Paper P5-3	***Translating Food Intakes Into Data***
	Betty P. Perloff, ARS

CONTINUING SURVEY OF FOOD INTAKES BY INDIVIDUALS (CSFII) METHODOLOGY: TRANSLATING FOOD INTAKES INTO DATA

Linda A. Ingwersen, Etta Sue Haggerty, Randy P. LaComb and Betty P. Perloff
U.S. Department of Agriculture - Agricultural Research Service
Food Surveys Research Group, Riverdale, MD

ABSTRACT

The method used by the Agricultural Research Service to process nearly 11,000 24-hour dietary recalls collected in the first year of USDA's Continuing Survey of Food intakes by Individuals 1994-96 (CSFII) contributed to the release of data about eight months after final data receipt from the contractor. This method involved several new facets, one of which was the use of a computer-assisted food coding system, Survey net, that made the survey food coding data base, the recipe data base, and the survey nutrient data base accessible with a keystroke. Other important aspects included the electronic transmission of the data, the approach taken to code new foods and mixtures that were not in the data base, the use of coding guidelines, the capability to modify survey recipes, and the updating of the data bases in Survey Net. All of these procedures increased the flexibility of the food coding process and were also responsible for the efficient translation of descriptions of foods and quantities into data.

USDA's Continuing Survey of Food Intakes by Individuals (CSFII) measures the kinds and amounts of foods eaten by Americans. Data from this survey are used to address a variety of economic, nutrition, and food safety issues. Information about foods people eat is collected through a series of open-ended probes which ask for details about the foods, their preparation, their ingredients, and the amounts consumed. These details must then be translated into data that are structured and easily manipulated by computer programs to make them effective for research and analysis.

BACKGROUND

In the CSFII, the food-to-data translation process preserves survey respondents' descriptions of foods and measures as closely as possible using a hierarchical food coding scheme established and maintained by USDA for over 30 years. These codes are used to estimate gram weights of the food portions consumed, to calculate nutrient amounts contributed by each portion, and to classify foods into food groups.

Although the food-to-data translation process changed over the years to take advantage of new data processing technology, it remained a significant bottleneck until the current survey began in 1994. Between the previous survey, which ended in 1992, and the 1994-96 CSFII, USDA developed new coding and related procedures to increase the efficiency of food survey data processing. As a result of this effort the food-to-data translation for the 1994-96 CSFII was very successful, contributing to the release of the data in record time.

Several significant factors led to the success of the food coding process. One factor was the ability to search the survey food coding database quickly and easily for foods. Once a food description was located, information about the food--the recipe, the nutrients, weights and measures--was accessed instantaneously. Other factors included increased flexibility of the food coding system within a structure comparable to past surveys; the development of new and quicker procedures for handling "unknowns;" the weekly electronic data transmissions from the

contractor; market checks conducted by field interviewers; and frequent food coding database updates. One aspect of the process which remained unchanged was USDA's use of a team approach to solve difficult coding issues and to decide when new foods were added to the database. This helped ensure consistency in the food coding database because decisions were made with a shared philosophy.

SURVEY NET

In 1990, USDA entered into a cooperative agreement with the University of Texas to explore and develop new procedures for coding and processing survey data. The University already had considerable experience using the survey food and nutrient databases, having begun to use them for their own research in 1986. Not only had they already prepared computer programs duplicating the recipe calculation methodology, but they also had developed routines for easily modifying ingredients within the survey recipes to provide nutrient calculations that were more precise for specific survey respondents. One of the most significant outcomes from this cooperative agreement was the development of Survey Net, a computer-assisted food coding and data management system, that incorporated features related to database searching capabilities, the handling of unknowns, and overall improvement of food coding and editing. For example, edit checks in Survey Net allowed many coding and reporting errors to be caught and corrected at the point of data entry. Survey Net operated on a multi-user computer network accessing a set of central databases. Survey Net was used by coders at Westat, Inc, the 1994-95 CSFII contractor, in Rockville, MD to enter food data, by supervisors to review and approve the entries, and by USDA to review the data and resolve any coding issues.

