CONTENTS FOR ATTACHMENTS 1 AND 2

CONTENTS FOR ATTACHMENTS 1 AND 2

CONTENTS FOR ATTACHMENTS 1 AND 2

CONTENTS FOR ATTACHMENTS 1 AND 2

ATTACHMENT 1

Documentation for the Continuing Survey
of Food Intakes by Individuals (CSFII)

1. CONTENTS OF THE FOOD COMMODITY INTAKE DATABASE CD-ROM

1.1 Tree diagram of files on CD-ROM

These files comprise the Food Commodity Intake Database for the 1994-96 Continuing Survey of Food Intakes by Individuals and its Supplemental Children's Survey (CSFII 1994-96, 1998)

Directory and File Names	Title or Description
[CD-ROM drive]:	Drive letter for CD-ROM
\readme.txt	Readme file
\document	Documentation files
\section1.wpd	Contents, tree diagram, data file list
\section2.wpd	Essential information
\section3.wpd	Methodology for translating food intakes into EPA food commodities
\section4.wpd	Data file characteristics and formats
\section5.wpd	Control counts
\appenda.wpd	Appendix A
\appendb.wpd	Appendix B
\appendc.wpd	Appendix C
\epa_comm	EPA Commodity Files
\csffcm.wpd	EPA procedures for assigning EPA cooked status, food form, and cooking method values to USDA food codes
\foodvoc.wpd	EPA Food commodity vocabulary
\intake	Food Commodity Intake Data Files
\comm9498.txt	Commodity intakes for the Continuing Survey of Food Intakes by Individuals (CSFII 1994-96, 1998)
\ffcm9498.txt	Commodity intakes by cooked status, food form, and cooking method for CSFII 1994-96, 1998
\smpl9498.txt	Sample person data for CSFII 1994-96, 1998
\Support	Supporting Data Files
\fc_comm.txt	CSFII 1994-96, 1998 foods linked to EPA commodities (grams of commodities per 100 grams food)
\fcdesc.txt	Food code descriptions
\fcincl.txt	Food code include statements (extensions of food code descriptions)
\fcscheme.txt	Food code outline
\moddesc.txt	Descriptions of recipe

modifications

1.2 Food Commodity Intake Database for the 1994-96 Continuing Survey of Food Intakes by Individuals its Supplemental Children's Survey (CSFII 1994-96, 1998)

Table of Database Files

Documentation Files

section1.wpd	Table of contents, File names
section2.wpd	Essential information
section3.wpd	Methodology for translating food intakes into EPA food commodity intakes
section4.wpd	Data file characteristics and formats
section5.wpd	Control counts
appenda.wpd	Appendix A
appendb.wpd	Appendix B
appendc.wpd	Appendix C

EPA documents

csffcm.wpd	EPA procedures for assigning EPA cooked status, food form, and cooking method values to USDA food codes
foodvoc.wpd	EPA Food commodity list

Intake Data Files

comm9498.txt	Commodity intakes for CSFII 1994-96, 1998
ffcm9498.txt	Commodity intakes by cooked status, food form, and cooking method for CSFII 1994-96, 1998
smpl9498.txt	Sample person data for CSFII 1994-96, 1998

Supporting Data Files

fc_comm.txt	CSFII 1994-96, 1998 foods linked to EPA commodities (grams of commodities per 100 grams food)
fcdesc.txt	Food code descriptions
fcincl.txt	Food code include statements (extensions of food code descriptions)
fcscheme.txt	Food code outline
moddesc.txt	Descriptions of recipe modifications

1.3 Summary descriptions of Food Commodity Intake Database Files

This CD-ROM contains the Food Commodity Intake Database. Summary descriptions of files included on this CD-ROM are presented below. Complete information on data file characteristics and formats are found in \document\section4.wpd.

1.3.1 Documentation files (files in directory \document)

Eight documentation files (5 documentation sections and 3 appendices) are included as WordPerfect files. They explain the methodology used to develop the Food Commodity Intake Database, data file characteristics and formats, and other information important for using this database.

1.3.2 EPA Commodity files (files in directory \epa_comm)

The Environmental Protection Agency (EPA) provided two WordPerfect files:

EPA Food Commodity Vocabulary - "foodvoc.wpd"

This file contains a list of EPA commodities used in this project. The list includes the name of each EPA commodity, its EPA code, and a description of the weight basis for the commodity.

PROCEDURES FOR CODING CSFFCM - "csffcm.wpd"

EPA assigned codes for cooked status, food form, and cooking method to each CSFII food code. This file discusses the procedures used by EPA to assign those codes, which were then used to create the "Commodity Intakes by Cooked Status, Food Form, and Cooking Method" file described below.

1.3.3 Food Commodity Intake Data Files (files in directory \intake)

These files include (a) two commodity intake data files that were created by translating the CSFII 1994-96, 1998 food intakes into EPA-defined food commodities, and (b) one file containing characteristics of sample persons in the survey.

COMMODITY INTAKE DATA - "comm9498.txt"

The Commodity Intake File contains one record per EPA-defined commodity, per person, per day. Each record contains the daily total intake of a commodity, aggregated from the foods reported by the sample person. If the sample person provided 2 days of intake data, records for daily averages of each commodity are also present. Data are presented as grams of commodity consumed per kilogram of body weight. Records are not present for commodities not consumed by the sample person.

COMMODITY INTAKES BY "COOKED STATUS, FOOD FORM, AND COOKING

METHOD" - "ffcm9498.txt"

This file further defines the CSFII commodity intake data by the EPA codes for "Cooked Status," "Food Form," and "Cooking Method" (CSFFCM). They contain one record per commodity per unique CSFFCM, per person, per day. Each record contains the daily total intake of a "commodity-CSFFCM" combination, aggregated from the foods reported by the sample person. Records of daily averages for each Commodity-CSFFCM combination are present, if the sample person provided 2 days of intake data. Data are presented as grams consumed per kilogram of body weight. Records are not present for Commodity-CSFFCM combinations not consumed by the sample person. For an explanation of "cooked status, food form, and cooking method" codes, see directory \epa_comm on this CD-ROM.

SAMPLE PERSON DATA - "smpl9498.txt"

The Sample Person data file contains demographic, sampling weight, and other respondent and household information. One record per responding sample person is present. Data in these files were extracted from the CSFII 1994-96, 1998 Household and Sample Person data records.

1.3.4 Supporting Data Files (files in directory \support)

Five data files provide supporting documentation for the translation of the CSFII 1994-96, 1998 intake data into EPA food commodities.

FOOD-CODE-TO-COMMODITY TRANSLATION FILE - "fc_comm.txt"

The Food-Code-to-Commodity Translation file documents the assignment of USDA food codes to EPA commodities as used in this translation of food intakes into commodity intakes. Data are expressed as the amount of each EPA commodity contained within a 100 gram portion of each specific CSFII food code. A separate record exists for each commodity assignment for a food code. For a discussion about the assignment of food codes to commodities, see section 3 on Methodology.

