Thompson, Dietrich and Elvehejem, 1950; Jay, 1984; AOAC, 1984; Sullivan and Carpenter, 1993
	Radio-isotopic		High	Casey *et al.*, 1982 ; Bates, 2000
Folates (folacin)	Microbiological	Responses to different vitamers may not be equal; total values only	Low	Wright and Phillips, 1985; AOAC, 1984; Shrestha, Arcot and Paterson, 2000
	HPLC	Not all vitamers measured properly	High	Finglas *et al.*, 1999; Vahteristo *et al.,* 1996
Pantothenic acid	Microbiological		Low	Bell, 1974; AOAC, 1984; Sullivan and Carpenter, 1993
	HPLC		High	Woollard, Indyk and Christiansen, 2000
Biotin	Microbiological		Low	Bell, 1974
	Isotope dilution		High	Hood, 1975
	Radiometric-microbiological		High	Guilarte, 1985
	Protein-binding radio-immunoassay		High	Bates, 2000
	HPLC		High	Lahély *et al.*, 1999

Notes: References selected provide detailed procedures, evaluations or reviews.
GLC = gas–liquid chromatography; HPLC = high-performance liquid chromatography.

studies (van den Berg *et al.*, 1996; Ndaw *et al.*, 2000) have shown that a common method for preparing the food samples can be used.

The food sample is hydrolysed with acid and then treated with takadiastase or a phosphatase. Some authors use a ion-exchange pre-column (Bognar, 1981). The extract is then oxidized with potassium ferricyanate to form the thiochrome; it is then analysed using a reversed-phase HPLC column and the thiochrome is measured fluorimetrically. The analyses are controlled using an external standard. A post-column oxidation can also be used. In the large collaborative study reported by van den Berg *et al.* (1996) variations between the different practices in a range of laboratories did not affect overall performance of the method. Microbiological results also showed good agreement with the results from the HPLC methods.

Riboflavin. The structure of riboflavin (vitamin B_2) is shown in Figure 7.9. It is found in foods as the free riboflavin or riboflavin-5'-phosphate (FMN) and as flavin adenine dinucleotide (FAD). The vitamin is very sensitive to light and UV radiation but relatively stable to heat and atmospheric oxygen. The analytical operations must therefore be carried out under conditions that minimize the exposure to light. The vitamin must be extracted from foods by treatment with acid and a suitable phosphatase enzyme. The riboflavin can be measured directly using fluorimetric methods, although many foods contain interfering substances and separation from these by HPLC is the preferred approach (Wimalasiri and Wills, 1985; Schüep and Steiner, 1988; Arella *et al.*, 1996). Reversed-phase HPLC separation using fluorescence detection is the method most commonly used. In the collaborative study reported by van den Berg *et al.* (1996) minor variations in local methods did not affect performance. Microbiological assay using *Saccharomyces carlsbergensis* and *S. uvarum* tended to give slightly higher results than the HPLC, as observed previously by Hollman *et al.* (1993).

Niacin. Niacin activity is due to nicotinic acid and nicotinamide (Figure 7.10). Both forms are stable to atmospheric oxygen, light and heat in the dry state and in aqueous solution. A number of bound forms have been found in cereals that are extractable by alkali but these are probably not bioavailable. Tryptophan is also metabolized to niacin and the total niacin activity must include the contribution from tryptophan (Paul, 1969).

Niacin can be measured microbiologically with *Lactobacillus plantarum* (AOAC Method Nos 960.46, 944.13 and 985.34 [Sullivan and Carpenter, 1993]). Colorimetric methods based on the Konig reaction using oxidation with cyanogen bromide and reaction with p-amino-benzoyl-diethylaminoethanol have also been used (AOAC Method Nos 961.14, 981.16 and 975.41 [Sullivan and Carpenter, 1993]), but the toxic nature of cyanogen bromide makes it difficult to recommend these for routine use.

An HPLC method has been proposed and seems to perform reasonably well (Finglas and Faulks, 1987). After acid hydrolysis the food sample is filtered, treated with alkali, autoclaved and microfiltered before reversed-phase HPLC and fluorescence detection. A simplified extraction protocol has been proposed (Lahély, Bergaentzlé and Hasselmann, 1999) and has been shown to perform well with a range of foods (Rose-Sallin *et al.*, 2001).

