Public Health Rev 1998; 26: 43–47

Development and maintenance of nutrient databases for national dietary surveys

Betty Perloff, BS
Jaspreet K. Chug Ahuja, MS

Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Riverdale 20737, MD, USA.

This paper was presented at the 3rd Abraham Conference on Preventive Nutrition: Dietary Assessment in Populations, 7–9 July 1997, The S. Daniel Abraham International Center for Health and Nutrition, Ben-Gurion University of the Negev, Beer Sheva, Israel.

The U.S. Department of Agriculture (USDA) operates a complex technical database system for processing and analyzing data collected in national food surveys. Responsibility for the system lies in the Food Surveys Research Group at the Beltsville Human Nutrition Research Center (BHNRC), Agricultural Research Service (ARS). It is composed of three technical databases: (1) a nutrient database which includes nutrient values, as well as factors for estimating changes during cooking to vitamins and minerals; (2) a food coding database which includes long and short food descriptions, brand names, measure descriptions, and volume–weight conversion factors; and (3) a recipe database, which includes all recipes and formulas used to calculate nutrient content of mixtures, as well as information about changes to the moisture and fat content of the mixtures during cooking.

The USDA technical database system for supporting food surveys has evolved over three decades, but it was last used for the Continuing Survey of Food Intakes by Individuals, 1994–96 (CSFII), popularly called "What We Eat in America." Addressing issues related to procedures, accuracy, and operational efficiency has been a priority for development and management of the system. Some of these issues are discussed below.

NUTRIENT VALUES

Nutrient values must be accurate and current for analysis of national dietary surveys. Because laboratory analyses of foods may require complicated and sometimes difficult procedures, database managers must assure that up-to-date methodologies, along with acceptable laboratory practices, have been used to generate the data. Inconsistencies in data must be evaluated to determine if they result from natural variations of foods, or from problems that may have occurred during sampling or laboratory analysis. When possible, especially for the most commonly consumed foods, analytical samples should be selected to provide nationally representative data. Procedures should be established for estimating missing values since analytical data may not be available for all foods. At USDA, the reliability of nutrient values are the responsibility of the Nutrient Data Laboratory in the Beltsville Human Nutrition Research Center, Agricultural Research Service. This laboratory works closely with the Food Surveys Research Group to produce a high-quality nutrient database for the national food surveys.

Protocols for prioritizing foods and nutrients for analysis are essential because of the expense involved in sampling and analyzing the food supply. Such protocols at USDA traditionally have taken into account the popularity of the food, the extent of existing data, the adequacy of analytical methodology, and the need for the data relative to public health concerns (1). When data for a nutrient are limited, it is essential to document the limitations. For example, when dietary fiber was first added to the USDA nutrient database, the fact that values were based on limited analytical data was noted in the published survey reports. As dietary fiber received priority for nutrient analysis, the database values were improved.

It is not unusual for national food composition tables to include data for raw forms of foods, but not cooked forms. Therefore, food consumption surveys may require that appropriate adjustments be made to the data to account for changes that take place in food during preparation and cooking. Procedures for estimating the nutrient content of mixtures should be established, and it is desirable to have a protocol for selecting recipes to represent mixtures typically prepared at home. The USDA system uses the retention factor method to calculate the nutrient content of recipes based on the nutrient content

of recipe ingredients (2). In this method, retention factors may be applied to the raw forms of foods to estimate percent retention of vitamins and minerals during cooking. Estimates of other types of changes that occur during cooking, i.e., gain or loss of moisture and fat, are also applied in the calculations. Powers and Hoover (3) compared the retention factor method with other recipe calculation methods commonly used in nutrient database systems. Some differences in results of various methods do exist, and database managers should document the methodology they chose for their systems.

FOOD DESCRIPTIONS

Data must represent foods as they are consumed, and descriptions must include the types of information asked about foods during data collection. For example, in the CSFII, respondents were asked if their coffee was made from ground coffee or from instant coffee, and descriptions for coffee items in the database reflect this information. Respondents usually know the type of coffee and can supply the answer. Other types of questions, such as variety of produce, may not be easily answered. Therefore, food descriptions do not include specific varieties, and nutrient values reflect either the most common varieties, or composites of several varieties.

Procedures must be established for coding foods when information reported by survey respondents is incomplete. For example, when a respondent cannot designate whether milk was whole, low fat, or nonfat, a "default" is used based on market data for the different types of milk. The proportion of each type of milk used in the default is documented in the recipe database, and a nutrient profile for the default is calculated using the recipe calculation program. The default recipe is updated each year based on the latest available market data.

MAINTENANCE OVER TIME

Provisions for updating the nutrient database and related information are critical if national surveys occur on a continuing basis. Special requirements are imposed on maintenance procedures if the analysis of trends in nutrient intakes over time is a goal, in order to differentiate

between changes that occur because of (i) improvements in food composition data or (ii) changes to the food products. Sometimes it may be necessary to apply data improvements retroactively to survey data before meaningful comparisons can be made. For example, the average iron intake originally reported from the USDA's 1977–78 Nationwide Food Consumption Survey was 12.7 milligrams per person per day. After new nutrient values were developed in the early 1980s based on improved procedures for measuring iron in meat, the 1977–78 survey data were reanalyzed and the average daily iron intake for that time period was calculated as 11.3 milligrams (4). Thus, comparisons with later surveys were based on comparable data.

DATABASE QUALITY CONTROL

Because of the many different types of technical data that are required to support a national survey, a comprehensive quality control plan is essential to ensure that errors are not inadvertently introduced into the databases. At USDA, relational database software is used to maintain and update all data files. After updates are performed, a series of checks are performed across all relational files to ensure that nothing has been inadvertently eliminated or changed. In addition, after all nutrient values are updated, a series of integrity checks are performed to confirm that new values are reasonable. More information about computerized quality control checks in the USDA database system may be obtained by contacting the authors.

AVAILABILITY

The survey methodology for the CSFII, including use of the technical support system, is described in the CSFII 1994–96 Design and Operation Report (5). The complete set of databases (nutrient database, food coding database, and recipe database) were released on CD-ROM along with results from the survey (6). Information about the latest data release is available on the internet at http://www.barc.usda.gov/bhnrc/foodsurvey/home.htm.

REFERENCES

1. Haytowitz DB, Pehrsson PR, Smith J, Gebhardt SE, Matthews RH, Anderson BA. Key foods: setting priorities for nutrient analyses. J Food Composition Analyses 1996; 9: 331–364.
2. Perloff BP. Recipe calculations for NFCS database. In: Murphy S, Rauchwarter D, eds. 1985. Proceedings of Tenth National Nutrient Data Bank Conference, p. 11–21. Springfield, VA: National Technical Information Service, Accession No. PB86-206281/HBF.
3. Powers PM, Hoover LW. Calculating the nutrient composition of recipes with computers. J Am Diet Assoc 1989; 89: 224–232.
4. Guenther PM, Perloff BP, Vizioli TL. Separating fact from artifact in changes in nutrient intake over time. J Am Diet Assoc 1994; 94: 270–275.
5. Tippett KS, Cypel YS, eds. 1997. Design and operation: The Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey 1994–96. U.S. Department of Agriculture, Agricultural Research Service, Nationwide Food Surveys Report. No. 96–1, 240 p.
6. U.S. Department of Agriculture, Agricultural Research Service. 1995 Continuing Survey of Food Intakes by Individuals and 1995 Diet and Health Knowledge Survey. Springfield, VA: National Technical Information Service, 1996, Accession No. BP97-500789 (CD-ROM).