NUTRIENT DATA BASES: AVAILABILITY, OPTIONS AND RELIABILITY

Betty Perloff
Consumer Nutrition Division
Human Nutrition Information Service, USDA
Hyattsville, Maryland 20782

The key parts to any Nutrient Data Base System are the computer, the programs (software), and the nutrient data base. The nutrient data base is the collection of nutrient values that is accessed by the programs as nutrient values are needed for calculations.

AVAILABILITY AND OPTIONS

Nutrient data bases currently available from the U.S. Department of Agriculture (USDA) either correspond to a USDA publication, or were developed to be used with the USDA Nationwide Food Consumption Survey. Agriculture Handbook No. 8 is the Department's standard reference on the composition of foods. A complete revision of the handbook was published in 1963 (Watt and Merrill, 1963). It is now undergoing another revision which is being completed and released in separate sections according to food groups. Nine sections have been released to date (U.S. Department of Agriculture, 1976-1982), and an additional 14 sections will be published over the next few years. The nutrients covered in the current revision and the earlier edition are shown in Table 1.

Table 2 lists computerized data sets of nutrient values (nutrient data bases) prepared at the Consumer Nutrition Division of USDA's Human Nutrition Information Service. The Nutrient Data Base for Standard Reference corresponds in both content and format to the revised sections of Handbook 8 and contains the largest number of nutrient values. As new sections of the handbook are completed, the data set is updated and a new version is released. Each version is identified by a release number and date. Until the revision is complete, releases of this data base will be supplemented with data from the earlier edition of the handbook. Of course, data from the earlier handbook are limited to the nutrients that were reported in that edition.

Data Sets 456-3 and 456 correspond to the 1963 edition of the handbook. They contain identical nutrient data, but in 456-3, data are expressed on the basis of 100-gram edible portions, and in 456, on the basis of household measures. Data Set 72-1 is an abbreviated data base corresponding to Home and Garden Bulletin No. 72, which is a food composition publication intended for consumers' use.

Imputed values have been included in Data Sets 456-3, 456, and 72-1 where blank spaces appeared in the publications. Imputed values have been included in the Standard Reference Data Base for most of the blank spaces in the revised section of the handbook. Exceptions are blank spaces for: spices and herbs; nutrients in food groups where analytical data were unavailable for all foods within the group (e.g., copper in dairy products); and fatty acids, cholesterol, phytosterols, and amino acids for baby foods and foods from plant products. Values for enriched flour and bread and other products made with enriched flour have been updated to reflect the current standards of identity.

The last two data bases listed on Table 2 were developed for the analysis of data from the USDA Nationwide Food Consumption Survey 1977-78. The Data Base for Individual Intake Surveys contains data on the 100 gram basis for those food items reported as being consumed by respondents in the individual phase of the survey. The Data Base for Household Food Use Surveys contains data for 1 pound as purchased for food items reported in the household phase of the survey. Cooking losses were deducted for vitamins in this last data base. Values were imput for both survey data bases where data were unavailable.

These six data bases are available on magnetic tape from the National Technical Information Service (NTIS) at the current cost of $140 per tape. Information about tape formats and instructions for ordering the tapes may be obtained by writing the Consumer Nutrition Division[1]. The USDA data bases do not include computer programs.

The Nutrient Data Base for Standard Reference is used in the Consumer Nutrition Division to create other, more specialized, data bases. Figure 1 shows the interrelationship between the Standard Reference Data Base and the Individual Survey Data Base. The large box represents our Nutrient Data Bank. The three circles within this box represent the three levels of data with the Bank: Level 1 (sometimes called Data Base 1) = individual values; Level 2 (Data Base 2) = preliminary summary; Level 3 (Data Base 3) = final summary.

The food items for which data appear on the Survey Data Base are foods that respondents consumed during the survey reporting period. The items are frequently composites, or recipes, of two or more foods from the Standard Reference file. For food items for which data summarization is incomplete, information may be drawn from the data bank at Level 2. If there are no data available for an item, values are imput or estimated. The food identification numbers on the survey data base are different from those on the Standard Reference Data Base because they were designed to allow for specific data processing needs for the survey.

There are also several commercial data bases available and a directory of these data bases is being prepared by a volunteer committee chaired by Dr. Loretta Hoover. The directory will contain information on all available nutrient data bases, other than USDA data bases, that the committee could locate and for which they could obtain a completed survey questionnaire. Dr. Hoover is making arrangements to have a copy of the directory mailed to each conference registrant, and additional copies may be obtained through the University of Missouri.

