

SURVEY NET: CODING AND MANAGEMENT OF CSFII FOOD INTAKE DATA

Randy P. LaComb, USDA, Amy Green, Westat, Inc., Linda Ingwersen, USDA
Randy P. LaComb, USDA, 4700 River Rd, Riverdale, MD 20737

KEY WORDS: Computer-Assisted Food Coding; Dietary Intake; Food Consumption; Data Management; Survey Net; Food Intake

Nationwide data about the foods Americans eat are derived from the Continuing Survey of Food Intakes by Individuals (CSFII), also known as the *What We Eat in America* survey, conducted by the U.S. Department of Agriculture's (USDA) Agricultural Research Service. These data are used to assess the nutritional adequacy of the American diet and to monitor food choices. They are also used to evaluate the impact of food assistance programs, study the effects of food fortification regulations, and estimate exposures to pesticide residues.

In the CSFII 1994-96, field interviewers asked participants about the kinds and amounts of foods and beverages they ate on the day before the interview. Using structured, open-ended questions about each food item, interviewers recorded responses on hard copy intake questionnaires. For example, when the participant reported milk, the interviewer asked, "What kind was it?" and then, "Was it liquid, dry, sweetened condensed or evaporated?" If the answer was liquid, the interviewer asked, "Did you add anything to it?" Interviewers also collected information on the time of day the food was eaten, the name of the eating occasion, such as breakfast, where the food was obtained, and whether it was eaten at home or away from home. During the three years (1994-1996) of the latest survey, interviewers collected information from over 15,000 participants for two separate 24-hour periods.

Collecting and coding detailed information about types and quantities of foods can be difficult. For example, it isn't enough to know that someone ate carrots. Additional information is needed about whether the carrots were fresh or cooked; whether they had been frozen or canned; the type of fat, if any, added during cooking; whether salt was added during their preparation; whether the carrots were eaten in combination with other foods, such as in a salad; and the type of dressing, if any, that was used. An estimate of the amount of every food that was eaten must also be obtained.

This information was coded using a structured food classification system. In CSFII 1994-96, over 31,000 intake questionnaires with approximately 448,000 foods were coded.

To be most useful, the survey data had to be collected, reviewed, coded and released as quickly as possible. In 1994, a new, computer-assisted food coding and nutrient

analysis system called "Survey Net" was used. Not only was Survey Net faster than methods used in the past, but its flexibility allowed the coding of new or modified foods that were reported, and its special editing and management features helped ensure accuracy of the coded data. The use of Survey Net contributed to the release of 1994 and 1995 data on CD ROMs in less than half the time it had taken in the past.

Developed with the University of Texas Health Science Center, Survey Net was specifically designed to improve data processing procedures for the CSFII 1994-96. During the initial planning stages, three goals were established for Survey Net. The first goal was to reduce the coding time; the second was to provide flexibility in the food coding system while maintaining a structure comparable to past surveys; and the third goal was to increase oversight of the food coding operations conducted by survey contractors.

Operating on a computer network with multiple users accessing a set of central databases, Survey Net was used by both USDA in Riverdale, Maryland and by the CSFII contractor, Westat Inc., in Rockville, Maryland. Survey Net has three databases: a food coding database, a pre-defined recipe database, and a nutrient database. The food coding database contains the food name, description, and units of measurement for 7,300 foods. Pre-defined recipes, including ingredients and their amounts, comprise the recipe database and are used to estimate the nutritional value of the food. The nutrient database contains nutritional profiles of foods.

Survey Net has three access levels. The basic level was used by Westat coders to enter foods and flag items requiring additional review. The intermediate level was used by supervisors to review and approve the data, after which the data were electronically transmitted to USDA. The highest level was used by USDA staff to perform final review and acceptance of the data. At each of the two higher levels, reviewers used a special field to document their approval of any items that may have been in question.

A special effort was made to make the Survey Net software as easy to use as possible. On each of the different screens used during coding and editing, a menu bar lists available functions for which a single key stroke usually accomplishes the task. The database files can be accessed conveniently within the system when needed. Coding options are presented on pop-up screens

throughout the data entry part of the program, and help screens are available to review the operation of various software features.

The food coding database contains 7,300 foods with descriptions of varying length. To code a food, all items that are possible matches to a sample person's food description are located and reviewed to select the closest match or determine if there is no match in the database. The "food search" feature helps to minimize the time required to locate foods. A "search term" may be partial or complete words or single or multiple terms. The words or terms used in the search can occur anywhere in a food's description. An average search, which involves searching over 70,000 words, takes less than one second.

When a single term is used, Survey Net reports the number of descriptions or items which contain the term as the first word in the description. For example, when "milk" is used as a search term, Survey Net reports 120 descriptions with milk as the first word. Single word searches are convenient when coders are not sure how a food is described in the database. Multiple search terms can be used to narrow the search. For example if the multiple term, "milk 2 %," is entered only 8 entries are presented. As coders became familiar with terms used in the food descriptions, they became more efficient using the multiple search terms.

Occasionally, foods were reported that could not be matched to descriptions in the coding database. These were called "unknowns." Sometimes they were new foods that had just entered the market place and sometimes they were unique recipes. Resolving unknowns was one of the most time-consuming aspects of coding and editing the food consumption data.