Figure 7.9 Structures of riboflavin (vitamin B_2)

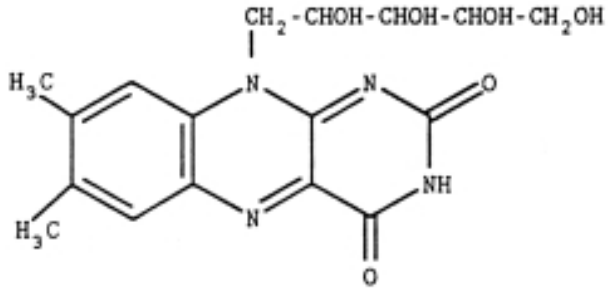


Figure 7.10 Structures of niacin and niacinamide {vitamin B_3)

Vitamin B_6. There are five compounds showing vitamin B_6 activity whose structures are shown in Figure 7.11: pyridoxamine, pyridoxine, pyridoxal and the corresponding phosphate esters.

Vitamin B_6 activity cannot therefore be measured using a method for a single substance. Microbiological assay using *Saccharomyces carlsbergensis* provides a measure of total activity (AOAC Method Nos 960.46, 961.15 and 985.32 [Sullivan and Carpenter, 1993]). The assay is carried out after an acid hydrolysis and hydrolysis of the phosphates enzymatically, and the same extraction procedures as for thiamin and riboflavin can be used (van den Berg *et al.*, 1996; Ndaw *et al.*, 2000). The acid hydrolysis also hydrolyses glycosides, which are present in plant foods and which may or may not be bioavailable to humans.

Comparison of HPLC and microbiological assay has indicated that further work is required (van den Berg *et al.*, 1996; Bergaentzlé *et al.*, 1995). Ndaw *et al.* (2000) used an extraction procedure without the acid hydrolysis stage and the HPLC method of Schüep and Steiner (1988) and the procedure performed well with standard materials.

Vitamin B_{12}. A group of complex structures possesses vitamin B_{12} activity (Figure 7.12). Classically it has been measured microbiologically with *Lactobacillus leichmanii.*

The levels of vitamin B_{12} in foods are very low and it is extracted with hot water or a buffer in the presence of potassium cyanide, which converts the vitamin into the cyano form (AOAC Method Nos 960.46, 952.20 and 986.23 [Sullivan and Carpenter, 1993]).

A number of sensitive methods have been developed for clinical use (Bates, 1997; 2000) using competitive protein binding and a range of radio-immunoassays, but these have not been evaluated in a range of foods.

Folates. The folates comprise a group of compounds related to folic acid (pteroyl-glutamic acid). Folic acid does not occur naturally in foods but is widely used in food fortification or as a supplement. Most of the naturally occurring folates are derivatives of 5,6,7,8-tetrahydrofolic acids and exist in the monoglutamate or polyglutamate forms. Their structures are summarized in Figure 7.13.

The biological activity of the forms differs and the ideal analytical nutritional procedure therefore should involve the measurement of the different vitamers.

Total folate values are best measured by microbiological assay using *Lactobacillus rhamnosis* (*casei*). Most organisms cannot use the polyglutamate forms, and deconjugation with a suitable enzyme (hog kidney, chicken pancreas, human plasma) is a preliminary stage in the analysis. The extraction is carried out in the presence of ascorbic acid to minimize

