# USDA'S NATIONAL NUTRIENT DATA BANK

**Betty Perloff**

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

The U.S. Department of Agriculture (USDA) has been preparing food composition tables for public use for almost 100 years. This process has involved the collection, evaluation, categorization, and summarization of hundreds of thousands of individual food composition values. Many factors complicate this process, such as the natural variability of foods, the dynamic nature of our food supply, the continual improvements in agricultural and food processing practices, and the lack of adequate and inexpensive methods for nutrient analysis. Despite the many challenges involved in the preparation of reliable food composition tables, they continue to increase in size and scope to meet the expanding needs of both the public and private sectors.

Approximately 15 years ago, USDA implemented the Nutrient Data Bank (NDB) system to automate the preparation of food composition tables. This system, which has undergone one major revision and is continually being refined and improved, is a major component of the government's National Nutrition Monitoring System. It is operated by USDA's Human Nutrition Information Service (HNIS) and is currently used to produce the data published in revised sections of, and supplements to, Agriculture Handbook No. 8. These data, which are also released in computerized form as updates to the USDA Nutrient Data Base for Standard Reference, are used in the special nutrient data base maintained by HNIS for analysis of dietary information collected in nationwide food consumption surveys. This paper discusses both the Nutrient Data Bank and this special Survey Nutrient Data Base. Also included is information about HNIS nutrient data accomplishments since the last conference and the Nutrient Data Bank electronic bulletin board, which was implemented last year.

## Nutrient Data Bank

Information on the nutrient composition of foods is collected from various sources. Data are furnished voluntarily by many food companies and trade associations. Government laboratories contribute data; for example, the Food and Drug Administration supplies the data its laboratory generates for the Total Diet Study. A considerable amount of data is generated by private or university laboratories under contract with the Human Nutrition Information Service. In addition, Nutrient Data Bank nutritionists regularly scan the scientific literature for food composition data.

Only data meeting established reliability criteria are entered into the NDB system. Questions must be answered satisfactorily about the food sample, the laboratory procedures and method of analysis, and the statistical validity of the values. For example:

Is the description of food sample complete?
Is the item representative of a food product in the food supply?
Was the sample selected and handled appropriately?
Was an acceptable method used for analysis?
Do data meet established criteria for statistical validity?

Laboratories generating data for HNIS under contract are required to demonstrate their ability to analyze standard reference materials accurately for selected nutrients before the contract is awarded and are required to practice strict quality control procedures while performing the contract. In some cases, contractors may be monitored periodically during the contract period to reconfirm their ability to measure a nutrient accurately. USDA's Nutrient Composition Laboratory, located within the Agricultural Research Service, cooperates with

HNIS to administer the testing and monitoring of laboratories.

The NDB is a complex operating system comprising approximately 150 different computer programs. It involves--

--entering and editing data,
--generating preliminary summaries,
--reporting data in different formats for review,
--applying special factors to produce values for calories, protein, retinol equivalents, and fatty acids,
--performing special processing routines for amino acids and fatty acids,
--calculating recipes,
--applying weighting factors when needed to account for market predominance of a food's subgroups,
--producing the representative nutrient profiles for individual food items that are released to the public and used for other research at USDA.

Included with each set of nutrient values entered into the NDB system are (1) a food code containing basic information about the food and (2) qualifier codes to completely identify the food and to indicate any conditions about the sample that may have affected its nutrient content. For example, a recent data entry for a sample of melba toast included the following codes:

| | | |
|---|---|---|
| Food code: | MRCF | Melba toast, wheat |
| Qualifiers: | K171 | Commercial product |
| | O492 | Flour, enriched, unbleached wheat (July 83) |
| | O239 | Flour, whole wheat |
| | P901 | Yeast |
| | P006 | Dextrose |
| | P341 | Oil, partially hydrogenated soybean |
| | R266 | Salt (sodium chloride) |
| | R134 | Color, caramel (burnt sugar) |
| | R257 | Propionate, calcium |

When data for different samples are summarized, a nutritionist determines which data sets can be combined by designating the specific qualifiers that cause an item to be included (or excluded) in a summary. Temporary summaries are run and analyzed to determine the extent of variation in nutrient content found among samples having either identical or different qualifiers.

As data travel through the Nutrient Data Bank system, they are closely monitored and reviewed at each stage by a NDB nutritionist. The main data bank stages are:

(1) Data base 1--the individual nutrient values.
Proximates, vitamins, and minerals are stored based on 100 grams of food, fatty acids on percentage of methyl esters, and amino acids on grams per gram of nitrogen.

