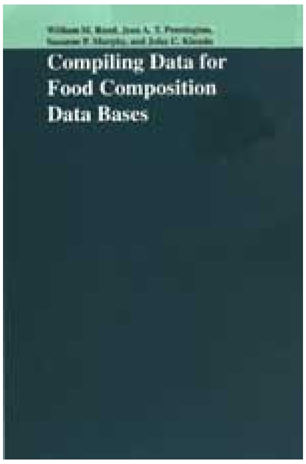


Compiling Data for Food Composition Data Bases

WILLIAM M. RAND, JEAN A. T. PENNINGTON, SUZANNE P. MURPHY, and JOHN C. KLENSIN

UNITED NATION UNIVERSITY PRESS

The views expressed in this publication are those of the authors and do not necessarily reflect the views of the United Nations University.

United Nations University Press
The United Nations University
53-70 Jingumae 5-chome, Shibuya-ku
Tokyo 150, Japan
Tel.: (03) 3499-2811. Fax: (03) 3406-7345.
Telex: J25442. Cable: UNATUNIV TOKYO.

Printed in Hong Kong

WHTR-15/UNUP-772
ISBN 92- 808-0772-2
United Nations Sales No. E.91.III.A.4
02000 P

Contents

Acknowledgements

The united nations university is an organ of the united nations established by the general Assembly in 1972 to be an international community of scholars engaged in research, advanced training, and the dissemination of knowledge related to the pressing global problems of human survival, development, and welfare . Its activities focus mainly on peace and conflict resolution, development in a changing world, and science and technology in relation to human welfare. The University operates through a worldwide network of research and post-graduate training centres, with its planning and coordinating headquarters in Tokyo, Japan.

The International Food Data Systems Project (INFOODS) is a comprehensive effort, begun within the United Nations University's Food and Nutrition Programme, 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. At present in many cases such data do not exist or are incomplete, incompatible, and inaccessible.

This volume is the third in a series that provide guidelines on the organization and content of food composition tables and data bases, methods for analysing foods and compiling those tables, and procedures for the accurate international interchange of the data. It presents an overview and guidelines for how information that results from the analyses of foods should be compiled into useful food composition tables.

Executive summary

Data on the composition of foods are essential for nutrition research, product development, nutrition education, trade of foods and food products between and within countries, and development of nutrition and agricultural policies by government agencies. Food composition data have been compiled into many data bases throughout the world. As the uses of those data increase, a larger number of individuals and organizations become involved in their compilation, and thus the need for guidelines on their gathering, formatting, and documentation increases. This document describes and presents recommendations for the procedures involved with compiling the values for food composition data bases and tables. Specifically addressed are the five major ways to obtain data on the nutrient content of foods:

- direct analysis based on analytical measurements,
- calculated as representative values (e.g., weighted means of several samples),
- gathered from other sources (e.g., taken from other tables or the literature),
- estimated from similar foods (e.g., substitution of data),
- estimated from ingredients (e.g., recipe calculations).

Two main themes of this manual are (1) the importance of careful definition of the data base, in terms of its ultimate use, before the actual process of compilation begins, and (2) the importance of careful documentation of the procedures used to obtain the data. The data gathering, manipulation, and estimation techniques associated with each data point in a data base should be clearly identified and described. In addition to fostering compatibility and consistency between data bases, adequate documentation will also give the users an

awareness of the extent and limitations of the data and a clear indication of necessary and potential improvements to the data base.

Compiling a data base is a complex endeavour requiring major effort; it is the goal of this document to provide guidelines to data base compilers to assist them in focusing their efforts, in the hope that future data bases will be more compatible, more consistent, and more useful to a wider audience.

William M. Rand is Professor of Biostatistics in the Department of Community Health of the Tufts University School of Medicine, Boston, Massachusetts, USA.

Jean A. T. Pennington is Associate Director for Dietary Surveillance in the Division of Nutrition, Center for Food Safety and Applied Nutrition, Food and Drug Administration, Washington, D.C., USA.

Suzanne P. Murphy is Assistant Research Nutritionist in the Department of Nutritional Sciences, University of California, Berkeley, California, USA.

John C. Klensin is Principal Research Scientist in the Department of Architecture, Project Coordinator for INFOODS, and Director of the INFOODS Secretariat at the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

The technical committee for this document consisted of the above and: G. Beecher, Nutrient Composition Laboratory, Department of Agriculture, USA L. Bergström, National Food Administration, Sweden I. M. Buzzard, Nutrition Coordinating Center, University of Minnesota, USA D. Feskanich, INFOODS Secretariat, Massachusetts Institute of Technology, USA F. N. Hepburn, Human Nutrition Information Service, Department of Agriculture, USA J. M. Holden, Nutrient Composition Laboratory, Department of Agriculture, USA L. W. Hoover, University of Missouri, USA B. Perloff, Human Nutrition Information Service, Department of Agriculture, USA G. J. Petot, Case Western Reserve University, USA W. Polacchi, Italy N. Rawson, Campbell Soup Company, USA W. A. van Staveren, Agricultural University, the Netherlands K. Yasumoto, Kyoto University, Japan

Preface

Public health nutrition activities; agricultural, nutritional, and epidemiological research; food industry and trade decisions; and government planning and policies concerning nutrition and agriculture all depend on accurate knowledge of what is in foods. Currently, these data are not always adequate for existing needs. Often they are incomplete, inaccurate, inconsistent, incompatible, or inaccessible. While there is much excellent information on food composition throughout the world, its ultimate utility could be increased by better communication and interchange of both information and ideas among countries.

INFOODS was formed in 1983, under a mandate of the United Nations University, to develop operational communication paths between the gatherers, the compilers, and the users of food composition data [70, 69]. As part of its activities, funded primarily by agencies of the United States Government (the National Cancer Institute, the Food and Drug Administration, the National Heart, Lung, and Blood Institute, and the Department of Agriculture), INFOODS was commissioned to prepare a series of guidelines on how to collect and analyse foods for nutrients and other substances, how to record and communicate food composition data, and how to use food composition data in research and practice. This document focuses on the issues involved in gathering together, and estimating where necessary, the specific data needed for a food composition table or data base. It should be useful to developers as well as users of food composition data bases, at both local and national levels.

A prime concern of INFOODS has been fostering formal and informal discussions to outline areas in which guidelines would be useful. Starting with the initial INFOODS meeting in Bellagio, Italy [73], and continuing through the initial regional meetings of EUROFOODS [102], ASIAFOODS [71], LATINFOODS [12], and OCEANIAFOODS [20] and annual meetings of the U.S. National Nutrient Databank Conference, there have been groups focusing on how one actually obtains and combines food composition data into a data base. In late 1986 an international meeting was held in Washington, D.C., to discuss the general problems of "missing data", i.e., how to deal with situations in which no analytic data exist. This meeting produced the general plan for the current document. A first draft of the document was circulated to those who had attended that meeting and to the various regional liaison groups; the current version incorporates their very helpful suggestions.

Part I the data base

Tables listing the components of specific foods have been published for over 150 years. Food composition tables have been developed by international agencies, governments, universities, industries, and individuals around the world [28, 37, 72]. Each of these tables differs from others in terms of the foods examined, the nutrients analysed, and the data presented. Many of these tables also differ in their methods of gathering and handling data. As food composition data assume more scientific, academic, and political importance, compatibility and consistency of analysis and presentation become more critical.

This part of the manual discusses specific aspects of data bases, stressing the importance of planning and documentation. Its primary emphasis is on the importance of careful definition of foods and nutrients and documentation of data sources and methods of data manipulation. The utility of food composition data depends on the information that accompanies them.

1. Data base considerations

Importance and status of food composition data

Food composition data are important to a spectrum of users ranging from international organizations to private individuals [11, 13, 36, 65, 100, 105].

At the international level:

- Food assistance programmes, both between and within countries, distribute foods in response to needs for specific dietary components such as energy, protein, and vitamin A.
- Epidemiologists correlate patterns of disease with dietary components to construct hypotheses of causality.
- Food trade as countries seek to import and export foods which are nutritionally important.
- At the national level:
- Governments monitor both nationally produced and imported foods to maintain their nutritional content and safety.
- Governments assess the nutritional status of their populations through food consumption studies.
- Agricultural researchers work to improve the food supply by developing new strains and cultivars, and devising new methods of cultivation, harvesting, and preservation.

At the regional level:

- Schools, hospitals, armies, and other institutions provide nutritionally adequate foods for those in their care.
- Food industries continually modify their products, changing levels of components to improve sensory appeal, improve nutritional value, or lower cost.
- The educational sector teaches people how to prepare and combine foods to provide or avoid certain food components.

At the individual level:

- Nutritionists examine diets of individuals to better understand their health and disease status.
- Dietitians counsel people with respect to their personal dietary needs and goals.
- Individuals may decide what to serve and eat on the basis of food composition.

Each of these activities requires accurate data on the composition of foods, and requires that these data be in a form that permits easy access, intelligent manipulation, and confident usage.

There are currently over 150 food composition tables in use around the world [28].

This number excludes most of the tables that exist for the United States [37] since, for the most part, these depend heavily on data from the United States Department of Agriculture (USDA). Because of the overlap between tables and inadequacies of documentation, an exact count of the number of unique foods for which there are data is difficult to determine, but they number in the thousands. Even so, the number of foods consumed is probably several times greater than the number for which analytic data exist.

Nutrient profiles of the foods that have been analysed are not always complete. Major food composition tables routinely contain data on 25 to 50 food components, but do not include all compounds that routinely occur in foods and are suspected of having biological activity. The biological activity of a nutrient in any given food derives both from the total profile of components in that food (and others with which it may be combined) and from the physical condition of the consuming individual. Nutrient bioavailability is one of the frontiers of modern nutritional research and an area not yet reflected in most food composition data bases.

Without information on their statistical distribution, the values available in food composition tables may be misleading. The amount of a nutrient in a food item has a probability distribution and cannot be adequately described by the results of a single analysis, or averages of analyses performed on samples drawn for convenience rather than for representativeness.

There should be better documentation of work in the areas of the generation and use of food composition data. There are few generally accepted standards and guidelines for gathering, aggregating, saving, identifying, manipulating, or using food composition data. One goal of this manual is to provide some of these guidelines so that future developers can avoid the deficiencies of the past.

Analytic and non-analytic data

Analytic data are values based on laboratory values, including those obtained by well-defined conversion factors and straightforward formulae. Thus, protein data calculated by multiplying nitrogen content by a constant are considered analytic data. By contrast, non-analytic data are values which involve either no chemical analyses (e.g., using a value of zero for the cholesterol in an orange because it is a plant product) or the use of analytic data with varying amounts of estimation involved (e.g., the calculation of the vitamin content of a stew from the vitamin content of its raw ingredients). Non-analytic data are often, but not always, less accurate than analytic data.

The terms "calculated" and "imputed" are often used for data that are not analytic, "calculated" implying more trustworthiness than "imputed". While many people distinguish between these two terms, there is little agreement on their precise meaning. To avoid conflict, we use the term "estimated" for those data which are not strictly laboratory values.

A data base should contain the most accurate and precise data that are available. To many, this suggests that the entries should be purely analytic data; however, there are several reasons why this is neither a practical nor a theoretical goal in food composition data bases.

Impractical Analyses

For most foods and many nutrients, there exist analytic procedures that produce reasonable data [83, 86]. However, these procedures differ in complexity and expense, and in the course of research, decisions will be made which effectively declare certain of the analyses impractical, prohibitively expensive, time-consuming, or labour-intensive. These decisions are made because the costs cannot be justified on the basis of the potential uses of the data-the lack of involvement of a particular food or nutrient with a health problem or the infrequency of consumption of a food by the population of interest-or simply because of cost-benefit trade-offs, e.g. where the choice is between analysing a number of foods for one nutrient at the same cost as that of completing a single analysis for another nutrient.

Multiple Measurements

If more than a single observation is available for a nutrient in a specific food, it may be necessary to summarize these "replicate" data. While certain data bases, such as those reflecting raw laboratory values, may include all the analytic data that exist as individual entries, most users need some indication of the most likely values for the nutrients of the food in question, as well as some indication of how variable the data are. In order to provide this information, it is necessary to manipulate the given data statistically to derive data base entries. A more complete discussion of these manipulations is contained in chapter 3.

Aggregated Foods

Frequently, data are needed for representative foods which do not in fact exist: one cannot determine the exact type of potato that a consumer purchases, since the consumer chooses from many different cultivars which are grown, harvested, and stored under a variety of conditions. What can be analysed are the various types or varieties of the food available to the consumer. Then these data can be combined into a single entry in a data base, an entry which does not represent any existing food, but can be used in certain well-defined situations. The appropriate procedures for combining data in this manner, producing weighted averages, are detailed in chapter 3. (Note that another alternative is to prepare and analyse a food sample that consists of specific proportions of different cultivars or brands and that this is occasionally done.)

Number of Different Foods

With limited resources for the compilation of food composition data bases and the current state of development of analytic techniques data base compilers can analyse all foods for all the nutrients that users require. The alternatives are using analytic data that already exist, using data on similar foods, and using recipe data for multiple ingredient foods.

It must be recognized that these various procedures all have certain inherent limitations such as potentially low accuracy or impressions of unrealistically high precision. These two independent aspects of data are critically important.

- The "accuracy" of the reported value of a nutrient in a given food is the closeness of that number to the "true" level of the nutrient in the food. This is sometimes called the "validity" or "representativeness" of the data, and is also referred to as the "lack of bias". Determining the "accuracy' of a set of data is not easy, since often the "true" value cannot be determined. The chemical analyst has a number of ways to deal with this. (See, for example, Piggot [66] or Stewart and Whitaker [87].)
- "Precision" indicates how close the data in a given set are to each other, or how close repeated measurements would be (or how "repeatable" a method of producing data is). Precision is independent of how close the data are to the true value. The variability of a nutrient in a given food complicates the precision of food composition data. Care must be taken to distinguish between the variability inherent in the food and the variability inherent in any method of obtaining the data.

Discrepancies between different estimated nutrient values reflect different sources of foods and nutrients, methods of estimation, modes of expression for the data, judgements of coders, or values for yields retentions, and water/fat changes during preparation. This situation can be improved by careful definition and documentation of estimation procedures.

Types of data bases

There are several different ways of classifying food composition data bases, just as there are different ways of classifying the quality and origins of the data in them. The most important distinction is between REFERENCE data bases and SPECIAL-PURPOSE or APPLICATION data bases. These two types of data bases (which are not totally distinct but rather represent two ends of a spectrum) reflect an important two-level structure in the community of data base compilers. As a generalization, some individuals and organizations seek to maintain and provide general storehouses of food composition data for diverse users to draw upon, and others build specific data bases tailored to specific uses (such as the analysis of consumption data from a particular survey). There are also groups that maintain a large, general-purpose data base as well as produce subsets of it for specific applications sometimes supporting themselves and the maintenance of the general data base by sale of the specific application data bases and services.

More specifically, REFERENCE data bases are those designed primarily to provide the raw material for the construction of other, APPLICATION data bases. Reference data bases may contain analytic data for specific food items (e.g., a particular cultivar of apples), analytic data for composite foods (e.g., apples by market share), data that have been compiled and aggregated from various sources (e.g., data on "apples" from various sources weighted by number of samples), or any combination of these. A key property of a reference data base should be that it contains complete information about sampling, sources, and methods used to produce data, in addition to complete descriptions of the nutrients and foods included. This information is essential for evaluating selecting, and combining the data into more application-oriented food tables or data bases. The standard examples of reference data bases are those of the FAO (Food and Agriculture Organization) [47, 46], INCAP (the Institute of Nutrition of Central America and Panama) [48], the United States [96], the United Kingdom

[59], and Japan [75]; however, the full set of data for reference data bases is not usually published, remaining instead in the files of the issuing organization.

Alternatively, an APPLICATION, or SPECIAL-PURPOSE data base, is one that is keyed to a specific application or problem. It generally does not contain extensive descriptive or primary data; however, it does need references to this information. An application data base may be a subset of a single reference data base (a specific set of foods or class of nutrients); it may draw data from several reference data bases (similar foods in different pans of the world); or it may supplement data from reference data bases with data gathered from other sources. Application data bases tend to be more compact and easier to handle and use than the general reference data bases. Often they are organized differently from the very general format required by the reference data base.

Data base development

The compilation of a food composition data base involves a number of different tasks which require clear, careful planning and integration. To minimize waste and optimize selection, the purpose of the data base must be defined before assembling the data. Its ultimate use dictates both the content (the data) and the form (organization and medium) of the data base.

Once the data needed have been specified, it is necessary to identify the source and location of the data. This is discussed in detail in chapter 4. Parallel with these considerations is the handling of the data, both logically and physically. The merging of the data into the data base requires funkier decisions and planning, more fully discussed in pan II.

For data bases expected to provide long-term use, it is essential to embed the effort in a strong, ongoing, institutional framework. This will provide the machinery for correcting entries and keeping the data base current by adding and deleting foods and modifying those nutrients whose data have changed. These changes include food reformulations by industry (based on consumer demand, regulatory changes including changes in fortification levels, etc.) as well as the appearance of new foods in the marketplace. Additionally, reported nutrient levels may change due to improved analytical methods.

One of the most important activities is documentation of what is done and why. This record, essentially an annotation of the data, is fundamental to most uses of the data base.

Finally, the compiler should make contact with individuals and organizations with previous experience in compiling food composition data bases. The individuals and organizations that produce data bases, as listed in the available directories and reports [28, 37, 55, 99, 103], will often provide advice and assistance to both compilers and users.