Figure 7.11 Structures of the most common compounds with vitamin B_6 activity

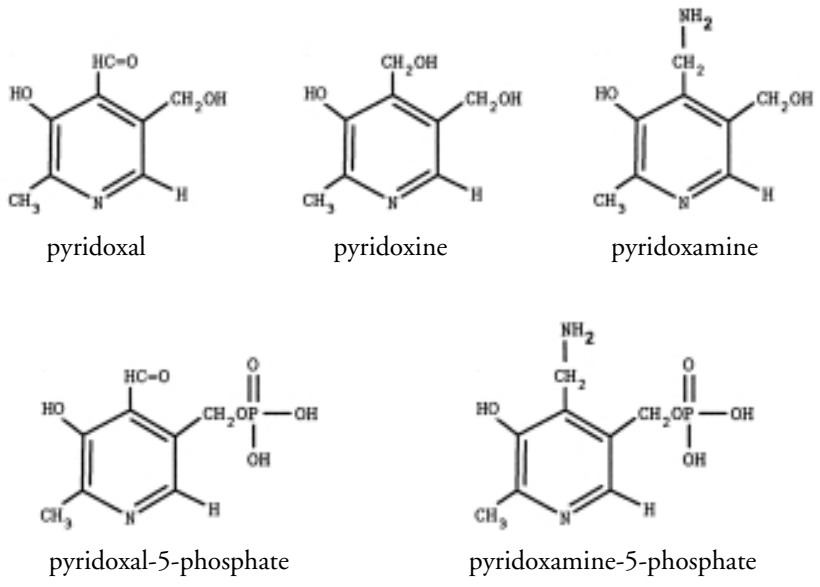
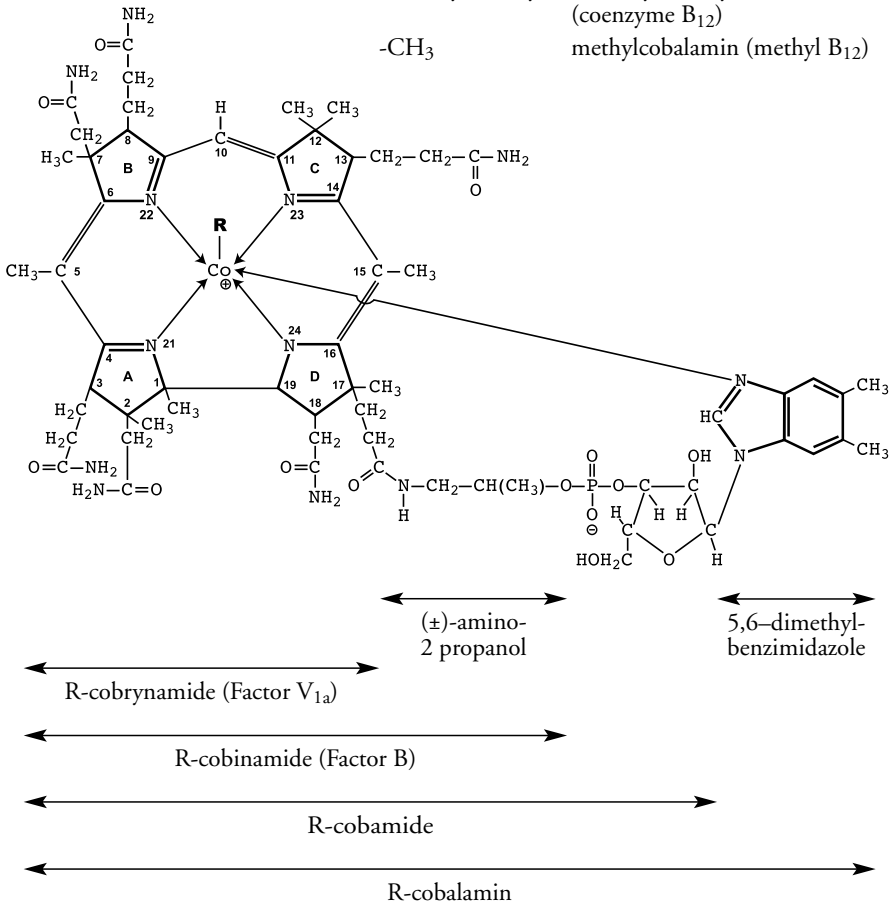


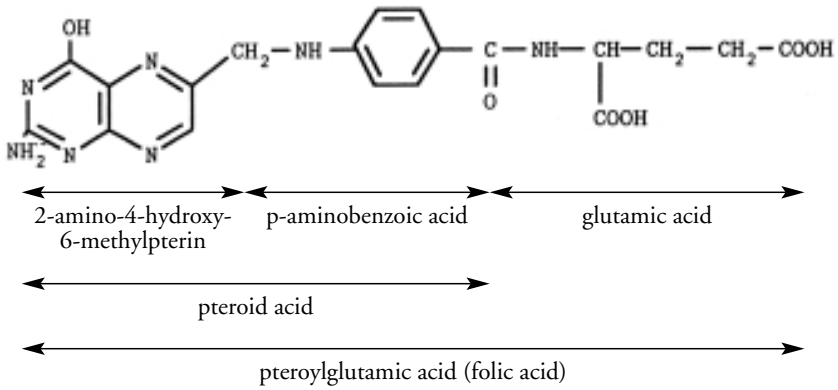
Figure 7.12 Structures of vitamin B_{12} and analogues