The different commercial sources listed in the directory offer different products and services--from data bases that are basically copies of the USDA data, to data bases with greatly expanded coverage. Most of the commercial data bases are accompanied by computer programs that can be used with their respective data bases. Some of these products may be purchased, some may be accessed through a remote terminal. Some of the sources will perform the data entry and computer processing for users who do not have access to a computer or a terminal. The directory summarizes the services that each source offers, and provides the name of a contact person and phone number for obtaining additional information.

[1]Consumer Nutrition Division, Human Nutrition Information Service, U.S. Department of Agriculture, Room 304, 6505 Belcrest Road, Hyattsville, MD 20782.

OPTIONS

Summarizing briefly the options that are available for using a computerized nutrient data base, a user may:

Purchase a data base and prepare the computer programs (the data base may be from USDA, or a commercial source);
Purchase both the data base and programs;
Contract to use a system through a remote terminal; or
Contract with a nutrient data base system for nutrient data calculations.

The choice depends on which option will suit specific needs. Regardless of the option chosen, accuracy of the data and the programs must be of primary concern to the user. Studies comparing nutrient data base systems have shown that different systems may give different results (Danford, 1981; Hoover, 1981). Causes of the differences may be the nutrient data, the computer programs, or coding differences.

There are several reasons why nutrient values may differ from one system to another, even though literature from most systems cite USDA as their data source. One reason is that data from USDA are being updated, and some systems may be using information that has been superseded. Also, systems may supplement their data bases with information from other sources, for food items or nutrients not covered by USDA. If two systems use different sources for supplementary data, the values may differ. Other systems may choose not to supplement their data bases and may instead perform calculations with the data that are present, flagging the calculation results when data for an item are missing.

RELIABILITY

The nutrient content of any food varies naturally from sample to sample, and conditions under which foods are grown or stored, as well as any processing or preparation treatments, may increase the variability. The only way to know the exact nutrient content of a particular food sample is to perform analytical measurements on that sample in a laboratory. Such measurements must be made for metabolic studies or other projects for which precise knowledge is essential. Agriculture Handbook No. 8 provides data on the nutrient content of foods that can be used when direct measurements are not practical.

The process of deriving nutrient values for Agriculture Handbook No. 8, including the current revision, differs as circumstances prescribe, but the goal always is to present values that are representative of products on a year-round, nationwide basis. For that reason, the data published in the handbook are usually referred to as "representative values". Users should not mistakenly interpret this term to infer that the values are not derived from analytical data. Except for fortified nutrients in some foods, all of the values in the handbook are based on analytical data. Most of them are arithmetic means of several individual values.

Values that are not themselves averages of actual analytical data are usually calculated in some particular way from the analytical values. Protein, for example, is not analyzed directly but is calculated from the average nitrogen value for each food. Calculations for some food mixtures are made using analytical values for the ingredients in the recipe, adjusting the total values for changes in moisture or fat content, and for losses in vitamins and minerals, that occur during cooking of the food.

Individual analytical values are collected by USDA from a variety of sources, primarily from the food industry, scientific literature, and government laboratories. Individual values for each food item are stored in USDA's computerized Nutrient Data Bank, organized into subgroups according to the unique characteristics of the individual foods sampled, analyzed or studied to determine the factors that affect each item's nutrient composition, and summarized to obtain the mean values reported in Handbook 8.

A critical factor in the reliability of nutrient data is the analytical methods that are used for measuring the nutrient levels in foods. Dr. Gary Beecher, Chief of USDA's Nutrient Composition Laboratory, recently prepared a chart showing the current status of nutrient methodology (reproduced following Figure 1). According to this chart, most nutrients in Handbook 8 have acceptable methods. Note, however, that some of the nutrients designated as having methods that give conflicting results do appear in the handbook. Ten components are designated as lacking suitable methods; none of these components appear in Handbook 8.

APPRAISING NUTRIENT DATA BASES

Most commercial data bases use one of the USDA data bases as a core, supplementing it with nutrient values from other sources to provide coverage beyond the coverage available from USDA. To evaluate a commercial nutrient data base, first obtain information to answer the following series of questions:

1. Data Base Core

   Which USDA data base is used as the core?

   What was the date of its release?

   Have any revised sections of Agriculture Handbook No. 8 been published since the release date, and if new sections have been published, have the revised values been incorporated into the data base?