FOOD CODE DESCRIPTIONS - "fcdesc.txt"

The Food Code Descriptions file provides descriptions for each food code used in the CSFII 1994-96, 1998. Descriptions are usually generic in nature except for certain breakfast cereals, infant formulas, and candies. The file includes descriptions up to 200 characters long, as well as abbreviated descriptions up to 60 characters long.

FOOD CODE INCLUDE STATEMENTS - "fcincl.txt"

Food Include Statements are names of specific foods associated with a particular generic food description in the Food Code

Descriptions file.

FOOD CODE OUTLINE - "fcscheme.txt"

The food code scheme identifies the meanings associated with the first, second, and third (and sometimes fourth) digits of the 8-digit food codes.

DESCRIPTIONS OF RECIPE MODIFICATIONS - "moddesc.txt"

In the CSFII, some database recipes were modified when a responding sample person supplied specific information about certain food ingredients. This file contains descriptions of those modifications.

1.4 Suggested citations, acquisition, contacts and disclaimers

Suggested citation for this database:

> U.S. Environmental Protection Agency, Office of Pesticide Programs and U. S. Department of Agriculture, Agricultural Research Service. 2000. Food Commodity Intake Database, Version 2.1. CD-ROM.

Suggested citations when identifying the original sources of data for this database:

> U.S. Environmental Protection Agency. 2000. The EPA Food Commodity Vocabulary, Master List dated June 15, 2000.
>
> U.S. Department of Agriculture, Agricultural Research Service. 2000. Continuing Survey of Food Intakes by Individuals 1994-96, 1998. National Technical Information Service, 5285 Port Royal Road, Springfield VA 22161;(703)487-4650. CD-ROM. NTIS Accession no. PB2000-500027.

Acquisition

Copies of the Food Commodity Intake Database may be purchased from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161; telephone 1-800-553-6847.

Contacts

U.S. Environmental Protection Agency(7509C)
Ariel Rios Building
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460
FAX: 703-305-0871

Disclaimers

Mention of trade names, commercial products, or companies in this

database is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture over others not mentioned.

The United States Department of Agriculture (USDA) prohibits discrimination in its programs on the basis of race, color, national origin, sex, religion, age, disability, political beliefs, and marital or familial status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (Braille, large print, audiotape, etc.) should contact the USDA Office of Communications at (202) 720-2791.

To file a complaint, write the Secretary of Agriculture, U.S. Department of Agriculture, Washington, DC 20250, or call (202) 720-7327 (voice) or (202) 720-1127 (TDD). USDA is an equal employment opportunity employer.

2 ESSENTIAL INFORMATION

The notes in this section provide information to help use this database effectively, although they do not substitute for reading the remainder of the documentation.

The Food Commodity Intake Database is based on:

(a) Food consumption surveys conducted by the U.S. Department of Agriculture (USDA)--the 1994-96 Continuing Survey of Food Intakes by Individuals and its Supplemental Children's Survey (CSFII 1994-96, 1998), and

(b) A list of commodities prepared by the U.S. Environmental Protection Agency (EPA)--the Food Commodity Vocabulary, dated June 15, 2000.

The Food Commodity Intake Database was developed for the purpose of estimating human exposures to pesticide residues through intakes of foods and beverages. For information about how to use these data for dietary risk assessment, users should contact the EPA Office of Pesticide Programs, Division of Health Effects, Ariel Rios Building, 1200 Pennsylvania Avenue, NW(7509C), Washington, DC 20460, FAX 703-305-0871.

Food commodity intakes are included for each sample person from the surveys. Intakes are expressed as grams of commodity consumed, per kilogram of body weight, per day. Two days of intakes, plus daily averages, are present for most sample persons.

Data files are also included which indicate how commodities were consumed in terms of their "cooked status, food form, and cooking method." Codes for these parameters were assigned to the USDA food codes by EPA.

The smallest amount of commodity consumption represented in this database is 0.000001 grams, per kilogram of body weight, per day. Amounts smaller than 0.000001 were set to this amount. This expression is not intended to suggest that data can be estimated this precisely. It was set to this small amount at the request of EPA to reduce the likelihood that small amounts would round to zero.

Water consumption is not included in this database at this time. Intakes of water, as defined in the EPA food commodity list, are being prepared by the EPA Office of Water.

A Food-Code-to-Commodity Translation File is included in this database to document how each USDA food code was assigned to the EPA commodities. Data are expressed as grams of commodity per 100 grams of food.

The amounts of commodities assigned to one USDA food code do not always equal 100. This occurs because (a) some food constituents were not assigned to commodities, for example, salt was excluded, and (b) conversions were sometimes necessary in order to express the commodities as defined by EPA; for example, rice as consumed was converted to the dry weight of the grain.

The smallest amount of a commodity identified in the Food-Code-to-Commodity Translation File is 0.001 grams per 100 grams of food. Commodity amounts lower than this limit, but greater than zero, were set to 0.001 grams. Three decimal places were selected for expression of the commodity amounts to accommodate an EPA requirement that the presence of commodities in very small amounts be represented in this database. It is not intended to suggest that data can be estimated this precisely.

Several other special requirements were requested by EPA to guide the translation of USDA food codes to commodities. These requirements are discussed in Section 3.

Users may observe some unexpected findings in the Food-Code-to-Commodity Translation File. For example, the following five commodities are present in small amounts in many foods: "Cassava," "corn, field, starch," "Potato, flour," "Rice, flour," and "Wheat, flour." This occurs because modified food starch is frequently listed as an ingredient in processed foods. When the source of the starch was not identified, it was assigned to those commodities. Other examples of unexpected findings can be found in section 3.8, Notes about reviewing the Food-Code-to-Commodity Translation File. Reasons for unexpected findings can be found in discussions on matching USDA food codes to EPA commodities (section 3.6.1), converting weights of foods and ingredients (sections 3.6.2), and special requirements for assigning foods and ingredients to EPA commodities (section 3.7).

Not all commodities were consumed.

Data users are encouraged to read section 3, Methodology for Translating the CSFII 1994-96, 1998 into intakes of EPA-defined commodities. Limitations of the data are discussed in section 3.9.

3 Methodology for translating the 1994-96 Continuing Survey of Food Intakes by Individuals and its Supplemental Children's Survey (CSFII 1994-96, 1998) into intakes of EPA-defined commodities

3.1 Food Commodity Intake Database

The Food Commodity Intake Database contains intake data and related information for approximately 548 commodities listed in the Environmental Protection Agency's (EPA) Food Commodity Vocabulary (16). The database is the result of cooperative efforts by the U.S. Department of Agriculture (USDA), Agricultural Research Service (ARS), and EPA's, Health Effects Division. It was developed for use in estimating human exposures to pesticide residues through intakes of foods and beverages. A discussion of EPA's use of food consumption data in assessing dietary risk from pesticides is provided in Appendix A.

The database contains several different types of data files (intake data, supporting data, documentation). A complete list of the data files is located in section 1.

The intake data are based on the 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII 1994-96) and its Supplemental Children's Survey (CSFII 1998) (8), conducted by ARS.