Content of the Data Base

The information in a particular data base depends on the expected uses of the data base. Its focus can be determined by answering questions about the data base: What foods should be included? What nutrients should be included? What son of precision, accuracy, and description of the data will be required? And what information, in addition to composition of foods, will be needed?

The information required must be weighed against the cost of obtaining that information. Thus, early in its planning, any food composition data base effort must come to grips with the economics of the project. Cost estimates, at least approximate, should be calculated for each type of information to be included, and the different classes of information should be ordered in terms of priority so that informed decisions can be made.

Foods for Inclusion

The most important criterion for selecting foods for inclusion is that of relevance to purpose. Data bases compiled for use with a consumption survey would include foods likely to be consumed by the population under study. Extensive site visits may be required to assess agricultural products and market availability as well as to determine food intake patterns at home and in restaurants. Researchers interested in the relationship of diet to heart disease would need a data base that includes foods which contained nutrients or ingredients suspected of having a role in the promotion or prevention of that disease. Those responsible for monitoring contaminants or toxins might be interested in a data base listing foods commonly consumed, or containing significant amounts of components thought to be of public health significance.

A principal concern in selecting foods for inclusion in a food composition data base is the contribution of the foods to the diet.

Foods such as bread, rice, or corn consumed in large amounts by the population or subpopulation of interest, as well as those that supply large amounts of specific nutrients, should be among the first considered for inclusion in a data base.

Another concern when deciding on foods for inclusion is the level of aggregation that is needed. For many purposes, very specific data are needed, such as the variety or cultivar name, where the food was grown, or the name of the manufacturer (e.g., when analysing or designing a specific diet). Often, however, data on generic or aggregated foods are required, such as a generic apple or a prepared steak and kidney pie, when analysing a food consumption survey in which specifics were not collected or not recalled. For example, in national tables the entry for apples may be an average of data for several different varieties of apples, weighted by their representation in the market. Market representation, in turn, reflects the frequency of consumers' choices. It is essential that data base compilers decide whether such aggregated foods are necessary and then carefully, and explicitly, define them in terms of individual foods for which data are available. Alternatively, where information on foods that are prepared as mixed dishes (such as steak and kidney pie) is needed, it is often most efficient to prepare the dish by following a "representative" recipe and analyse that dish for the required food components. This has definite economic advantages over going into homes and shops, selecting different preparations, and running multiple analyses; however, it is important that some site visiting be done to ensure that the dish resulting from the chosen recipe is similar to what is actually eaten. This entire area of recipe variability and validation needs much further effort.

Nutrients for Inclusion

A data base should contain information on the food components that potential users of the data base will need. Some data bases, such as those compiled for particular research studies, will have their nutrients completely specified in advance. Other data bases will be intended for more general use, and will therefore contain a wide range of nutrients. However, all data base compilations have economic constraints that preclude listing all food components. Selections must be made considering the specific nutrients for which there are identified needs, information on additional components that might expand the user community and extend the utility of the data base, and the projected cost of including each nutrient. Criteria for inclusion are the importance of individual components and the specificity and adequacy of the analytic methods.

Importance of individual components. The primary need of users of food composition data bases is for data on components that affect human health. This includes the proximate nutrients, as well as specific other components, such as fatty acids and trace minerals, that are related and relevant to some distinct area of concern. Some components are included in food composition data bases because they describe the food (e.g., water), or may be useful in checking the other data (e.g., ash).

Specificity of analytic methods. In recent years, new analytic methods have been developed which permit the separation, identification, and measurement of individual vitamins and compounds which were previously aggregated under single nutrient labels. The compiler must be aware of such significant changes in methods and the research requirements for data on specific forms of nutrients. The number of different definitions and relationships between nutrients can cause confusion unless there is careful definition of exactly what is intended. For example: "vitamin A activity" may include activity from beta-carotene [22], and "niacin" sometimes includes activity from tryptophan, while "fibre" and even "carbohydrate" have a number of alternative definitions. The data compiler must be very careful to determine exactly what to include, and define this in unambiguous terms. The problem is compounded by the fact that many nutrients are reported in several different forms which have different meanings (e.g., it is not possible to unambiguously calculate retinol equivalents for vitamin A from International Units), while others are merely different conventions (e.g., the conversion from kilocalories to joules involves only multiplication). (See the INFOODS guidelines on these topics [26, 93] for further discussion.)

Adequacy of analytic methods. Adequate analytic methods are required to obtain accurate and precise data; however, the current state of food analysis is, overall, very uneven. For some nutrient and food combinations, accurate and precise methods exist, while for other combinations, methods are non-specific, inaccurate, tedious, or expensive [8, 83, 86]. Moreover, good methods used carelessly without adequate quality control will generate poor data. Even when appropriate analytic methods are used on the same sample in different laboratories, there may be a wide spread in values [35]. Foods which are main sources of nutrients in a region should be analysed, and the methods used should be subject to quality control. The compiler must often choose to include or exclude certain nutrient values on the basis of whether adequate analytic methods exist or were employed.

Specific Form of the Data for Inclusion

Modes of expression. Nutrient values are frequently expressed per 100 grams of edible portion or per "common household measure". If household measures are used, the gram weight of the measure should be included. For specialized data bases other modes are often used, such as grams of amino acid per gram of protein.

Accuracy and precision of the data. The accuracy of nutrient measurements is a function both of the representativeness of the sample to the population of given food items and of the accuracy and precision of the analytic technique employed. Precision is primarily a function of how many samples are taken and how carefully (and which) analyses are conduced. Both accuracy and precision are directly related to cost and effort expended in design and execution of the sampling plan (for accuracy) and in numbers of replicates (for precision). The compiler must make the decision as to how accurate and precise the data should be on the basis of the uses to be made of them. For example, data to support analysis of large consumption surveys intended to estimate mean intakes require less precision than data to support analysis of individual diets where intake extremes are of interest.

Numbers for Inclusion

The numbers (data) included in a data base depend on the needs of the user. At a minimum, for each nutrient and food, there should be some measure of average value, some measure of the variability of the data, and some indication of the number of sample points upon which these statistics were based and how these data points were manipulated to arrive at the given statistics. The amount of a specific nutrient in a specific kind of food is variable, and the task is essentially one of summarizing the potentially variable data, i.e., the statistical distribution of the values. This topic is more fully discussed in chapter 3.

Special Conventions

In the planning stage, it is important to choose conventions of data recording that are consistent with the desired data base. Two important conventions are the number of decimal places to be used for each nutrient and the distinctions between and notations used to represent zero levels of a nutrient, trace levels of a nutrient, and nutrient data that are missing or not available.

The number of significant figures reported for the level of a nutrient should not mislead the user into believing that the data are more precise than they actually are.

Thus, the decision on the number of significant figures to be retained and displayed involves the consideration of the variability of the data themselves. In general, a change in the last significant digit of a statistic, such as the average level, should be of the same order of magnitude as the standard error of that statistic.

There are very important differences between data that are missing, values that are very small, and values that are zero. The data base must clearly distinguish between these situations, and the differences must be preserved as the data are collected.

- The term MISSING should be used when analytic data are not available and there is reason to believe that some amount of the substance is present or when there is no relevant basis on which to predict whether or not it is present. If a value was estimated, the value should be reported with appropriate history, rather than either being identified as missing or treated as an analytic value.
- The term TRACE indicates that chemical analysis (or an equivalent process) indicates a value (i.e., high confidence that the component is present), but that value is below the measurement accuracy of the method. TRACE is sometimes used when analysis

indicates a measured and interpretable value that is nonetheless believed to be nutritionally insignificant.
- ZERO is used when actual measurements, or estimates based on measurements, were performed and detected nothing. In particular, if available qualitative methods reliably indicate the presence or absence of the food component, regardless of its quantity (at least within rational limits), and those methods indicate that the component is not present, then the value should be considered ZERO. It should be recognized, however, that even ZERO is not immutable as ideas and methods evolve.

For some purposes, it is possible to effect a further collapse into two categories, ZERO and MISSING, by imputing a value of ZERO to TRACE. This should be done only where the distinction is unambiguously unimportant. MISSING and ZERO, however, must always be kept distinct, with the numeral "0" never used to represent MISSING.

The above distinctions do not suggest specific symbols to be used in food tables and data bases. The usual alternatives are alphabetic and numeric codes. An argument often made against the use of alphabetic characters is that they may make computer processing of the data more complicated. However, with the current computer technology, this should not be a major concern. The obvious alternative strategy is the use of negative, or large positive, integers. This has two serious disadvantages: scaling procedures to convert units based on a global default may distort these values, and the user of numeric values to encode these indications may accidentally use them as levels, for example, by including them in an average level.

Ancillary Information

Food composition data are used in conjunction with a variety of other types of data. To enhance its utility, a food composition table will often contain sets of ancillary information, as separate tables or text. For example:

- The British tables [59] include standardized recipes.
- The Indian tables [25] include a large section on dietary principles and practice and an appendix giving the names of foodstuffs in various Indian languages (the text is in English).
- The Chinese tables [16] include a table on the effect of cooking on vitamin losses and an appendix on daily dietary allowances.
- The German tables [80] provide data on digestibility factors.
- In the United States, the USDA publishes, separately from its data tables, food yields for different stages of preparation [49] and nutrient retention factors [97]. Additionally, Pennington [62] includes tables of recommended dietary allowances.

A data base compiler will have to decide what, if any, ancillary information to include, and then plan the gathering and integration of this information so that it will be most useful to the users. An important aspect is organizing the data files so that there is natural correspondence between the ancillary data and the actual composition data (e.g., the same units, the same names, etc.), with sufficient cross-referencing.

Information on Quality

Information on the sources of data does not necessarily provide information on the quality of those data. It has been proposed that food composition data bases should include a code indicating the "quality" or "confidence" of each data point generated by the compiler. Although data quality is very difficult to define, research has been conducted on the form and utility of quality codes for food composition data (e.g., references 21 and 81).

Work in this area is difficult for two fundamental reasons. First, quality is not one dimensional. It should at least include aspects of accuracy, precision, and representativeness. Second, reporting data quality is sometimes dependent on the use to be made of the data. In particular, the question of the representativeness of data can only be evaluated in the context of "representative of what?" and therefore differs from application to application.

Work done thus far suggests that it is possible to derive quality codes in certain situations. Analytic data can be ranked according to method used, sample handling, and quality control. Values derived, directly or indirectly, from analytic values can also be ranked, but additional, less objective, factors must be considered:

- the validity of methods used to combine data (such as mean values of multiple species of a food item, or data from multiple regions);
- the validity and availability of adjustment factors for yields, retention's, refuse, and ingredient proportions; and
- the similarity of matched foods when data are estimated by substitution.

It is important that the criteria for evaluation be clearly presented so that the user can assess the data included in the computation of the ranking. For examples of tables of nutrient values which provide confidence codes see Schubert et al. [81] for selenium and Exler [21] for iron. The ultimate utility of these rankings (as translated into confidence or quality codes) is obvious; for example, a user analysing diets could Bag nutrient totals that included data with confidence codes below a selected limit. However, much work is needed before such codes can be widely used.

Format

A vital factor in food composition data bases is arrangement of the data for easy access. The presentation of food composition data is necessarily a function of how the data will be used.

In REFERENCE data bases (see page 6), all the data and information are preserved, although they are often not readily accessible. Usually a subset of these data bases is available for distribution, with the additional data and information (replicate observations, exact sources of data, etc.) remaining in the files and available on special request.

In SPECIAL-PURPOSE data bases (see page 7), the data are often presented with only pertinent information attached; the very nature of the application obviates the need for complete information. Such a data base is sometimes embedded in a computer program or system, but often exists only in a printed version, with an abbreviated introduction. Of paramount importance in SPECIAL-PURPOSE data bases are references to the location of more complete information.

Most food composition data bases are presently available as printed documents. This medium permits a variety of organizational plans and a very flexible range of information that can be included. The usual organization includes an extensive introduction that describes and discusses the contents, the major table with varying amounts of annotation, additional tables of nutrients for which limited data exist, and various indexes and supplementary tables.

Food composition data bases which are made available on electronic media (currently tapes or diskettes) tend to have less information than do printed tables, due, in part, to the perceived restriction of having the data arranged linearly, making various forms of annotation (e.g., footnotes) awkward. Partly in response to this, INFOODS developed its interchange scheme [44], which captures all the available information.

Printed tables tend to arrange the composition data in matrix fashion with the rows indicating the foods, the columns indicating the nutrients, and the cells (the intersection of a food and a nutrient) containing the numeric data. This format leads to difficulties when the number of nutrients increases to more than can be easily contained on a single page, or when there are nutrients for which there are few data. Consequently, other tables (e.g., those from Germany [79, 80] and the United States [96]) present one page per food, with nutrients in rows and attributes (e.g., mean value, number of samples, variances) in columns, thus providing easily located data with more analytic detail possible.

Documentation

Careful and complete documentation of procedure is required throughout the compilation of a food composition data base. This information can inform the potential users of the extent and limitations of a given data base, enabling them to assess its suitability, understand its optimal use, and identify areas for addition or improvement. The utility of any data base will be greatly enhanced by complete documentation.

The information accompanying food composition data should describe the foods and nutrients included, define the precise meaning of the numbers and symbols used, and record how the data were obtained.

Description of Foods

Any food composition data base must contain or reference descriptions of the foods that are included. These descriptions must be sufficient for user identification of the entries. It is important, especially for reference data bases, that there be as much information as possible, to permit potential users to identify foods and decide if they are pertinent. (It is recognized that complete identification is impossible; the descriptions of foods are culturally determined, and pathways for global exchange of this information have not yet been adequately tested.) Ideally, this descriptive information should include common names, other names, scientific name, (reference to) recipe, preservation method, food source, growing location, conditions of growing and use, the condition of the food when encountered, etc. (See Truswell et al. [93].)

Description of Nutrients

Complete description of the reported nutrients is essential so that a user can judge the appropriateness of the data base. This information must include sufficient detail (or references to such detail) to permit duplication of the analyses; thus, documentation of sample preparation, method, etc. is of paramount importance. Klensin et al. [44] provide specific terminology and guidelines in this area.

Description of the Data

The data should be completely described, in terms of origin and manipulation (including the steps in that manipulation). Aggregated data, and data from composite foods, should be described in terms of what they represent. (See chapter 3.) Additionally, it is useful to include a description of the evolution of the data base, including why, how, and with what success the data base was developed.

Verification/Validation

The compiler should attempt to ensure the validity of the data base. Validity is a complex concept and ideally requires that:

- the numbers correctly represent the levels of nutrients in the foods analysed;
- the foods analysed are representative of the foods listed;
- the foods and nutrients are carefully and precisely identified;
- the data are identified as to their origin; and
- the data are not represented as being more precise than they really are.

It is important to review the data for consistency and to verify any numbers that appear incorrect. Data can be examined in several different ways to identify and verify numbers which are beyond expected ranges. While these checks will often find inaccurate analyses or non-representative samples of foods, they are particularly useful in identifying transcription errors and mistakes made in the units of expression. These validity checks fall into three classes:

- Checks between or across nutrients (within foods). These involve checking specific relationships between certain nutrients (e.g., the sum of carbohydrate components [starch, sugars, fibre] must not exceed total carbohydrate) as well as examining the absolute levels of nutrients (e.g., on a weight basis, the protein levels of meat, fish, and poultry rarely exceed 30%).
- Checks between or across foods (within nutrients). This verifies that "similar" foods have "similar" levels of nutrients (e.g., calcium levels of various milks should be similar).
- Checks of the whole data base. The third level of checks calculates the nutrient content of specific meals, diets, or both and compares the results to known values. A detailed procedure has been developed by Hoover and Perloff [38].

Data base validation is an area that needs further effort. Although the basic procedures are relatively simple and easily performed by computers, there is no consensus on approaches.

Maintenance

The maintenance of a food composition data base consists, at minimum, of correcting mistakes. Additionally, there may be a need to modify data or add or delete foods. Note that any new data or changes in the data base should be checked for validity as indicated above. An important aspect of maintenance is to note and justify any change to a data base, and to carefully identify the different editions of the data base. Users should identify a particular edition of a data base that is used, so that possible or necessary corrections can be made when a newer edition is available. Several papers are available concerning the maintenance and management of food composition data base systems [11,14, 30, 43, 65].

Part II Gathering the data

There are five categories of procedures for obtaining the numerical data for a food composition data base. These are discussed in detail in chapters 2 through 6.

Analysing foods. Given ample resources, the preferred method of obtaining data for a food composition data base is to analyse,or have analysed, samples of the particular foods for the desired nutrients. Chapter 2 contains an outline of the considerations involved in analysing food samples, and a listing of the subsequent calculations involved in determining nutrient levels. A much more extensive discussion and guide is provided by Greenfield and Southgate [26] and earlier by Southgate [84].

Calculating representative data. Multiple data points representing a single nutrient in a food must be combined into specific data entries. Chapter 3 describes the considerations and techniques used to calculate representative values.

Data from other sources. If satisfactory data already exist for the foods and nutrients of interest--in the literature, or in the files or data bases of others--these should be used. Chapter 4 discusses the concerns involved in searching for and evaluating these data.

Estimation from data on similar foods. If data cannot be obtained by analysis, or satisfactorily found elsewhere, it may be possible to estimate the needed data on the basis of data for a similar food. This can be a difficult procedure and must be done with great care. Chapter 5 discusses the considerations involved in these situations.