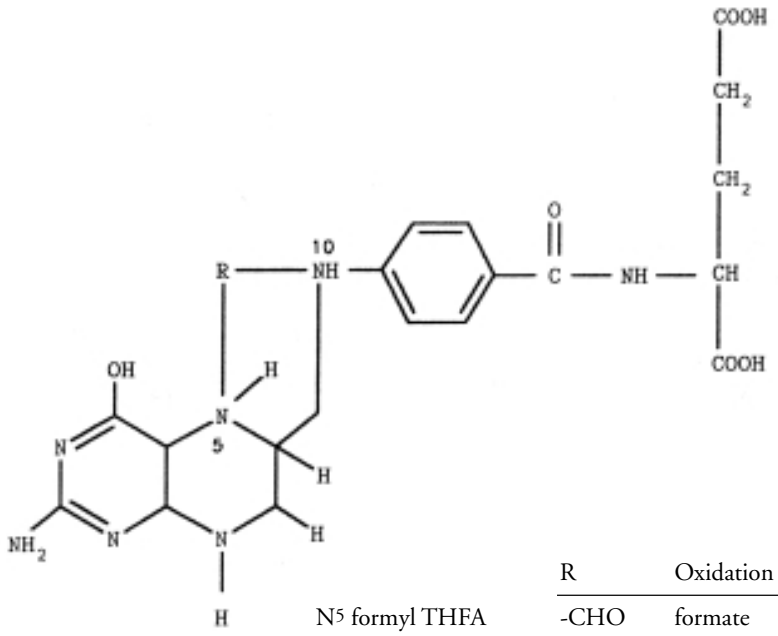
-R	Permissive name
-CN	cyanocobalamin (vitamin B_{12})
-OH	hydroxocobalamin (vitamin B_{12a})
$-H_2O$	aquocobalamin (vitamin B_{12b})
$-NO_2$	nitritocobalamin (vitamin B_{12c})
5'-deoxyadenosyl	5'-deoxyadenosylcobalamin (coenzyme B_{12})
$-CH_3$	methylcobalamin (methyl B_{12})



Source: Modified, with permission, from Brown, G.M. & Reynolds, J.J., *Annual Review of Biochemistry*, 32: 419-62. © 1963 by Annual Reviews Inc.; reproduced with permission from Shils, M.E. & Young, V. (1988) *Modern nutrition in health and disease.* 7th ed. Philadelphia, PA, USA, Lea & Febiger.

Figure 7.13 Structures of folacin (folates)





	R	Oxidation state
N^5 formyl THFA	-CHO	formate
N^{10} formyl THFA	-CHO	formate
N^5 formimino THFA	-CH=NH	formate
$N^{5,10}$ methenyl THFA	>CH	formate
$N^{5,10}$ methylene THFA	$>CH_2$	formaldehyde
N^5 methyl THFA	$-CH_3$	methanol

oxidation. The extract is treated with a combination of protease, lipase and amylolytic enzymes, which improve the efficiency of extraction. The different conjugase enzymes give similar performances. At one time it was assumed that the measurement of folate before and after deconjugation would give values for "free" folate and total folates. The organisms respond to varying extents to the glutamate derivatives and the concept is flawed. The conditions for the microbiological assay were studied by Phillips and Wright (1982, 1983), Wright and Phillips (1985) and Shrestha, Arcot and Paterson (2000); these procedures give satisfactory quantitation.

Separation of the different folate vitamers using HPLC techniques is now widely used (Finglas *et al.*, 1999) and some databases give values. Intercomparison studies have shown that values for 5-methyl tetra-hydrofolate showed reasonable agreement, but the agreement with other vitamers was not satisfactory (Vahteristo *et al.*, 1996). Subsequent studies on the standardization of the HPLC methods have shown that while it is possible to measure the 5-methyl form with reasonable confidence, the other vitamers are still not measured properly by existing methods that use fluorimetric detection. A kit is available for folic acid and an evaluation has been published by Arcot, Shrestha and Gusanov (2002).

Pantothenic acid. The structure of pantothenic acid is given in Figure 7.14. Pantothenic acid in the free form is unstable and extremely hygroscopic. It is usually present bound to proteins or in the form of salts. Only the dextro- form is active. The classical method is microbiological using *Lactobacillus plantarum* as the test organism (Bell, 1974; AOAC Method Nos 960.46 and 945.74 [Sullivan and Carpenter, 1993]). The food is extracted with water and where the food is rich in fats these are best removed before analysis. The aqueous extract is usually autoclaved and the pH adjusted with acid and alkali to around 6.8. The mixture, after incubation overnight, is heat-treated to stop growth and growth is measured turbidometrically.

