Chapter 7: Food Intake Data Processing
By Linda A. Ingwersen and Betty P. Perloff

Survey Net

The efficiency of food and nutrient intake survey data processing for CSFII 1994–96 was facilitated by the creation and use of Survey Net, a computer-assisted food coding and data management system. As mentioned in an earlier chapter, Survey Net was developed cooperatively by ARS and the University of Texas-Houston School of Public Health. It was tailored specifically to the questions, quality control needs, and data processing needs of the survey. A general-use version of the software, the Food Intake Analysis System, was also developed to provide other research organizations with user-friendly access to the ARS survey food coding and nutrient databases (University of Texas-Houson School of Public Health 1996).

Survey Net operates on a computer network in which multiple users access a set of three central databases. These include (1) the Food Coding Database, which contains food descriptions and food measures with their corresponding gram weights, (2) the Recipe Database, which includes predefined recipes for mixtures in the Food Coding Database, and (3) the Survey Nutrient Database (see "Technical Support Files" below).

Survey Net features include the abilities to quickly search the Food Coding Database to locate reported foods, review recipes of mixtures in the Recipe Database to assist in making selections, and modify existing recipes to provide selections that match reported foods more closely. It allows the user to copy coded foods (single or multiple lines of data) when the same food or foods are consumed either more than once by the same person or by more than one person within a household. Survey Net provides easy selection of appropriate portion sizes and requires no manual calculations, such as determining cubic inches for a portion described with dimensions. Also, the ability to code foods eaten in combinations (for example, individual food items within a sandwich) was expanded over past surveys and improved. A notepad is available for each 24-hour food intake where comments and questions can be written to coding supervisors and USDA to flag unusual or uncertain situations. In addition, when a questionable item is reviewed and approved by either coding supervisors or USDA, a special code signals acceptance and eliminates the need for further review.

Survey Net's method for handling "unknowns"--foods that are new on the market and do not yet exist in the Food Coding Database or foods that cannot be matched

exactly to the database--is one of the most important features of this software because of the frequency with which new or unique foods are reported in dietary surveys. Survey Net allows these foods to be recorded by food coders into a central file of unknown foods where they are automatically assigned temporary food codes for tracking their use.

Several quality checks occur automatically during food coding to catch the most common types of data entry errors, therefore reducing postentry checks and corrections. These include checks on extremely large or small portion weights and other logic checks related to codes for eating occasion, time of day, foods eaten in combination, and foods eaten in the home or away from home.

Food Coding and Editing Process

During the CSFII 1994–96, Survey Net was used by Westat for food coding and data review. At ARS, it was used to edit and finalize the food intake information. Westat's food coders used Survey Net to match descriptions of foods eaten by sample persons (SP's) to foods listed in the Food Coding Database. Coders entered partial or complete words or phrases from the SP's descriptions of foods to retrieve food codes containing the same terms. Once a matching food description was found and selected, Survey Net provided a list of common household measures (such as 1 cup or 1 small piece) appropriate for that food. Coders selected the measure corresponding to the SP's description of the amount eaten. When descriptions of foods or quantities not present in the Food Coding Database were encountered, they were entered as "unknowns" for ARS to resolve later.

A recipe modification feature of Survey Net allowed coders to view the predefined recipes listing ingredients and amounts for every food code in the Food Coding Database and to modify the recipes to match more closely the foods eaten by sample persons. Recipes were modified primarily by deleting or substituting ingredients. Modified recipes were numbered for reference purposes and included with the Recipe Database.

For CSFII 1994–96, there were three main purposes for recipe modifications: to record the specific type of fat, to record the type of milk, and to record the dilutions of foods. Recipes for foods such as vegetables, eggs, pasta, rice, and hot cereals were modified to reflect the type of fat (such as oil, margarine, margarine spreads, or butter) used in cooking. Recipes for foods such as puddings, soups, and beverages were modified to reflect the type of milk (such as whole, 2 percent, 1 percent, or skim) used in their preparation. Some foods commonly modified for

type of fat and type of milk were scrambled eggs and omelets, and macaroni and cheese. Recipes for foods such as soups, infant formulas, and beverages were modified to reflect dilutions with amounts of milk or water that differed from label directions. For example, the survey recipe for orange juice was modified if an SP reported that one can of frozen concentrate was mixed with four cans of water, instead of three cans of water.

