

Data Sources, Conventions, and Terminology

David B. Haytowitz, Human Nutrition Information Service, U.S. Department of Agriculture, Hyattsville, Maryland

USDA develops a number of different nutrient data bases. Some are generated as machine-readable counterparts of published food composition tables. Some data bases are generated for use in the Department's food consumption surveys. Others are generated for special purposes as needed. With a number of different data bases to choose from, it is frequently difficult to determine which is best suited for one's own projects. This paper will attempt to sort out the information on the various data bases so that the user can make an informed choice of the data base that will best meet his needs. A report describing these data bases, along with price and ordering information, is available from our office (1).

USDA Nutrient Data Base for Standard Reference

The USDA Nutrient Data Base for Standard Reference is the machine-readable version of Agriculture Handbook No. 8 (AH-8). The current version available is release 10, which includes all AH-8 sections 1-21, data from the first three supplements (1989, 1990, and 1991), and the new data on fresh pork (AH-8-10). The complete data set is available on both the Nutrient Data Bank Bulletin Board and Internet.

Future releases will contain data from future supplements to AH-8 and from those sections which are completely revised. Each year a new supplement is produced containing updated data on foods in the 21 sections plus new data not previously published. The 1992 supplement will be available shortly. A data set containing the information in the 1992 supplement will also be placed on the bulletin board. Work has begun on revising the Nutrient Data Bank System, our in-house system for developing the Standard Reference Data Base. We are also looking into ways to speed up the release of new data through the bulletin board and other means.

When possible, values are supplied for those nutrients where blanks appear in the printed handbook. In future handbook revisions, the values once included only in the machine-readable versions will also be available on the printed page. The Standard Reference Data Base uses the NDB (Nutrient Data Bank) numbers that appear at the bottom of each page of AH-8 to identify food

items. The first two digits denote the food group (1-21), while the last three digits indicate a specific food item within the group. These numbers are not necessarily in numerical order in the printed volumes of AH-8.

This data base is available on both diskettes and magnetic tape. Diskettes are formatted for IBM-compatible PC's and are available in both double and high density for both 3-1/2" and 5-1/4" diskettes. An update, containing all of the new information, is available for those users who wish to update an earlier release they have on their systems. For example, the update to release 10 will contain those data added to release 9 to create release 10.

Release 10 of the Standard Reference Data Base contains data on 5,245 foods for the nutrients given in table 1.

The data base contains a 20-character description with nutrient values for each food item on the same file. The coding manual in a separate file has full descriptions. The coding manual also contains weights and descriptions corresponding to the column E, F, and G headings on the AH-8 pages.

An abbreviated version, containing fewer nutrients, but the same number of foods, is also available. The nutrients in this data set are listed in Administrative Report No. 378 (1).

Data Set 72-2

Data set 72-2 contains the data published in Home and Garden Bulletin No. 72, "Nutritive Value of Foods." This publication, originally published in 1965 and revised several times, was last revised in 1991 primarily to incorporate changes in the cholesterol content of eggs. A complete revision of this publication is anticipated within the next couple of years.

The data, which are expressed only in terms of common household units, are based on the USDA Nutrient Data Base for Standard Reference. For sections not published at the time, data were taken from unpublished data in USDA's National Nutrient Data Bank. The data set contains data on 961 food items arranged by food groups. The nutrients included in this data set are shown in table 2. The printed publication includes an index. A description of each item is also included in the data file.

Proximates:

Water, protein, total fat, carbohydrate (by difference), crude fiber, total dietary fiber (when available), ash, and energy (both in kilocalories and kilojoules).

Minerals:

Calcium, iron, magnesium, phosphorus, potassium, sodium, zinc, copper, and manganese.

Vitamins:

Ascorbic acid, thiamin, riboflavin, niacin, pantothenic acid, vitamin B₆, folate, vitamin B₁₂, vitamin A (both IU and RE), and tocopherol (when available).

Lipids:

Total saturated, total monounsaturated, total polyunsaturated, and individual fatty acids, cholesterol, and plant sterols.

Amino acids:

Tryptophan, threonine, isoleucine, leucine, lysine, methionine, cystine, phenylalanine, tyrosine, valine, arginine, histidine, alanine, aspartic acid, glutamic acid, glycine, proline, and serine

Table 1 — Nutrients in USDA Nutrient Data Base for Standard Reference.