   If the release date precedes 1977, have values for B vitamins in enriched bread and flour, and foods using enriched flour, been updated to reflect the current standards of identity for those products? (The standards were revised in 1975. USDA data bases released in 1977 and thereafter were changed accordingly.)

   If the release date precedes 1982, have values for iron in the enriched flour and bakery products been updated? (New standards for iron became mandatory July 1, 1983. The revised iron levels could be used voluntarily by manufacturers before the effective date, and the 1982 USDA data releases reflected this revision.)

2. Supplementary Data

What, if any, data have been added to the data base?

What are the sources for the supplementary data?

Do personnel selecting the supplementary data have a food science, dietetics, or similar background?

What precautions are taken to ensure reliability and accuracy of the supplemental data?

3. Nutrient Coverage

Are any nutrients included in the data base that are not included in Handbook 8? If additional nutrients are included, how complete is their coverage?

Are any of the nutrients in the data base included in revised sections of Handbook 8 but not in the earlier edition, i.e., B_6, folacin, pantothenic acid, magnesium, zinc? Their coverage will be incomplete until all sections of the handbook revision are published. How complete is their coverage in the data base for those food groups that have not yet been revised--that is, how many of the food items actually have values present for these nutrients?

Are any nutrients which have questionable analytical methods included in the data base? (Biotin, choline, molybdenum, vitamin K, cobalt, heme-iron, nonheme-iron, silicon, tin, or vanadium). If any of these nutrients are included in the data base, how complete is their coverage? (Remember, data for these nutrients are very limited and their analytical methods need further study.)

If dietary fiber is included in the data base, is it identified properly? (Dietary fiber consists primarily of cellulose, hemicellulose, lignin, pectin, and gums. Different methodologies, however, measure different combinations of these components. The neutral detergent method has become increasingly popular in the last few years, but it measures only the insoluble components: cellulose, some of the hemicellulose, and lignin. These data should be identified as insoluble dietary fiber, or as neutral detergent fiber. They do not represent the total dietary fiber in foods containing soluble components.)

4. Management

Is information about the data base core, past updates, and inclusion of supplementary data well-documented and readily available?

Is the data base manager familiar with Agriculture Handbook No. 8, including the status of the current revision?

Once answers are assembled for the above questions, they can be used to judge the status of the data base, the completeness of the nutrient coverage, and whether sufficient effort is being made by the data base management to provide a high quality product.

RESPONSIBILITIES OF A DATA BASE OWNER

Once a data base is selected and purchased, new responsibilities are acquired if the data base is to be used over a period of time. The data must be updated as new releases from USDA are made. If data are to be added from other sources, criteria for data selection should be established. All updates and additions should be completely documented. Whenever the data base is used in a research project, the name, release number, and release date of the USDA core data base, as well as information about updates and added data, should be included in the research report.

Responsibilities also extend to ensuring the accuracy of computer programs that are written for use with the data base. Any program that you purchase, that you prepare yourself, or that is prepared for you by someone else should be checked against a set of hand calculations.

It is also important for a system's users to understand the calculating procedures that are employed by the system. Our office has cooperated with the University of Missouri at Columbia to develop a model[2] (Hoover and Perloff, 1981) that can help a user review a nutrient data base system. The model has three main parts. The first part contains questions for which any user should know the answers relative to any system that he/she is using. Second, there is a series of five computer processing tasks that must be completed using the nutrient analysis system. These include: updating the data base, calculating a simple recipe, reporting baseline data, converting nutrients to various portion sizes, and computing a dietary record. The third part of the model is an interpretation guide to facilitate evaluation of the results from the computing tasks; it includes reference data calculated from USDA sources for comparison with the computer output, as well as several questions which should be answered during the computer output evaluation. The model was prepared as a prototype and may not suit all purposes, but individuals are free to utilize the methodology to prepare a more specialized tool for their own purposes.

CONCLUSION

Many decisions face potential nutrient data base users. The first step toward becoming a responsible user is to recognize that all nutrient values, if not direct analyses of the particular samples under question, are representative values, and some are more reliable than others. The next step is to accept the responsibility for the reliability of the nutrient data base being used, and to take the necessary precautions to ensure that it is current and accurate.

---

[2]"Model for Review of Nutrient Data Base System Capabilities" may be ordered from Department of Human Nutrition, Foods and Food Systems Management, College of Home Economics, 217 Gwynn Hall, University of Missouri-Columbia, Columbia, Missouri 65211. Price $7.50.

