WHAT IS A NUTRIENT DATA BASE?

Grace J. Petot

Case Western Reserve University

Each year the list of nutrient data bases and systems grows longer. As more food composition information becomes available, it becomes even more practical to computerize it. As we continue to make comparisons among the various data bases we become more confident about the resulting calculations. A computerized food composition table, a nutrient data base, is the heart of the data base system; therefore it is essential that the compilers provide descriptive information and documentation for users. It is then the user's responsibility to become aware of the integrity of the data and any limitations in their use.

In today's discussion I will start with some definitions, describe the elements of a nutrient data base, identify the related types of data required and conclude with a list of issues still confronting us.

DEFINITIONS

A *data base* may be defined as "a collection of interrelated data stored together without harmful or unnecessary redundancy to serve one or more applications in an optimal fashion; the data are stored so that they are independent of programs which use the data; a common and controlled approach is used in adding new data and in modifying and retrieving existing data in the data base" (1).

A *nutrient* is a "substance obtained from food and is used in the body to promote growth, maintenance, and/or repair" (2).

For purposes of our discussion, we will consider that the primary elements of a nutrient data base are foods and any constituents of foods as consumed, which we will call nutrients. All other data included are related to these primary elements and are used for identification, documentation and applications. The size of the data base may or may not be related to the degree of specificity of the information included and will be determined by the user.

THE DATA

Foods

A reference data base, such as USDA Data Base I, includes a very large number of foods with unambiguous descriptors. These descriptors include food group, species, variety, maturity, season of year, geographical location, soil conditions, processing or preparation method, form, cut, size, brand name, packaging, etc. Not all of this information is necessary for all applications. This has led to the development of a series of data sets of varying smaller sizes. As this kind of aggregation and reduction occurs, the naming of many food items in common, easily interpreted names has presented problems. Food identification systems have been developed, such as the Factored Food Vocabulary being used by the Food and Drug Administration (3). The Nutrition Coding Center at the University of Minnesota is developing a hierarchical food identification system with coded descriptors to facilitate interactive food identification dialogue (4).

The selection of which foods to include is dependent on the users' needs. Since the capabilities of hardware at affordable prices have increased, the size of the database has become of lesser concern. However, a large number of foods, in some cases, may not be the optimal choice as I am sure my colleagues will point out today.

Nutrients

Nutrient information may be obtained from a wide variety of sources. In many countries, the

national food composition table, compiled under governmental direction, serves as the primary source. In the United States, USDA Handbook 8 is this source. These data are available in printed and machine readable forms. Additional sources of nutrient data from USDA include specialized data sets used for food consumption surveys, Home and Garden Bulletin 72, Provisional Tables published as nutrient data become available prior to inclusion in a revised Handbook 8 section and articles published in the research literature.

Other primary sources of nutrient information include food composition tables from other countries, published journal articles and manufacturers' information. With the advent of nutrition labeling in the 1970's, most manufacturer's have generously made results of their laboratory analyses available. Some data base compilers have funded independent laboratory analyses of food products.

Among the many secondary sources of data are tables published in books, such as Pennington and Church's Food Values of Portions Commonly Used, or machine readable data sets compiled independently.

In spite of the seemingly great availability of food composition data, we do not yet have information for all nutrients for all foods and all forms of foods. Therefore, calculated and imputed nutrient values are often used. USDA has used these methods and has carefully documented them in the explanatory information in each Handbook 8 revision volume. All imputed values are flagged in the printed and machine readable tables. Any calculation and imputations made by a data base compiler must be precisely documented.

Nutrient values for recipes or formulated foods are frequently included in nutrient data bases. The validity of these calculations will depend on the accuracy of yield factors applied to ingredients, to total recipe yield, and to consideration of nutrient retention.

Most nutrient data bases use nutrient values for a standard weight of food, e.g., 100 grams. This allows for uniform validation checks in a systematic manner. A decision must be made on the number of decimal places to include. This may depend on available computer space, operating system capabilities, and desired level of precision related to use of the data.