Biotin. Biotin is found in foods as the free vitamin and bound to protein. Figure 7.15 shows the structure of the vitamin. The classical method is microbiological using *Lactobacillus plantarum* (Bell, 1974; AOAC Method No. 960.46 [Sullivan and Carpenter, 1993]). An HPLC method has also been described (Lahély *et al.*, 1999). Preliminary extraction with acid followed by papain treatment is required to extract the vitamin from the food. The HPLC

Figure 7.14 Structure of pantothenic acid

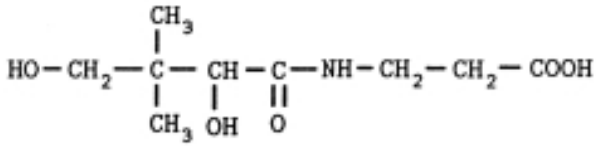
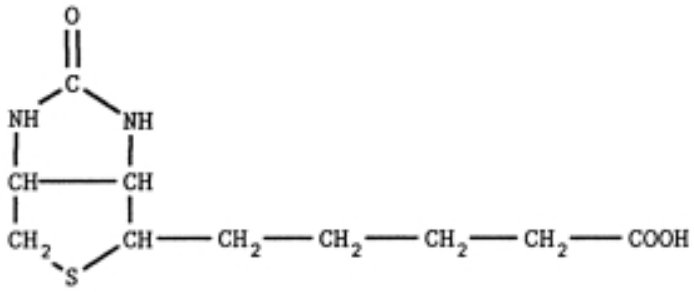


Figure 7.15 Structure of biotin



method uses a reversed-phase separation, post-column derivatization with avidin-fluorescence 5-isocyanate and fluorescence detection.

Radio-assays using the specific binding protein have also been described (Bates, 2000).

Bioactive food components

Pennington (2002) has published a comprehensive review of food composition databases for bioactive food components, including flavonoids, tannins, allyl sulphides, capsaicin, indoles, lignans, monoterpenes, phenolic acids, plant sterols and probiotics, categorized by food and by compound, and available as an annotated bibliography of over 400 pages on individual components (Pennington, 2001). Given the number and diversity of these components, it is not possible to review the methods for all of them (Speijers and van Egmond, 1999). This section therefore focuses on methods for measuring flavonoids, isoflavonoids, lignans and total antioxidant activity in view of the fact that these have been the subject of much interest in recent years. Methods for plant sterols were reviewed earlier in this chapter.

Flavonoids. A rapid method based on reversed-phase HPLC with UV detection was developed by Hertog, Hollmann and Venema (1992) for the quantitative determination of five major flavonoid aglycones (quercetin, kaempferol, myricetin, luteolin and apigenin) in freeze-dried vegetables and fruits, after acid hydrolysis of the parent glycosides. More recently Merken and Beecher (2000) published a gradient HPLC method with photodiode array detection for 17 prominent monomeric flavonoid aglycones representing all of the five common classes of flavonoids.

Phytoestrogens. The main plant compounds with known or suspected estrogenic activity are lignans, isoflavones, coumestans and resorcyclic acid lactones (Price and Fenwick, 1985). The modes of estrogenic action are discussed by Clarke *et al.* (1996). The major isoflavonoids are genistein, daidzein, formononetin, biochanin A and glycitein. Genistein, daidzein and glycitein

occur in foods as their glycosides, all of which are biologically inactive. The free aglycones are formed by metabolic action of the human gut microflora, although this hydrolysis varies considerably from person to person (Xu *et al.*, 1994). The total bioactivity is represented by the analysis of aglycones; however, this potential activity is represented by analysis of the conjugates and aglycones separately. The most active plant estrogen known is coumestrol (a coumestan); zearalenone is a potent resorcyclic acid lactone formed as a secondary metabolite of fungal species, mainly *Fusarium* (and is thus regarded as a contaminant). The lignans matairesinol, secoisolariciresinol, pinoresinol and isolariciresinol are potent phytoestrogens and are precursors of the mammalian lignans, enterolactone and enterodiol.

Given the very large number of plant compounds with estrogenic activity and the question of whether to analyse both the conjugates and the free forms or only the aglycones (after hydrolysis), many methods of analysis are in existence and there is little agreement on which method is best. No method is available to separate and quantify all bound and free compounds of interest in this category. Probably the most comprehensive method for the aglycones is the isotope dilution gas-chromatographic–mass spectrometric method of Adlercreutz and coworkers (Mazur *et al.*, 1996), which analyses daidzein, genistein, biochanin A, formononetin, coumestrol, secoisolariciresinol and matairesinol, but not glycitein, as silyl derivatives. The method is expensive and needs access to mass spectrometry (MS). Another comprehensive method for foods that analyses daidzein, genistein, biochanin A, formononetin, coumestrol, secoisolariciresinol and matairesinol, but not glycitein, uses an HPLC-MS method originally developed for plasma and urine (Horn-Ross *et al.*, 2000; Coward *et al.*, 1996; Horn-Ross *et al,*. 1997; Barnes *et al.*, 1998).