LITERATURE CITED

Danford, D.E., 1981. Computer applications to medical nutrition problems. JPEN 5:441-446.

Hoover, L.W,, 1983. Computerized nutrient data bases: I. Comparison of nutrient analysis systems. J. Am. Dietetic. A. 82:501-505.

Hoover, L.W. and B.P. Perloff, 1981. Model for Review of Nutrient Data Base System Capabilities. Columbia: University of Missouri-Columbia Printing Services.

U.S. Department of Agriculture, 1976-1982. Composition of Foods: Raw, Processed, Prepared. USDA Agriculture. Handbook No. 8. Revised Sections: 8-1, Dairy and Egg Products; 8-2, Spices and Herbs; 8-3, Baby Foods; 8-4, Fats and Oils; 8-5, Poultry Products; 8-6, Soups, Sauces, and Gravies; 8-7, Sausages and Luncheon Meats; 8-8, Breakfast Cereals; and 8-9, Fruits and Fruit Juices.

Watt, B.K. and A.L. Merrill, 1963. Composition of Foods...raw, processed, prepared. Rev. USDA Agriculture Handbook No. 8.

Table 1. Nutrients in Agriculture Handbook No. 8

	1963 edition	Current revision
Proximates:	Moisture	Moisture
	Energy	Energy
	Protein	Protein
	Total lipid	Total lipid
	Carbohydrate	Carbohydrate
	Crude fiber	Crude fiber
	Ash	Ash
		Dietary fiber, insoluble *
Minerals:	Calcium	Calcium
	Iron	Iron
	Magnesium *	Magnesium
	Phosphorus	Phosphorus
	Potassium	Potassium
	Sodium	Sodium
		Copper *
		Manganese *
		Zinc
Vitamins:	Ascorbic acid	Ascorbic acid
	Thiamin	Thiamin
	Riboflavin	Riboflavin
	Niacin	Niacin
	Vitamin A	Vitamin A
		Pantothenic acid
		Vitamin B_6
		Folacin
		Vitamin B_{12}
		Tocopherol *
Lipids:	Cholesterol *	Cholesterol
	Oleic acid *	Oleic acid
	Linoleic acid *	Linoleic acid
	Total saturated fatty acids *	Total saturated fatty acids
		Total monounsaturated fatty acids
		Total polyunsaturated fatty acids
		Other fatty acids
		Phytosterols *
		Amino acids

*Limited data provided.

Table 2. USDA computerized nutrient data bases

USDA Nutrient Data Base for Standard Reference, release 3, 1983 *
Data Set 456-3, release 4, 1983 +
Data Set 456, release 4, 1983 +
Data Set 72-1, release 2, 1982
USDA Nutrient Data Base for Individual Intake Surveys, release 1, 1980
USDA Nutrient Data Base for Household Food Use Surveys, release 1, 1980

* Release 3 includes information from revised sections 8-1 through 8-9
\+ Updated values from revised sections 8-1 through 8-9 have replaced original data for those food groups