Proximates:

Water, protein, total fat, carbohydrate (by difference), and energy

Minerals:

Calcium, iron, phosphorus, potassium, and sodium

Vitamins:

Ascorbic acid, thiamin, riboflavin, niacin, and Vitamin A (IU and RE)

Lipids:

Total saturated fatty acids, total monounsaturated fatty acids, total polyunsaturated fatty acids, and cholesterol

Table 2 — Nutrients in Data Set 72-2.

A unique four-digit number is assigned to each item.

The data set is available on both diskettes and magnetic tape. Diskettes are formatted for IBM compatible PCs. It is also available on the Nutrient Data Bank Bulletin Board.

Other Nutrient Data Sets

HNIS has published a number of summaries of other nutrient and food components of interest to researchers. Data sets for these are all available on the Nutrient Data Bank Bulletin Board. Among these are "Sugar Content of Selected Foods," Home Economics Research Report No. 48. This data set contains data on total sugar, monosaccharides and disaccharides, starch, and carbohydrate for 522 foods. Data sets based on provisional tables include those on vitamins D and K. These data sets contain 165 and 109 items respectively. The vitamin D data set presents data in both micrograms and International Units. A data set on selenium corresponding to the recently published provisional table has recently been added. These data sets also include the appropriate NDB number as a cross-reference. As new provisional tables are produced, data sets will be made available on the bulletin board.

USDA Nutrient Data Base for Food Consumption Surveys

A separate data set is created for each survey. The first data set, Release 1 of the USDA Nutrient Data Base for Food Consumption Surveys was developed for use in the 1977-78 Nationwide Food Consumption Survey. It contains data on 15 nutrients. Releases 2.0 and 2.1 were developed for the 1985 Continuing Survey of Food Intakes by Individuals. Release 2.0 was used for the first set of data collected in the 1985 survey (Wave 1, core monitoring group), while release 2.1 contains about 500 additional food items and covers the complete 1985 survey. These data were also used for Hispanic HANES. Release 3.0 was developed for the 1986 Continuing Survey of Food Intakes by Individuals. It has not been released, but is available to researchers requesting it from our office. Release 2.0 and subsequent releases contain the 30 food components listed in table 3.

Release 4.0 was developed for use in the 1987-88 Nationwide Food Consumption Survey and contains data on 6,237 food items. Release 5, developed for the 1989 continuing survey, is also available and contains data on 6,659 food items. Both are available on the Nutrient Data Bank Bulletin Board. Each food item is identified by a 7-digit code used in USDA food consumption surveys. The nutrient file contains a 51-character description of each item. A separate code book with full description and weights used in survey coding accompanies the nutrient files. One useful

part of the code book is the "include" statements, which list those foods items similar in nutrient

Proximates:

Water, protein, total fat, carbohydrate alcohol, total dietary fiber, and energy

Minerals:

Calcium, iron, magnesium, phosphorus, potassium, and sodium, zinc, and copper

Vitamins:

Ascorbic acid, thiamin, riboflavin, niacin, vitamin B₆, vitamin B₁₂, Vitamin A (IU, RE, and carotene), and Vitamin E (alpha-tocopherol equivalents)

Lipids:

Total saturated fatty acids, total monounsaturated fatty acids, total polyunsaturated fatty acids, and cholesterol

Table 3 — Nutrients in USDA Nutrient Data Base for Food Consumption Surveys.

content to the title food for the code.

A "salt in cooking" code is used to distinguish between two records, one for items with salt added and one for the same item with no added salt. The code is used when the meal preparer has a choice of adding salt. On the bulletin board file, to save space, these two records are combined into one that reports both sodium values. The "fat in cooking" code is used to access the nutrient records calculated using fats or oils other than the one designated in the recipe for a particular item. For example, if butter was designated in the recipe, an alternate nutrient profile is calculated for the food cooked in margarine as well as several other cooking fats and oils.

Data Sets Used to Create the USDA Survey Nutrient Data Base Primary Nutrient Data Set for Food Consumption Surveys

The Primary Nutrient Data Set for Food Consumption Surveys (PDS) contains the nutrient data used to create the Survey Nutrient Data Base. The PDS is based primarily on the USDA Nutrient Data Base for Standard Reference. Values are added to the PDS when a food item reported by a survey respondent is either not in the Standard Reference Data Base or is missing nutrients. Future releases of the PDS will contain data from future releases of Standard Reference and other foods as reported by the survey respondents. The PDS contains data on approximately 3,300 foods and uses a 5-digit code to identify them. Food items taken from the Standard Reference use the same NDB numbers as the Standard Reference Data Base, while a unique number is assigned to each food added to the PDS. The PDS contains the same nutrients as the Survey Nutrient Data Base (table 3). A 20-character description is part of the nutrient file, while a longer description is available in a separate file. A code indicating the source of each data record is also part of the nutrient data file. The source codes are given in table 4. The date a value was added to the data base is also part of the nutrient file.