Foods combined and consumed as one unit by SP's were considered to be "combinations." Combinations were often instances of one food being added to another, such as sugar to coffee, margarine to toast, or gravy to potatoes. Also, foods with separate ingredients, such as salads and sandwiches, were considered combinations and each ingredient was coded separately. In other instances, multi-ingredient mixtures that lacked a close match in the Recipe Database were considered combinations and coded as precisely as possible using multiple food codes. Combinations were designated as 1 of 11 combination types by coders. The combination types were beverage, cereal, bread/baked product, salad, sandwich, soup, frozen meal, ice cream/frozen yogurt, vegetable, fruit, and other mixtures. All foods within each combination were linked by a sequence number.

Westat electronically transmitted all coded intakes to ARS on a weekly basis. All entries in each intake that required review or resolution by ARS were highlighted in Survey Net's food summary screens. These included all "unknowns," newly created recipe modifications, and notepad entries of questions and explanations of coding decisions. Feedback was provided to Westat on reviewed intakes.

As the final step in Survey Net processing, the nutritive value of each food eaten was calculated using the weight of the food and data from the Survey Nutrient Database. Where recipes had been modified, nutritive values reflected those modifications.

Technical Support Files

The three databases (the Food Coding Database, the Recipe Database, and the Survey Nutrient Database) that make up the technical support files were used in processing data for the CSFII.[1] Fifteen updates of the technical support files were provided to Westat during the survey, therefore allowing data to be processed using current information.

---

1. The ARS databases are also used with the National Health and Nutrition Examination Survey III conducted by the National Center for Health Statistics, U.S. Department of Health and Human Services.

Food Coding Database

The Food Coding Database contained over 7,300 food codes, each denoting a complete description of the food and, if relevant, the preparation method. Each food code consists of eight digits used to classify foods into groups for study. The first digit in the food code identifies one of nine major food groups:

1. milk and milk products
2. meat, poultry, fish, and mixtures;
3. eggs
4. legumes, nuts, and seeds
5. grain products
6. fruits
7. vegetables
8. fats, oils, and salad dressings; and
9. sugars, sweets, and beverages.

The second, third, and sometimes fourth digits of a food code identify increasingly more specific subgroups within the nine major food groups. The remaining digits are used for identification of particular foods within a numerical sequence. Examples of code numbers and descriptions are provided below:

Code Number	Complete Food Description
28141010	Chicken, fried, with potatoes, vegetable, dessert (frozen meal, large meat portion) (include Banquet Extra Helping Fried Chicken Dinner and Swanson Hungry Man Fried Chicken Dinner)
53105260	Cake, chocolate, devil's food, or fudge, with icing, coating, or filling, made from home recipe or purchased ready-to-eat (include chocolate, devil's food, or fudge, not specified as from home recipe, from mix or bought ready-to-eat; Jack-in-the-Box Double Fudge Cake)

SP's varied in their knowledge of foods, as well as in their ability to recall or describe foods eaten. Therefore, the descriptions of foods provided varied from very specific to very general. Also, SP's could not always provide details regarding food preparation (such as the method of cooking or whether the food was cooked with or without fat), the original form of the food (such as fresh, frozen, dry, or canned), or the ingredients in a mixture.

Generally, foods reported with complete descriptions were assigned codes that preserved the identity or name of the food and the amount of detail specified. However, if the description of a food was general, such as bread, juice, or beef, a "not further specified" (NFS) code was assigned (see "Recipe Database" below). In other cases, foods were reported with descriptions that lacked only one detail. These foods were placed in codes that provided as much detail as given and noted the one lacking detail as "not specified" in the code description, for example, "chicken breast, fried, no coating, not specified as to skin eaten."

In preparation for the CSFII 1994–96, the Food Coding Database used for CSFII 1989–91 was expanded to capture important food processing and nutritional information that was to be collected during the food intake interviews. Some expanded food groups were vegetables; infant formulas; baby foods; margarines, spreads, and butters; fast-food sandwiches; and home-prepared soups. Ethnic foods and new foods reported during the course of the survey or during NHANES III were added to all food groups.