Survey Net used three complex databases--the food coding database, the recipe database, and the survey nutrient database. The food coding database contained food names and descriptions for over 7,000 food codes, as well as typical household food measures (such as cup, tablespoon, slice) and gram weights appropriate to those measures. The pre-defined recipe database contained ingredients and amounts for food mixtures for the purpose of calculating the nutritional content of foods. The USDA Nutrient Data Laboratory supplied the nutrient values for ingredients in foods and nutrient profiles for the survey codes were calculated from the individual ingredients in the recipes and their nutrient values. The Food Coding scheme used for CSFII evolved from a set of codes originally developed for a Nationwide Food Consumption Survey conducted in 1965 and used in several USDA surveys over the past three decades. Foods were categorized into 9 major food categories and over 200 subcategories. Each food has been assigned an 8-digit numeric code with the 1st digit representing one of the 9 major groups and the first 3 digits representing the subgroups. New foods that were reported in the survey were added within the structure of this coding scheme.

CODING FOODS

To code a food, all items that were possible matches with a sample person's food description were located and reviewed to select the closest match, or to determine that the food was missing from the database. The Food Search feature minimized the time required to locate and select food codes. A "search term" of either partial or complete words was entered and Survey Net displayed all the food descriptions in the food coding database that contain matching terms. Up to 10 different terms could be entered in any order. Single word searches, such as “milk,” were convenient when coders were not sure how a food was described in the database. Multiple search terms, such as “milk 2%,” were used to narrow the search eliminating descriptions that were not relevant. An average search, which involves searching over 70,000 words, took less than one second.

Many foods were consumed as mixtures of individual foods, and selecting codes for mixtures was especially problematic because of the many different variations of recipes and products. Since food mixtures in the food coding database have a pre-defined recipe in the recipe database, recipes were accessed to aid in making the closest selection during the food search. For

example, if a respondent reported eating quick cooking cream of wheat made with 2% milk, the coder would first check the pre-defined recipe and find that the cereal was made with water and, therefore, was not a perfect match. However, recipes were modified to match more closely the food eaten by sample persons. In this case, the coder replaced the water with 2% milk. Modified recipes were numbered automatically by the program and saved in a central data file that was accessed by all coders. The modification feature was one of the most important new features included in Survey Net. It increased the flexibility of the existing food classification system, providing the ability to be more specific about the foods people consumed without increasing the size and maintenance requirements of the food coding database. It also provided a more accurate interpretation about the nutritional value of the foods consumed. For example, a cup of cream of wheat made with 2% milk contains more than twice the calories and more than thirteen times the fat than when made with water.

Another feature which enhanced the flexibility of the food code structure was the ability to link foods together which were eaten in combination, such as foods combined to make a unique sandwich. The number of types of combinations was increased since previous surveys. The "combination code" identified when specific foods were consumed together as one food item and indicated the type of combination, e.g. sandwich. The individual foods within the combination were coded with their own separate food codes and amounts. This feature has provided useful information for studying food habits.

UNKNOWNS

One of the most time consuming aspects of coding and editing food consumption data over the years has been dealing with "unknowns"--foods or measures which were new and did not exist in the database, or foods or measures that could not be matched exactly to the database. This occurred frequently because new foods are continually entering the market, new package sizes of existing foods are frequently introduced, and foods are always being combined in new and different ways. Even though extensive technical databases on foods and portion size weights was maintained, many items were still reported which could not be matched with existing entries.

The processing of food data in previous surveys required extensive record keeping of food coding questions and answers. Each description of an uncodable food was written up on a paper form, a photocopy was made of the respondent’s recorded intake page, and then both the form and the photocopy were sent to USDA. Processing of an intake stopped until an answer was received. Food coding specialists at USDA researched the food, made decisions on how it should be coded, and returned the answered form to the contractor. This process sometimes took several weeks. In the meantime, a new food on the market might be reported again several times before the answer was received by to the contractor. A special feature of Survey Net, the Unknown Foods File, finally eliminated this burdensome paper request trail. The Unknown Foods File was searched and updated by all coders. When a new or unusual food was encountered for the first time, a coder entered a description of the food in the Unknown Foods File. Each entry was automatically assigned a number which served temporarily as a food code. Until USDA described how an unknown was to be handled, subsequent encounters of the food were selected from the Unknown Foods File and the temporary food code placed in the output record.