Recipe File for USDA Survey Nutrient Data Base

The recipe file contains the component records and their proportions used to calculate the USDA Survey Nutrient Data Base. It contains recipes for all items on the Survey Nutrient Data Base. Approximately half of these are single item records, while the remainder are of varying complexity.

The recipe file is composed of a header record and a number of component records. The header record contains the name of the food item and the recipe yield. The yield factor indicates, when

appropriate, the amount of moisture or fat gained or lost during preparation. If fat is gained or lost, the type of fat is also indicated. The component records contain the name and ID number of each component. The ID numbers refer to the 5-digit ID used in the Primary Nutrient Data Set. The ID number can also refer to another 7-digit code in the Survey Data Base. There is a code to trigger the calculation of additional values for the "fat in cooking" or "salt in cooking" records described above. If needed, a code indicating the appropriate set of retention factors is also part of the component record. The retention factors are described in greater detail below. The weight or proportion of each component is included, along with a household measure for documentation.

USDA Table of Nutrient Retention Factors

The USDA Table of Nutrient Retention Factors contains the retention factors used in recipes to calculate values for the Survey Nutrient Data Base. Additional nutrients and additional food categories were added to this table to match the nutrients and foods in the Survey Nutrient Data Base. These factors are reviewed periodically by NDRB staff. The most recent review was for release 3, which is available on the bulletin board. This file contains retention factors for 16 minerals and vitamins currently used in the Survey Nutrient Data Base. Each set of retention factors is referenced by a four-digit computer code, which is used in the recipe file to access the factors. The relationship between these data sets is shown in figure 1.

Any questions regarding these data bases can be answered by contacting HNIS at (301) 436-8491.

- 1 — Analytical data from Standard Reference
- 2 — Analytical data added for Survey
- 3 — Data from 1963 Handbook (Removed as new sections are added)
- 4 — Imputed data from Standard Reference
- 5 — Label claim data from Standard Reference (Primarily Breakfast Cereals, AH-8-8)
- 6 — Imputed data added for survey
- 7 — Assumed zero
- 8 — Label claim data added for survey (Primarily Breakfast Cereals)

Table 4 — Source Codes for Primary Nutrient Data Set.