Commodities listed in the Food Commodity Vocabulary were selected and defined by EPA. Because the survey data are expressed in terms of foods as eaten, conversions of the data were sometimes necessary in order to express them in terms of commodities as defined by EPA. For example, consumption of cooked rice needed to be expressed in terms of the EPA commodity, "Rice, dry weight of grain." Data conversions and other steps used to develop the commodity intake data are referred to, collectively, as a translation. The data sets and processes used in the translation are identified below. For information about how to use these data for dietary risk assessments, users should contact the EPA Office of Pesticide Programs, Division of Health Effects, Ariel Rios Building, 1200 Pennsylvania Avenue, NW (7509C), Washington, DC 20460.

3.2 Continuing Survey of Food Intakes by Individuals (CSFII)

The CSFII is a series of U. S. Department of Agriculture surveys designed to measure the kinds and amounts of foods eaten by Americans. The CSFII 1994-96 was conducted between January 1994 and January 1997 with a target population of noninstitutionalized individuals in all 50 states and Washington, D.C. In each of the 3 survey years, data were collected for a nationally representative sample of individuals of all ages. The CSFII 1998 was a survey of children ages 0 through 9 which was a supplement to the CSFII 1994-96. It used the same sample design as the

CSFII 1994-96 and was intended to be merged with CSFII 1994-96 to increase the sample size for children. The merged surveys are designated as CSFII 1994-96, 1998.

In the CSFII 1994-96, 1998, dietary intakes were collected through in-person interviews using 24-hour recalls on 2 nonconsecutive days. A total of 21,662 individuals provided data for the first day; of those individuals, 20,607 provided data for a second day. Over 11,000 of the sample persons represent children up to 18 years of age.

Several sets of sampling weights are available for use with the intake data, and all sets have been included in this database. Using appropriate weights, data can be analyzed in the following ways: (1) one year alone, i.e., 1994, 1995, 1996, or 1998; (2) 3 years combined for the CSFII 1994-96; or (3) 4 years combined for CSFII 1994-96, 1998. It is recommended that all 4 years be combined in order to provide an adequate sample size for children. For an explanation of sample weighting, see section 5 of the documentation on the survey CD-ROM (8). Selections from this section are also included here for easy reference (\document\appendc.wpd).

Complete results from the surveys in microdata form are available on CD-ROM (7,8). Documentation on the CD-ROMs includes survey instruments, detailed descriptions of survey and data processing methods, sample design, and calculation of weights. The survey instruments, as well as a Design and Operation Report for CSFII 1994-96 (5), can also be found on the Food Surveys Research Group web site: http://www.barc.usda.gov/bhnrc/foodsurvey/home.htm.

Several sections of the documentation from the CSFII 1994-96, 1998 CD-ROM are included here for easy reference. Section 3, a discussion of survey methodology, is located in Appendix B; excerpts from sections 5 and 6 on sampling weights and statistical notes are located in Appendix C in \document.

Three different data files (record types) from CSFII 1994-96, 1998 were used during the translation of the survey data to EPA commodities:

1. Record Type 15 (Household Level Data)
2. Record Type 25 (Sample Person Data)
3. Record Type 30 (Food Intake Data)

In addition, information was used from the following three ARS technical databases (7a,7b,7c):

1. Food Coding Database
2. Recipe Database
3. Survey Nutrient Database

These databases are maintained by ARS for processing and analyzing food consumption data, and the versions used for a survey are released along with the survey data. Primarily, databases from the CSFII 1994-96 CD-ROM were used for this translation (7c). However, if a food was reported for the first time during 1998, it may not appear on the earlier databases. For those foods, entries can be found on the CSFII 1994-96, 1998 CD-ROM (8).

A total of 5,831 different foods were reported as consumed in the CSFII 1994-96, 1998. Each food is identified by a unique 8-digit food code in the CSFII data (Record type 30). An additional 3,913 variations of these foods were also reported and identified by a unique 6-digit modification code. Appendix B, section 3.3.3 (\document\appendb.wpd) contains a discussion about food codes and descriptions. A complete list of food codes and modification codes are included in the directory \support.

3.3 EPA list of food commodities

EPA provided ARS with the EPA Food Commodity Vocabulary for this project. It is a list of over 540 commodity items for use in estimating human exposure to pesticide residues through food consumption. The list, which includes foods, beverages, and water, is included as a WordPerfect document on this CD-ROM (\epa_comm\foodvoc.wpd). It lists the name of each EPA commodity, its EPA code, and a description of the weight basis for the commodity. For example, the code for the commodity "Apple, juice" is 11000100 and the weight basis is "weight of juice at single strength (or standard dilution)." Many foods have separate entries and codes on the list to represent baby foods. For example, "Apple, juice, baby food," has a separate entry with a unique code, 11000101.

A discussion by EPA of the basis for selecting the various commodities is included in the EPA Food Commodity Vocabulary. In general, items on the list (other than water) can be categorized by the following different types:

Commodity type	Example
Food (no further description)	Blackberry
Food with distinguishing characteristic	Apple, peeled fruit
Food part	Turnip, tops
Food, processing specific	Peach, dried
Food component	Milk, fat
Food consumed as baby food	Carrot, baby food

Some foods, for example, peach, may be included on the list as several different commodity types:

EPA Code	Commodity Description
12002600	Peach
12002601	Peach, baby food
12002610	Peach, dried
12002611	Peach, dried, baby food
12002620	Peach, juice
12002611	Peach, juice, baby food

3.4 Translating CSFII to EPA food commodities

Four major steps were involved in translating the CSFII 1994-96, 1998 to EPA commodities. 1) Many of the foods reported in the survey were mixtures, and the first step was to define those foods in terms of their ingredients and amounts. 2) Next, each food or ingredient was assigned to its appropriate EPA commodity, or commodities, including converting amounts to meet EPA's weight basis requirements when necessary. The assignments were documented in a Food-Code-to-Commodity Translation File. 3) The Food-Code-to-Commodity Translation File was then used in conjunction with the food intake data from CSFII 1994-96, 1998 to produce commodity intake estimates. During this step, each record type 30 (food intake data record) was translated to represent consumption of specific EPA commodities. 4) Finally, consumption of each EPA commodity was totaled into a daily amount for each sample person and expressed in terms of grams consumed per kilogram of body weight. Two-day averages of commodity intakes were calculated for sample persons providing two days of food intake.

The lower limit for representing intake of a commodity was 0.000001 grams per kilogram of body weight per day. Any daily amount less than .000001 grams, but greater than zero, was set to this lower limit of daily consumption on the commodity intake data files. The purpose of expressing the commodity intake data to six decimal places was not to imply precision to that level. Expression to six decimal places was requested by EPA in order to ensure that intakes of very small amounts would not be overlooked. See section 3.7.20 for discussion of EPA's special requirement to include commodities consumed in very small amounts. See section 1.3 for a description of the Food Commodity Intake Data file.

Additional details regarding preparation of the Food-Code-to-Commodity Translation file are provided below.