Calculations for multi-ingredient foods. Foods which contain more than a single ingredient (sometimes called recipe foods or mixed dishes) require procedures for estimation of the nutrient content of the dish or final product based on the composition of the ingredients. As in the estimation of data from similar foods, such procedures can be complex and inexact and should be used with care. They are discussed in chapter 6.

These five categories cover the basic techniques used to produce the numbers for food composition data bases. If none of these approaches is appropriate or satisfactory, then the data must be left as "missing" and carefully labelled as such. It is especially important that missing data never be represented by zero.

In general, the preferred method for data acquisition is by analysis, while estimation of data by analogy with similar foods or from the composition of ingredients is not preferred; however, a nutrient level estimated from the contributions of several ingredients may well be better than an analysis that was poorly done. Moreover, the term "quality" is not necessarily equivalent to "accuracy", but must include some aspect of "suitability for purpose". (See page 12.)

2. Analysing foods

Given unlimited resources, direct analysis of foods is the preferred method of obtaining food composition data for specific foods of interest. However, analysis has the major drawbacks of time and expense involved. Anyone contemplating this approach must be aware of the magnitude and complexity of such an effort and the practical experience and assistance required to do it properly. There are many references on this topic; the most relevant are the AOAC manuals [29, 104] and the INFOODS manual *Guidelines for the Production, Management, and Use of Food Composition Data [26]. Others are Stewart and Whitaker's Modem Methods of Food Analysis [87], Southgate's Guide Lines for the Preparation of Tables of Food Composition [84], Oiso and Yamaguchi's Manual for Food Composition Analysis [57], Aurand et al., Food Composition and Analysis [7], Jacob's The Chemical Analysis of Foods and Food Products [40], and Pearson's The Chemical Analysis of Foods* [60].

Generating analytic data for use in a food composition data base consists of several logically distinct steps: sampling foods representative of the population of foods of interest, preparing the samples and performing the analyses, and recording and preparing the final data. Each of these steps needs careful planning, documentation, and execution, whether done within the data base compiler's own laboratory or in another. A detailed plan of work must be developed by the compiler with the assistance of experts in the various areas and with the main users of the table. In addition, there must be frequent review and discussion between the compiler and the analytic laboratory to ensure that there is no disparity between the information that is wanted and the information that is provided.

Obtaining foods

A sample is defined (statistically) as one or more items taken from a population in a way to ensure that they are "representative" of that population. In food composition work, the population of interest would be a type of food; sampling describes how examples of that food are collected for analysis. Definition of the food must be complete enough to determine whether the aspects of brand names, cultivars, growing region, preservation, form, etc., should be considered in drawing the sample. (See the discussion in chapter 5 on how foods differ.) After defining the food, it is then necessary to collect a sample of food from across the spectrum of those factors. The number of individual items to be drawn is a function of the precision which is desired. The care with which the sample is taken, in terms of how well it represents the population of interest, is a major factor in the accuracy of the data. Additional discussion of extensive sampling protocols may be found in Greenfield and Southgate [26] and Holden et al. [34]. Some sampling plans are specified by national regulations; for example, the sampling plan required by the US Food and Drug Administration for nutrition labelling is found in reference 56.

Adequacy of analytical methodology

Choice of method is perhaps the most important aspect of analysing foods, although a good method will not compensate for a bad sampling plan. With the latter, one measures the wrong foods very well. The following represents only an outline of the considerations in choosing an analytic method. The INFOODS manual *Guidelines for the Production, Management, and Use of Food Composition Data* [26] discusses this topic in detail.

The methods chosen must be generally ACCEPTED, RELIABLE, and PRACTICAL. Even some methods that have been generally accepted in the past are now understood to produce inaccurate and imprecise data (e.g., calorimetric analysis for cholesterol). These methods should be avoided despite their past popularity. Reliability includes such attributes as accuracy, precision, specificity, and detectability (sometimes, although incorrectly, called sensitivity).

The accuracy and precision of an analytic method must be addressed separately. (See page 10.) The accuracy of a method should be verified by use of standardized or certified reference materials if these are available. Day-to-day accuracy must be monitored by analysis of quality control material. It is important that variability due to the method be minimized so that food composition data users can assess the variability of the nutrients in the food. Since these two sources of variation cannot be separated by examining a single value, or even a set of values, analysts should publish estimates of analytical error, and those estimates should be available when evaluating data.

The specificity of a method refers to its ability to measure only the specific nutrient for which it is being used, while the detectability, or limit of detection, of a method is how little of the nutrient need be present before it can be measured; both of these aspects should be considered in terms of the uses that will be made of the data.

The practical aspects of analytic methods include costs (space, equipment, consumable supplies, personnel time, and training), dependability, and safety. These tend to be country- and laboratory-specific, and often very real determinants in the choice of methods.

Calculations to derive nutrients

Because of the difficulty and cost of directly measuring some nutrients, many data base nutrient values are calculated from other analytic measurements. This procedure is used by virtually all compilers for some nutrients (such as energy and protein) and by some compilers for various other nutrients. Some of the calculations most commonly performed are briefly described below. Fuller details are contained in Greenfield and Southgate [26]. One of the complexities of discussing nutrients is that often the same name for different components is used in different food composition data bases (see below). An INFOODS listing of the "nutritionally distinct" compounds appears in *Identification of Food Components for INFOODS Data Interchange* [44].

Carbohydrate

The term carbohydrate, as used in food composition data bases, includes a number of different carbohydrate species. While for many purposes measurement of total carbohydrate is not useful, it commonly appears in tables and is used to calculate energy values. However, there are several different ways that total carbohydrate is calculated: the two basic methods are "by difference" and "by sum". Additionally, tabulated values are either "available" or "total" carbohydrate, depending on whether dietary fibre has been included. Difficulty arises when these different calculations are indistinguishably labelled as carbohydrate.

In the United States, values for carbohydrate are usually calculated "by difference". For each food, the value for carbohydrate is estimated as follows (all values are per 100 g):

Carbohydrate = 100 g - water - protein - fat - ash - alcohol

This method includes indigestible carbohydrate in the carbohydrate component, as well as any other components of a food item that are not measured as water, protein, fat, ash, or alcohol. This method yields only an approximation of carbohydrate.

Other data bases use carbohydrates "by sum", where individual carbohydrate components (sugars, starches, oligosaccharides, and dietary fibre) are measured and summed. There are various ways of doing this. (See Paul and Southgate [59] for a description of the British procedure for measuring "available" carbohydrate and expressing it in terms of monosaccharides.) Carbohydrate values calculated by these different methods may not be comparable.

EXAMPLE: Carbohydrate content of raw carrots (all values are per 100 g):

- By difference (data from the United States [96]):

100 g carrots - 87.8 g water - 1.0 g protein - 0.2 g fat - 0.9 g ash = 10.1 g carbohydrate.

- Available carbohydrate by difference (data from the United States [96, 98]):

100 g carrots - 87.8 g water - 1.0 g protein - 0.2 g fat - 0.9 g ash - 3.2 g dietary fibre = 6.9 g carbohydrate.

- Measured as available carbohydrate (and expressed as monosaccharides) (from Paul and Southgate [59]):

5.4 grams carbohydrate.

Protein

Since it is relatively easy to measure the nitrogen content of a food (by the Kjeldahl method) and fairly difficult to measure protein, it is usual to use nitrogen values as a basis for reported protein values, converting to protein either by a general factor of 6.25 g protein per gram of nitrogen or by more specific factors for different food types (e.g., Jones [41] and the German tables [80]). Since the amount of nitrogen varies among amino acids, and amino acid composition varies among foods (there is also nonprotein nitrogen in many foods), the food-specific factors will give a more precise estimate of the protein content; however, this is still an approximation.

EXAMPLE: Protein content of almonds (data from China [16]):

- Using the general factor:
  6.25 x 4.06 g nitrogen = 25.4 g protein per 100 g edible portion.
- Using the specific factor for almonds (5.18):
  5.18 x 4.06 g nitrogen = 21.1 g protein per 100 g edible portion.

In a daily food pattern over- and under-estimations due to using a general factor will tend to level out. Even with vegetarian diets a general factor versus specific factors differs less than 5% [101].

Energy

Metabolizable energy values are usually calculated, since the standard method for measuring the energy content of a food (bomb calorimetry) does not reflect the energy actually available to the body. The Atwater system [6, 51] uses factors to estimate available energy from the protein, fat, carbohydrate, and alcohol components of food items. Errors in the analyses of the four components, or in the determination of the factors, will be reflected as errors in the energy content. The general Atwater factors are used by some data base developers:

Energy (kcal) = (4 kcal/g x g protein) + (4 kcal/g x g carbohydrate) + (9 kcal/g x g fat) + (7 kcal/g x g alcohol)

The specific Atwater factors reflect variations among food groups and should provide more accurate estimates of the true available energy. These specific energy factors are now commonly used to calculate energy values. A comparison of the effect of the two methods on the energy value of several foods can be found in Merrill and Watt [51]. In the USDA tables [96], this is documented by including a listing of the factors that have been used for the individual foods in each section (food group).

Different methods of determining any of the energy components can clearly affect the value given for the total energy content of a food item. For example, the British food tables [59] express carbohydrate as available monosaccharides. The developers of these tables chose to use the general factors of 4 kcal/g for protein, 9 kcal/g for fat, and 7 kcal/g for alcohol, but to use 3.75 kcal/g as the factor for carbohydrate since there is less energy in the monosaccharide forms. (See the British tables for further discussion.)

Recently, the Atwater system was reviewed by the Federation of American Societies for Experimental Biology [3], who concluded "the Atwater system provides estimates of metabolizable energy within the limits of accuracy for measuring food intake and also within the predictive limits of food composition tables". A general review of the methods used in different countries to calculate energy values can be found in Perisse [63].

EXAMPLE: Energy content of raw apples:

- Using general factors (4 kcal/g protein, 4 kcal/g fat, 4 kcal/g carbohydrate) and data from the United States [96]:
  4 x 0.19 g protein + 9 x 0.36 g fat + 4 x 15.25 g carbohydrate = 65 kcal per 100 g.
- Using specific USDA factors (3.36 kcal/g protein, 8.37 kcal/g fat, 3.60 kcal/g carbohydrate) and data from the United States [96]:
  3.36 x 0.19 g protein + 8.37 x 0.36 g fat + 3.6 x 15.25 g carbohydrate = 59 kcal per 100 g.
- Using available carbohydrate and general factors and British data 159] (4 kcal/g protein, 9 kcal/g fat, and 3.75 kcal/g carbohydrate):
  4 x 0.3 g protein + 9 x 0.0 g fat + 3.75 x 11.9 g available carbohydrate = 46 kcal per 100 g.

An additional complexity with energy values is that they are alternatively expressed in either kilojoules or kilocalories. Conversion from one to the other may be performed either on the final value (4.184 kJ per kcal) or on the individual components (17 kJ per g protein, 37 kJ per g fat, 16 kJ per g carbohydrate, and 29 kJ per g alcohol). It should be recognized that, because of rounding, these two methods sometimes give slightly different values.

EXAMPLE: Energy value of dark beer in kilojoules, using data from the German tables [80]:

- Conversion:
  Energy = 4.184 kJ/kcal x 37.3 kcal = 156 kJ per 100 g.
- Summation:
  Energy = 5.9 kJ for protein + 47 kJ for carbohydrates + 101.5 kJ for alcohol = 154 kJ per 100 g.

Dietary Fibre

Total dietary fibre can be determined directly or calculated from its various fractions. Often these fractions are also included in the data base (e.g., insoluble fibre or neutral detergent fibre, and soluble fibre [pectins, gums, soluble hemicelluloses]), since these fractions have known physiological importance. Alternatively, one of the fractions may be calculated as the difference between total fibre and the other fraction (e.g., soluble fibre as the difference between total and insoluble fibre). The multiple definitions and methods for determining dietary fibre values analytically make these types of calculations complex, and care must be used to ensure they are done correctly. If dietary fibre is of interest to the compiler of a food composition data base, a food chemist with expertise in this area should be consulted. See Lanza and Butrum [45], Stasse-Wolthuis [85], and Trowell et al. [92] for reviews of the complexities concerning dietary fibre. A recent collection of data from the United States appears in reference 98. Dietary fibre values have been available for many years in the British [59], German [79, 80], and other tables.

There is no accepted method for calculating dietary fibre from crude fibre. The ratios of crude to dietary fibre are highly variable among food items, and even among diets.

Vitamin A

Vitamin A is an example of a nutrient that has multiple chemical forms, each of which has differing biological activity. In such a case it is often necessary to assign different availability factors to these different forms in order to calculate the total nutrient value. Commonly, total vitamin A activity is expressed in retinol equivalents (RE) as a function of retinol, beta-carotene, and other carotenoids. The recommended factors are one for retinol, one-sixth for beta-carotene, and one-twelfth for other carotenoids.

Vitamin A activity (RE) = µg retinol + 1/6 µg beta-carotene + 1/12 µg other carotenoids

Some data bases contain the individual components which permit calculation of the total value directly, while others contain only the total (with indication of the factors used). The specific values desired must be chosen by the compiler on the basis of the purposes for which the data base will be used.

Some food composition tables continue to report vitamin A in international units (IUs). This system does not adjust for absorption of the different forms of vitamin A and cannot be directly converted to RE for many foods. Since requirements are now expressed in retinol equivalents, use of the older units is no longer appropriate if the data base will be used to estimate dietary adequacy.

Vitamin E

Vitamin E is another example of a nutrient with a number of different forms (the tocopherols and tocotrienols), each with different biological activity. Here the choice is to measure and present the major forms of interest, to calculate and present a total vitamin E activity in "alpha-tocopherol equivalents", or both. Unfortunately, it is not possible to calculate alpha-tocopherol equivalents accurately if the mix of isomers is unknown, and further, the validity of these factors for all types of food is not known. In mixed diets, beta-tocopherol has approximately one-half the activity of alpha-tocopherol, while gamma-tocopherol has only one-tenth the activity. The 1989 edition of the US Recommended Dietary Allowances [54] provides additional details.

Niacin

Analytic values for the niacin content of a food item are usually expressed as mg of preformed niacin. However, dietary guidance is usually offered in terms of mg of niacin equivalents, since some tryptophan (an essential amino acid) can be converted to niacin. The amount that can be converted depends on the total protein and energy in the diet, on the mix of amino acids in the diet, and on the nutritional status of the individual. However, a figure of 1 mg niacin per 60 mg tryptophan is considered average, and some data base compilers calculate niacin equivalents as the sum of preformed niacin and onesixtieth of the tryptophan in a food item.

EXAMPLE: Niacin equivalents in dinner rolls made with baking powder:

Niacin equivalents = 2.7 mg preformed niacin + 1/60 x 90 mg tryptophan = 4.2 mg per l00 g.

If no value for tryptophan is available, it is sometimes assumed that approximately one per cent of protein is tryptophan [58].

3. Calculating representative data

Often a data compiler has several data points which must be summarized into a single value for the data base: multiple values representing a single nutrient in a single food. This arises when one has analyses for each of a number of different samples of the same food; when one has data from several sources, each described same food; and when one must combine data on different food items into an entry for a "generic" food (e.g., one needs an entry for apples and has individual data on each of several different varieties).

Multiple measurements provide valuable information both on the level of a nutrient in a food and on the variability of that nutrient in different samples of the food. While it is important for the results of the individual analyses to be maintained in some reference data base, for many purposes summary statistics of these data will be required. Whatever statistics are presented, it is important to realize that the numbers included in the data base are not directly analytic but are produced from the analytic data by certain manipulations, and that these manipulations must be carefully and completely documented.

Preliminary examination of the data

Before any calculations are made, it is important to examine the data to see whether they can be considered to represent a single biological population. There are two pertinent questions: Are there errors in the data? Could the data have come from two or more distinct foods? In addition, an inspection of the data should give insight as to what sort of a statistical distribution might best describe the sample (could it be symmetric about some mean value or perhaps skewed?). While these questions are very difficult to answer with only a moderate sample size, they are very important.

Errors

It is possible that some data points may have resulted problems of sample collection, laboratory procedure, or transcription. Standard statistical tests exist for single, potentially errant values [27]. While these tests tend to be quite conservative (retaining data when in doubt), it is recommended that the compiler follow them, leaving suspected data points in a set unless they can be statistically rejected or independent evidence can be found to explain their deviation.

Two or more populations

There are two basic situations in which the question of underlying multiple populations arises: when examination of the data suggests that there are two or more populations represented in data that are all identified as being the same food, and when the compiler knows, from external evidence, that there are two or more populations which must be aggregated.

Apparent Multiple Populations

In examining data which all have the same identification, it often appears that there are two (or more) distinct populations. No statistical or graphical procedure that examines the data from a single collection can ever completely assure the compiler that there are, or are not, multiple populations, and sophisticated judgement must ultimately be combined with whatever procedures are used. If there are multiple populations, the decision as to whether or not to split them may depend on such external considerations as project objectives as well as on the data. Separating multiple populations is normally the better choice in reference data bases because it is feasible to combine them later, while separation of aggregated summary values is typically impossible.

Statistical determination of whether multiple populations are present in a data set, and where the boundary between them falls, is an active area of statistical research. In general, the data should be examined visually or by preliminary statistical tests. Suitable graphical procedures are discussed by several authors [15, 17, 94]. It is often helpful to use known parameters such as cooking method, cut of meat, preservation state (frozen vs. fresh), part of plant, and other relevant variables for graphing. If distinct populations are suspected as a result of these examinations, the circumstances of data collection and the possible nutritional significance of distinct subsets should be investigated as part of deciding whether to assume a single population or to begin the often tedious search to identify the subsets and their boundaries. Without clear-cut guidance from the data themselves, the compiler should work with the data as if they were derived from a single population of food.