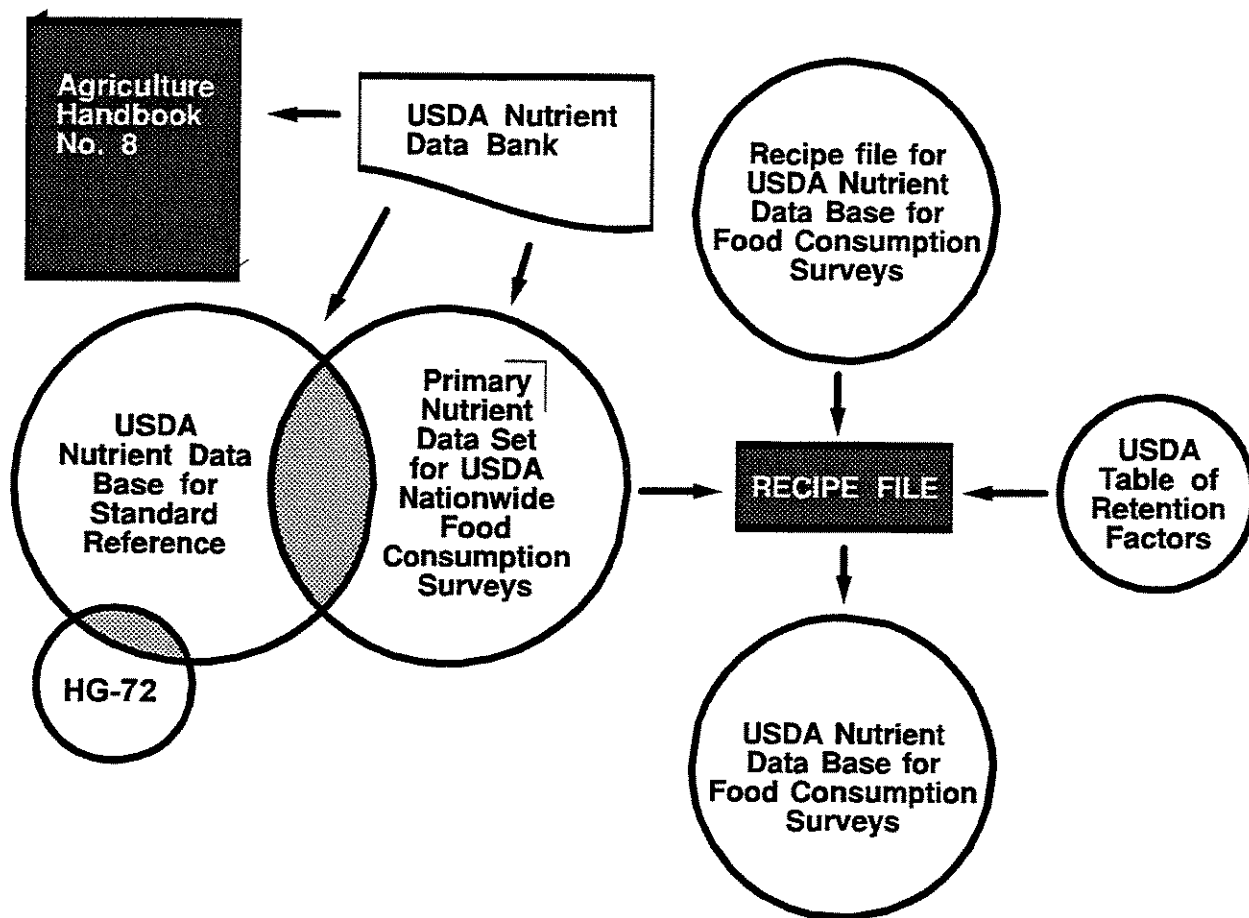


Figure 1 — Relationships Between Data Bases Used to Create the USDA Nutrient Data Base for Food Consumption Surveys

Nutrient Data Bank Bulletin Board

In addition, a number of the data bases described here are available on the Nutrient Data Bank Bulletin Board. The information on the bulletin board is available via telephone and through Internet.

To access the bulletin board directly you will need the following: Computer, modem, communication program, and telephone line. Although the bulletin board can be accessed from any computer, files for downloading are compressed, and made self-extracting using a program that runs on an IBM-compatible PC. Therefore users of other machines will not be able to uncompress the files but will still be able to read all bulletins. Users of other machines, as well as users of IBM-compatible machines, who have access to the Internet will be able to access the data through Internet. The data files on Internet are not compressed. Instructions on accessing the data through the Internet will be described later.

The telephone number for the bulletin board is (301) 436-5078 and operates at either 1200 or 2400 baud. The line settings are N-8-1. Although these line settings are commonly the default values, you may have to change these on your communication program.

First you will be asked to identify yourself. If you have called the bulletin board before, you will be asked to provide your password. If this is your first call you will be asked to select a password and register. The purpose of registering is to obtain some information to be used by the bulletin

board program in communicating with you. These include graphics, transfer protocols, etc. If any of the selections don't work correctly, they can always be changed by selecting the utilities option from the main menu. Other functions such as bulletins and file submenu can be accessed from the main menu.

In 1991, HNIS took another step to broaden access to nutrient data. In cooperation with the University of Maryland in College Park, nutrient data were made available through the Internet system. The Internet is a network of regional networks connected by a high-speed "spine" maintained by the National Science Foundation and others and is capable of transferring data at speeds of over 1 megabyte per second. Work is under way to increase the speed of data transfer even more. To access nutrient data over the Internet, type

telnet info.umd.edu

at your system prompt. The logon ID is info. There is no password. In the Info system select Government, then US, followed by NutrientData.

Accessing the bulletin board, either through the telephone or through Internet, was described in greater detail in the proceedings of last year's conference (2).

References

1. U.S. Department of Agriculture. 1993. Machine-Readable Data Sets on Composition of Foods and Results from Food Consumption Surveys. Administrative Report No. 378. Hyattsville, MD. 48pp.
2. Haytowitz, D.B. and Klensin, J.C. 1993. Electronic Dissemination of Nutrient Data via Bulletin Boards and Internet. 17th National Nutrient Data Bank Conference Proceedings. In press.

Mention of commercial products in this publication is solely for identification purposes and does not constitute endorsement by the U.S. Department of Agriculture over other products not mentioned. Mention of brand names is necessary to report factually on available data. USDA neither guarantees nor warrants the standard of the products, and the use of brand names implies no approval of the products to the exclusion of others which may also be suitable.