Nutrient nomenclature and units from different sources frequently differ and it is wise to establish a single standard such as those used for the current Recommended Dietary Allowance or USDA Handbook 8. Conversion factors may be included as related nutrient information in the data base to reduce calculation errors. Additional related information for each nutrient may include the source or reference, sampling method, analytical method, number of samples, variability with median, standard deviation or data quality codes, and date of inclusion of data or revision. Nutrient data for one food item may be obtained from several different sources; therefore it is important that source information be included for each nutrient. Maintenance of the nutrient data is an ongoing process and requires quality control procedures to maintain integrity. Most of these can be automated in a data base management system. The following are examples of such procedures:

- The sum weight of protein, fat, carbohydrate, alcohol and ash to equal standard weight, e.g., 100 grams
- Energy factors applied to protein, fat, carbohydrate and alcohol to equal total calories
- Range checks for each nutrient and food group for reasonableness of values
- Duplicate entry of manually entered nutrient data
- Sum of nutrient components to equal total weight of nutrient, e.g., total fatty acids not greater than total fat

Missing nutrient values should never be represented as zeros. Just as imputed and calculated values are documented, these should be documented so that a retrieval system will make these known to the user.

WHAT IS A NUTRIENT DATA BASE?

Related Information

A third critical element in any nutrient data base is a systematic identification of foods and nutrients which allows for translation of food names and nutrients into a machine readable form. Alphabetic and numeric codes, or a combination of both may be very simple or quite complex. A hierarchical code may include food item descriptors which begin with a food group and range through more specific levels of description. A decision tree approach may be used based on logical queries by the user. Both of these approaches may be made invisible to the user (4).

Weight-volume equivalent factors, nutrient retention factors and yield factors are usually included. Other information related to foods may include color, texture, use frequency, cost, menu group, etc. User needs and practicality will determine the extent of additional data to include.

DATA BASE MAINTENANCE

A nutrient data base may be static or dynamic in nature. Most static data bases have been compiled for specific purposes. New foods come into the market place, reformulations occur, nutrient fortification regulations change, new nutrient information becomes available and consumer consumption behaviors change continuously. It is therefore necessary to keep a nutrient data base current. As this is done, data validations should be made and documented. If at all possible, a data base management team should have members or consultants who have knowledge of food science, food composition, nutrients and computer data base structure. Awareness of users' needs and ongoing dialogue with users is also necessary. The proceedings of nutrient data bank conferences contain a wealth of information for data base compilers and users.

CURRENT ISSUES

Since the first data bank conference in 1976, we have been identifying, discussing and resolving issues. We have come a long way since then and recognize that the mutual support fostered by these conferences has led to greater knowledge in the community of users and compilers.

Work yet to be done includes:

- Food nomenclature conventions
- Common conventions for food descriptors
- Nutrient retention
- Food yields
- Weight - measure equivalents (density factors)
- Missing nutrient values
- Recipe calculations
- Timely availability of data
- Standard analytic procedures
- Variability among data bases
- Standardized documentation procedures

CONCLUSION

A nutrient data base is an organization of food composition values, related information and supporting documentation. Its size may range from being large and complex to very small and specialized. It may be static or dynamic. The discussions which follow will elaborate on applications and selection of a nutrient data base to meet specific needs.

REFERENCES

1. Martin, J. *Computer Data-Base Organization.* Page 19. Prentice-Hall, Inc., Englewood Cliffs, NJ 1975
2. Whitney, E.N. and Cataldo, C.B. *Understanding Normal and Clinical Nutrition.* Page 4. West Publishing Co., NY 1983
3. Pennington, J.A.T. Food composition data base features. Proceedings of Tenth National Nutrient Data Bank Conference, San Francisco, CA U.S. Dept. of Commerce, NTIS. Springfield, VA 1985
4. Arab, L. Coding and entry of food intakes. Proceedings of Eighth National Nutrient Data Bank Conference, Minneapolis, MN. U.S. Department of Commerce, NTIS. Springfield, VA 1983
5. Buzzard, I.M., Ellis, L.B.M. and Asp, E.D.H. A Microcomputer-based interactive model for collection and coding of dietary data. Proceedings of Eighth National Nutrient Data Bank Conference, Minneapolis, MN. U.S. Dept. of Commerce, NTIS. Springfield, VA 1983

SOURCES OF NUTRIENT COMPOSITION OF FOODS

1. USDA Publications
2. USDA - Reports Published in Journals
3. Laboratory Analyses - Published in Journals
4. Directly from Industry
5. Tables Published in Books
6. Independently Compiled Data Bases
7. Labels On Food Products
8. Own Laboratory Analyses
9. Calculations of Recipes - From One or More Data Sources
10. Imputed - Varying Criteria