Food Chemistry 43 (1992) 289–293
Analytical Methods Section

Trace elements: databases and food composition compilations

D. A. T. Southgate

AFRC Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA

While it is nutritionally desirable to include values for trace elements in compilations of food compositional data such as nutritional databases, their inclusion does create special problems for the compilers.

These relate to the availability of data for many of the trace elements that conform to the criteria for data quality that many users of the databases require.

The natural variability in all concentrations of many trace elements in foods effectively imposes limits on the predictive accuracy of any database and it is ideally desirable to have sound measures of statistical variability in order to maximise the benefits of including values for trace elements.

INTRODUCTION

Intuitively from the viewpoint of a nutritionist the ideal nutritional database or food compositional compilation should be as comprehensive as possible. There are many users of these compilations who would like to see a complete coverage of those trace elements which have established roles as essential nutrients, and those concerned with studies of the relation between diet and health and disease see advantages in extending the coverage to include those trace elements which have toxic properties.

This paper will be restricted to the discussion of the nutrients and the issues that compilers of databases have to address in meeting the requirements of users will be considered first. Second will be considered the performance of such compilations and the factors that determine their predictive accuracy in use.

In this paper the term database will be used for any compilation of food compositional data, both printed and computerised, because the principles are common to both types of compilation. When there is a need to distinguish between printed and computerised databases the adjectival qualifiers will be used.

The compilers have three principal questions to address: which nutrients to include?; which food items to include? and, finally, how to acquire the data?

The first two questions are very closely linked to the users' requirements and it is important to recognise that the usefulness of compositional data bases, although they are an information source in their own right, depends on the way they perform. The uses of nutritional compositional databases fall principally into two categories: Nutritional Analysis, where they are used to calculate nutrient intakes from data on food consumption, and Nutritional Synthesis, where they are used to design dietary mixtures that have prescribed nutrient compositions. These two types of use are carried out at several levels, each of which has specific constraints depending on the level at which food consumption is measured and the predictive accuracy that is required. The focus will be on the individual level because this is the more important as regards trace elements.

Food Chemistry 0308-8146/92/$05.00 © 1992 Elsevier Science Publishers Ltd, England. Printed in Great Britain

TRACE ELEMENTS DESIRABLE IN A NUTRITIONAL DATABASE

The elements that most nutritionists would like included are given in Table 1. This includes those elements for which there are specific recommendations and those which are known to be nutritionally essential, either because deficiency states are known in man, or they have been demonstrated experimentally in animals. The table gives the nutrients for which values are given in the US National Research Council's recommendations (1989). It is important to recognise that many nutritionists believe that recommendations should be moving towards the concept of optimal levels of intake based on functional indices of status, rather

Table 1. Candidate trace elements for inclusion in databases (Values for adult male)

	Reference nutritional intake (per day)	Safe and adequate (per day)
Iron	8·7 mg	
Zinc	9·5 mg	
Copper	1·2 mg	1·5–3·0 mg
Manganese		2–5 mg
Chromium		50–200 μg
Molybdenum		75–200 μg
Selenium	75 μg	
Iodine	140 μg	
Fluoride	—	1·5–4 mg

National Research Council, 1989.

than levels of intake to prevent deficiency states which are rarely seen in humans (Department of Health, 1991). Most workers would think that this list of candidates for inclusion would be a reasonable target. In fact most current compilations include the elements at the top of the list.

COVERAGE OF FOOD ITEMS

The very large number of different foods eaten, reasonably regularly, by a developed population such as in the UK creates one of the most difficult problems for the compiler and is one reason which his work is never complete. The fourth edition of *The Composition of Foods* contains about 1000 food items (Paul & Southgate, 1978) and the process of revision will almost certainly expand this number to probably 5000 foods; however, a typical supermarket may have upwards of 15 000 items and many of these are ingredients. If one considers how many different combinations of ingredients are used in the preparation of the mixed dishes which form a substantial portion of our diet, then probably 100 000 food items would be needed in a comprehensive database. The magnitude of the analytical task that this would involve will be clear to all, so we have to live with the position that databases will only be able to cover a limited range of food items. This has important implications with regard to the performance of databases. The size of existing databases is sufficient to create problems when one wishes to include values for a 'new' nutrient, that is, one not included in the present database.

COMPILER'S REQUIREMENTS FOR ANALYTICAL METHODOLOGY

One cannot include values for any nutrient for which a satisfactory analytical method does not exist. A compiler's requirements are more demanding than this; one needs to know that the element can be measured accurately in a wide range of food matrices. It is possible to include interim values while analytical methodology is developing but from experience this can create difficulties, some of which have nothing to do with science! Stewart (1980) compiled a list of the state of the art for the analysis of most nutrients. It is the author's impression that some progress has been made in resolving the problem areas, principally due to the increased numbers of collaborative studies and the growing recognition that quality assurance in the analytical laboratory is essential. The expanding range of Certified and Standard Reference Materials has been important in this improvement (Wolf, 1985; Southgate, 1987). At the present time the major problem areas are the demonstration of the applicability of the methods to a wide range of matrices coupled with intermethod comparisons. The ideal is to have two or three methods, based on different principles, giving analytically comparable values which provide reassurance regarding systematic bias.