(2) Data base 2--summarized values.
Items from Data base 1 are summarized based on designated food codes and qualifiers.

(3) Data base 3--final nutrient profiles for each food item.
When records for this data base are generated, fatty acid and amino acid data are converted to grams per 100 grams of food and some food subgroups are weighted to improve representativeness of the average values. For example, when data for eggs were summarized, we adjusted for seasonal variation by applying equal weighting factors to samples analyzed during winter and summer. Data base 3 items are sometimes combined through a recipe program to create nutrient profiles for food mixtures.

Each of these three segments operates in two environments: test and production. The test environment is used for review and analysis of the data. At the Data base 1 level, data editing takes place in "test"; only when data are "clean" do they pass into the production environment. From the Data base 1 production environment, data for different samples of the same food item may be averaged into Data base 2 "test" records. When the Nutrient Data Bank nutritionists are satisfied that a test summary is accurate, it moves into Data base 2 production and is available for further processing into Data base 3 records. Again, "test" records are created in Data base 3 and reviewed before they are released into "production."

Foods and nutrients are not all treated the same by the system. First, the types of information included as qualifiers depends on the food group. For example, for either animal or plant products we code (1) portion analyzed; (2) whether the sample was raw, processed, or prepared; (3) any processing or preparation techniques that had been applied; and (4) preservation technique if applicable.
In addition, for fruits and vegetables we include--

--form, i.e., cut, sliced, etc.
--length and conditions of storage.
--if canned: type of pack and sample (drained solids, liquids, or total can contents).
--if raw: peeled or unpeeled, ripe or immature; also variety, growing location, and color, if appropriate.
--if dried: sulfured or nonsulfured.

However, information about meat includes--

--type of tissue, i.e., muscle, separable lean, separable fat, organ meat, or total edible.
--amount of trimming.
--if cooked: degree to which cooked.
--grade of animal.

Most of the values issued from the data bank are averages. These are reported in Handbook 8 tables with numbers of samples and standard errors. Some of the values are derived from other values as described below.

*Protein.* Nitrogen, not protein, values are entered into the data bank. After the final summary for a food item, a special factor is applied to the average nitrogen value to obtain the protein value. Factors used to calculate protein from nitrogen are based on research that determined the nitrogen content of proteins in different types of foods. The factor used for each item is printed in Agriculture Handbook No. 8, (AH-8).

*Carbohydrate.* The carbohydrate value is determined by difference, i.e., 100 minus the sum of protein, moisture, fat, and ash.

*Food Energy.* The energy value of a food item is calculated by applying a set of factors to its protein, fat, and carbohydrate values. This procedure for estimating the metabolizable energy of foods was developed primarily by W.O. Atwater and associates and dates back to the 19th century. The specific factors used for different foods within each AH-8 section are listed in the section's introductory text.

General factors of 4, 9, 4 kilocalories per gram of protein, fat, and carbohydrate, respectively, were developed to estimate available energy in average mixed diets. Substantial differences in energy estimates can result if general factors are applied to specific foods instead of the specific factors. For example, energy for whole grain wheat flour is calculated using the specific factors of 3.59 kilocalories per gram of protein, 8.37 kilocalories per gram of fat, and 3.78 kilocalories per gram of carbohydrate. The result is 339 total kilocalories. Using the general factors of 4, 9, and 4, the estimate would be 362 kilocalories.

*Fatty acids.* Data for fatty acids frequently require special processing within the data bank system. Before values

for similar food samples are averaged, they are processed through a special routine that adjusts for missing values when samples do not have complete fatty acid patterns. Then, using special factors, the values are adjusted to account for the lipid constituents that are not fatty acids.

**Survey Nutrient Data Base**

One of the major uses of food composition data is analysis of food consumption survey results. HNIS is responsible for providing the food composition data used for nationwide dietary surveys in the Federal government's National Nutrition Monitoring System, and a special nutrient data base was designed and is maintained by HNIS for this purpose. Three nationwide surveys use this data base:

(1) The large Nationwide Food Consumption Survey (NFCS), which is conducted every 10 years by HNIS.

(2) The Continuing Survey of Food Intakes by Individuals (CSFII), a smaller survey conducted by HNIS to track consumption between the larger decennial surveys.

(3) The National Health and Nutrition Examination Survey (NHANES III), conducted by the Department of Health and Human Services' National Center for Health Statistics (NCHS).

