

## Data Sources, Conventions, and Terminology

---

**David B. Haytowitz**

*Human Nutrition Information Service*

*U.S. Department of Agriculture*

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.

In addition, a number of the data bases described here are available on the Nutrient Data Bank Bulletin Board. To access the Bulletin Board, you will need the following: PC, modem, communication software, and a telephone line. The telephone number for the bulletin board is (301) 436-5078. The information on the bulletin board is also available to users of Internet and can be accessed by entering:

telnet info.umd.edu

at your system prompt. For more information on accessing the bulletin board, either by phone or through Internet, see the article entitled "Electronic Dissemination of Nutrient Data via Bulletin Boards and Internet" elsewhere in these proceedings.

### **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 9, which includes AH-8 sections 1-17, 20, 21 & the 1989 Supplement. Release 10 will be available later this year and will include the remaining sections of AH-8: Baked Products (AH-8-18), Snacks and Sweets (AH-8-19), the 1990 and 1991 Supplements, plus new data on Pork Products (AH-8-10). The complete data set will be available on both the Nutrient Data Bank Bulletin Board and Internet.

Release 10 contains data from all of the sections of AH-8; future releases will contain data from future supplements to AH-8 and from those sections which are completely revised. For example, Breakfast Cereals (AH-8-8) is scheduled for a complete revision, as were the sections on Beef Products (AH-8-13) and Pork Products (AH-8-10). Other sections may also be completely redone, depending on the food group and availability of data and staff time to complete the work. Data from earlier releases are removed as new releases are issued. With the development of Release 10, almost

## Data Sources, Conventions, and Terminology

---

all of the data retained from the 1963 edition of AH-8 have been removed. The few remaining items will be removed as those items are updated in future supplements.

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 in numerical order in 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 new information, is available for those users who wish to update an earlier release they have already loaded into their systems. For example, the update to release 10 will contain those data added to release 9 to create release 10.

The Standard Reference Data Base contains data on 5,144 foods for the nutrients given in table 1.

## Data Sources, Conventions, and Terminology

---

### 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 B<sub>6</sub>, folate, vitamin B<sub>12</sub>, 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:

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

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 the same nutrients as Agriculture Handbook No. 456, is also available.

### Dataset 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 to primarily 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

## Data Sources, Conventions, and Terminology

---

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), and energy

**Minerals:**

Calcium, iron, phosphorus, potassium, and sodium

**Vitamins:**

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

**Lipids:**

Table 2 - Nutrients in Data Set 72-2.

A unique four-digit ID 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 for downloading.

### 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. Both data sets also include the appropriate NDB number from the Standard Reference data base 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

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, and 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.

<p>Proximates:</p> <p>Water, protein, total fat, carbohydrate, alcohol, total dietary fiber, and energy</p> <p>Minerals:</p> <p>Calcium, iron, magnesium, phosphorus, potassium, sodium, zinc, and copper</p> <p>Vitamins:</p> <p>Ascorbic acid, thiamin, riboflavin, niacin, vitamin B<sub>6</sub>, vitamin B<sub>12</sub>, Vitamin A (IU, RE, and carotene), and Vitamin E(alpha-tocopherol)</p>
--

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

Release 4.0 was developed for use in the 1987-88 Nationwide Food Consumption Survey and contains data on approximately 6237 food items. It is 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 along with 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 content to the title food for the code.

A "Salt in cooking" code is used to distinguish between two records for items with salt added and records for the same items with no added salt. It is used when the meal preparer has a choice of adding salt. On the bulletin board file, to save

## Data Sources, Conventions, and Terminology

---

space, these two records are combined into one that reports both sodium values reported. 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 Dataset 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 additional foods as reported by survey respondents. The PDS contains data on approximately 3,300 foods and uses a 5-digit code to identify them. Food items taken from 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 are in 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.

- 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

Table 4 -Source Codes for Primary Nutrient Data Set.

### **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 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.

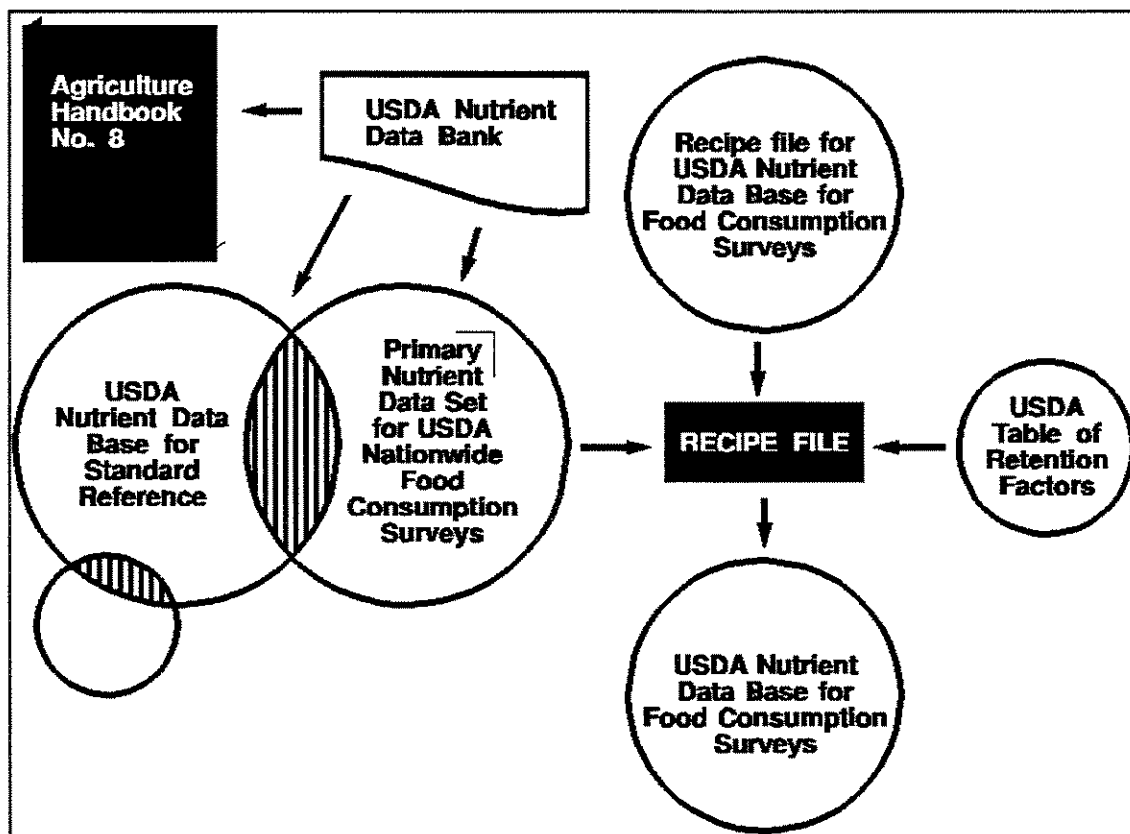


Figure 1 - Relationships Between Data Bases Used to Create the USDA Nutrient Data Base for Food Consumption Surveys