Multiple Different Populations

A data base compiler is sometimes confronted with two (or more) different sets of data (e.g., from different sources) which may or may not represent the same food. If the variables under consideration have reasonably normal (Gaussian) distributions, this is a straightforward statistical question which can be answered by use of Student's l-test, or Hotelling's T-squared test, or some form of analysis of variance [2, 82]. It is recommended that the results of these tests be evaluated through visual inspection of the data sets, with the final decision resting with food specialists rather than with statisticians.

Data Aggregation

Often, it is necessary to aggregate data for several different brands or varieties of a food in order to produce a single entry in a data base, perhaps to produce data for a "generic" food. The standard procedure is to calculate a "weighted" average, a procedure which permits the component data sets to have differing importance in determining the ultimate result. Thus, if one had two kinds of potatoes and wanted an average value that could be used with results from a consumption survey where the respondents did not know which type of potato they consumed, one would want the data to reflect the probability of consuming one type over the other. This can be resolved by 'weighting" the data by the consumption pattern or the market share.

The question of variability is more complex, with either a "pooled" or an "overall" variability to be calculated. The question of exactly what to do depends fundamentally on the ultimate purpose of the data base, and at the present time no general recommendation can be made except to carefully document what has been done.

Shape of the Data Distribution

Given that a single population can be assumed, it is very useful to characterize the shape of the distribution. Here the compiler must decide if the data seem to have come from a standard normal (Gaussian) or bell-shaped curve, or if another shape is underlying it (usually skewness). This is an important consideration in the choice of descriptive statistics. While there are statistical tests for normality, these are not powerful when used with small to moderate data sets. Until further research is done into the general questions of distribution of nutrients, data sets which do not look obviously non-normal should be considered normally distributed, but considerable caution should be exercised.

Summary statistics

The aspects of a data set that are commonly reported in a summary of the data are the number of samples, the "central" value of the data, and the variability of the data. The actual statistics which best summarize these aspects of the data currently represent an open area of food composition data research, an area which is only now beginning to be explored. The recommendations given below represent suggestions which are quite general and should be expected to be revised and expanded as more effort is focused on this area. These statistics, or equivalent distribution information, are most important for reference data bases and ideally should be supplied with any data base from which a user may need to reinterpret, rather than simply use, the data values.

Number of Samples

Every cell of a data base should have an assigned value that represents the number of independent observations which have gone into the calculation of the summary statistics.

This is straightforward if the compiler has data that represent a single population. In cases where the compiler is working with data that have already been summarized (e.g., means for several different samples), assigning values to cells becomes complicated. Because of the variety of different possible situations, each must be handled separately, and the end users must be provided with access to the details. It is recommended that the sample size indicated in the data base should reflect the total number of data points contributing to the statistics, and that information should be included to describe what this number actually represents.

Central Value

If the compiler has only a few data points, or the distribution appears to be "not different" from the normal (Gaussian), the single "representative" number used to denote the level of a nutrient in a single food should be the arithmetic mean (the usual average). If the distribution is skewed, then the median (the value that lies in the middle of the sample) or the geometric mean (assuming a log-normal distribution) should be presented, and this should be explicitly noted. It is rare that there will be sufficient data points to empirically determine the shape of the distribution with any confidence. When the data come from several distinct sources (laboratories or cultivars), the average of these averages should be calculated. This may be done by assuming that each data source is equivalent to the others, or by weighting the component averages. Weights are often assigned on the basis of the sample sizes, thus assuming that sources with more samples are better than those with fewer samples. Alternatively, a relative contribution, such as one based on market availability, can be used to determine weights.

Variability

Many food composition data users require information on nutrient variability. This is useful in a general sense to know how variable a nutrient source is and in a specific situation to know how extreme a value could be. For example, one might wish to estimate how much of a food must be provided to ensure that a certain nutrient level would likely be achieved in a hospital feeding programme. In the case of a known normal distribution, variability can be summarized by the standard deviation (together with the mean). However, this is often inappropriate when dealing with nutrients and foods that have not been well studied. Until such research is carried out, consideration should be given to providing estimates of extreme percentiles. If there are too few sample points available to compute a reliable standard deviation, or if a skewed distribution is suspected, mid-spread or fourth-spread, or the closely related interquartile range [32] should be considered, as should the alternative, used in the Chinese tables [16], of simply listing the data points. As estimators of population spread, these alternatives are less sensitive to extreme points than the range of a nutrient, especially when there are a small number of samples.

The calculation of standard deviation is straightforward if the individual data points are available. If a compiler is working with individual component means and standard deviations, then the appropriate action depends on both what is available and what is wanted. As in the case where all the individual data points are considered to provide equal information (all samples from the same population are equally important), a "pooled" standard error should be calculated. If standard deviations are available for samples which are essentially different (as in the case of differing samples for a generic food), then both a "between" and a "within" standard deviation should be calculated.

4. Data from other sources

Obtaining data from other sources, sometimes called "borrowing" data, refers to using data originally generated or gathered by someone else. This is the most frequent way of obtaining data for many special-purpose data bases, with the usual sources being the large reference data bases (such as those of the USDA [96], the United Kingdom [59], etc.). One problem with the data of others is that they are often incompletely described. However, borrowing of data is not only justified but essential when analyses are impractical (i.e., allocation of resources is not justified) and "good" data are available elsewhere. See Jacob [39] for an overview of this problem from the point of view of some social scientists.

Given the decision that certain data are needed and cannot be generated de novo, the two basic tasks facing the compiler are finding and evaluating the data.

Where to find food composition data

Potential sources of nutrient data include published food composition tables and data bases, journal articles, books and book chapters, proceedings from meetings, project reports, unpublished laboratory reports, university theses, reports from food producers, and materials produced for consumers. Data are generated and compiled by international and national agencies, academic groups, clinical/medical groups, food industries, trade associations, and private groups, usually for specific studies relating to the nutritional properties of foods or as part of efforts to create food composition data bases, either reference or special-purpose. (See Schakel et al. [78].)

International agencies, such as the FAO,collect food composition data in order to judge the quality of international food trade, and the potential needs and resources of various areas of the world. Government agencies conduct food composition analyses for regulatory purposes (e.g., to determine the nutritional quality of foods and to determine compliance with label claims), to develop data bases for use in dietary surveys evaluating the dietary status of population groups, and frequently to provide food composition data to professional groups and consumers. Academic groups generate food composition data as part of research efforts in food science and nutrition, while clinical/medical groups generate food composition data for analyses of patient diets and for nutrition intervention studies. The food industry analyses its products to determine their nutritional quality and to develop nutritional labelling claims. Trade associations generate food composition data to provide information for constituents and consumers. Private laboratories generate food composition data under commission for other groups.

Schakel et al. [78] provide a detailed listing of various sources of food composition data. *The Nutrient Data Bank Directory [37], Computer Programs and Databases in the Field of Nutrition* [31], and the INFOODS *International Directory of Food Composition Tables* [28] list many of the major tables and data bases in existence, and provide first steps for finding already compiled data. Journal, book, and proceedings references may be researched through computer retrieval, using keywords for titles, to assist in locating appropriate references. Food industries and trade organizations can be contacted directly, by post or by telephone. Listings of these industries and organizations (e.g., in the United States: *the Thomas Register* [90], *the Million Dollar Directory* [18], and the *Trade Names Dictionary* [106]) may be found in some libraries. It is more difficult to obtain project reports and unpublished laboratory data. Contact with individuals known to be involved with such work, as well as with government agencies, may give access to some of this information. INFOODS and its various regional liaison groups are also good potential sources of such information.

In some countries certain information about nutrients can be found on food labels, provided by food packagers and manufacturers. Nutrition information on food labels is generally under the regulation of a government agency seeking to ensure the safety and quality of the food supply. National laws may indicate that some label information is mandatory and some is voluntary, with the form and content of the mandatory information often being quite rigidly defined. As a result, the specific information provided by a food label usually varies both between and within countries.

In dealing with label information, manufacturers should be contacted to determine the specifics of the information, for example, if the values are based on analyses done by the manufacturer or if they are derived from a data base. It should be noted that label data for multi-ingredient foods may be calculated by the manufacturer from data base values of ingredients, and that label values may represent upper or lower estimates or be variously manipulated to be in compliance with national laws. Original data should be requested from the manufacturer with as much documentation as possible. If original data cannot be obtained, label data may be used as a "last resort" and subsequently updated or replaced when "better" data become available.

Label information is especially useful for identifying foods, describing foods, determining weight and other measures of serving sizes, identifying ingredients of multi-ingredient foods, and estimating nutrient values. Additionally, label information for multi-ingredient foods can sometimes be used to reconstruct recipes, and from this information the levels of nutrients not included on the label can be estimated. (See chapter 6.)

Evaluation of data from various sources

Once sources of relevant data are found, these data must be evaluated for suitability to the purpose at hand, and for their fit with the other data being assembled. Based on the discussion in chapter 1, the data must be examined according to the criteria of food identification, nutrient identification, data manipulation, and data quality.

Food Identification

A difficult task facing the food composition data base compiler is to determine if the data found in other sources are representative of the foods that are needed. While the detailed information required for such a determination is generally not available, it is still necessary to attempt to match the food for which data are being borrowed as closely as possible to the food required for the data base.

As an example, one may want to present nutrients for the food "fried calf. A search through the available sources may find nutrient values for four foods described as:

cooked beef liver
fried liver
calf liver, cooked
beef, yearling liver, fried

It is possible that these descriptions pertain to the same food or that, even though the foods are somewhat different, the nutrient values may be similar enough to warrant "borrowing" the data. This is a decision that ultimately the compiler must make on the basis of the projected use of the data base, although it is recommended that advice be sought from experts in individual food areas.

The effort to match food names encounters a plethora of problems due to the cultural importance of food. Foods and their names change from place to place, both between and within countries. The name "tortilla" is applied to a variety of foods throughout Latin America, while a variety of names is applied to carbonated beverages in the United States (e.g., cola, pop, tonic, soda). Cuts of meat vary around the world, as do parts of foods that are considered edible (e.g., tops and bottoms of beets and the rind of cheese). As one seeks to incorporate data from specific foods into a table it is essential that one know precisely what foods and parts of foods were analysed.

Nutrient Identification

When borrowing data from others, one must first determine the basis of the measurement, whether it is per purchased weight, per hundred edible grams, per wet or dry weight, etc. Some of these bases can be easily converted into one another (e.g., per pound to per kilogram), while others need special information (e.g., amino acids reported as milligrams of nitrogen per gram of total nitrogen can be converted to milligrams of nitrogen per 100 g edible portion only if total nitrogen content is known).

Beyond the basis of measurement are the forms and units of the nutrients. There are many nutrients which informally share the same name, yet are quite different for the chemist and are measured in different ways. Of special concern are dietary fibre, carbohydrate, protein, vitamin A, vitamin E, and niacin. (See chapter 3 and, for additional discussion, the INFOODS manual on nutrient identification [44].) The compiler of a food composition data base must completely specify the forms and units of the nutrients desired for the data base, and then carefully examine data obtained from other sources to ensure that they are expressed in, or can be converted to, the desired units.

Data Manipulation

There are two levels of data representation of concern to the food composition data base compiler: what the numbers that are presented represent, in terms of the distribution of the values for the nutrients that were originally measured, and how food samples that were analysed relate to the food described. In borrowing data from other sources, one must be sure that the reported data were manipulated in the same way as the data desired; for example, if means are to be used in the projected table, then means must be sought in other data.

To judge what the data represent in terms of a food, it is essential to go beyond the description of the food and explore the sampling method. Some information about the type of sample (e.g., locale, colour) may be inherent in the food name (e.g., Maryland steamed crab, green chili); however, information about the sampling scheme, number of samples, etc. may not be available from some data bases. Information of this type is generally more available from journal articles, books, proceedings, and unpublished reports. Such information is essential to determine the comparability of a food from a primary source to the food in the data base.

Data Quality

Perhaps the most important aspect of data borrowing, and the most difficult to judge, is that of "quality". This is discussed in chapter 1 in the context of determining the general quality of a data point. In the specific question of how to judge the quality of a borrowed data point, there are three factors involved: documentation, consistency, and reputation.

Documentation

Evaluation of food composition data relies on the information that accompanies those data; data unaccompanied by descriptive information (what they are and how they were gathered) are not, in general, useful. Documentation should include a description of the food, details of the nutrients and how they were determined, and description of the sampling and the data manipulation. (See chapter 1 for further discussion of these topics.) It is the presence and content of such documentation that provides the basis for evaluation of the data.

Consistency

Data obtained from other sources must be examined to determine how "reasonable" they are, in terms of both absolute and relative consistency. (See the discussion of validity checks in chapter 1.) Consistency checks are of three types:

- absolute checks (e.g., amounts of protein, carbohydrate, fat, alcohol, ash, and water should add up to the total weight of the food being analysed; calories from vegetables and fruits [except dried fruits] rarely exceed 150 kcal per 100 g;if all amino acids are analysed, they should add up to total protein [but note that protein estimated from nitrogen may overestimate total protein]);
- relative checks-i.e., nutrient data on the same foods from different sources should agree within expected variation;
- checks of overlapping data: If only a subset of the required nutrients are to be gathered from the external source, it is likely that other nutrients for the food item will be included in both the data base being developed and the external source. Thus, these nutrients in common can be compared to see if the external source food is really comparable to the food in the data base. For example, if moisture content is included in the external data, this value should be checked against the moisture content that is already in hand.

These consistency checks must be performed with an awareness of the inherent variability in foods, so unusual data are not automatically excluded.

Reputation

An important component of quality, and the most difficult to define, is the reputation of a food composition data source for performing analyses well and preparing and providing good data. If a data source is well known for quality work, then the values that it produces may be used with more confidence, even if they differ from what is expected. However, mistakes are possible even by the most reputable data sources. Questionable data should be brought to the attention of those who generated the data to rule out possible reporting errors (e.g., transcription or other inadvertent errors that could occur in preparing the data for publication). Data that are not from a known source must be examined more carefully and documented more rigorously. Unfortunately, many compilers are not well acquainted with many of the sources of data, which makes the assessment of reputation difficult.

5. Estimation from Data on Similar Foods

Data are often needed for a food or nutrient that cannot be generated or found in the literature. In some of these cases it is possible to estimate the needed data on the basis of data from a "similar" food. It may be a biologically similar food (e.g., a different variety of apple, or cabbage for brussels sprouts), or a different form of the same food (e.g., raw for cooked). The basic problem is the choice of a food which is "close" to the food of interest; this requires an intimate knowledge of the foods and nutrients, and experience in matching them.

Occasionally, a food is found which can be considered so similar to the food needed that data can be substituted directly, with no adjustment necessary for the nutrient values. More often, however, the two foods differ enough that some adjustment, or calculation, is involved in the substitution process. These adjustments attempt to correct the data for the differences between the substituted food and the food of interest, and it must be realized that the resultant data are approximate.

Estimating food composition data from data on similar foods must be done with care, and only when other options are not feasible. No hard and fast rules can be given for these estimating procedures, only general guidelines. Documentation must be provided to indicate which values have been estimated and how they were estimated. This chapter discusses the major factors involved in judging the similarities and differences between foods, the specific adjustments that can be made to correct for estimated differences, and the special situation in which it is appropriate to assume the total lack of a nutrient in a given food.

How foods differ

Knowledge of how foods differ is important for judging a food's suitability for substitution. The personnel involved in these decisions must have a detailed knowledge of food composition, and, in particular, the effect of various factors on the nutrients of interest. The following sections discuss some of the more important factors that can affect the nutrient composition of foods. Further discussion of this important topic, and additional references, can be found in the work of Harris and Karmas [42], Pennington[61], Bender [9], and Breen [10].

Biological Relationships

The genetic similarity of food items is perhaps most important in determining nutrient similarity, although genetic similarity does not guarantee nutrient similarity. Since common names of foods vary from place to place, many food composition tables list the scientific names of the food items. This is not a universal solution, since the relationship between common name and scientific name is not entirely consistent. Not only does the consumer sometimes use the same name for several different foods, but scientific names may not be helpful for complex dishes. Polacchi [67] discusses the complex situation with regard to fish. Moreover, scientific names are not universally unique. For example, the 1981/82 German tables [79] place pears and apples in the same genus while the British and US tables separate them. There is the additional problem that scientific names give little assistance with mixed dishes, which are a major component of diets worldwide. Finally, it is not a general truth that closely related foods are similar in nutrient content. There are extreme variabilities among cultivars (e.g., of rice and corn). (See also Meyer et al. [52].) Often a nutrient data base developer must rely on verbal descriptions of a new food item and make assumptions regarding its genetic similarity to known food items.

EXAMPLES: USDA [96] uses amino acid data from standard yams for the related Hawaiian mountain yam.

Goddard and Matthews [24] have reported on five sweet potato cultivars in North Carolina: the carotene content varied from 2,175 IU to 3,950 IU per 100 g, almost a twofold difference.

Crowing Conditions

For genetically similar foods, much of the observed nutrient variability is related to the conditions under which the foods are grown or raised. The condition of the soil, the amount of rainfall and sunlight, the altitude, and the average and extreme temperatures all affect the nutrient content of plant crops. Similarly, the environmental conditions of animals during their growth, in addition to what they are fed and how they are handled, affect the nutrient content of the products derived from them. While it is sometimes possible to infer information about foods from the same geographic area, determination of the similarity of foods raised in different parts of the world is a very difficult procedure.