The current operating version of the Survey Nutrient Data Base contains almost 6,000 items. Foods have been organized to facilitate summarization of dietary intake data by food groups. Each food has a 7-digit numeric code in which the first position represents one of nine major food groups. Positions 2 and 3 represent subgroups. For example, all foods with "5" in the first position are grain products or mixtures in which grain is the major ingredient, those beginning with "51" are yeast breads or rolls, and those beginning with "512" are whole-wheat yeast breads or rolls.

The nutrient values for each item are given for 100 grams of the edible portion of the foods. These values are translated into the amounts of nutrients in the portions consumed using information from a computerized "Manual of Food Codes for Individual Intake," which provides a complete description of each food item in the data base, as well as information on weights of common portions for each food.

The current data base includes data for energy and the 28 food components listed below. The nutrient profiles are developed by the nutrient data specialists at HNIS who are responsible for the Nutrient Data Bank.

Energy
Moisture
Protein
Total fat
Total saturated fatty acids
Total monounsaturated fatty acids
Total polyunsaturated fatty acids
Carbohydrate
Total dietary fiber
Alcohol
Cholesterol
Calcium
Iron
Magnesium
Phosphorus
Sodium
Potassium
Copper
Zinc
Ascorbic acid
Thiamin
Riboflavin
Niacin
Vitamin B-6
Vitamin B-12
Vitamin A (IU)
Vitamin A (RE)
Vitamin E
Folate

**Survey Nutrient Data Base System.**

The Standard Reference Data Base, which is the output from the Nutrient Data Bank system, is the main source of data for the Survey Nutrient Data Base. A special information system has been developed to generate or update the Survey Nutrient Data Base, and data needed for the survey are transferred from the Standard Reference Data Base into the Primary Data Set (PDS) within this system.

Values are added to the PDS for the nutrients missing from the Standard Reference Data Base, such as carotene. Complete nutrient profiles are added as necessary for missing food items. If analytical data are not available for the missing items, the values are imputed from other forms of the foods or estimated from data for similar foods.

In addition to the PDS, the other major components of the Survey Nutrient Data Base system are the recipe file, a table of retention factors, and the computer program that updates the data base.

*Recipe File*. In the recipe file, each survey food code is linked to one or more PDS items through a set of recipe codes. For food mixtures, recipes contain information about changes in moisture or fat that occur during cooking, as well as information about the recipe ingredients, i.e., names, identification numbers, weights, and codes for accessing retention factors if applicable.

*Table of Nutrient Retention Factors. This data set contains factors for calculating retention of 18 vitamins and minerals during cooking. The factors are organized into food categories, and each category is assigned a code for computer access.*

*Computer Program.* The computer program uses the recipe file to determine which sets of nutrient values from the PDS will be used for each item on the survey data base. For one-component recipes, i.e., direct links between the survey and PDS, nutrient values are moved directly from the PDS to the survey data base. For multicomponent recipes, the values are calculated from data for the ingredients. Nutrient retention factors are accessed and used in the calculations as appropriate.

The entire data base is updated before each USDA food consumption survey. The first step in the process is to update the PDS with the most current version of the Standard Reference Data Base. During the course of a survey, when new food items are reported by survey respondents, special forms are submitted to USDA. An interdepartmental food code committee, with representatives from both HNIS and NCHS, reviews these forms and decides what new food items should be added to the data base.

The benefits of the Survey Nutrient Data Base system go beyond the obvious ability to quickly update the data base. Another benefit of particular significance is the ability to convert data on the consumption of mixed foods into data on the consumption of PDS items. This conversion is possible by using the recipe file which designates the amount of each food component in mixed food items. Consumption data for the PDS items serve two important functions. They permit studying nutrient contribution at the primary food level, and they also provide a mechanism for evaluating the nutrient values on the PDS.

For example, using food consumption data from the 1985 CSFII and the recipe file in the survey data base system, amounts of mixed foods consumed were converted into amounts of ingredients consumed. Consumption data for foods at this primary food level were then aggregated. For each nutrient in the data base, the minimum number of PDS food items required to reach 80 percent of the total intake of the nutrient was determined. For most nutrients, 100-200 foods account for 80 percent of the consumption of the nutrient.

The nutrient values on the PDS for these foods were then evaluated to determine the greatest needs for nutrient analyses. Based on this evaluation, HNIS conducted a project to improve the analytical nutrient base for the survey. Generally, the items analyzed were frequently consumed foods and foods that were major contributors of nutrients for which intakes were below the RDA in the 1985 CSFII. Analysis of vitamin E was emphasized, since its analytical base was not as strong as those for other food components. A similar project began last year, and two more are planned.