THE ACQUISITION OF DATA

There are two basic strategies for getting the data one needs (Southgate, 1974). First, in what might be called the direct method, one can commission an analytical programme to obtain the data directly. The second approach, the indirect method, is to compile the data from the literature including both published and unpublished sources.

Direct analytical method

This involves identifying the food items for which one would like data. To do this one needs to establish priorities because one cannot analyse every food. Sampling and analytical protocols have to be set up and implemented. This approach has several advantages in that one has control over the identity of the food item, one can ensure that the sampling is representative and that the analytical methods used are sound and conducted within a proper quality assurance programme (Greenfield & Southgate, 1985).

The main disadvantages of this approach are that it is very demanding of resources and is time-consuming. It is possible to argue that it is also wasteful in that it does not take advantage of the large numbers of analytical results available from a variety of sources.

The indirect method; compilation from the literature

This, although superficially less demanding of resources, requires detailed literature searching to identify sources of data. Once these have been archived the data are then evaluated against a series of criteria (Southgate, 1974).

The first series of criteria relate to the identity of the sample analysed and whether the sample taken for analysis was representative of the food item that it is desired to include in the database. This is especially important for many trace elements as the levels are dependent on the soil where the plant food was grown and the agronomic practices used in its cultivation. In the case of foods of animal origin the type of husbandry and the sources of the feeds may be critical. The second group of criteria relate to the preparation that the sample received before analysis since this will affect the levels of some trace elements.

The analytical methods themselves need to be evaluated giving close attention to the quality assurance procedures in use.

Finally the modes of expression of the data have to be scrutinised and converted into the modes of expression that have been agreed for the database.

It is extremely difficult to apply these criteria to much of the literature and it is often difficult to know whether the reported differences between values for nominally the same food are analytical artefacts or represent the true range of concentrations likely to exist. The remedy lies in the hands of the scientific community; journal editors are often under pressure to constrain authors who wish to publish all the details that these criteria require and we need archival arrangements to make these data available in other ways.

Pragmatic approaches

In practice most compilers of databases use a combination of the two approaches. In doing this they work against a decision tree of the type shown in Fig. 1. The first essential is to establish whether the values for the food in question are available in the literature, to assess the values against the criteria and then if the data are not available or of doubtful quality, and the food is sufficiently important, to set up the appropriate sampling and analysis.

THE PERFORMANCE OF DATABASES

While the compilers can do their utmost to ensure that a database contains data that meet the most rigorous criteria, in many cases pragmatic factors intervene to restrict the coverage of nutrients and food items and the acceptance of data whose provenance is not as well established as the compiler would like. These factors mean that all databases do not meet the ideal specifications of users; in particular they are never comprehensive in respect of food items and they are rarely complete in respect of nutrient values. Furthermore the fact that foods are biological materials showing natural variations in composition imposes a limit on their predictive accuracy. All these factors combine to affect the performance of the database.

As has been said earlier, databases are most commonly used in two types of ways: nutritional analysis where the database is effectively used to translate data on food consumption into nutrient intakes, and nutritional synthesis where the database is used to derive dietary formulations that will have a specified nutrient content.

Performance in nutritional analyses

This is at the present time the major mode of use; the user expects the database to give a reliable estimate of the trace element content of a diet. The level of accuracy required is only rarely stated but most people would expect the calculated value to be comparable

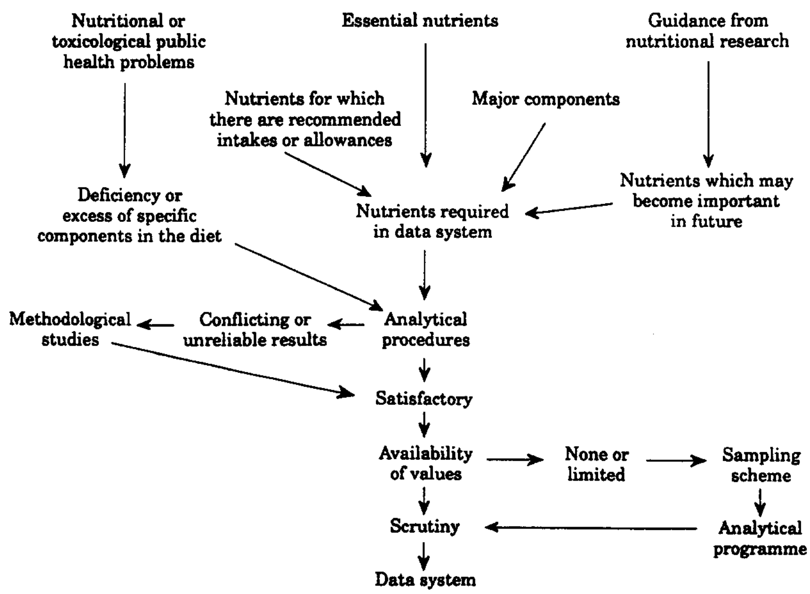

Fig. 1. Decision tree for obtaining nutrient data.