3.5 Defining mixtures in terms of ingredients

During processing of CSFII intake data, recipes had been identified for many of the food mixtures for the purpose of estimating nutrient content. These recipes appear on the CSFII CD-ROMs in the Recipe Database (7c). A discussion of the Recipe

Database can be found in Appendix B, section 3.3.5 (\document\appendb.wpd). The recipes are considered "representative," meaning they were not exact for every sample person, nor were they constructed for the purpose of estimating consumption of ingredients. Nevertheless, the recipes had been carefully developed and refined over many years for USDA food consumption surveys; and prior to the CSFII 1994-96, 1998 they had been reviewed and updated to reflect the food marketplace at that time.

These recipes were used as the starting point for the translation project. EPA identified several situations where changes to the recipes would provide a broader representation of commodity sources. Such changes were incorporated when ingredients were assigned to the commodities. These situations are discussed below in Section 3.7, "Special Requirements for Assigning Foods and Ingredients to EPA Commodities."

When a recipe ingredient was also a mixture (e.g., bread crumbs), its ingredient amounts were estimated for this project using procedures described in Appendix B, section 3.3.5 (\document\appendb.wpd). Ingredient amounts in brand-name products were estimated based on label information using procedures previously described by Marcoe and Haytowitz (1).

During CSFII 1994-96, 1998 food coding, a recipe could be modified to match more closely the foods eaten by sample persons. There were three main purposes for recipe modifications: to record the type of fat (for example, butter); the type of milk (for example, skim milk); and the amount of water or milk used to dilute foods (for example, condensed soup). Modified recipes are included in the recipe database and were used in this translation.

3.6 Assigning foods and ingredients to EPA commodities

3.6.1 Matching USDA food codes to EPA commodities

Most CSFII food codes were matched to one or more commodity items on EPA's list. Food codes which represented mixtures were assigned based on their recipe ingredients. Ingredients which were not assigned to commodities included water, which was assigned to commodities by EPA's Office of Water, and several ingredients which were not included on EPA's list: salt, leavening, artificial sweeteners, pectin, added nutrients, and some forms of alcohol.

Commodity assignments for brand-name foods were based on ingredients listed on labels from the marketplace at the time of the CSFII 1994-96, 1998. CSFII food codes frequently represented more than one brand of a product. Assignments were based on ingredient listings of national brands, and frequently

represented a blending of ingredient lists for two or more brands.

Modified food starch is an ingredient in many processed foods. When the type of modified food starch was not specified on the label, it was assigned as 95% to "Corn, field, starch," and the remaining 5% was assigned in equal amounts to the following commodities: "Cassava," "Potato, flour," "Rice, flour," and "Wheat, flour." Maltodextrin was assigned to commodities in the same proportions as modified food starch. Soya lecithin was assigned to "Soybean, oil." Carrageenan, alginate, and agar were assigned to "Seaweed."

3.6.2 Converting weights of foods and ingredients

As stated above, conversions were made as needed to conform to the weight basis of the commodity as defined on the EPA list of commodities. For example, the EPA commodity "Corn, field, syrup" is based on the weight of the syrup; therefore, when corn syrup solids was an ingredient within a mixture, the weight was converted to the syrup form (100 grams of corn syrup solids = 131 grams of corn syrup). Factors used in this type of adjustment are based on differences between total solids of the consumed foods or ingredients and total solids for the commodities, and were derived by dividing the percent total solids of the food or ingredient by the percent total solids of the corresponding commodity. Values for total solids were derived from the 1994-96 Survey Nutrient Database (7b) as follows: 100 - percent moisture content = percent total solids.

Since grain commodities are defined as the dry weight of grain or flour, they frequently required weight adjustments to account for the amount of moisture which is absorbed during cooking. For example, rice absorbs moisture when it is steamed, so the weight of cooked rice was adjusted downward to eliminate the weight of the water.

Sometimes grain ingredients, such as flour, actually lose moisture during food processing. For example, ready-to-eat breakfast cereals may lose moisture during processing, so the weight of flour ingredients were adjusted upward to approximate their original weights before the moisture loss.

Examples of how weights were adjusted to account for differences in total solids between foods as consumed and foods as defined in EPA's commodity list:

Example 1:

Food or ingredient = Rice, cooked
Total solids = 31.56 percent
EPA-defined Commodity = Rice, dry weight of grain
Total solids = 88.38 percent
Conversion factor = 31.56/88.38 = 0.357

Example 2:

Food or ingredient = Orange juice concentrate Total solids = 42.15 percent
EPA-defined Commodity = Orange juice (weight of single strength juice)
Total solids = 11.90 percent
Conversion factor = 42.15/11.90 = 3.542

When the weight basis of a commodity was identified as "edible portion as consumed," the amount was assigned according to the form in which the food was consumed. For example, if a carrot was consumed raw, it was assigned according to its raw weight. If it was consumed cooked, assignment was made according to the cooked weight. If the carrot was consumed cooked, but it appeared in the recipe ingredient list as a raw carrot, its weight was adjusted to the cooked form before assignment. Adjustment factors embedded within the recipes were used for this type of adjustment. These factors were based on data from Agriculture Handbook No. 102, Food Yields Summarized by Different Stages of Preparation (2); Agricultural Handbook No. 8, Composition of Foods...Raw, Processed, Prepared (6), or unpublished data from ARS food laboratories.

3.6.3 CSFII Food-Code-to-Commodity Translation File

A CSFII Food-Code-to-Commodity Translation File was prepared to identify the commodity assignments (excluding water) for each food code from the CSFII 1994-96, 1998. Some food codes have more than one set of commodity values on this database. This occurs when a modification was made to the food code's corresponding recipe during the survey. (See section 3.2 above.) To identify one set of commodity values in the database, use the 8-digit food code in conjunction with the 6 digit modification code. The translation file is included on this CD-ROM in \support.

In this file, information is expressed as the amount in grams of each commodity per 100 grams of a CSFII food. However, the amounts of all commodities associated with one food code, plus modification code, may not total exactly to 100 grams. This occurs for several reasons: 1) Commodities are not always expressed in the same forms as foods are consumed; for example, the commodity form for white rice is "dry form of grain." As a

result, 100 grams of food code 56205010, "Rice, white, cooked," translate to only 35.7 grams of the EPA commodity, "Rice, white, dry form of grain." See section 3.6.2 above for examples of how weights of foods and ingredients were adjusted during the translation. Additional examples of adjustments include: 100 grams of frozen orange juice concentrate translate to 354 grams of single strength orange juice, 100 grams of corn syrup solids translate to 131 grams of corn syrup; 100 grams of dry egg powder translate to 393 grams of raw egg; and 100 grams of cooked macaroni translate into 37.9 grams of wheat flour. Anytime grains, concentrated juices, corn syrup solids, or dry egg powder are identified as an ingredient within a CSFII food, the total of commodities for the CSFII food code will likely add to more or less than 100 on the Food-Code-to-Commodity translation file. 2) Water as a commodity was excluded from this database. For example, the weight of water used to reconstitute canned condensed soups is not represented in this data file. When water is an ingredient in a CSFII food, commodities will add to less than 100 on the Food-Code-to-Commodity Translation file. 3) Some food constituents were not assigned to commodities: salt, leavening, artificial sweeteners, pectin, added nutrients, and some forms of alcohol. Anytime these foods are ingredients in a CSFII food, commodities will add to less than 100 on the Food Code-to-Commodity Translation file.