Expansion of the data base to include additional food components is also planned. The addition of fatty acids is in progress and work on selenium should begin later this year.

A great deal of interest has been shown in the Survey Nutrient Data Base as a resource for other research purposes. The University of Texas School of Public Health in Houston began using all of the components of the Survey Nutrient Data Base system--the recipe file, the Primary Nutrient Data Set, the Table of Nutrient Retention Factors, and the recipe calculation program--for their research projects in 1987. Because of the demand for this system for USDA-sponsored and other research, we are working with the university to make their software available to other organizations. This joint system will maintain the integrity of the USDA data base but will allow users to add additional foods and nutrients. It also provides options of using the USDA recipes, modifying the USDA recipes, or entering new recipes.

The Survey Nutrient Data Base, which is an essential component of the National Nutrition Monitoring System, will be kept current to meet the needs of both CSFII and NHANES III. New foods will be added when they are reported on either survey, and updated nutrient values will be incorporated periodically. Data for priority nutrients will be added to the data base as they are identified and as research is completed to provide the necessary values.

**Nutrient Data Accomplishments and Plans**

Since the last conference, Agriculture Handbook No. 8-20, covering cereal grains and pasta, was completed and released; Handbook No. 8-13, covering beef products, was revised and re-released; and the first supplement to the AH-8 series was released. The beef section was revised because more external fat is now trimmed from beef for the retail market than when the original beef section was prepared in 1986. The supplement was issued to update selected foods within previously published sections of the handbook. It contains replacement pages for foods that have had data updates and insertion pages for foods that are being added to the handbook.

One new version of the USDA Nutrient Data Base for Standard Reference was released. Data from Agriculture Handbook No. 8-17 on lamb, veal, and game; Agriculture Handbook No. 8-20 on cereal grains and pasta; Agriculture Handbook No. 8-21 on fast foods; and the 1989 supplement were incorporated into data base Release No. 8. Release No. 9 will include the new data on beef from the revised section. It has been completed at HNIS and should be available from the National Technical Information Service shortly.

Other projects that have been completed during the past year are the comparison of wild and cultivated fish, which will be discussed in August at the American Chemical Society meeting; the selenium project which will be reported elsewhere in these proceedings; the dietary fiber collaborative study; and expansion of the Primary Data Set in our Survey Nutrient Data Base system to include individual fatty acids.

Projects underway or planned to begin in the near future include Agriculture Handbook No. 8-19 on snacks and sweets, Agriculture Handbook No. 8-18 on baked products, Agriculture Handbook No. 8-22 on mixed dishes, the 1990 supplement to Agriculture Handbook No. 8, revision of Agriculture Handbook No. 8-10 on pork products, a provisional table on vitamin D content of foods, and a revision to the provisional table on vitamin K content of foods.

**Bulletin Board**

We began operating an electronic bulletin board in the summer of 1989 to provide a ready source of information about our current nutrient data publications and to provide an easy mechanism for the public to obtain our latest nutrient data files. Anyone with a computer, a telephone line, a 1200- or 2400-baud modem, and communications software can access this bulletin board. It operates 24 hours a day and the only cost to the user is for the phone call.

The bulletin board phone number is 301-436-5078. Once into the bulletin board, you may read bulletins, transfer data files from our computer to your computer, or leave a comment for the system operator. Currently the system contains 13 bulletins. They include announcements about our latest food composition data publications and computer data sets, descriptions of these items and instructions for ordering them, Nutrient Data Bank Conference announcements, and other relevant information about food composition data. The data files that can be down loaded include the latest updates to the USDA Nutrient Data Base for Standard Reference, data from a table on the sugar content of selected foods, data from Home and Garden Bulletin No. 72, and the Dietary Analysis Program developed jointly by HNIS and the Extension Service. We will be adding additional files on nutrients of special interest as they are developed.

**Summary**

In summary, food composition tables are undergoing continual improvement and expansion through USDA's operation of the National Nutrient Data Bank. Data are collected from various sources, but all data in the system have passed a rigorous evaluation. The system is currently used to update Agriculture Handbook No. 8 and its corresponding computer file, the USDA Nutrient Data Base for Standard Reference. A major use of these data is analysis of nationwide food consumption surveys, and HNIS maintains a special nutrient data base for this purpose. HNIS recently implemented the Nutrient Data Bank electronic bulletin board to provide a ready source of information about nutrient data releases and a convenient mechanism for providing updates to users of the Standard Reference Nutrient Data Base.