EXAMPLES: Paul and Southgate [59] show great differences in the carotene and vitamin D content of milk depending on whether it was obtained in the summer or winter.

Holden et al. [33] show that selenium values for homegrown foods in the United States are two to ten times those for foods bought commercially.

Maturity and Portion or Part

The nutrient composition of plants and animals changes during their lifetime, and therefore the maturity of a plant crop at harvest or the age and status of an animal at slaughter may greatly affect its composition. In addition, the part of the food item consumed is an important factor; generally, different parts of the plant or animal differ considerably in nutrient content. Cuts of meat vary tremendously, as do parts of plants (e.g., roots vs. leaves). Furthermore, between and even within different cultures, the portions considered edible may vary (e.g., apple peels and cores, cheese rinds). Thus, direct substitution may not be appropriate unless all these factors agree. Confusing the issue is the fact that the descriptions included in food composition tables are often culturally biased and may not be detailed enough to allow careful evaluation of similarity.

EXAMPLES: The vitamin A activity in tomatoes more than doubles (from 64 RE to 133 RE per 100 g) as they ripen from the green to the red-ripe stage of maturity [96], while the distribution of carotenoids changes from primarily beta-carotene in the green and pink stages to primarily lycopene in the red-ripe stage [77].

"Turnips *(Brassica rapa)*" may refer to either the root or the leaves. The former contains only a trace of vitamin A, while the latter are a concentrated source of this vitamin.

The fat content of meat depends very much on its maturity at slaughter. Furthermore, while fat is a prized component of meat products in many countries, in others it is routinely trimmed and discarded. In general, it is rarely appropriate to directly substitute prepared meat products across cultures, even though the items have the same apparent name.

Processing and Preparation

Processing and preparation are general terms used here to refer to any changes made to a food between harvest and consumption. Once the plant is harvested or the animal is slaughtered, changes in nutrient values begin to occur immediately. If consumption is delayed, a variety of processes can be used to preserve the food until it can be consumed, to make it more appetizing, or both. Almost all of these processes affect the nutrient content of a food, with different effects depending on the food, nutrient, and process. Finally, the preparation of a food prior to consumption is a further source of nutrient modification and variation. Some of these processes are of great commercial interest and have been examined, but more work remains to be done. Food processing and preparation procedures include heating or cooking, canning or bottling, freezing, drying, fermenting, irradiating, packaging, storing, separating mechanically, adding chemicals, cutting, grinding, and mixing.

These techniques have various effects on the nutrient content of foods, primarily as a result of mechanical separation, moisture addition or removal, heat, oxidation, acidity change, light exposure, and the addition of food additives. The effects depend on both the specific foods and specific nutrients involved. See Rechcigl [74], Karmas and Harris [42], Tannenbaum [88], and Bender [9]

Mechanical Separation

Mechanical separation includes milling, husking, peeling, trimming, and other mechanical activities that separate out a portion of the foods for discarding or repackaging. It is well known that nutrients are not usually uniformly distributed throughout a food, and differentially removing a portion of the food may change its nutrient composition.

EXAMPLE: The Japanese tables [75] show that in the processing of brown rice to "wet/milled" rice the iron content drops from 1.1 mg to 0.5 mg per 100 g while the thiamin level drops from 0.54 mg to 0.12 mg per 100 g.

Moisture Addition or Removal

Water is used to prepare and cook various foods such as rice, oatmeal, legumes, pasta, and vegetables. Water may also be added to rehydrate foods such as condensed soup or fruit juice concentrates. The addition of water dilutes the nutrient content of foods on a weight basis. Nutrient loss may occur if water-soluble nutrients migrate from the food and are discarded when excess water is drained.

Dehydration and the concentration of foods is a major way that foods are preserved (both commercially and in the home) since dry foods are less vulnerable to microbial contamination. The specific procedure by which the moisture is removed can be critical to changes in nutrient levels, with potential losses due to heat, as well as to removal of some of the water-soluble vitamins with the water. A primary effect of moisture removal, from the point of view of the table compiler, is the increase in the concentration of most nutrients on a weight basis.

EXAMPLE: In the USDA data base [96] (all per 100 g edible portion), raw apricots have 86.4 g water, 14 mg calcium, and 10.0 mg ascorbic acid while dehydrated apricots have 7.5 g water, 61 mg calcium, and 9.5 mg ascorbic acid. Therefore, drying concentrates the calcium but destroys much of the ascorbic acid.

Heat

Heat is intrinsic to the preparation of many foods (roasting, frying, baking, boiling, etc.) as well as to many commercial processes of food preservation (blanching, pasteurization, sterilization, etc.). It induces changes in many nutrients, both degrading their utility to humans (e.g., destroying vitamins directly as well as increasing rates of oxidation) and enhancing it (e.g., by denaturing some destructive enzymes). Most important in the prediction of the results of heat on nutrient values are the temperature and time exposed, facts often not recorded in the descriptions of foods.

EXAMPLE: In the British tables [59], spaghetti has an almost sevenfold increase in moisture content (10.5 to 71.7 g per 100 g) when it is boiled. This cuts to a third the concentration of dry solids (from 89.5 to 283 g per 100 g) and of all nutrients in the dry solids. However, due to degradation and cooking losses, there are disproportionate decreases in some nutrients. For example, thiamin decreases to 7% of its original value (0.14 to 0.01 mg) and riboflavin to 17% (0.06 to 0.01 mg), considerably less than the 33% predicted by dilution alone.

Oxidation

Oxidation, resulting basically from exposure to air, is an important phenomenon of many food preservation and preparation procedures (e.g., storage and shredding). Oxidation results in the destruction of some vitamins (e.g., vitamin A, vitamin C, folic acid).

EXAMPLE: In the USDA tables [96] there are data for frozen onions, both whole and chopped. While most of the nutrients are virtually unchanged by chopping, ascorbic acid levels differ by a factor of two (whole onions contain 8.0 mg per 100 g edible portion; chopped onions contain 3.3 mg per 100 g edible portion).

Acidity Change

A change in the pH of a food can trigger a number of changes in the level of its nutrients by enhancing or inhibiting chemical reactions. Alkaline conditions are used to extract proteins, and also accompany the use of alkaline baking powders. Particularly important with high pH (alkaline) processes are the increased destruction of certain vitamins (thiamin, ascorbic acid, and pantothenic acid) which are protected by acidity.

EXAMPLE: Tannenbaum et al. [89] note that the rate of ascorbic acid oxidation reaches a maximum at pH 4, declines to a minimum at pH 2 and then increases with increasing acidity.

Light Exposure

Exposure of food to light has the effect of accelerating the oxidation process and specifically degrades riboflavin and vitamin C.

EXAMPLE: Bender [9] notes that, in two hours of exposure to bright sunlight, bottled milk lost as much as 50% of its riboflavin.

Food Additives

Food preservation and preparation often involve adding substances to modify (e.g., enzymes), ferment (e.g., yeast), preserve (e.g., antioxidants), or enhance (e.g., spices) the food product or to do some combination of these. Additionally, metallic contamination may be an inadvertent side effect of food processing and preparation. Some of these additions increase nutrient levels (e.g., fermentation may initiate synthesis of vitamins); some protect nutrients (e.g., sequestering agents bind to minerals and prevent their enhancement of oxidation); and others degrade or destroy nutrients (e.g., bleaching of flour can promote oxidation of nutrients). While much research has gone into the effects of these additions, it is such a large area that relatively little is known about additive interactions and the quantitative effects of these additions.

EXAMPLES: Tannenbaum et al. [89] point out that sulfite, commonly used to prevent browning fruits and vegetables, protects ascorbic acid but is detrimental to thiamin.

Due to its commercial importance, the fermentation of cheese has been studied extensively. Dworschak [19] notes that during ripening of cheese some of the B vitamins increase while others deteriorate, in part depending on the specific fermentation process.

Many countries require fortification of various products, and both the types of additives and levels of fortification must be considered when substituting nutrient values for food items from different countries. Many countries iodize salt; some add vitamins A and D to milk.

Similarity of Ingredients

For multi-ingredient foods, the type and proportions of the ingredients affect nutrient content. For example, the amount of salt added to a product will obviously affect the sodium and chloride content of the food, and, similarly, the amount and type of fat added will affect the energy content and lipid profiles of a food. The amount of water added will affect the concentration of all nutrients. When many ingredients are involved, it becomes very complex to estimate the similarity of mixed dishes unless exact recipes are known. In most cases, it is more accurate to develop recipes for multi-ingredient dishes than to assume that dishes are similar enough to directly substitute one for another. (See chapter 6.)

The similarity between foods cannot be judged without careful assessment of the above factors. However, it must be emphasized that close agreement on all of the above does not ensure that the nutrient values for two apparently "similar" foods are similar.

Adjustments for differences between foods

There are many publications on the effects of processing and preparation on the nutrient content of foods in specific processing/food/nutrient situations [9, 42, 74, 88]. While quite detailed predictive equations have been derived for a few of these situations [91], in general it is not possible to predict the nutrient changes that will occur in unstudied situations [76]. However, because of the need for food composition data, tables of approximate factors for US foods have been compiled for three important aspects of food preparation [49, 97]:

- removal of inedible parts (where a refuse or yield factor corrects for removal of a portion of the food);
- gain or loss of water and fats during cooking; and
- destruction of nutrients during food preparation.

Weight Adjustments for Inedible Parts

Foods are commonly expressed in terms either of the "as purchased" weight of the food or of the "as consumed" weight of the food, or, in some cases, in both ways (e.g., the

Chinese tables [16]). Moreover, many food items have multiple stages of preparation, and therefore multiple stages at which they may be consumed (e.g., few persons would consume the bones in meat, but the fat of meat may not be trimmed by all A data base developer may have to interconvert nutrient data between these various forms; the mechanism for this is the "refuse factor", the ratio, expressed as a percentage, of the weight of the "inedible" portion of a food to the total weight of the food as purchased. It is important to realize that this is only a weight adjustment since the nutrient content of the inedible portion of a food often differs significantly from the content of the edible portion.

EXAMPLE: Table 1 of the 1963 edition of Agriculture Handbook No. 8 [95] gives nutrient values of foods in 100 g edible portions as consumed. Table 2 in this same publication gives nutrient values in the edible portion of one pound of food as purchased, with an additional column labelled "Refuse". Using this factor, the nutrient data of the two tables can be easily converted from one to the other. For example, item number 1411, green olives, is listed as containing 20% of its weight as pits; thus an energy value of 338 calories per 100 g edible portion converts first to 1,536 calories per pound edible portion (338/0.22, converting from grams to pounds) and then to 1,228 calories per edible portion of one purchased pound (80% of 1,536) since only 80% of a pound of green olives is edible (the 20% refuse subtracted).

Given a food/process-specific refuse factor, a data base user or compiler can convert between "as purchased" and "edible portions". Many tables contain such a factor for their entries (e.g., the USDA tables [96] give both sets of values and the factors for many entries; the FAO African table [46] uses "g inedible per 100 g purchased", while the Indian [25] and Chinese [16] tables use "g edible per 100 g purchased"). Additionally, within the United States there are a number of other sources of such information: Agriculture Handbook No. 102 (*Food Yields Summarized by Different Stages of Preparation*) [49], Agriculture Handbook No. 8 [96, 95], and the American Home Economics Association *Handbook of Food Preparation [4]*.

There are two important assumptions upon which refuse factors are based:

- the proportion removed is not a function of the size of the food (e.g., the shells of both large and small eggs are about 12% of their weight), and
- the percentage of a food declared refuse or inedible is constant from sample to sample of the same food. (Note that Agriculture Handbook No. 102 [49] does give ranges of many of its refuse factors.)

Given full recognition of the oversimplification of these assumptions, refuse factors will permit food composition data users to convert nutrient calculations from "as purchased" to "edible portion".

Adjustments for Losses and Gains During Preparation

Often, in preparation of foods for consumption, there is a loss or gain of water, a loss of fat, or both. (When water is gained, it may be included as an ingredient in a recipe, and similar conventions can be used for fat gain.) Volatile components other than moisture may be lost during preparation, but these are seldom measured and thus are included in the moisture loss.

If cooking time is short, or the container is tightly covered, the water loss due to evaporation may be small; however, if cooking time is long and the container is not covered, the loss can be large. Evaporative losses can also occur without cooking, if a food item is left standing uncovered for some time. If the assumption is made that only water is lost (or added), nutrient ratios will not change, but their density will. These weight losses are traditionally summarized by a "yield factor", the weight of the prepared item divided by the weight of the unprepared item. A major compilation of yield percentages of losses and gains appears in Agriculture Handbook No. 102 [49]. These factors can be used to predict the nutrient contents of prepared foods from the values in the same foods raw.

EXAMPLE: Agriculture Handbook No. 102149] lists the "yield" from cooking carrots as 92%. Selected nutrient values for 100 g raw carrots are: energy 43 kcal, protein 1.03 6 calcium 27 mg, and ascorbic acid 93 mg Use of the yield factor predicts:

43 kcal / 0.92 = 47 kcal energy, 1.03 g / 0.92 = 1.12 g protein, 27 mg / 0.92 = 31 mg calcium, and

93 mg / 0.92 = 10.1 mg ascorbic acid, all per 100 g of cooked carrots.

These numbers compare favorably with the entries for cooked carrots for energy (measured value of 45 kcal), protein (measured 1.09 g), and calcium (29 mg), but not at all with the measured ascorbic acid value of 23.

As this example shows, yield percentages provide only rough estimates of the cooking losses that occur, and further, one must be aware that the cooking process may affect certain nutrients directly (e.g., ascorbic acid is sensitive to heat and, further, may be leached into the cooking water).

To estimate the nutrient content of an item cooked in water, with the water drained off and discarded, it is necessary to estimate the nutrient value of the discarded liquid. It is likely that some water-soluble vitamins and minerals will be lost into the discarded liquid. These losses may be estimated from measurements made on similar foods. Drained and undrained nutrient values are given in many food composition tables. Once these losses are applied, the values of all nutrients must be adjusted to reflect the new yield of the food item.

If fats (drippings) are drained, the nutrient value of the fat can be roughly estimated from yield factors in the same fashion as above. The situation is more complex since both water and fat losses must be considered, and due to the extremely high caloric density of fat, small errors can have a large impact on nutrient values. Thus, since these estimates are often crude, they must be used with great care.

EXAMPLE: The USDA tables [96] give 57.54 g fat and 8.66 g protein in 100 g raw bacon. For broiling, Handbook No. 102 [49] gives a yield factor of 29%; this 71% loss is separated into losses from drippings (49%) and from volatiles (22%). Assuming that the drippings are all fat, this gives, after cooking 100 g of bacon, 8.5 g (57.54 - 49.0) of fat in 29 g of cooked bacon. This converts into 29.5 g (8.54 / 0.29) per 100 g. By comparison, the protein conversion is a single step since there is no loss of protein during cooking: 8.66 g protein per 100 g in raw bacon / 0.29 (the overall yield factor) = 29.0 g protein per 100 g in broiled bacon. Comparison of these values with measured USDA values shows that the fat calculation is not accurate (49.24 g fat reported in cooked bacon) while the protein value is quite acceptable (30.45 g reported). The difficulty arises in assuming that the drippings are all fat; clearly, some is water loss, but one cannot estimate how much from Handbook No. 102.

Loss of Nutrients During Preparation

Because of the special interest in vitamin and mineral losses during food preparation, the USDA has published a table of "Nutrient Retention Values" [97]. This lists percentages of retention of nine vitamins and nine minerals for a number of foods and cooking methods, based on standard cooking times and temperatures. These factors are averages of a wide range of possible values and reflect food preparation practices in the United States. They may be used when more specific data are unavailable. As with the factors above for refuse and yield, these values are approximations and may not be appropriate in all situations. Such values are available in many countries. The most appropriate values can be obtained by comparing the cooking methods used to develop these factors with the habitual cooking methods in a region.

In many cases, the retention factors will include losses due to heating and losses due to draining. If there is also evaporative loss, a yield factor must be applied in addition to the retention factors.

EXAMPLE: The USDA tables [96] give the folacin level in raw onions as 19.9 µg. The retention factor of folacin is 70% for general "preparing and draining" of root vegetables, which predicts 13.93 µg per 100 g (19.9 x 0.70) folacin in cooked onions. Since the cooking of 100 g of onions yields 95 g cooked and drained onions, the final predicted folacin level is 13.93 / 0.95 or 14.7 µg per 100 g cooked onions. This compares with the measured value of 12.7 µg.

This example, like the previous ones, is presented both to illustrate the calculations necessary to apply these factors, and to show the potential inaccuracies of these methods of estimating nutrient contents of prepared foods. In most cases, data from similar foods modified by factors of refuse, yield, and nutrient retention must be regarded as a temporary measure to obtain entries in a food composition data base until more specific foods are analysed, more extensive and accurate theoretical procedures are developed, or both.

Assumed zero: a special case

Most data base developers estimate some nutrient values by assuming they are zero (i.e., not present in any detectable amount in the food item). Often these are decisions based on logic alone. For example, if there is no fat in a food item, there is obviously no saturated fat; if there is no carbohydrate, there is no sucrose. Or there may be a biological basis for this assumption (e.g., certain nutrients such as vitamin B-12 do not appear in plant products, while others such as dietary fibre do not appear in animal products). If these estimated zeros appear in a data base, an explanatory annotation should accompany them.