with, and ideally the same as, an analysed value for that diet. There are several factors that limit the agreement or the predicted accuracy. The estimates of food intake are usually the most important source of error or uncertainty for the major nutrients and are also important for trace elements but here other factors become more important. Missing values are of major concern particularly when the database is being extended to include a trace element. Missing food items are, and, for the reasons discussed above, will remain a continuing problem and the pragmatic solution of using the data for a related food is one approach. Missing values for a trace element create more difficult problems with a computerised database. When using a printed table the absence of data can be seen; in a database the datafile is hidden and unless the missing value is flagged one can obtain data for intakes that are underestimates and sometimes gross underestimates. Some notable papers have been written because the authors were unaware of this problem. A neutral pragmatic approach is to assign an average trace element concentration to the weights of missing foods but the only really sound aproach is to carry out new analyses.

The final limitation is one alluded to before: the limitations posed by the natural variations in the composition of foods (Paul & Southgate, 1988). This sets a limit on the predictive accuracy of a database and this limit can only be assessed if data are available on the natural variations in composition. Databases are therefore only capable of providing estimates of the probable range of nutrient intakes from a dietary mixture. Since the compilers have striven to provide representative data one would expect calculated values to approach analytical values as the numbers of samples increase, so that with more individual studies over longer periods the agreement should improve.

Those users whose work requires high accuracies in measured intakes should follow the maxim that 'there is no substitute for analysis'. Other users may be content with semi-quantitative estimates; often rankings of intakes of individuals may suffice. These users, provided that the missing value issues are recognised, can be reasonably happy with nutritional databases.

Performance in nutritional synthesis

In this type of use, missing food items do not present a major problem. However there is a virtually infinite range of foods that can be combined to meet a specific nutritional prescription and most programs have to include constraints based on food preferences or normal food patterns and a missing food can cause minor annoyance; missing values for nutrients can make the food combinations rather more restricted than necessary, because when a trace element intake is specified, foods without values will not be selected. However, at the present stage of development, most synthesis packages are focused on macronutrient specifications and on those micronutrients for which the existing databases are reasonably complete. Some researchers would also like reassurance that when they have set the macronutrients they have not produced imbalances. The accuracy of the prediction is usually less critical for one would be trying to set intakes within a range.

Databases as an information source

There is one other type of use that we must consider. If one is using the database as a source of information, for example, in preparing a nutritional label, missing values are a disappointment; however, in making judgements about the validity of a value for a nutrient one needs to know the natural variation in the concentration of the nutrient in the food being interrogated. The concentrations of micronutrients, especially those of the trace elements, vary over a wider range than those of many macronutrients and therefore, as far as the trace elements are concerned, this type of use introduces a requirement that is only rarely seen in food compositional databases, namely estimates of natural variance in composition. As has been said earlier, this is critical when using the database in an analytical mode because it determines the predictive accuracy one can expect from the database.

TRACE ELEMENTS IN DATABASES: FUTURE DEVELOPMENTS

What is required if we are to achieve the aim of a comprehensive coverage of trace elements?

Firstly, it is essential to expand the coverage so that values are available for the majority of foods that are consumed. Missing values need to be clearly flagged to preclude underestimation or misuse.

The analyses that this will involve must be conducted with proper recognition of the importance of quality assurance. The details that the compilers need to evaluate the data must be available, either in the primary published source, or via an archive of original data.

In order to use the database effectively it is important to have estimates of the natural variability in the concentrations of the trace elements (and other nutrients as well!) because these are essential for estimating the predictive accuracy of calculations made from the database.

Finally it would be very useful if the numerical database were linked to an information system which gave the user indications of the major factors causing variations in the concentration of micro and other nutrients.

REFERENCES

Department of Health (1991). *Dietary Reference Values for Food Energy and Nutrients for the United Kingdom.* Report Health and Social Subjects No. 41, HMSO, London.

Greenfield, H. & Southgate, D. A. T. (1985). A pragmatic approach to the production of good quality food composition data. *Asean Food J.*, **1**, 47–54.

National Research Council (1989). *Recommended Dietary Allowances 10th edn.* National Academy of Sciences, Washington, DC.

Paul, A. A. & Southgate, D. A. T. (1978). *McCance and Widdowson's The Composition of Foods 4th edn.* HMSO, London.

Paul, A. A. & Southgate, D. A. T. (1988). Conversion into nutrients. In *Manual for Methodology for Food Studies*, ed. M. E. Cameron & W. J. van Staveren. Oxford Medical, Oxford, pp. 121–42.

Southgate, D. A. T. (1974). *Guidelines for the Preparation of National Tables of Food Composition*. S. Karger, Basel.

Southgate, D. A. T. (1987). Reference materials for improving the quality of nutritional compositional data for foods. *Fresenius Z. Anal. Chem.*, **326**, 660–4.

Stewart, K. K. (ed.) (1980). *Nutrient Analysis of Foods: State of the art for routine analysis.* Association of Analytical Chemists, Arlington, VA, pp. 1–19.

Wolf, W. R. (1985). *Biological Reference Materials: availability, uses and needs for validation of nutrient measurements.* John Wiley, New York.