In previous surveys, an unknown could not be coded immediately. Survey contractors sent a request to USDA for resolution and set aside the intake questionnaire containing the unknown. USDA researched the food, made a decision, and replied to the contractor who completed the coding for that intake questionnaire. The process sometimes took several weeks, during which the same unknown might appear again on several more intake questionnaires. Keeping track of these unknowns, and the intake questionnaires containing them, was one of the overriding reasons for the development of Survey Net. A great deal of planning went into the design of the system related to the procedures, screens, and file management for these situations.

In Survey Net an unknown foods file residing on the computer network could be searched and updated by all coders. In the CSFII 1994-96, when a new or unusual food was encountered for the first time, a coder selected the food category of the unknown and entered a description into the file. This file was a shared network file that could be searched, updated and used by all coders. Each entry into the file was automatically assigned a number which served

as a temporary food code. Until USDA decided how an unknown was to be handled, subsequent encounters of the food could be coded with the temporary code in the unknown foods file.

Each week when coded intake questionnaires were transmitted to USDA, the unknown foods file was also sent. USDA identified, researched and decided how each unknown should be handled and then edited the intake records using Survey Net. An unknown was usually resolved by either selecting a code from the food coding database for a similar food, by creating a recipe modification, or by adding a new food code to the database. USDA removed foods from the unknown foods file as decisions were made and sent an updated version of the file to Westat weekly throughout the survey. About 9,400 unique unknowns were created during the survey and they were used 14,711 times.

Westat coders found the "copy feature" in Survey Net to be a time-saver. This feature allowed the coder to select and copy single or multiple lines of data from the same intake or from another intake within a household. This feature allowed a meal such as a spaghetti dinner to be coded once and then copied as many times as needed for other household members.

In food consumption surveys, many foods are reported as mixtures, such as lasagna which is a mixture of pasta, tomato sauce, and cheese, and may also contain meat and vegetables. A variety of different pastas, cheeses, meats or vegetables may be used when lasagna is prepared. Selecting codes for mixtures can be problematic because of the many variations of recipes and products. In Survey Net, recipes could be modified to more closely match the food eaten by the participant. For example, if a participant reported eating instant oatmeal made with whole milk and the pre-defined recipe in the recipe database was made with water, the coder could modify the recipe by replacing water with whole milk. Like an unknown food, the recipe modification was accessible to all coders because it was stored in a shared network file.

The "recipe modification" feature in Survey Net increased the flexibility of the food classification system. And, without increasing the size and maintenance requirements of the food coding database, it allowed coders to accurately code the specific information reported by participants. For example, a cup of instant oatmeal made with milk contained 102 more calories, 197 more milligrams of calcium and 5.8 more grams of fat than did instant oatmeal made with water. Throughout the survey, over 4,600 different modifications were used about 16,500 times.

To enter the amount of food eaten, Survey Net provided several different measure descriptions. The amount could be entered directly as weight or with common units such as cup; small, medium or large item;

drink box; or as a shape with dimensions. All hand calculations by coders were eliminated to save time and reduce error. The database contained over 31,000 weights linked to measure descriptions.

"Reasonableness checks" were automatically performed each time a portion size was entered. If the amount was above or below a pre-established range for each food, a warning message alerted the coder to check the amount they had entered.

In addition to the reasonableness checks, Survey Net had a series of "automatic edit checks" that identified a variety of common recording or coding errors such as 2 foods from breakfast eaten at different times of the day. When edits revealed discrepancies, coders were alerted to review the entries and make corrections. Data were flagged for review by the supervisor or USDA.

One of the most frequently used features of Survey Net was the "notepad" which coders used to document their choice of a certain code or amount when the choice was not clear cut. The supervisor reviewed the note, responded to it and alerted USDA to review the item if needed. Notepad entries with supervisor and USDA responses were given to coders as feedback and were used for documentation of coding problems and their resolutions.

Databases were maintained outside of Survey Net using Paradox, a commercial database management system. They were continually updated and transferred to Survey Net approximately every two months. A data conversion program served as a pipeline between data files in Paradox and in Survey Net. This feature was implemented so that file structures could change in one system without having an impact on the other system.

Survey Net used a common database file format so that commercial software programs could be used to manipulate the files. This eliminated the need for additional custom programming and allowed USDA and Westat to generate special reports using files created in Survey Net including those for unknown foods, recipe modifications, and notepads.

Westat employed the equivalent of 7½ full-time coders with knowledge of food and food preparation and 2 coding supervisors with degrees in food and nutrition. At USDA, 3.4 full time equivalents and one supervisor, all with backgrounds in nutrition or home economics, worked as a team to review and edit the data and to maintain the food coding database.

Survey Net was an important tool for processing the data collected in the CSFII 1994-96. The 3 initial goals for developing the system were met. First, food coding and editing times decreased, primarily due to Survey Net's effective word search, unknown foods feature, copy foods feature, automatic calculation of quantities, built-in edits and electronic data transfer. Second, food coding flexibility increased through the recipe modification feature. And third, oversight of the food coding operation increased

through enhanced communications provided by the Survey Net's notepads, edit tracking features and overall monitoring capabilities. Data were processed, reviewed and released to the public approximately 8 months after the final data receipt from the contractor--a record time for nationwide food consumption surveys.