Alternatively, if a nutrient is present in trace amounts, the data base compiler must decide how to represent that information, and must avoid misleading the user into believing that "trace" is equivalent to "zero". If data are "not available", the compiler must ensure that the user will not assume a zero by default. See page 10 for more on this topic.

6. Calculations for multi-ingredient foods

Mixed dishes or multi-ingredient foods represent the majority of items in diets worldwide. These include not only foods prepared in the home but also foods prepared in restaurants, by food vendors, in institutions such as hospitals, schools, and the military, and by the food industry. To enable dietitians, nutritionists, and epidemiologists to evaluate the role of these foods in the health of individuals, there is a need for composition data on these foods. Obtaining and using data on the content of multi-ingredient foods presents a number of inherent difficulties, primarily because of the abundance and diversity of these kinds of foods. Many mixed dishes, as prepared for consumption, are variable and poorly defined, differing from kitchen to kitchen, day to day, around the world. Analytic data do not exist for most of these foods, and accurate estimation of their nutrient content is perhaps impossible. However, such data are needed, and are routinely being estimated and used. For example, in the food industry, predicted nutrient content of proposed new product formulations is critical to decisions regarding further work on products.

This chapter provides guidelines for estimating the nutrient levels of multi-ingredient foods based on the nutrient levels of the ingredients. For simplicity, *a multi-ingredient food is defined as a food with two or more ingredients.* In addition to including standard mixed dishes such as curries, stews, casseroles, salads, and many dessert items, the term "multi-ingredient food" is used for simple mixtures such as foods prepared with the addition of water (e.g., reconstituted condensed milk or gelatin dessert made from dry powder), foods prepared with the addition of fat (e.g., sauteed or fried foods), and foods which have added sauces, gravies, or toppings (e.g., asparagus with hollandaise sauce or ice cream with chocolate syrup, whipped cream, nuts, and cherries).

The procedure for calculating the nutrient content of a multi-ingredient food starts from a recipe-a list of ingredients and a description of how they are combined-and the nutrient content of the ingredients. It is possible that some of the nutrient values for some ingredients may be missing and must be obtained by methods described above in this document (i.e., from another data base or by analogy with a similar ingredient). It is also possible that an ingredient item may itself be a multi-ingredient food (e.g., bread or a sauce) and may require calculation from a recipe with its own individual components. Types of recipes are discussed below.

Given a recipe and data on the ingredients, the problem is then how to combine these data. Since many multi-ingredient foods involve the processing of foods in ways which change their nutrient content (at least on a per weight or volume basis), calculations often employ the factors described in the previous section:

- refuse factors to correct for weight changes due to removal of inedible portions during preparation, cooking, or both;
- corrections for weight (yield factors) and nutrient levels due to changes in water and fat during preparation, cooking, or both; and
- nutrient retention factors to correct for nutrient (primarily vitamin and mineral) losses or gains during preparation, cooking, or both.

Guidelines for recipe calculations are given below. It must be stressed that calculation of the nutrient content of multi-ingredient foods from the nutrient data of the ingredients are estimates and not meant to replace nutrient values obtained by laboratory analysis.

Calculating the values is an intermediate solution until adequate analytic data become available. A major deficiency of recipe calculation is the lack of information on how foods and components interact. It is obvious that they do, and such interactions are especially important in mixed dishes; however, such information does not currently exist. For many foods, recipe calculation may be the only cost-effective way to obtain nutrient data. Analytic data for multi-ingredient foods are not generally as available as they are for single foods, with data especially lacking for ethnic and regional dishes, many variations of homemade and restaurant-made foods, and the numerous varieties of industrially prepared canned, frozen, and packaged entrees. Because of the difficulties and costs of analyses of mixtures of different food types, and limitations on resources to adequately sample these mixtures, calculated values that rely on a broad base of representative samples may actually be more accurate than values derived from laboratory analysis of one or two samples of the prepared food.

Types of recipes

It is conceptually useful to distinguish between recipes for simple combinations which require only mixing of ingredients and then correction for weight or volume, and recipes which require more extensive modifications of the data through estimation of losses or gains of water, fat, vitamins, and minerals. A third class of recipes includes those which require estimation of the amounts of the ingredients, and occasionally the nature of the ingredients themselves.

Simple Combinations

Some recipes require only the addition of the nutrients of the specified quantities of ingredients, followed by adjustment of the weight or volume of the final food, in order to express the nutrient levels per standard amount of the food (such as 100 g or household measures). In these cases the nutrient values should be for the ingredients in an "as consumed" form (i.e., cooked if the ingredients are cooked and containing no refuse). Examples include:

- coffee prepared from instant powder with added sugar and cream;
- baked potato with sour cream and chives;
- tomato soup prepared from condensed soup and whole milk. (Note that heating may modify the nutrient content in this situation.)

The nutrient levels of multi-ingredient foods that are prepared with the addition of water (such as milk reconstituted from powder) are also of this type, and can be estimated quite well if water (e.g., tap water, well water, bottled spring water, distilled water, mineral water) and its associated nutrient levels are included in the data base. If a data base does not include water as a food item, the nutrient contribution from water can be considered zero when calculating the nutrient content of a multi-ingredient food; however, this must be noted, since in some areas and for some nutrients the contribution from water is important. The weight contribution of water must, of course, be considered when expressing the nutrient levels per unit of multi-ingredient food.

A special aspect of these recipes is that, while the weight of ingredients add directly, volume often does not, and special care must be taken to ensure that results expressed on a per volume basis are correct. For example, one cup of powdered milk plus three cups of water yields significantly less than four cups of fluid milk; likewise, 1/2 cup mayonnaise added to one cup chopped apples, Y. cup raisins, and Y. cup chopped walnuts yields about 1.3 cups salad.

Occasionally, simple combinations may be used to calculate the nutrient content of a food from its component parts. In this case, the components are treated as "ingredients". For example, the nutrient composition of a chicken leg may be calculated from the nutrient composition and proportions of meat, skin, and separable fat.

Recipes Which Involve Nutrient Changes

Many recipes specify the combination of "as consumed" ingredients, but then require additional cooking which modifies the levels or densities of the nutrients of the foods involved. Here it is necessary to apply the various factors to correct for these changes due to water or fat loss or gain. Examples of these multi-ingredient items include:

- refried beans, made from boiled beans, lard, and spices which are fried in fat; and
- shepherd's pie, made from cooked minced (ground) beef, boiled onion, boiled potato, water, milk, margarine, salt, and pepper which is baked for 25 minutes to brown.

Recipes become more complicated when starting with the raw ingredients. These recipes require adjustment for refuse and yield during preparation and cooking as well as nutrient losses and gains. Consider, for example, the above shepherd's pie when data are available only for raw meat and vegetables.

Recipes for Which Ingredient Amounts Must Be Estimated

Occasionally, the amounts of ingredients are unknown and must be estimated. If partial nutritional data are available for the recipe, these may be used to estimate ingredient proportions, which in turn may be used to estimate the missing nutrient data. This procedure may be necessary in the case of proprietary food mixtures for which only some data on content (e.g., from the package label) are available, in order to obtain estimates for the other nutrients. These estimates are usually less certain than the usual recipe estimates.

In the following example the zinc content of a product estimated on the basis of information available about proximate nutrients. A more complex example (for a chocolatecoated ice cream bar) is given by Posati [68].

EXAMPLE: A compiler wishes to estimate the zinc content of canned corned beef hash and has only the manufacturer's information for proximate nutrients. The two main ingredients are potatoes and corned beef, but the ratio is not known. However, it is known that the product contains 11 g carbohydrate and 8 g protein per 100 g edible portion. Assuming all the carbohydrate is from the potatoes, and knowing that cooked potatoes have approximately 15 g carbohydrate per 100 g, a proportion of 11/15 or 73% potatoes (27% corned beef) can be assumed. This assumption can be checked by calculating the protein content. The corned beef component contains 25 g protein/100 g, which would contribute about 7 g/100 g to the hash (0.27 x 25). The protein in potatoes (about 2 g/100 g) would contribute about 1 g, giving the correct total of 8 g/100 g for the hash. Given this information, the zinc content of the hash can be estimated if the zinc content of corned beef and of cooked potatoes is known.

Another illustration of the procedure is the use of known ratios to estimate amounts of individual fatty acids and amino acids. Well-defined patterns of fatty acids for different classes of foods can be used to calculate specific fatty acid levels based on known total lipid content, and, similarly, patterns of amino acids can be used to calculate specific amino acid levels based on known total nitrogen content.

EXAMPLE: The amino acid pattern of milk can be used to estimate the individual amino acids of cream given the nitrogen content of cream and milk, and the assumption that cream has the same amino acid pattern as does milk. From Paul and Southgate [59], whole milk has 510 mg of methionine per gram of nitrogen and single cream (21% fat) has 0.376 g of nitrogen per 100 g. From this it is estimated that the methionine content of cream is 0.376 x 510 = 192 mg per 100 g.

Recipe calculation guidelines

The detailed procedure for calculation of the nutrient content of a multi-ingredient food is shown in figure 1. The major steps are selection of the recipe, data collection for the nutrient content of the ingredients, adjustment of the content of each ingredient for effects of preparation, summation of ingredient composition, final weight (or volume) adjustment, and determination of the yield and final volumes.

This procedure for recipe calculation incorporates the following considerations which must be recognized by the user.

- The procedure assumes that data are available on each of the ingredients which are then combined and, if necessary, prepared. If a recipe calls for one or more ingredients which must be prepared from recipes (e.g., a sauce or a stock), the calculation procedure must first be applied to each of these ingredients separately, and then to the final product using the results of the initial recipes as ingredients.
- This procedure recommends nutrient adjustment of individual ingredients before their combination, with a final adjustment for weight. This is a slight distortion of what often happens when ingredients are combined and then prepared. However, the work which has produced the various factors (e.g., Agriculture Handbook No. 102 [49]) follows this simplified model, and, moreover, the variability of nutrients in foods and their responses to cooking will tend to overwhelm any inaccuracies which derive from these simplifications.
- This procedure produces ESTIMATES of nutrient content of multi-ingredient foods. The preferred method of determining the nutrient content is chemical analysis.

1. Select or develop appropriate recipe.
2. Collect weight and nutrient content data for each ingredient.
3. Correct ingredient nutrient levels for weight of edible portions where appropriate.
4. Correct ingredients for effects of cooking:
either

- if data for cooked ingredients are available, use yield factors to adjust from raw to cooked weights

or

- if data for cooked ingredients are not available, use data for uncooked ingredients applying yield factors to adjust for weight changes and retention factors for nutrient losses or gains during cooking.

5. Sum weights of ingredients to get weight of recipe.
6. Sum nutrient values of ingredients to obtain nutrient value of recipe.
7. Adjust recipe weight and nutrient levels to reflect changes in fat/water when whole mixture is cooked; make any additional refuse adjustments; apply retention factors if available for whole recipe.
8. Determine the quantity of prepared food produced by the recipe.
9. Determine the final values per weight (e.g., per 100 9), volume (e.g., per cup), or serving portions as desired.

Fig. 1. Guidelines for calculating the nutrient levels of recipe foods

The guidelines in figure 1 are intended to be applicable for the most complicated situation encountered, where the user has available a detailed description of ingredients, data on the nutrient values of the ingredients, and an adequate description of how those ingredients are to be combined.

While final values should be rounded to reflect the precision of the calculation, the intermediate values (e.g., the weights of the individual ingredients) should be used in the precision to which they are calculated. This is preferred to rounding values and using them in further calculations, since numerical errors due to rounding tend to accumulate, especially when the rounded values enter into multiplicative calculations as in the discussion that follows [5]

Select Appropriate Recipe

Review possible recipes for the food in question and select one that is appropriate (e.g., representative of the food habits of the population or subgroup of interest). The selection of the recipe should consider the source of the food (home, restaurant, industry, or institution) and whether the recipe might be altered by regional or local food practices. Ethnic or regional variations may be needed. Sources fat recipes include cookbooks, food package instructions, and suggested recipes on food packages. (Note that recipes may also be deduced from food ingredient labels (see page 51), and assistance from the food manufacturer may be obtained for more quantitative information.) In some cases, it may be necessar Garher Available to obtain the food and weigh or measure the pans (e.g., a pepperoni pizza consisting of crust, tomato sauce, cheese, and pepperoni, or a fast food hamburger consisting of meat, bun, sauce, pickle, lettuce, and tomato).

Gather Available Data on Ingredients

Determine the weight in grams of each ingredient, noting the recipe yield if provided. Depending on the recipe, some ingredients may be raw and others cooked. Some data bases have weight conversion factors for the household portions of foods. In addition, there are several references which may be useful for determining the gram weights of household portions of foods [1, 23, 62, 96].

Correct Ingredients to Edible Weight

For ingredients with inedible parts or handling losses during preparation, determine the weight of the edible portion and adjust nutrient levels to this weight. (See page 44 ff.) For example, consider a recipe that calls for one kilogram of raw lamb shoulder. Agriculture Handbook No. 102 [49] indicates that raw lamb shoulder contains, by weight, 17% bone, 27% trimmable fat, and 8% cutting waste. Thus, the edible weight of 1 kg lamb shoulder is 480 g (100% - 17% - 27% - 8% = 48%).

Adjust Ingredients for Cooking

If the recipe involves cooking, the ingredients must be adjusted for cooking effects. For ingredients with nutrient data available for the cooked state, use the paragraph "Cooked ingredients" below. For ingredients with data for the raw state only, use "Uncooked ingredients".

Cooked ingredients. Ingredients with nutrient data for the cooked state require correction for the effect of cooking on the weight (or volume) of the dish. Thus, if the recipe calls for a specific weight of raw meat to be added and cooked, and data exist for the nutrient content of cooked meat, one must apply a yield factor to the amount of raw meat needed in order to determine the amount of cooked meat that will result, and then calculate the nutrient content of that amount. For example, if data were available on cooked lamb, one would still have to correct the above 480 g of edible portion of raw lamb shoulder for the weight loss during cooking. Agriculture Handbook No. 102[49] gives a 32% cooking weight loss and thus the 480 g raw meat becomes 326 g (0.68 x 480 g) of cooked meat. The nutrient content of the lamb is then based on 326 g.

Uncooked ingredients. If nutrient values are available only for raw ingredients, it is necessary to calculate the yield (change of total weight) of the ingredients due to preparation and cooking, the aired effect of the change of weight on nutrient density and content, and the effect of actual loss or gain of nutrients due to cooking. For example, 1 kg of raw lamb shoulder, as purchased, contains 480 g of edible meat. The thiamin content of raw lamb is 0.165 mg per 100 g edible portion [96]; 480 g contains 0.79 mg thiamin. During cooking, the level of retention of thiamin is 75% [97] and the weight loss (assumed to be all water) is 32%, which produces 326 g of meat containing O.59 mg thiamin. (These calculations predict that cooked lamb contains 0.18 mg of thiamin per 100 g.)

Sum Weights of Ingredients

Sum the weights of all prepared ingredients to get the weight of the finished dish.

Sum Individual Nutrient Values

Sum the total nutrient contents of all prepared ingredients to obtain the total amounts of nutrients in the final dish.

Adjust Recipe Weight and Nutrient Levels

Adjust the recipe weight and nutrient level for the changes in water and fat that occur when the whole dish is cooked. If additional retention factors and corrections for refuse for the whole dish are available, they should also be incorporated at this point. See Agriculture Handbook No. 102 [49] and Merrill, Adams, and Fincher [50] for possible factors.

Determine Recipe Yield

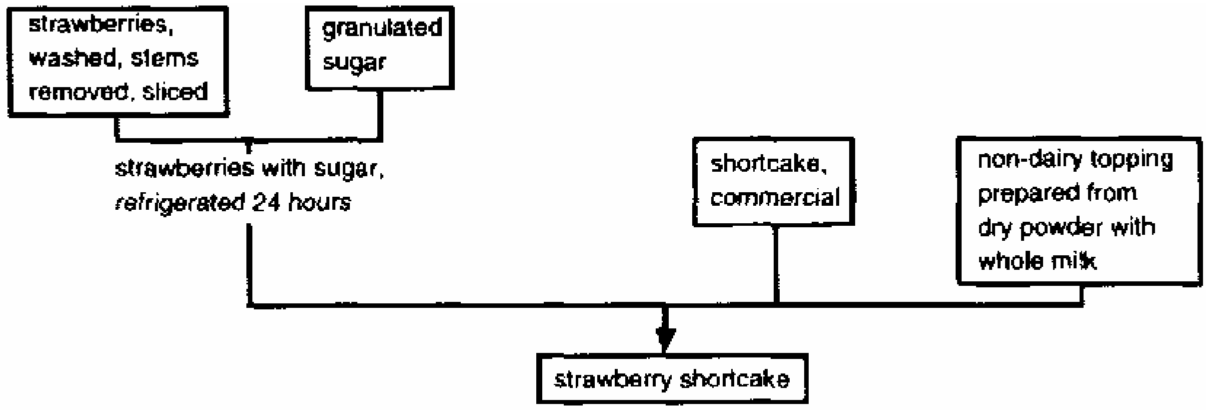
Determine the recipe yield as the total weight or volume produced or as the number and weight or volume of servings. Yields may be determined by:

- using the yield given by the recipe,
- adding the volumes of the ingredients in prepared form (in some cases the volumes may not be additive),
- measuring container size,
- preparing the recipe and determining volumes and number of servings.