The food intake data coded at Westat was transmitted electronically to USDA weekly. The weekly transmission was used to generate a list of the new items coders added to the Unknown Foods File. It included the household identification number and the interviewer’s initials who collected the information. This list was used to request market checks for additional information on certain unknowns. Interviewers who collected the information were asked to conduct the market checks. For the first year, about 185 market checks were requested from field interviewers who then purchased the food items and sent the labels to USDA within two weeks. Information on the label such as nutrient content or nutrient claims, the ingredient list, and the package and item weight assisted in decisions on handling unknowns. To resolve unknowns, USDA either used existing survey food codes by incorporating additional information into the food code's description, such as

a brand name, or by adding new food codes to the food code database.

During the course of the 1994 survey, over 152,000 foods were coded and about 4,500 unknown foods were sent to USDA and resolved. There were two major categories of unknowns--food unknowns and amount unknowns. If the description provided by the sample person did not entirely match the survey food code description, if the sample person provided incomplete or conflicting information, or if a well-known brand was not listed in the food coding database, the food item was sent to USDA as an unknown. The expansion of the market place with new or modified foods since the previous survey, CSFII 1989-91, contributed a large number of unknowns. Some revisions to the wording of codes were required to accommodate computerized word searches for the food coding database. For example, search terms such as “diet” were added to sodas that were previously described as “sugar-free” and “coke” was added to “soft drink, cola type.” The number of unknowns in the next survey year, 1995, was 2891--one-third less than 1994. It is believed this decrease was primarily attributable to the updates to the database that clarified code descriptions and that added new market products introduced since the previous survey. Also, with a year's experience, Westat coders were more knowledgeable about the database and more skilled in conducting word searches for appropriate codes.

RESOLVING UNKNOWNS

A team approach was used to provide resolutions for unknown foods. Food coding specialists with a background in home economics or nutrition worked independently on different sets of unknowns. Various available resources were used when proposing coding options--label information from market checks, food encyclopedias, recipe books, manufacturers, restaurant and grocery chain managers, and routine consultation with the staff of the Nutrient Data Laboratory of USDA. After independently researching unknowns in about 30 intake questionnaires, three specialists met to determine the best solutions to the most difficult issues. Survey Net was used to search the databases, resolve unknowns and edit the intake records. Resolving problems with a team proved to be effective and interesting, but it was also an effort because teamwork required members:

- to listen to others
- to consider various opinions and to compromise
- to evaluate the pros and cons of proposed answers
- to use common sense
- to accept the decisions of the team
- to realize the best answer is the goal.

Individuals were rewarded with the personal satisfaction of resolving problems as a team, and the opportunity to learn about foods and food habits. Above all, however, the team approach ensured consistency in how coding problems were resolved and in which foods were added to the database. This consistency ensured that the survey data can be used for meaningful analytical research.

During team discussions certain questions, such as "Why was this food sent as an unknown?,” helped the team make decisions. If the food was in the food coding database but not selected, the team tried to determine if the coder misunderstood the database or perhaps needed additional guidance on appropriate search terms or survey code descriptions. In fact, comments entered by food coders in the notepad--a pop-up window in Survey Net--were the basis for many clarifications in the wording of survey codes. Other questions asked were: "Is this a new food and if so, is it produced by a manufacturer with a large distribution?" "How does the reported food compare to the food description of an existing survey food code?" "How much of the food was eaten by each household member?" "Was it a major part of the diet?" "Who reported it?" Small quantities, such as a tablespoon or so, consumed of an unknown were not considered as important as larger quantities and thus not as much time was spent on them. However, a food which was a major contributor to the days intake--such as infant formula fed to a baby--received serious consideration. Foods produced by a manufacturer with country-wide distribution were likely to be reported again and were incorporated in the database. Finally the following questions were asked--"Does our resolution retain the identity of the food, match the ingredients, and also have a reasonable nutrient

match? What pieces of information will be lost or saved with each proposed answer?"