Isoflavones and coumestrol. For the USDA–Iowa State University Isoflavones Database (2002), the reference method adopted was the linear gradient method of Murphy *et al.* (1997), which separates daidzein, genistein, glycitein and their conjugates in soy-based infant formulas. Hutabarat, Greenfield and Mulholland (2000) have published a rigorously validated isocratic HPLC method for genistein, daidzein, formononetin, biochanin A and coumestrol (but not glycitein), while King and Bignell (2000) have published an HPLC method for daidzin, genistin, glycitin and their aglycones. A collaborative trial published by Klump *et al.* (2001) led to a recommendation to adopt as first action AOAC Method No. 2001.10 for the determination of isoflavones in soy and selected foods containing soy. This method uses reversed-phase liquid chromatography to separate and measure genistein, glycitein and daidzein and their glucosides, and also produces values for total isoflavones expressed as aglycones.

Lignans. Meagher *et al.* (1999) measured isolariciresinol, pinoresinol, secoisolariciresinol and matairesinol using HPLC with photodiode array detection, and Liggins, Grimwood and Bingham (2000) have published a GC-MS method for the determination of matairesinol, secoisolariciresinol and shonanin in foods as trimethylsylyl derivatives.

Total antioxidant activity. There is growing interest in ways to represent the total antioxidant

Table 7.13 Energy value of some constituents of food[a]

Constituent	*kcal/g*	*kJ/g*[b]
Protein	4	17
Fat	9	37
Available carbohydrate as monosaccharide equivalent	3.75	16
Available carbohydrate (as weight, by difference)	4	17
Total carbohydrate	4	17
Monosaccharide	3.75	16
Disaccharide	3.94	16
Starch and glycogen	4.13	17
Ethyl alcohol	7	29
Glycerol	4.31	18
Acetic acid	3.49	15
Citric acid	2.47	10
Lactic acid	3.62	15
Malic acid	2.39	10
Quinic acid	2.39	10

Notes:
References selected provide detailed procedures, evaluations or reviews.
[a] Individual countries may have additional factors defined within food regulations.
[b] Conversion factor: 1 kcal = 4.184 kJ; the kJ equivalents have been rounded to two significant figures (Royal Society, 1972).
Source: Adapted from Paul and Southgate (1978).

activity of foods. A number of methods have been used but no standards exist and at this stage the inclusion of values for total antioxidant activity in foods in databases is not recommended. The topic is fully reviewed by Frankel and Meyer (2000).

Energy

The gross energy content of a food may be determined experimentally with a bomb calorimeter (Brown, Faulks and Livesey, 1993). An adiabatic bomb calorimeter is preferred for precise measurements, but the ballistic bomb calorimeter (Miller and Payne, 1959) gives a precision that is adequate for most nutritional studies. The values obtained using an adiabatic bomb calorimeter are corrected for the heat generated from the oxidation of nitrogen and sulphur in the food. The calorimeters are usually calibrated using benzoic acid as a thermo-chemical standard.

The values obtained are the gross heats of combustion and are not the values used in nutritional sciences and food composition databases; for these purposes, metabolizable energy is used. This is the energy that is available for use in metabolism by the body. Metabolizable energy values are calculated using energy conversion factors (Atwater and Bryant, 1900; Southgate and Durnin, 1970; Merrill and Watt, 1973; Allison and Senti, 1983) for the protein, fat, carbohydrate and alcohol contents. Recently, Livesey (2001) has argued that a better system for calculating the energy values of food would be the net metabolizable energy system (Blaxter, 1989).

Recently, the contributions from dietary fibre, polyols and oligosaccharides have been widely discussed (Livesey, 2001; FAO/WHO, 1998), but most databases do not yet use the energy conversion factors for these components.

In many countries, Le Système International d'Unités (or International System of Units [SI]) (BIPM, 1998, 2003) is used to express the energy values of foods and diets, using the Joule (J) (work): 1 kcal is equivalent to 4.184 kJ (thermochemical equivalent) (Royal Society, 1972). When expressing the energy value of foods, no more than three significant figures should be used. Whichever system of calculation is chosen for energy, it should be clearly indicated.