The smallest amount of a commodity that was included in this database was 0.001 grams per 100 grams of the CSFII food. Commodity amounts lower than this limit, but greater than zero, were set to 0.001 grams. Three decimal places were selected for expression of the commodity amounts to accommodate an EPA requirement that the presence of commodities in very small amounts be represented in this database. See section 3.7.19 for discussion of EPA's special requirement to include commodities consumed in very small amounts.

3.7 Special requirements for assigning foods and ingredients to EPA food commodities

In addition to converting amounts of foods and ingredients to account for weight bases specified in the EPA food commodity list, several special requirements were established to guide the assignment of foods and ingredients to commodities. These requirements, which are listed below, were based either on descriptions of foods within the EPA food commodity list, or on communications between EPA and ARS.

3.7.1 Alcoholic beverages

Rum was assigned to the commodity "Sugar, cane, sugar," adjusted for the difference in total solids.

Beer was assigned 9.6 percent to the commodity "Barley, flour"

0.7 percent to "Hop," and 4.1 percent to "Rice, flour." The remainder was considered non-commodities, or was water and was excluded from this database.

Wine was assigned to the commodity "Grape, wine and sherry," except rice wine was assigned to the commodity "Rice, white, dry form of grain," adjusted for the difference in total solids.

Other forms of alcohol were not required by EPA to be assigned to commodities.

3.7.2 Baby foods

Commodity codes are specific for commercial baby foods. All ingredients from food codes identified as baby foods were assigned to commodities specified as "from baby food." CSFII food codes representing commercial baby foods are listed in Table 1.

3.7.3 Chips

The potato portion of regular potato chips was assigned to the commodity, "Potato, chips." The potato from restructured potato chips was assigned to "Potato, dry." Corn from corn chips was assigned to the commodity, "Corn, field, meal." Apple from apple chips was assigned to "Apple, dried." Banana from banana chips was assigned to "Banana, dried." The fat ingredients from all these types of chips were assigned to commodity oils as described in the section 3.7.6 on fats and oils.

3.7.4 Coffee and tea

Commodities for coffee and tea are:

Coffee, roasted bean
Coffee, instant
Tea, dried (leaf)
Tea, instant

Prepared coffee and tea were assigned to their dry commodity forms, based on amounts typically recommended for preparation. Coffee not specified as brewed or made from instant was assigned 90% "Coffee roasted bean," and 10 percent "Coffee, instant" (4). Tea not specified as brewed from leaves or made from instant was assigned 86 percent "Tea, dried (leaf)" and 14 percent "Tea, instant" (13). Herbal teas were assigned to the commodity, "Herbs, other."

3.7.5 Egg, dried

Dried egg ingredients (whole, yolk, or white) were assigned to the commodities for their fresh forms, adjusted for differences

in total solids.

3.7.6 Fats and oils

3.7.6.1 Vegetable oils

When CSFII food descriptions represented one specific brand name, oil ingredients listed on the labels were assigned to their corresponding commodities; for example, coconut oil was assigned to the commodity "Coconut, oil."

When a label listed more than one type of oil, for example, coconut and/or palm oil, and no information was available on the proportion of each oil in the food product, assignment was made according to their relative proportions in the food supply. When labels from more than one brand were used to identify ingredients, all identified oils were used in the assignments.

Information on relative proportions of certain oils in the food supply was used in situations as described below. These relative proportions are based on unpublished data from the USDA Center for Nutrition Policy and Promotion's Food Supply Database. Their data were based upon information from the USDA's Economic Research Service and Foreign Agricultural Service (3,10,14).

Percent	Oil
74.6	Soybean
8.8	Corn
6.6	Cottonseed
3.7	Rapeseed (canola)
3.6	Olive
2.4	Peanut
0.3	Sunflower
0.001	Safflower
0.001	Sesame

When specific oil ingredients could not be determined, assignments were based on composites from the above list of oils, based on relative proportions of each oil. Judgments were made for each recipe regarding the most appropriate oils to include, based on information from popular and ethnic cookbooks, and consultation among food and nutrition specialists.

Two composites were used in assigning shortening to oil commodities. One was for commercial foods (soybean and cottonseed oils) and the other for home-prepared foods (soybean and rapeseed oils). Proportions of each oil within the composites were based on the above food supply data. For foods

that could be either commercial or home prepared, assignments were made 50 percent to each composite. Oil sources for shortening were based on information from the Nutrient Data Laboratory(9).

3.7.6.2 Margarine and Butter

In most instances, the oil ingredients in margarine and margarine-like spreads were assigned to commodities according to the following proportions (3):

Percent	Oil
81.7	Soybean
10.2	Corn
5.7	Rapeseed (canola)
0.8	Cottonseed
0.8	Safflower
0.8	Sunflower

However, when labels of manufactured foods identified a specific type of margarine as an ingredient, e.g., corn oil margarine, the oil ingredient was assigned to the specific oil source.

In the Survey Recipe Database, recipes frequently contained either margarine or butter as a "default" ingredient (specific source of fat not known by the survey respondent). In those cases, the butter or margarine ingredient was divided into 70 percent regular margarine and 30 percent butter before assignment to EPA commodities, based on the proportion of those foods in the food supply (3). Recipe modifications, when the respondent specifically named margarine or butter as the ingredient, were not changed.

Butter was assigned to milk commodities as discussed in section 3.7.13.

3.7.7 Fish

EPA commodities for fish include:

Fish, freshwater finfish
Fish, freshwater finfish, farm raised
Fish, saltwater finfish, tuna
Fish, saltwater finfish, other
Fish, shellfish, crustacean
Fish, shellfish, mollusc

Species of fish and seafood were assigned to their appropriate EPA commodities. Catfish was assigned 100 percent to the commodity, "Fish, freshwater finfish, farm raised." Trout was assigned 50 percent to the commodity "Fish, freshwater, finfish,"

and 50 percent to "Fish, freshwater finfish, farm raised" (9). No other species reported in the CSFII 1994-96, 1998 were assigned to the "farm raised" category.

3.7.8 Fruit

Fruit juices and nectars were assigned to commodities for their respective juices, when such commodities existed, adjusted to the weight of single-strength juice. When juice was not listed as a separate commodity for a fruit, juice was assigned to the regular form of the fruit. Sweeteners from sweetened juice and juice drinks were assigned to the appropriate commodities for carbohydrate sweeteners.

Fruits consumed dried were assigned to commodities for the dried forms. Re-hydrated or stewed dried fruits were required by EPA to be assigned to the succulent forms.

The recipe for food code 62101050, "Fruit mixture, dried (mixture includes three or more of the following: apples, apricots, dates, papaya, peaches, pears, pineapples, prunes, raisins)," was assigned to all 9 commodities listed in the description, in equal proportions.

Sweeteners in canned fruit were assigned to the appropriate commodities for carbohydrate sweeteners.