Codes for cooked vegetables were made specific as to the original form of the vegetable before cooking, such as fresh, canned, or frozen. Food weights were keyed to the original form of the vegetable.

Codes for margarines, spreads, and butter were made specific as to form (stick, tub, liquid) and salt content (unsalted or not), and many brand names were specified within this food group.

In CSFII 1989–91, infant formula codes were specific as to brand (such as Enfamil and Similac), but in 1994–96 they were also specific as to the original form of the formula before preparation--made from dry powder, made from liquid concentrate, or purchased ready to feed.

Identification by brand names was widespread in the Food Coding Database. Several types of survey codes were brand specific in the description of the code or in the weights provided. Codes were unique to a particular brand if warranted, such as for breakfast cereals that differ in fortification levels, or they encompassed several brands of similar foods, such as cheese crackers. When appropriate, measures and their gram-weight equivalents were specified by brand.

The guidelines used to decide if a new code was needed for a brand name food were the same as for other foods. A new code was created for one or more of the following reasons: if no code existed for a food similar to the food reported, if the reported food contained either sizable amounts or intentionally reduced amounts of

one or more nutrients, if the food was likely to be reported again, or if the form or type of food was of special interest to data users. Special effort was made to incorporate ethnic foods and foods modified to be lower in fat, sodium, or sugar. Another area of expansion was in the Food Coding Database's list of food measures and their corresponding weights in grams. Food measures and gram weights were examined for consistency by a weights and measures team that included members from ARS and the National Center for Health Statistics of the U.S. Department of Health and Human Services. Cubic-inch weights of many meats and fluid-ounce weights of beverages were reviewed and revised if necessary. Dimensions were added to the measure description for many fresh fruits and vegetables. New foods and ethnic foods were prepared and weighed in a USDA food laboratory and added to the database. Brand-specific and household measures were also added to the list as needed. There are presently over 30,000 weights for measures of foods in the Food Coding Database.

Recipe Database

The purpose of the Recipe Database was to provide information for use during generation of the Survey Nutrient Database. It contained a recipe entry for each unique food code in the Food Coding Database. These entries included ingredients and their amounts, as well as information for determining changes in nutrients that might occur during cooking. Foods that are not mixtures, for example, whole milk, were represented as single ingredient recipes. Ingredients were identified with codes linking them to the primary data set of nutrient values (see “Primary Data Set” under “Survey Nutrient Database”). The Recipe Database also serves as public documentation for how nutrient values were calculated for each survey food code. Recipes are considered representative, meaning they are not exact for every SP, nor were they developed to determine the intake of specific food ingredients. A variety of popular, regional, and specialty cookbooks were consulted in constructing representative recipes. Recipes for many of the commercially available mixtures were estimated from labels (Marcoe and Haytowitz 1993). An extensive review of the Recipe Database was conducted before CSFII 1994–96 began. Recipes were evaluated for current culinary practices or, if the recipe represented a commercial item, ingredient formulations were assessed for relevancy in the 1994–96 market.

Many recipes for foods in the database were reviewed and revised to permit easy modification of ingredients. This was accomplished by expanding recipes consisting of one or two ingredients to multiple ingredients. For example, the recipe for a cheese omelet that had been two ingredients (scrambled egg and cheese) became five ingredients (egg, cheese, milk, fat, and salt). The separate

ingredients for the omelet recipe allowed for the type of milk, fat, and cheese in the recipe to be modified to match information supplied by the SP.

In preparation for the CSFII 1994–96, recipes for NFS food codes were reviewed. These codes were used when SP's were unable to provide further detail about a food. For example, the "Milk, NFS" code was used when SP's did not give the fat content of the milk they drank. The recipe for "Milk, NFS" was a composite of whole milk, 2-percent milk, 1-percent milk, and skim milk in proportions that reflected milk production statistics. Recipes for other NFS codes were based on composites, as for milk, or they were based on the form of food most frequently consumed in the food group in question. For example, the recipe for "Bread, NFS" was white bread.