Determine Values to Be Used in the Food Composition Data Base

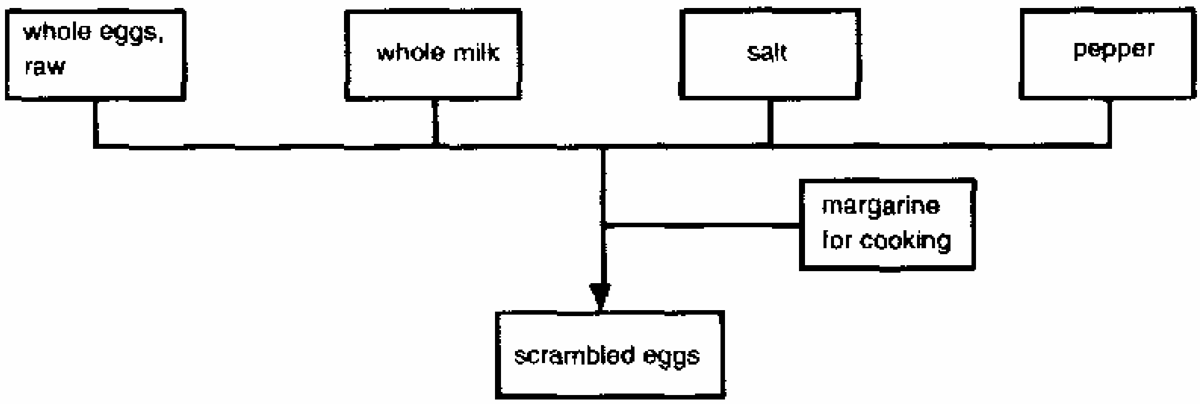
Divide the nutrient totals by the total weight, volume, or number of servings of the dish. For example, if the recipe instructions indicate that the recipe yields four one-cup servings, divide the recipe weight and nutrient levels by four to determine the weight and nutrient levels of a one-cup serving.

Figures 2 through 6 and examples 1 through 4 display use of these guidelines in the estimation of the nutrient levels of various types of multi-ingredient foods.



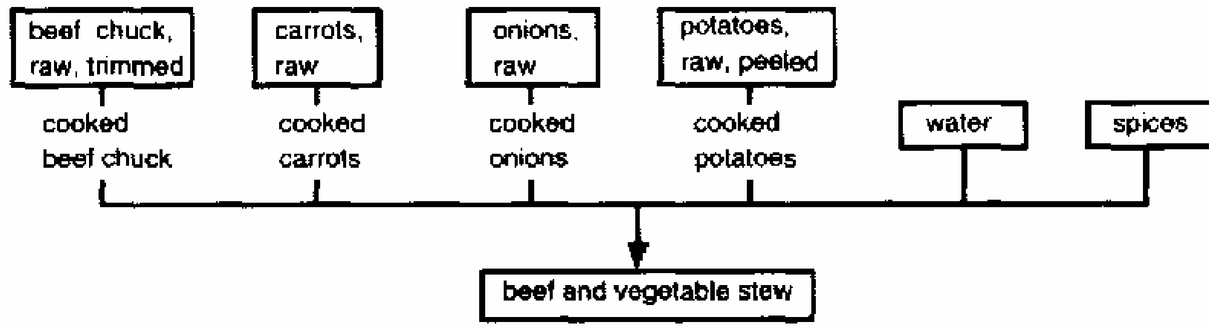
Apply nutrient retention values for holding sliced strawberries overnight in refrigerator.
Sum weights of ingredients and sum nutrient levels of ingredients.
Divide sums by yield of recipe to determine weight per serving and nutrients per serving.

Fig. 2. Application of guidelines to strawberry shortcake



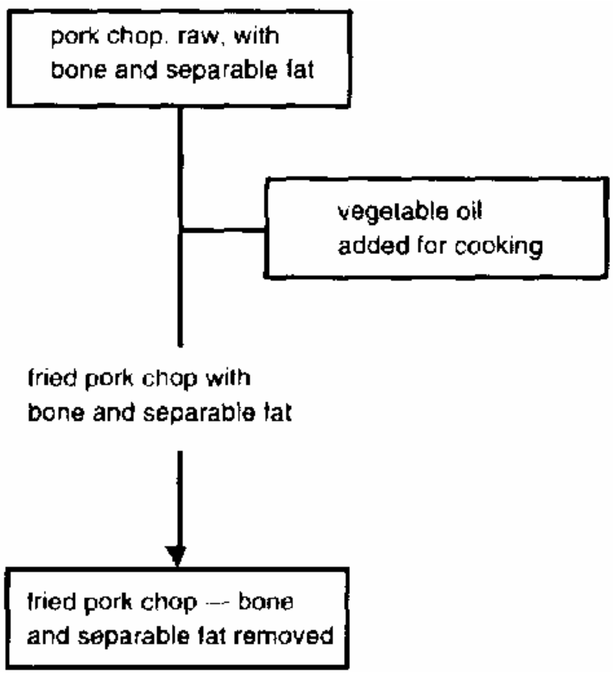
Add nutrient levels for specified quantities of eggs, milk, salt, pepper, and margarine.
Apply factor for water loss during cooking and readjust nutrients to this new weight.
Apply retention factors for nutrients lost during cooking.

Fig. 3. Application of guidelines to scrambled eggs



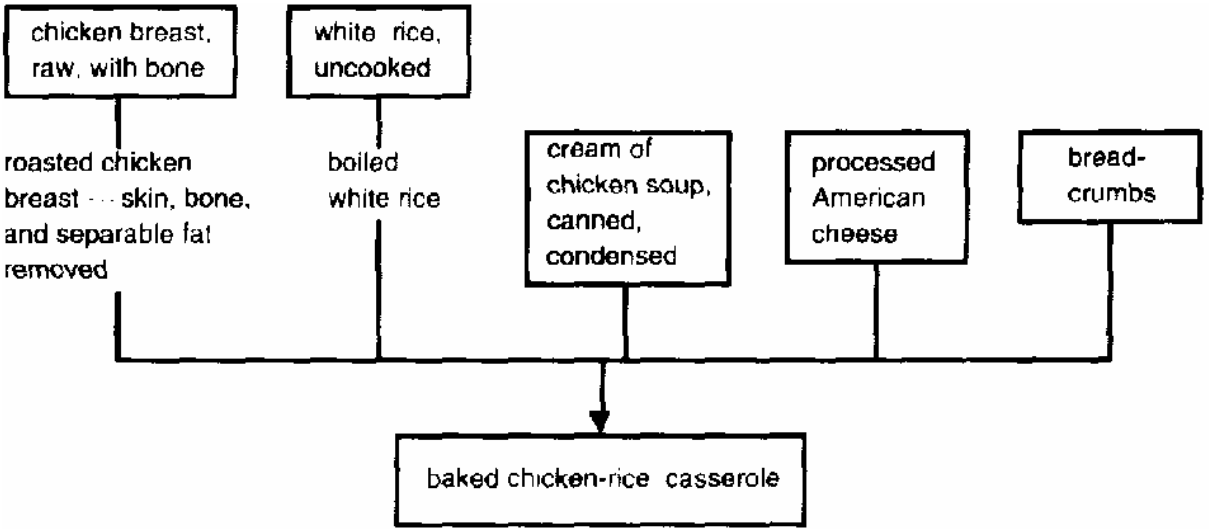
Use yield factors to determine quantities of cooked ingredients from recipe quantities of raw ingredients.
Sum ingredient weights.
Sum nutrient values of cooked ingredients, water, and spices.
Adjust recipe weight for water loss during cooking.
Apply retention factors for nutrient losses during stewing.

Fig. 4. Application of guidelines to homemade beef and vegetable stew



Apply factor for decrease in water weight during cooking.
Apply factor for increase in weight and also in total fat, SFA, PUFA, cholesterol, and vitamin E from vegetable oil during cooking.
Apply nutrient retention factors for nutrient losses during cooking.
Adjust weight for removal of bone and separable fat.
Adjust tat, energy, SFA, PUFA, and cholesterol for removal of separable fat.

Fig. 5. Application of guidelines to fried pork chop



Apply refuse factor for bone and skin of chicken.
Apply yield factors for raw chicken and uncooked rice.
Apply water-loss and nutrient changes with skin loss of chicken.
Apply water-gain changes for cooking rice.
Apply nutrient-retention factors for cooking rice and chicken.
Sum weights and nutrient values of ingredients.
Apply water-loss factor for cooking casserole. Apply nutrient-retention factors for cooking casserole.

Fig. 6. Application of guidelines to chicken-rice casserole

EXAMPLE 1: Mixed Vegetable Salad

3 tomatoes, cut in wedges
1 cucumber, cut in thin slices
1 green pepper, cut in strips
6 green onions, sliced
6 radishes, sliced
1 head lettuce, shredded
1/2 cup French salad dressing

Prepare vegetables and toss together with dressing in salad bowl. Makes six servings. This recipe is an illustration of a simple combination of raw ingredients. For the purposes of demonstration, it will be assumed that all ingredients are present in the data base available to the user, and that the data base includes the weight of each food as purchased, a refuse factor (if applicable), and the nutrient content per 100 g of edible portion of food. The steps involved in calculating the nutrient content of the mixed vegetable salad are outlined below.

Step 1: Calculate the weight of each ingredient.

tomatoes: 1 medium tomato = 135 g
refuse factor (removal of stem) = 2%
135 g x 3 = 405 g
405 B x 0.98 = 396.9 g
cucumber: 1 small cucumber = 175 g
refuse factor (removal of ends) = 3%
175 g x 0.97 = 169.75 g
green pepper: 1 green pepper = 90 g
refuse factor (removal of stem and core) = 18%
90 g x 0.82 = 73.8 g
green onions: 6 small green onions (bulb only) = 30 g
radishes: 10 medium radishes = 50 g
refuse factor = 10%
50 g x 6/10 = 30 g
30 g x 0.9 = 27 g
lettuce: 1 head = 567 g
refuse factor = 5%
567 g x 0.95 = 538.65 g
French dressing: 1 cup = 250 g
250 g x 0.5 = 125 g

Step 2: Calculate the weight of the total recipe and the weight of an individual serving. The weights of the ingredients are added together for a total recipe weight of 1,361 g. Since the total recipe makes six servings, the weight of an individual serving is 227 g.

Step 3: Calculate the nutrient content of the total recipe, the nutrient content per 100 g of salad, and the nutrient content of an individual serving.

The nutrient contents of the individual ingredients are added together to get the nutrient content of the total recipe. These values are divided by 13.61 to get the nutrient content per 100 g of salad, and they are divided by 6 to get the nutrient content per serving.

Step 4: If possible, calculate the weight per unit volume of salad.

It is possible to find conversion factors to convert the weights of the individual ingredients into volume measurements. For example:

1 cup sliced cucumber = 105 g
1 cup green pepper cut in strips = 100 g
1 cup sliced radishes = 115 g
1 cup chopped lettuce = 55 g
1 cup sliced green onions = 100 g

However, it is not possible to determine a total recipe volume by simply adding the volume measurements of the individual ingredients, because the smaller ingredients and the salad dressing wit fill the space between the larger ingredient pieces. A volume could be estimated, or the recipe could be prepared and measured.

EXAMPLE 2: Scrambled Eggs with Onions

2 large eggs
1/6 cup whole milk
1/8 teaspoon salt
1/4 cup chopped raw onions
2 teaspoons oil

Add milk and salt to eggs and beat with a fork. Fry onions in the oil. Pour egg mixture into frying pan with the onions, and stir mixture with a fork while cooking until it solidifies. Makes one serving.

For purposes of illustration, it is assumed that the available nutrient data base includes nutrient values for the raw ingredients, but not for the cooked ingredients. In this case, adjustments must be made for water loss and nutrient loss during cooking. The steps involved in calculating the nutrient content of the scrambled eggs are outlined below.

Step 1: Add nutrient levels for the specified quantities of ingredients.

The nutrients in the raw eggs, whole milk, salt, raw onions, and oil are added together.

Step 2: Readjust quantities of those nutrients that are lost during cooking.

Nutrients might be lost in evaporation or destroyed by heat during cooking. For example, it is estimated that 15% of the thiamin, 5% of the riboflavin, and 5% of the niacin in eggs are lost during cooking. For milk, 10% of the thiamin and 25% of the ascorbic acid are lost. Cooked onions are estimated to have 20% less ascorbic acid, 15% less thiamin, and 30% less folacin than raw onions. These factors should be applied to the nutrient levels in the raw ingredients to determine the nutrient levels in the cooked recipe.

Step 3: Determine weight of recipe before cooking.

The weight of the specified quantity of each raw ingredient is determined. A refuse factor is applied to the weight of the egg to calculate the weight without the shell. The weights are added together to get a total weight for the recipe.

eggs: 1 large egg = 57 g
refuse factor (shell) = 11%
57 g x 2= 114 g
114 g x 0.89 = 101.46 g
milk: 1 cup = 244 g
244 g x 1/6 = 40.66 g
salt: 1 tsp - 5.5 g
5.5 g x 1/8 = .688 g
onions: 1 cup chopped = 170 g
170 g x 0.25 = 42.5 g
oil: 1 tsp = 4.53 g
4.53 g x 2 = 9.06 g
total weight = 19437 g

Step 4: Determine weight of recipe after cooking.

In cooking scrambled eggs, weight loss occurs from evaporation. This loss is estimated to be about 8%. No fat is lost in cooking. Therefore, the final recipe weight would be 19437 g x 0.92 = 179 g.

Step 5: Determine the nutrient levels of the recipe per 100 g and per serving.

The calculated nutrient levels represent the nutrient content of one serving, since the recipe as stated is one serving. The nutrient levels must be divided by 1.79 to determine the nutrient content of 100 g of scrambled eggs.

EXAMPLE 3: Broiled Pork Chop

In this illustration, it is assumed that the available food composition table has an entry for a raw pork chop, but no entry for a cooked pork chop. The raw pork chop weighs 151 g with the bone and some visible fat. The bone is 21% of the total weight of the raw pork chop.

Step 1: Determine the decrease in weight during cooking.

It is estimated that 32% of the weight of the pork chop is lost during broiling. Therefore, the final weight of the broiled pork chop will be 151 g x 0.68 - 102.68. Step 2: Subtract nutrient loss in fat drippings.

Of the 32% loss in weight during broiling, it is estimated that 9% is fat drippings and 23% are volatiles. Therefore, the amount of pork fat lost during cooking will be 151 g x 0.09 = 13.59 g.

The nutrient content of 13.59 g of pork fat should be subtracted from the nutrient content of the raw pork chop.

Step 3: Determine nutrient loss due to cooking.

Cooking destroys some of the nutrients in the raw pork chop. For example, it is estimated that 30% of the thiamin, 20% of the niacin, and 25% of the calcium are lost in broiling. These factors should be applied to the nutrient levels in the raw pork chop to determine the nutrient levels in the cooked pork chop.

Step 4: Determine the weight of the edible portion of the broiled pork chop.

After cooking, the bone is estimated to be about 18% of the total weight of the pork chop. It is assumed that all visible fat is eaten. The weight of the edible portion of the pork chop will then be 102.68 g x 0.82 = 84 g.

EXAMPLE 4: Githeri (Kenya)

maize, dry, raw: 2 Kimbo tins (1 kg size)
kidney beans, dry, raw: 1 Kimbo tin (1 kg size)
water: 3 large Kimbo tins (2 kg size)

(Note that Kimbo tins are volume measures.)

Combine dry maize and kidney beans with water and cook the mixture for one or more hours.

For purposes of illustration, it is assumed that the nutrient data base includes nutrient values for the raw ingredients per 100 g.

Step 1: Add nutrient levels for the specified ingredients.

Convert volume measures (Kimbo tins) to gram equivalents, and multiply the nutrient values per 100 g by the gram weight of the ingredient divided by 100. For example, for energy (kcal):

maize: 355 kcal/100 g x 1,080 g per tin x 2 tins = 7,668 kcal
beans: 327 kcal/100 g x 1,000 g per tin x 1 tin = 3,270 kcal
total kcal for the recipe = 7,668 + 3,270 = 10,938

Step 2: Readjust quantities of those nutrients that are lost during cooking.

Since githeri is commonly cooked for one or more hours, there will be destruction of heatlabile vitamins; some losses may exceed 50%.

Step 3: Determine weight of recipe before cooking.

maize = 1,080 g per tin x 2 tins = 2,160 g
beans = 1,000 g per tin x 1 tin = 1,000 g
water = 2,500 g per tin x 3 tins = 7,500 g
total raw weight = 10,660 g

Step 4: Determine weight of recipe after cooking.

The average cooked weight of githeri is approximately 75% of its raw weight:

0.75 x 10,660 g = 7,995 g cooked weight

Step 5: Determine the nutrient levels per 100 g.

Divide all nutrient totals by (7,995/11)0) to give nutrients per 100 g. For example, for energy: 10,938 kcal (from step 1) / (7,995/100) = 137 kcal/100 g.

Limitations

There are few studies comparing the calculated and analytic nutrient values for recipe foods; there are even fewer studies comparing the results of calculated recipes using different data bases. Both types of studies are necessary to determine the accuracy (validity) of these suggested guidelines. Differences between calculated and analytic results may reflect the effects of different types of cooking equipment (household, restaurant, industrial, or institutional), surface area of food contact exposure, length and temperature of cooking, and volume of product on the yield factors, retention factors, and water/fat, loss/gain factors. In addition, the assessment of the accuracy of the factors in current use is not adequately understood.

Differences in results as calculated by two or more data base systems may reflect different methods of estimation or calculation, choices and sources of foods and nutrients in the data base, and judgements of coders, as well as different values for yields, retentions, and water and fat changes. The uses of recipe calculations in a computerized system is discussed by Perloff [64].