Unknowns were resolved in basically four ways:

1. With one existing survey code. This was the resolution for about 70 percent of the unknowns. In these cases information, such as brand name, was added to the food description in the food coding database.
2. With more than one existing survey code. In these cases the foods were tied together by a combination code to indicate that the food was eaten as a single item such as a salad, a sandwich, a mixture, an addition to a beverage or cereal, etc. About 18 percent of unknowns were handled in this manner.
3. With recipe modifications. About 7 percent of unknowns were resolved using recipe modifications.
4. With the creation of new codes for items that did not fit in the database. This occurred for 4 percent of the unknowns.

NEW CODES

New codes were created when: 1) no code existed for a food similar to the food reported; 2) sizable amounts of nutrients were present in the food or the food was modified to have reduced amounts of certain nutrients; 3) the food was likely to be reported again; or 4) the form or type of food was of interest to researchers. Among the nine major food groups, more codes were added to the grain products and mixtures group than to other food groups. Vegetables, meat mixtures and milk products followed grain products. A special effort was made to incorporate ethnic foods and those foods modified to be lower in fat, sodium, or sugar in the food coding database.

Good management of resources required a balance be maintained between the amount of employee time spent in efforts to resolve the unknown with the result of that effort. Sometimes, this meant that procedures changed. For example, at the beginning of the survey, the entire team of five members plus the supervisor met together for all resolution discussions, but this took too much time. Later, after team members were familiar with the problem-solving approach, discussions with two team members plus the supervisor were held. The supervisor attended all meetings and was responsible for providing a summary of the review to Westat as feedback to their coders.

UNKNOWN AMOUNTS

The quantification of amounts of foods eaten and matching those amounts to the database were also important. Portion size estimation guides--standard measuring cups, measuring spoons, a ruler, thickness sticks--were used by sample persons to estimate quantities of foods they ate. A one-pint measuring cup was used to actually measure the volume of soup and cereal bowls, juice and milk glasses, and coffee mugs used by the sample person. An advantage of conducting the interview in the kitchen was the proximity of eating utensils and food labels. There were about 32,000 gram weights in the database linked to measure descriptions (cup, slice, bar, small/medium/large item) associated with specific foods. These weights were based on information from manufacturers and food labels, reference books, contracts with food laboratories, food associations, calculations, similar foods, or actual weighings at ARS. Foods continued to come out in new sizes--bite-size cookies, snack size puddings, and king-size candy bars--that needed to be added to the database. Unknown weights were entered into Survey Net and transmitted to USDA where they were resolved along with the food unknowns.

These amount measures and weights worked well most of the time. However, sometimes a sample person did not recall the amount of a food eaten or the quantity given was incomplete or ambiguous. When this occurred, a measure called "Quantity Not Specified," or QNS, was used for the amount eaten. The development of the QNS measure, with a corresponding gram weight designation, was begun at USDA prior to the 1977-78 Nationwide Food Consumption Survey. In the process of establishing QNS measures and weights for typical foods, a team reviewed food

service manuals, actual consumption data, marketing data, USDA references such as Handbook 8 and Handbook 456, dietary guidance recommendations and the size of foods in the market place. The goal in the survey was to use information provided by the sample person first and only use QNS as a last resort. Well-trained interviewers and conscientious sample persons have helped limit the use of QNS. In CSFII 1994, QNS measures were used for about 1.7% of all foods reported. The five foods in 1994 with the most instances of QNS were mayonnaise, mustard, coffee, frozen French fries, and mayo-type salad dressing. Three of these foods--mayonnaise, mustard, and mayo-type dressing (such as Miracle Whip) were spread on foods which make them difficult to quantify. Coffee may be consumed frequently throughout the day, perhaps making it difficult to recall the quantity for each occasion. Reported consumption of foods will be compared to QNS amounts and adjustments made to QNS values in the future, if necessary.

CONCLUSION

The translation of food intake information into data that can be used to address various economic, nutrition, and food safety issues is a critical part of food survey processing. Dedicated and hard-working staff at USDA and Westat, along with applications of new technology, improved the efficiency of this operation for the 1994-96 CSFII. A major component of this operation was Survey Net, a custom-built food coding and data management system. Survey Net features included access to centralized food databases, efficient search capabilities, recipe modifications, coding of foods eaten in combination, the ability for coders to record details about unknown foods and amounts quickly, and built-in editing routines. At USDA, the team approach to resolving food coding issues and maintaining food databases established over 30 years ago was retained to ensure consistency within CSFII and comparability to past surveys.