Survey Nutrient Database

The Survey Nutrient Database has been maintained since 1985, specifically for use with nationwide food surveys (Perloff et al. 1990). Its source of nutrient values is the primary data set of nutrient values maintained in the ARS Nutrient Data Laboratory (see "Primary Data Set" below). A new version of the Survey Nutrient Database is prepared to represent each year of the survey.

For CSFII 1994–96, the Survey Nutrient Database included values for food energy and the following nutrients and food components: protein, total fat, 19 individual fatty acids, total saturated fatty acids, total monounsaturated fatty acids, total polyunsaturated fatty acids, cholesterol, total carbohydrate, dietary fiber, vitamin A (as international units and as retinol equivalents), carotenes, vitamin E, vitamin C, thiamin, riboflavin, niacin, vitamin B^6, folate, vitamin B^{12}, calcium, phosphorus, magnesium, iron, zinc, copper, sodium, potassium, alcohol, and moisture (water).

The Survey Nutrient Database contained two files of nutrient values: (1) the survey nutrient values, set 1, which included data for each unique survey food code from the Food Coding Database, and (2) the survey nutrient values, set 2, which was identical to set 1 with the following exception: in recipes where salt was considered an optional ingredient, it was removed from the recipe before the nutrients were calculated.

Both set 1 and set 2 of the survey nutrient values were used during the last step of Survey Net processing when the nutritive value for each consumed food was calculated. If the SP indicated salt was used in cooking the food, or if he or she did not know, data were selected from set 1. If salt was not used, data were selected from set 2.

The amount of salt as an ingredient changed for many foods in 1994–96 as a result of two procedural changes in development of the Survey Nutrient Database. First, the weight used for a teaspoon of salt in recipes was revised from 5.5 grams to 6.0 grams, based on a recommendation from the Morton Salt Company. Second, the amount of salt as an ingredient was adjusted in the recipes for cooked vegetables and meats to correspond to the amounts used in the Third National Health and Nutrition Examination Survey III (NHANES). The latter adjustment, which was made to increase comparability between CSFII and NHANES, raised the amount of salt in some recipes and lowered it in others. Overall, the two types of changes resulted in higher sodium values for many foods in the nutrient database.

Primary Data Set. The primary data set of nutrient values is maintained by the ARS Nutrient Data Laboratory in support of the National Nutrition Monitoring and Related Research Program (U.S. Department of Health and Human Services and USDA 1993). The main source of data for the primary data set is the USDA Nutrient Data Base for Standard Reference (NDBSR). Release 11 (USDA–ARS 1996) of NDBSR was used for preparing the primary data set for the CSFII 1994–96 Survey Nutrient Database. Unpublished data collected by the ARS Nutrient Data Laboratory were also used as needed, especially for new products such as low-sodium and low-fat foods. As the survey was conducted, data for new foods were added as they were reported by SP's. The number of foods in the primary data set after processing CSFII 1994–96 was 3,055.

Several changes were made to the primary data set between CSFII 1991 and the 1994–96 survey. Vitamin C values in sausages and luncheon meats were changed to zero to reflect the industry practice of adding sodium erythorbate as an antioxidant instead of sodium ascorbate. Sodium values in soups that were major contributors of nutrients in CSFII 1991 were evaluated by market checks and company data and were updated if needed. All infant formulas were evaluated and updated as necessary. Nutrient values for margarines and spreads were updated to reflect the predominant oils currently being used based on information provided by the Institute of Shortening and Edible Oils and the National Association of Margarine Manufacturers. Values for added vitamin E in breakfast cereals were reviewed and updated when necessary. Many changes also occurred in the primary data set as a result of updates made to the USDA Nutrient Data Base for Standard Reference; these updates were primarily in bakery products, breakfast cereals, infant formulas, and canned vegetables.