The calculation of nutrient levels of multi-ingredient foods from recipes by the above scheme is, in general, an intermediate solution for a need that should eventually be solved by chemical analysis, or by more accurate estimation techniques yet to be determined. As previously mentioned, it may not be cost-effective to analyse many multi-ingredient foods. Furthermore, because of the extreme variability in many of these foods, analysis may not be the sole answer. In either case, the results from these calculations should be regarded, and annotated, as approximations.

References

1. Adams CF. *Nutritive Value of American Foods in Common Units.* Agriculture Handbook No. 456. Washington, DC: US Department of Agriculture, 1975.

2. Afifi AA and Azen SP. *Statistical Analysis: A Computer Oriented Approach,* 2nd ed. New York: Academic Press, 1979.

3. Allison RG and Senti FR. *A Perspective on the Application of the Atwater System of Food Energy Assessment.* Bethesda, Md., USA: Life Sciences Research Office, Federation of American Societies for Experimental Biology, 1983.

4. American Home Economics Association. *Handbook of Food Preparation,* 8th ed. Washington, DC: American Home Economics Association, 1980.

5. Armitage P and Barry G. *Statistical Methods in Medical Research,* 2nd ed. Oxford: Blackwell Scientific Publications, 1987, p. 15.

6. Atwater WO and Bryant AP. The availability and fuel values of food materials. *Conn Agr Expt Sta 12th Ann Rpt*, 1900, pp. 73-110.

7. Aurand LW, Woods AK, and Wells MR. Food Composition and Analysis. New York: Van Nostrand Reinhold, 1987.

8. Beecher GR and Vanderslice JT. Determination of nutrients in foods: Factors that must be considered. In Stewart and Whitaker, reference 87, pp. 29-55.

9. Bender AK. *Food Processing and Nutrition.* London: Academic Press, 1978.

10. Breen WM. Selection of raw products: Factors to be considered. In: *Proceedings A New Market ... Nutrition and Quality*. Washington, DC: The Food Processors Institute, 1983.

11. Bressani R. The data required for a food data system. *Food Nutr Pull* 5(2):69-76, 1983.

12. Bressani R, ed. Memorias de la primera reunion sobre tables de composicion de alimentos LATINFOODS. *Arch Latinoam Nutr* 37(4):607-826, 1987.

13. Bruce Å and Bergström L. User requirements for data bases and applications in nutrition research. *Food Nutr Bull* 5(2):24-29, 1983.

14. Buzzard IM and Feskanich D. Maintaining a food composition data base for multiple research studies: *The NCC food table*. In Rand et al., reference 72, pp. 115-122.

15. Chambers JM, Cleveland WS, Kleiner B. and Tukey PA. *Graphical Methods for Data Analysis*. Belmont, Calif., USA: Wadsworth International Group, 1983.

16. Chinese Academy of Medical Sciences. *Table of Food Composition*. Beijing, 1982.

17. Cleveland WS. *The Elements of Graphing Data.* Monterey, Calif., USA: Wadsworth Advanced Books and Software, 1985.

18. Dun's Marketing Services. *Million Dollar Directory. America's Leading Public & Private Companies*. Series 1988. Parsippany, NJ., USA: Dun's Marketing Services, Inc., 1988.

19. Dworschak E. Effect of processing on nutritive value of food: Fermentation. In Rechcigl, reference 74, pp. 63-76.

20. English R and Lester IH, eds. *Proceedings of the First OCEANIAFOODS Conference*. Canberra: Australian Government Publishing Service, 1987.

21. Exler J. *Iron Content of Food.* Home Econ Res Rpt No. 45, revised. Washington, DC: Consumer Nutrition Division, Human Nutrition Information Service, US Department of Agriculture, 1983.

22. FAO. *Requirements of Vitamin A, Iron, Folate, and Vitamin B12.* Report of a Joint FAO/WHO Expert Consultation. Rome: Food and Agriculture Organization, 1988.

23. Fulton L, Matthews E, and Davis C. *Average Weight of a Measured Cup of Various Foods.* Home Econ Res Rpt No. 41. Washington, DC: Consumer Nutrition Division, Human Nutrition Information Service, US Department of Agriculture, 1977.

24. Goddard MS and Matthews RH. *Current knowledge of nutritive values of vegetables.* Food Technol 33:71-73, 1979.

25. Gopalan C, Rama Sastri BV, and Balasubramanian SC. *Nutritive Value of Indian Foods*. Hyderabad: National Institute of Nutrition, Indian Council of Medical Research, 1984.

26. Greenfield H and Southgate DAT. *Guidelines for the Production, Management, and Use of Food Composition Data*. In preparation.

27. Hawkins DM. *Identification of Outliers*. London: Chapman and Hall, 1980.

28. Heintze D, Klensin JC, and Rand WM. *International Directory of Food Composition Tables*, 2nd ed. Cambridge, Mass., USA: International Network of Food Data Systems, 1988.

29. Helrich K, ed. *Official Methods of Analysis of the AOAC*, 15th ed. Arlington, Va., USA: Association of Official Analytical Chemists, 1990.

30. Hepburn FN. *Managing food composition data at the national level*. In Rand et al., reference 72, pp. 109-114.

31. Hildebrandt DM. *Computer Programs and Databases in the Field of Nutrition*. Seattle, Wash., USA: University of Washington, 1985.

32. Hoaglin DC. Letter values: A set of selected order statistics. *In Understanding Robust and Exploratory Data Analysis,* ed. by Hoaglin DC, Mosteller F. and Tukey JW, pp. 33-57, esp. p. 38. New York: John Wiley and Sons, Inc., 1983.

33. Holden JM, Hardison NW, Salbe AD, Howe SM, Taylor PR, and Longnecker MP. Selenium (Se) composition of self-selected diets and retail and homegrown foods in a seleniferous area of South Dakota. FASEB J (*Abstracts*) 2(6), 1988.

34. Holden JM, Lanza E, and Wolf WR. Nutrient composition of retail ground beef. *J Agric. Food Chem.,* 34(2):302-308, 1986.

35. Hollman PCH and Katan MB. Bias and error in the determination of common macronutrients in foods: Interlaboratory trial. JAm Dietetic Assoc 88:556-563, 1988.

36. Hoover LW. *Concerns of users of nutrient data bases.* In Rand et al., reference 72, pp. 103-108.

37. Hoover LW, ed. *Nutrient Data Bank Directory*, 7th ed. Columbia, Mo., USA: University of Missouri, 1988.

38. Hoover LW and Perloff BP. *Model for Review of Nutrient Data Base System Capabilities* Columbia, Mo., USA: University of Missouri, 1981.

39. Jacob H. *Using Published Data: Errors and Remedies*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-042. Beverly Hills, Calif., USA, and London: Sage Publications, 1984.

40. Jacobs MB. *The Chemical Analysis of Foods and Food Products*, 3rd ed. New York: Krieger, 1973.

41. Jones DB. Factors for converting percentages of nitrogen in foods and feeds into percentages of protein. USDA Circular 183. Washington, DC: US Department of Agriculture, 1941.

42. Karmas E and Harris RS, eds. *Nutritional Evaluation of Food Processing*, 3rd ed. Westport, Conn. USA: AVI Publishing Co., 1988.

43. Klensin JC. *Systems considerations in the design of INFOODS.* In Rand et al., reference 72, pp. 212-223.

44. Klensin JC, Feskanich D, Lin V, Truswell AS, and Southgate DAT. *Identification of Food Components for INFOODS Data Interchange*. Tokyo: United Nations University, 1989.

45. Lanza E and Butrum RR. A critical review of food fiber analysis and data. *J Am Diet Assoc* 86:732-743, 1986.

46. Leung W-T W. *Food Composition Table for Use in Africa*. Rome: FAO, and Washington: US Department of Health, Education, and Welfare, 1968.

47. Leung W-T W. Butrum RR, Chang FH, Rao MN, and Polacchi W. *Food Composition Table for Use in East Asia*. Rome: FAO, and Washington, DC: US Department of Health, Education, and Welfare, 1972.

48. Leung W-T W and Flores M. *INCAP-ICNND Food Tables for Use in Latin America*. Guatemala: INCAP, and Washington: National Institutes of Health, 1961.

49. Matthews RH and Garrison YJ. Food *Yields Summarized by Different Stages of Preparation.* Agriculture Handbook No. 102. Washington, DC: US Department of Agriculture, 1975.

50. Merrill AL, Adams CF, and Fincher LJ. *Procedures for Calculating Nutritive Values of Home-prepared Foods: As Used in Agriculture Handbook No. 8, rev. 1963.* ARS 62-13. Washington, DC: US Department of Agriculture, 1966.

51. Merrill AL and Watt BK. *Energy Value of Foods--Basis and Derivation*. Agriculture Handbook No. 74. Washington, DC: US Department of Agriculture, 1955.

52. Meyer B. van Oosten-van der Goes HJC, van Staveren WA, and West CE. *Missing values in european food composition tables and nutrient data bases: Preliminary results of a survey.* Food Sciences and Nutrition. 42F(1):29-34, 1985.

53. Murphy SP and Rauchwarter D, eds. *Proceedings of the Tenth National Nutrient Databank Conference*. Washington, DC: US Department of Commerce, National Technical Information Service, 1985.

54. National Academy of Sciences. *Recommended Dietary Allowances*, 10th ed. Washington, DC: National Academy of Sciences, 1989.

55. NORFOODS. *Nutrient Calculation Systems in the Nordic Countries*. Uppsala: The Swedish National Food Administration, 1988.

56. Nutrition labeling of food. Code of Federal Regulations 21.101.9(e). Washington, DC: Office of the Federal Register, National Archives and Records Administration, 1989. (Available from: Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, USA.)

57. Oiso T and Yamaguchi K. *Manual for Food Composition Analysis*. Tokyo: Southeast Asian Medical Information Center, 1985.

58. Orr ML and Watt BK *Amino Acid Content of Foods*. USDA Home Economics Research Report No. 4, 1968 (Available from: Superintendent of Documents, US Government Printing Office).

59. Paul AA and Southgate DAT. McCance & Widdowson's *The Composition of Foods*. London: HMSO, 1978.

60. Pearson D. *The Chemical Analysis of Foods*. New York: Academic Press, 1976.

61. Pennington JAT. *Dietary Nutrient Guide*. Westport, Conn., USA: AVI Publishing Co, 1976.

62. Pennington JAT. *Bowes and Church's Food Values of Portions Commonly Used,* 15th ed. Philadelphia, Pa., USA: Lippincott, 1989.

63. Périssé J. *Heterogeneity in food composition table data*. Food Nutr 1:14-17, 1983.

64. Perloff BP. *Recipe calculations from the NFCS data base*. In Murphy and Rauchwarter, reference 53, pp. 11-21.

65. Petot GJ. *Managing a nutrient data-base system: Meeting users' needs and expectations*. In Rand et al., reference 72, pp. 123-133.

66. Piggott JR, ed. *Statistical Procedure in Food Research*. New York: Elsevier, 1987.

67. Polacchi W. *Standardized food terminology An essential element for preparing and using food consumption data on an international basis*. Food Nutr Bull 82(2):66-68, 1986.

68. Posati LP. *Procedures for imputing values*. In Murphy and Rauchwarter, reference 53, pp. 124-133.

69. Rand WM and Klensin JC. *Problems and status of food composition data*. In Glaeser PS, ed. Computer Handling and Dissemination of Data. Amsterdam: North Holland, Elsevier Science Publishers: 1987, pp. 400 403.

70. Rand WM, Klensin JC, and Campbell JA. *INFOODS: A world-wide effort to organize data on the composition of foods*. J. Can Diet Assoc 49(3):158-163, 1988.

71. Rand WM, Stuckey A, Valasevi A, and Tontisirin K, eds. *Proceedings of the First ASIAFOODS Conference*. Bangkok: Prayurawong Co., Ltd., 1985.

72. Rand WM, Windham CT, Wyse BW, and Young VR, eds. *Food Composition Data: A User's Perspective.* Tokyo: United Nations University, 1987.

73. Rand WM and Young VR. Report of a planning conference concerning an International Network of Food Data Systems (INFOODS) meeting held in Bellagio, Italy, January 30February 5, 1983. *Am J Clin Nutr* 39:144 151, 1984.

74. Rechcigl M Jr, ed. *Handbook of Nutritive Value of Processed Food - Vol I : Food for Human Use*. Boca Raton, Fla., USA: CRC Press, 1982.

75. Resources Council, *Standard Tables of Food Composition in Japan*. Tokyo: Resources Council, Science and Technology Agency, 1986.

76. Saguy I, ed. *Computer-Aided Techniques in Food Technology*. New York: Marcel Dekker, 1983.

77. Salunkhe DK, Jadhav SJ, and Yu MH. *Quality and nutritional composition of tomato fruit as influenced by certain biochemical and physiological changes*. Qual Plant-PI Fds Hum Nutr 24(1/2):85-113, 1974.

78. Schakel SF, Sievert YA, and Buzzard IM. *Sources of data for developing and maintaining a nutrient database*. l Am Diet Assoc 8$(10):1268 1271, 1988.

79. Scherz H and Kloos G. Souci, Fachman, Kraut: Food Composition and Nutrition Tables 1981/82 Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1981.

80. Scherz H. Kloos G, and Senser F. *Souci, Fachman, Kraut: Food Composition and Nutrition Tables 1986/1987*. Stuttgart: Wissenschaftliche Verlagsgesellschaft mbH, 1986.

81. Schubert A, Holden JM, and Wolff WR. *Selenium content of a core group of foods based on a critical evaluation of published data*. J Am Diet Assoc 87(3):285-299, 1987.

82. Snedecor GW and Cochran WG. *Statistical Methods*, 7th ed. Ames, Iowa, USA: Iowa State University Press, 1980.

83. Southgate DAT. *Availability of and needs for reliable analytical methods for the assay of foods.* Food Nutr Bull 5(2):30-39, 1983.

84. Southgate DAT. *Guide Lines for the Preparation of Tables of Food Composition*. Basel: S. Karger, 1974.

85. Stasse-Wolthuis M. *Influence of dietary fibre on cholesterol metabolism and colonic function in healthy subjects*. Wld Rev Nutr Diet 36:100 140, 1981.

86. Stewart KK. *The state of food composition data: an overview with some suggestions*. Food Nutr Bull 5(2):54-68, 1983.

87. Stewart KK and Whitaker JR. *Modem Methods of Food Analysis*. Westport, Conn., USA: AVI Publishing Co., 1984.

88. Tannenbaum SR. *Nutritional and Safety Aspects of Food Processing*. New York: Marcel Dekker, Inc., 1979.

89. Tannenbaum SR, Young VR, and Archer MC. *Vitamins and minerals*. In Fennema OR, ed. Food Chemistry. New York: Marcel Dekker, Inc., 1985.

90. *Thomas Register of American Manufacturers and Thomas Register Catalog File*, 78th ed. New York: Thomas Publishing Society, 1988.

91. Thompson DR. The challenge in predicting nutrient changes during food processing. *Food Technol* 36:97-115, 1982.

92. Trowell H. Southgate DAT, Wolever TMS, Leeds AR, Gassull MA, and Jenkins DJA. Dietary fibre redefined (letter to the editor). Lancet 1: 967, 1976.

93. Truswell AS, Bateson D, Madafiglio K, Pennington JAT, Rand WM, and Klensin JC. INFOODS *Guidelines for Describing Foods*. In preparation.

94. Tukey JW. *Exploratory Data Analysis.* Reading, Mass., USA: Addison-Wesley, 1977.

95. US Department of Agriculture. Composition of Foods: Raw, Processed, Prepared. Agriculture Handbook No. 8. Washington, DC: Science and Education Administration, US Department of Agriculture, 1963.

96. US Department of Agriculture. *Composition of Foods: Raw, Processed, Prepared* Agriculture Handbook No. 8, revised, 8-1 through 8-17, 8-20, and 8-21. Washington, DC: Science and Education Administration, US Department of Agriculture, 1976-1989.

97. US Department of Agriculture. Provisional Table on Percent Retention of Nutrients in Food Preparation. Washington, DC: Human Nutrition Information Service, US Department of Agriculture, 1984.

98. US Department of Agriculture. Provisional Table on the Dietary Fiber Content of Selected Foods. Washington, DC: US Government Printing Office, 1988.

99. US Department of Agriculture. *Administrative Report No. 378*. Washington, DC: Human Nutrition Monitoring Service (HNIS), US Department of Agriculture, 1989. Updated frequently and available upon request from HNIS.

100. Vanderveen JE and Pennington TAT. Use of food composition data by governments. *Food Nutr Bull* 5(2):40-45, 1983.

101. van Staveren WA. Personal communication. December 1989.

102. West CE. Eurofoods: towards compatibility of nutrient data banks in Europe. *Ann Nutr Metab* 29(S1), 1985.

103. West CE. Inventory of European food composition tables and nutrient database systems. Uppsala: The Swedish National Food Administration, 1989.

104. Williams S, ed. *Official Methods of Analysis of the AOAC*, 14th ed. Arlington, Va., USA: Association of Official Analytical Chemists, 1984.

105. Windham CT, Hansen RG, Wyse BW, and Sorenson AW. Uses of nutrient data bases for identifying nutritional relationships to public health and nutrition education in the United States. *Food Nutr Bull* 5(2):46-53, 1983.

106. Wood D, ed. *Trade Names Dictionary*, 5th ed. Detroit, Mich., USA: Gale Research Co., 1986.