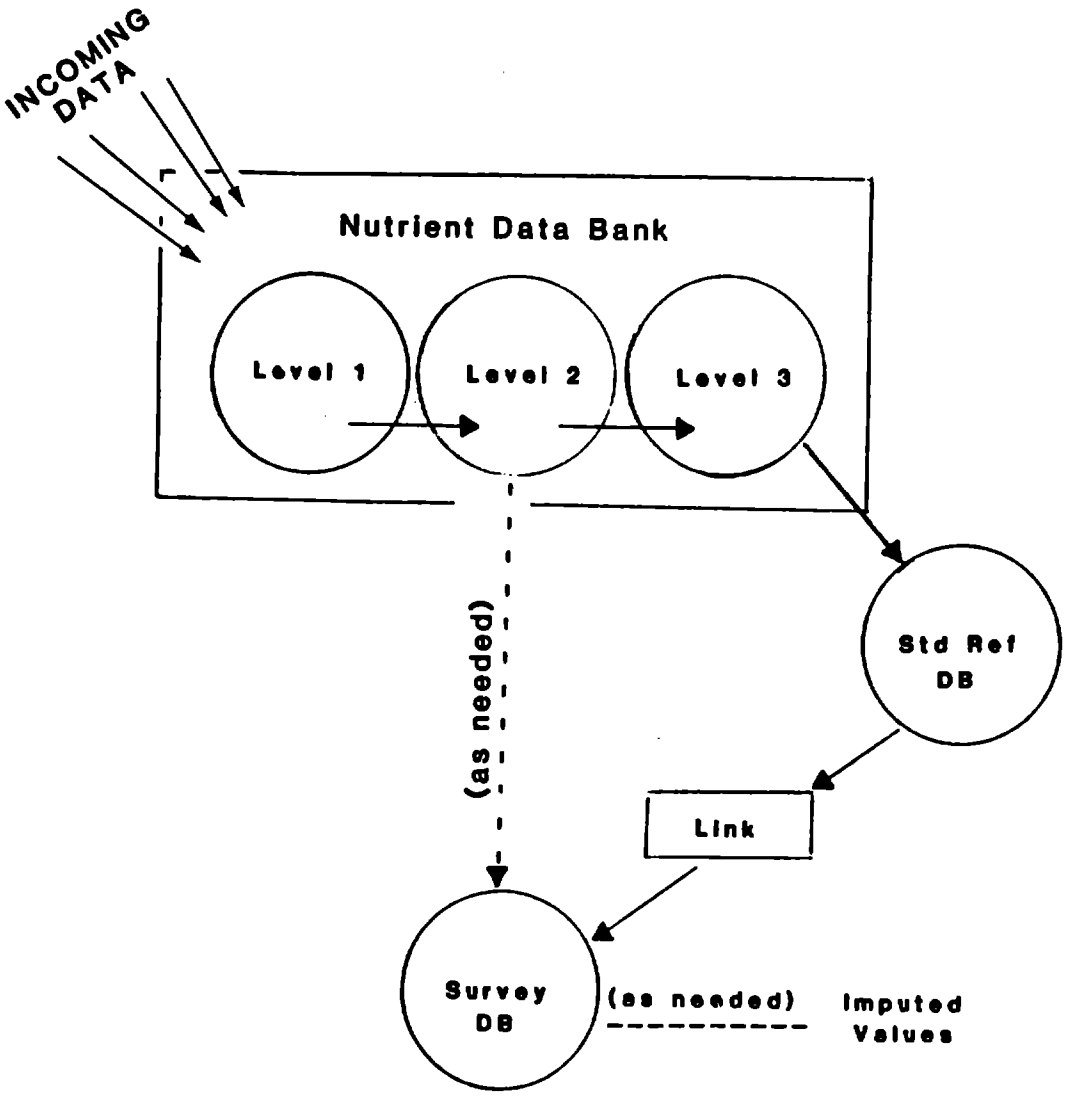
INCOMING DATA
Nutrient Data Bank
Level 1
Level 2
Level 3
(as needed)
Std Ref DB
Link
Survey DB
(as needed)
Imputed Values

Figure 1

STATE OF DEVELOPMENT OF METHODS FOR NUTRIENTS IN FOODS

Nutrient Composition Laboratory
ARS, USDA
Beltsville, MD 20705
July 1983

Nutrient category	State of Methodology[a/] Adequate	Substantial	Conflicting	Lacking
Carbohydrates, fiber and sugars		Individual sugars	Fiber Starch	
Energy			Food energy	
Lipids		Cholesterol Fat (total) Fatty acids (common)	Sterols Trans-fatty acids	
Minerals/Inorganic nutrients	Calcium Copper Magnesium Phosphorus Potassium Sodium Zinc	Iron (total) Selenium	Arsenic Chromium Fluorine Iodine Manganese	Cobalt Heme-iron Molybdenum Nonheme-iron Silicon Tin Vanadium
Proteins and amino acids	Nitrogen (total)	Amino acids (most)	Amino acids (some) Protein (total)	
Vitamins		Niacin Riboflavin Thiamin Vitamin B-6	Vitamin A Carotenes Vitamin B-12 Vitamin C Vitamin D Vitamin E Folacin Pantothenic acid	Biotin Choline Vitamin K

a/ Description of methodology states

Factors	Adequate	Substantial	Conflicting	Lacking
Accuracy	Excellent	Good	Fair	Poor
Speed of analysis	Fast	Moderate	Slow	Slow
Cost per analysis	Modest (<$100)	Modest to high	High	?
Development needs	---	Method modif. Extraction proc. Applications	Method develop. modif. Extraction proc. Applications	Method develop. Extraction proc. Applications

From: Beecher, G.R. and Vanderslice, J.T. Determination of Nutrients in Foods. Chapter in Proceedings of Symposium on Modern Methods of Food Analysis. Editors K.K. Stewart and J.R. Whitaker. Westport: AVI, 1984 (in press).