Most of the values for major contributors of nutrients were supported by laboratory analyses (Matthews 1991). Nutrient values not available from laboratory analyses were imputed by Nutrient Data Laboratory nutritionists from

data for other forms of the food or similar foods (Gebhardt 1992). For each value in the primary data set, a source code indicated whether the value was analytical or imputed. Values for carotenes were those used by ARS in arriving at the values for total vitamin A and were not solely beta-carotene. Also, the values for vitamin E (quantified as alpha-tocopherol equivalents) were based on somewhat limited analytical data.

The state of analytical methodology for measuring nutrients in foods was evaluated by Beecher and Matthews (1990), and they reported that adequate methodology for folate was lacking. Although the microbiological method approved by the Association of Official Analytical Chemists International applied only to foods that contain the free forms of the vitamin, data generated by ARS for use in food composition databases were obtained by a modified method using enzymes to release bound forms. Research on determining the folate content of high-protein and high-carbohydrate foods indicated that additional improvements in methodology were needed (Martin et al. 1990).

Recipe calculations. Entries in the Recipe Database identify the primary data set items used to derive the survey nutrient values, set 1 and set 2. As mentioned in the Recipe Database discussion, some survey food codes had a one-to-one correspondence with items in the primary data set and were represented by single ingredient recipes, such as the following:

Survey food code:
111-12110, Milk, cow's fluid, 2% fat
Recipe ingredient:

PDS Number	PDS item	Amount
1079	Milk, 2% fat, with vitamin A	100 grams

However, many survey food codes required multiple ingredients, for example:

Survey food code:
423-01010, Peanut butter sandwich
Recipe ingredients:

PDS Number	PDS item	Amount
16098	Peanut butter	24.0 grams
18069	Bread, white	52.0 grams

The retention factor method (Powers and Hoover 1989) was used for calculating the nutrient content of recipes to generate values for the Survey Nutrient Database (Perloff 1985). Factors for calculating moisture and fat changes were stored in each recipe. Factors for estimating vitamin and mineral losses were stored in a separate data file, the nutrient retention factors file, which was accessed during the

recipe calculation procedure. The presence of special codes in the recipe entries indicated when the retention factors were used.

References

Beecher, G.R., and R.H. Matthews. 1990. Nutrient composition of foods. *In* M.L. Brown, ed., Present knowledge in nutrition, 6th ed., pp. 430–443. International Life Sciences Institute Press, Washington, DC.

Gebhardt, S.E. 1992. Imputing nutrient values. *In* 17th National Nutrient Data Bank Conference Proceedings, pp. 143–153. International Life Sciences Institute Press, Washington, DC.

Marcoe, K.K., and D.B. Haytowitz. 1993. Estimating nutrient values of mixed dishes from label information. Food Technology 47:69–75.

Martin, J.I., W.O. Landen, Jr., A.M. Soliman, and R.R. Eitenmiller. 1990. Application of a tri-enzyme extraction for total folate determination in foods. Journal of the Association of Official Analytical Chemists 73:805–808.

Matthews, R.H. 1991. Current HNIS nutrient data research. *In* S.P. Murphy, ed., 16th National Nutrient Databank Conference Proceedings, pp. 129–132. The CBORD Group, Ithaca, NY.

Perloff, B.P. 1985. Recipe calculations for NFCS database. *In* S. Murphy and D. Rauchwarter, eds., Proceedings of Tenth National Nutrient Data Bank Conference, pp. 11–21. National Technical Information Service, Springfield, VA.

Perloff, B.P., R.L. Rizek, D.B. Haytowitz, and P.R. Reid. 1990. Dietary intake methodology II: USDA's nutrient database for nationwide dietary intake surveys. Journal of Nutrition 120:1530–1534.

Powers, P.M., and L.W. Hoover. 1989. Calculating the nutrient composition of recipes with computers. Journal of the American Dietetic Association 89:224–232.

University of Texas-Houston School of Public Health. 1996. Food intake analysis system, version 3.0. Houston, TX.

U.S. Department of Agriculture, Agricultural Research Service. 1996. USDA nutrient database for standard reference, release 11 (SR11). <http://www.nal.usda.gov/fnic/foodcomp.>

U.S. Department of Health and Human Services and U.S. Department of Agriculture. 1993. Ten-year comprehensive plan for the national nutrition monitoring and related research program. Federal Register 58:32751–32806.