In the CSFII 1994-96, 1998, yogurt items containing fruit were coded under umbrella codes described as "Yogurt, fruit variety." The kind of fruit was not specified. For this project, the fruit ingredients were assigned in equal proportions to the commodities "Strawberry," "Blueberry," "Cherry," "Peach," "Banana," and "Raspberry." The yogurt portion was assigned to milk as described in section 3.7.13.

In CSFII ice cream and ice milk were coded as either "chocolate," or "flavors other than chocolate." For ice cream or ice milk coded as "flavors other than chocolate," an amount of .001 grams per 100 grams of food was assigned to each of the following commodities: "Strawberry," "Blueberry," "Cherry," "Peach," "Banana," and "Raspberry."

3.7.9 Gelatin, dry

Dry gelatin as an ingredient in nonmeat foods was assigned in equal proportions to the commodities "Beef, meat byproducts" and "Pork, meat byproducts." As an ingredient in beef-containing foods, e.g., beef broth, dry gelatin was assigned to "Beef, meat byproducts" only. As an ingredient in pork-containing foods, it was assigned to "Pork, meat byproducts" only.

3.7.10 Legumes

Sprouted or cooked legumes (dry beans and peas) were assigned to their dry forms.

Beans described as white beans were assigned 50 percent to the commodity "Bean, navy" and 50 percent to the commodity "Bean, great northern."

3.7.11 Meat

3.7.11.1 Beef. EPA commodities for beef are:

Beef, fat
Beef, kidney
Beef, liver
Beef, meat
Beef, meat byproducts
Beef, meat, dried

Each beef item, other than kidney, liver, and byproducts, was separated into two components. The total fat component, as specified in the Survey Nutrient Database (7c), was assigned to the commodity "Beef, fat." After subtracting total fat, the remainder of an item was assigned to the commodity "Beef, meat."

Beef kidney, liver, and byproducts were assigned to their respective commodities, which included both their fat and nonfat components. See section 3.7.11.4 regarding kidney. Beef byproducts are defined in the EPA list of commodities, located in the \epa_comm.

Dried chip beef and beef jerky, including both fat and nonfat components, were assigned to the commodity "Beef, meat, dried."

Veal, bison, and buffalo items were included in beef commodities.

3.7.11.2 Game. Bear, caribou, deer, frog, and squirrel were assigned to the EPA commodity, "Meat, game." Other game meats (armadillo, beaver, elk, groundhog, moose, snake, opossum, and raccoon) listed in the description of this commodity were not reported in CSFII 1994-96, 1998.

3.7.11.3 Goat. EPA commodities for goat are:

Goat, fat
Goat, kidney
Goat, liver
Goat, meat
Goat, meat byproducts

No goat kidney, liver or meat byproducts were reported in the CSFII 1994-96, 1998. Other goat items were separated into two components. The total fat component, as specified in the Survey

Nutrient Database (7b), was assigned to the commodity "Goat, fat." After subtracting total fat, the remainder was assigned to the commodity "Goat, meat."

3.7.11.4 Kidney. The CSFII 1994-96, 1998 did not distinguish between different types of kidney. Kidney was assigned 56 percent to "Beef, kidney," 43 percent to "Pork, kidney," and 1 percent to "Sheep, kidney" (4).

3.7.11.5 Meat, not further specified. Some foods were described as containing meat without being specific for the type of meat, for example, "Meat Pie, Not Further Specified." The Survey Recipe Database entries for those foods usually contained no more than one meat ingredient. For most of those items, the meat ingredient was assigned as 56 percent beef, 43 percent pork, and 1 percent lamb. Recipes for processed meats and pizzas were not changed. These proportions were based on USDA food supply data for 1991-1995 (4).

For foods described as containing meat and/or poultry, see section 3.7.14.5.

3.7.11.6 Pork. EPA commodities for pork are:

Pork, fat
Pork, kidney
Pork, liver
Pork, meat
Pork, meat byproducts
Pork, skin

Pork products, other than pork kidney, liver, byproducts, and skin, were separated into two components. The total fat component, as specified in the Survey Nutrient Database (7b), was assigned to the commodity "Pork, fat." After subtracting total fat, the remainder of the pork was assigned to the commodity "Pork, meat." Pork includes both fresh and cured forms. Pork kidney, liver, skin, and byproducts were assigned to their respective commodities. See section 3.7.11.4 regarding Kidney. Pork byproducts are defined in the EPA commodity list in \epa_comm.

3.7.11.7 Rabbit. The commodity "Rabbit, meat" includes both fat and nonfat components. It also includes both wild and domestic forms of rabbit.

3.7.11.8 Sheep. EPA commodities for sheep are:

Sheep, fat
Sheep, kidney
Sheep, liver
Sheep, meat

Sheep, meat byproducts

In CSFII 1994-96, 1998, food descriptions for sheep products were identified as "lamb (including mutton)." These products, other than kidney and byproducts, were separated into two components. The total fat component, as specified in the Survey Nutrient Database (7b), was assigned to the commodity "Sheep, fat." After subtracting total fat, the remainder was assigned to the commodity "Sheep, meat." See section 3.7.11.4 regarding Kidney. Sheep byproducts are defined in the EPA commodity list, located in \epa_comm.

3.7.12 Meat and Poultry, Processed

CSFII foods and ingredients for processed meats were assigned to commodities in proportions that reflected the amount of meat, fat, organ meats, and other ingredients allowed by the USDA meat and poultry standards (11, 12, 12a, 12b).

3.7.13 Milk

EPA commodities for milk are:

Milk, fat
Milk, nonfat solids
Milk, water
Milk, lactose

Foods assigned to milk commodities included dry, fluid, and concentrated forms of milk, cheese, yogurt, cream, and butter. Goat milk was included, but not soy milk. The milk items were divided into fat, nonfat solids, and water based on the 1994-96 Survey Nutrient Database (7b) and were assigned to the appropriate commodities. The nutrient database entries for total fat and moisture were used to determine proportions of the commodities "Milk, fat" and "Milk, water." "Milk, nonfat solids" was the remainder after subtracting total fat and moisture. For example, commodity assignments for 100 grams of whole milk were 3.34 grams to "Milk, fat," 87.99 grams to "Milk, water," and 8.67 grams (100-(3.34 + 87.99)) to "Milk, nonfat solids."

Lactose identified as a separate ingredient in a commercial baby food was assigned to the commodity "Milk, lactose." Lactose as an ingredient most often appeared in infant formulas. The lactose from non-baby-foods and lactose inherent in milk and milk products was included in the commodity "Milk, nonfat solids."

3.7.14 Poultry

3.7.14.1 Chicken. EPA commodities for chicken include:

Chicken, fat

Chicken, meat
Chicken, skin
Chicken, liver
Chicken, byproducts

The total fat components of both chicken flesh and chicken skin were assigned to the commodity "Chicken, fat." The nonfat component of chicken flesh was assigned to "Chicken, meat." The nonfat component of chicken skin was assigned to the commodity "Chicken, skin." Chicken neck was assigned to "Chicken, byproducts."

The ratio of flesh to skin for different poultry pieces was obtained from Agricultural Handbook No. 8 (6). The amount of total fat present in different poultry pieces was based on the 1994-96 Survey Nutrient Database (7b).

3.7.14.2 Turkey. EPA commodities for turkey are the same as for chicken (fat, meat, skin, liver, and byproducts). Assignments of fat and nonfat components of turkey flesh and skin were made the same as for chicken. The commodity, "Turkey, byproducts," included turkey neck. For turkey items described as dark meat, a portion was assigned to turkey byproducts to represent turkey neck.

3.7.14.3 Poultry other than chicken and turkey. EPA commodities for "Poultry, other" are the same as for chicken (fat, meat, skin, liver, and byproducts). Assignments of fat and nonfat components of the poultry pieces were similar to assignments for chicken. Dove, duck, partridge, pheasant, pigeon, quail, and squab were assigned to the "Poultry, other" commodities. Emu, goose, guinea hen, and ostrich were also listed by EPA as "Poultry, other," but those foods were not reported in the surveys.

3.7.14.4 Mixtures described as containing chicken or turkey. In the CSFII 1994-96, 1998 several food mixtures were described as containing either chicken or turkey, for example, "Chicken or turkey teriyaki." The Survey Recipe Database entries usually contained chicken, but not turkey, as an ingredient for those foods. Therefore, for most of these foods the chicken ingredient was assigned to commodities for both chicken and turkey in the proportions of 77 percent chicken and 23 percent turkey. This distribution was based on chicken and turkey food supply data for the years 1990-1994 (3).

3.7.14.5 Mixtures described as containing meat and/or poultry. In the CSFII 1994-96, 1998 several food mixtures were described as containing either meat or poultry, for example, "Tamale with meat and/or poultry." Assignment of meat and/or poultry ingredients in most of these foods were handled in two steps. First, they were divided into 50 percent meat and 50 percent

poultry. Then, meat was divided into 56 percent beef, 43 percent pork, and 1 percent lamb. (See section 3.7.11.8, Meat, not further specified.) Poultry was divided into 77 percent chicken and 23 percent turkey (3).

3.7.15 Spices and herbs

The EPA Food Commodity Vocabulary lists some spices and herbs as separate commodities. Others are grouped together in "Spices, other" and "Herbs, other." Complete listings of spices and herbs included in these commodities can be found in the Food Commodity Vocabulary's Appendices B and C in \epa_comm\foodvoc.wpd. Spices and herbs usually were not included in CSFII recipes, but were added before commodity assignments were made.

3.7.16 Sweeteners, carbohydrate

EPA commodities for carbohydrate sweeteners are:

Beet, sugar, molasses
Beet, sugar
Corn, field, syrup
Honey
Maple sugar
Maple syrup
Sorghum syrup
Sugarcane, molasses
Sugarcane, sugar

Sugar and molasses were not identified by source (sugarcane vs. sugar beet) in the CSFII 1994-96, 1998. Therefore, sugar and molasses were assigned to the commodities "Sugarcane, sugar" and "Beet, sugar," according to the proportions of 56 percent sugarcane and 44 percent sugar beet. These proportions reflect an estimate of refined sugar use in the continental United States from 1990-94 (15). Sugar included table sugar, confectioner's sugar, light brown sugar, dark brown sugar, raw sugar, and caramelized sugar.

Dry corn syrup solids, dextrose, and fructose were assigned to the commodity of "Corn, field, syrup," adjusting from the dry forms to liquid based on differences in total solids. Sugarcane syrup was assigned to the commodity "Sugarcane, sugar," adjusting from the liquid to dry form.

Corn syrup, maple sugar, maple syrup, sorghum syrup, and honey were assigned directly to their respective commodities. No adjustments were necessary.

3.7.17 Vegetables

Dried vegetables and vegetable juices were assigned to commodities for their respective dried or juice forms when such commodities were listed. If dried vegetables or vegetable juices were not present in the commodities list, they were assigned to their respective fresh vegetable forms. No adjustments were made to their weights.

3.7.18 Water

The EPA Food Commodity Vocabulary includes the following entries for water:

Water, dilution, source not specified
Water, tapwater- direct (drinking)
Water- indirect (cooking)
Water, bottled water
Water, commercial beverage

Water in foods and beverages is being assigned to the above categories by the EPA Office of Water. The information is not a part of this database at this time.

3.7.19 Commodities not consumed

Several commodities on EPA's list were not consumed at all during the CSFII 1994-96, 1998 and do not appear in the intake database. Table 2 lists these commodities. Several commodities were included with food codes for similar foods. For example, the description for food code 53304000 is "Pie, blueberry, two crust (include huckleberry)." In this latter situation, EPA requested that a small portion of the food code be assigned to the "included" commodity, if the commodity was not covered in any other food-code-to-commodity translation. Since huckleberry had received no other assignment, a small portion of food code 53304000 was assigned to the commodity "Huckleberry." In accordance with EPA's request, assignment was 0.001 grams of "Huckleberry" per 100 grams of pie (the smallest amount the Food Code-to-Commodity Translation File could accept).

3.7.20 Special requirement to include commodities consumed in very small amounts

EPA requested that data processing procedures be established to ensure that commodities consumed in very small amounts be represented in the resulting database files. While overstating precision is a concern, the possibility of small amounts rounding to zero was an even greater concern to EPA. Therefore, data in the Food-Code-to-Commodity Translation File are presented to 3 decimal places, allowing an estimate as small as 0.001 to be

included. Data in the commodity intake file are presented to 6 decimal places, allowing intake estimates as small as 0.000001 gram per kilogram of body weight to be included. These expressions are not intended to suggest that data can be estimated to these small amounts with such precision.

3.8 Notes about reviewing the Food-Code-to-Commodity Translation File

When reviewing the Food-Code-to-Commodity Translation File, some unexpected findings may be observed. Reasons for these findings can be found in the above discussions on matching USDA food codes to EPA commodities (section 3.6.1), converting weights of foods and ingredients (section 3.6.2), and special requirements for assigning foods and ingredients to EPA commodities (section 3.7).

Examples of unexpected findings are listed below:

1. The following five commodities are present in small amounts in many foods: "Cassava," "Corn, field, starch," "Potato, flour," "Rice, flour," and "Wheat, flour." This occurs because modified food starch is frequently listed as an ingredient in processed foods. When the source of the starch was not identified, it was assigned to these commodities.

2. "Barley, flour," is found in small amounts in many foods. This occurs when malt was listed as an ingredient.

3. "Beef, byproducts" and Pork, byproducts" are found in non-meat foods. Presence of these commodities in non-meat foods usually results from the presence of gelatin as an ingredient in the food. For example, gelatin is an ingredient in marshmallows. Therefore, when marshmallows are used in other foods, meat byproducts occur in the commodity assignments for those foods.

4. "Beef, byproducts" are found in ground beef and in foods containing ground beef as an ingredient. Because the Standards of Identity for ground beef allow for inclusion of beef cheeks, a small proportion of ground beef was assigned to beef byproducts.

5. Small amounts of "Soybean, oil" may be found in foods not expected to have a fat ingredient. This results from the presence of soya lecithin as an ingredient.

6. Many foods were assigned to nine different oil commodities (per EPA requirement), including small amounts of "Olive, oil" and "Sesame, oil." This occurs because a composite of oils was used when the specific type of oil was unknown, such as in home prepared foods (see section 3.7.6).

7. Fruit juices are present in concentrates at greater than 100 percent. This occurs because, upon assignment to a commodity,

the juice concentrate was adjusted to the corresponding weight of the reconstituted product. Thus, 100 grams of orange juice concentrate translates to 354.2 grams of the commodity "Orange, juice."

3.9 Limitations of the Food Commodity Intake Database

Users of the Food Commodity Intake Database should understand the nature of food survey data, as well as the assumptions required for translating survey food codes into EPA-defined commodities. This information should be considered in the interpretation of analyses based on the Food Commodity Intake Database.

--The food intake surveys were based on ability of sample persons, with assistance of probing by trained interviewers, to recall the foods they ate during the 24-hour period prior to the day of their interview.

--Amounts of foods consumed were estimated, involving two steps. First, the sample persons provided a description of the volumes, using standard survey measurement aids or label information. Then, the volumes were converted to grams using a database of weight/volume relationships.

--Ingredient amounts for commercial products were estimated. Ingredients and amounts for other food mixtures were based on representative recipes, usually from popular cookbooks but from regional or specialty cookbooks when necessary. The recipes were not specific for each sample person.

--Sometimes ingredients were proportioned among commodities based on food supply information, for example, oils and sweeteners (see section 3.7 above).

--Body weights were self reported. Missing body weights were assigned default values (see section 4.5.2).

3.10 References

(1) Marcoe, K.K., and D.B. Haytowitz. 1993. Estimating nutrient values of mixed dishes from label information. Food Technology 47(4):69-75.

(2) Matthews, R.H., and Y.J. Garrison. 1995. Food Yields Summarized by Different Stages of Preparation. U.S. Department of Agriculture, Agricultural Research Service, Agriculture Handbook No. 102.

(3) Putnam, J.J. and J.E. Allshouse. 1996. Food Consumption, Prices, and Expenditures, 1996: Annual Data 1970-94. U.S. Department of Agriculture, Economic Research Service, Statistical Bulletin No. 928.

(4) Putnam, J.J. and J.E. Allshouse. 1997. Food Consumption, Prices, and Expenditures, 1970-95. U.S. Department of Agriculture, Economic Research Service, Statistical Bulletin No. 939.

(5) Tippett, K.S., and Y.S. Cypel (eds.). 1998. 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.

(6) U.S. Department of Agriculture, Agricultural Research Service. 1976-92. Composition of Foods...Raw, Processed, Prepared. Agriculture Handbook No. 8, Revised Sections 1-22.

(7) U.S. Department of Agriculture, Agricultural Research Service. 1998. 1994-96 Continuing Survey of Food Intakes by Individuals and 1994-96 Diet and Health Knowledge Survey. National Technical Information Service, 5285 Port Royal Road, Springfield VA 22161;(703)487-4650. CD-ROM: Accession no. PB98-500457.

(7a) ----. Food Coding Database for the 1994-1996 Continuing Survey of Food Intakes by Individuals. On CD-ROM: 1994-96 Continuing Survey of Food Intakes by Individuals and 1994-96 Diet and Health Knowledge Survey. NTIS Accession no. PB98-500457.

(7b) ----. Survey Nutrient Database for the 1994-1996 Continuing Survey of Food Intakes by Individuals. On CD-ROM: 1994-96 Continuing Survey of Food Intakes by Individuals and 1994-96 Diet and Health Knowledge Survey. NTIS Accession no. PB98-500457.

(7c)----. Survey Recipe Database for the 1994-1996 Continuing Survey of Food Intakes by Individuals. On CD-ROM: 1994-96 Continuing Survey of Food Intakes by Individuals and 1994-96 Diet and Health Knowledge Survey. NTIS Accession no. PB98-500457.

(8) U.S. Department of Agriculture, Agricultural Research Service. 2000. Continuing Survey of Food Intakes by Individuals 1994-96, 1998. National Technical Information Service, 5285 Port Royal Road, Springfield VA 22161;(703)487-4650. CD-ROM. NTIS Accession no. PB2000-500027.

(9) U.S. Department of Agriculture, Agricultural Research Service, Nutrient Data Laboratory. 1998. Unpublished Data. Personal communication between Nutrient Data Laboratory staff and Food Surveys Research Group staff.

(10) U.S. Department of Agriculture, Economic Research Service. July 1995. Oil Crops Yearbook.

(11) U.S. Department of Agriculture, Food Safety and Inspection Service, Inspection Operations, Regulatory Programs. 1996. Food Standards and Labeling Policy Book.

(12) U.S. Department of Agriculture, Food Safety and Inspection Service. 1998. Animal and animal products: Approval of substances for use in the preparation of products. Code of Federal Regulations, Title 9, Pt. 318.7.

(12a)----. Animals and animal products: Definitions and standards of identity or composition. Code of Federal Regulations, Title 9, Pts. 319.1-319.881.

(12b)----. Animals and animal products: Poultry products inspection regulations, Subpart P, Definitions and standards of identity or composition. Code of Federal Regulations, Title 9, Pts. 381.155-381.174.

(13) U.S. Department of Agriculture, Foreign Agricultural Service. 1995. Tropical Products: World Markets and Trade, Circular Series FTROP 3-95, page 21, Table: "Tea: Sales by U.S. Retail Stores, 1995 with Comparison".

(14) U.S. Department of Agriculture, Foreign Agricultural Service, Oilseeds and Products Division, Import and Export Data. 1998. Personal communication between Foreign Agricultural Service and Center for Nutrition Policy and Promotion staff.

(15) U.S. Department of Agriculture, National Agricultural Statistics Service. 1997. Agricultural Statistics 1997, page II-20, Table 2-33.

(16) U.S. Environmental Protection Agency. 2000. The EPA Food Commodity Vocabulary, Master List dated June 15, 2000.

Table 1. CSFII food codes representing commercial baby foods

Food code or food code prefix	Foods
117----- *	Infant formulas
133-----	Milk desserts
14220000	Cottage cheese, with fruit
20000070	Meat, not specified as to type
20000090	Meat sticks, not specified as to type of meat
217-----	Beef
228-----	Pork
234-----	Lamb or veal
247-----	Poultry
2518----	Variety meats
276-----	Meal, poultry, or fish mixtures with nonmeat items
34------	Eggs
53203050	Cookie, fruit
53203100	Cookie
53242250	Cookie, teething
54350000	Crackers
54408100	Pretzel
578-----	Cereals
585-----	Grain mixtures
67------	Fruits, fruit juices, and fruit mixtures
76------	Vegetables and mixtures mostly vegetable

* All food codes with “117" in the first three digits.

Table 2. EPA food commodities not appearing in the intake database

Acerola
Bean, broad, succulent
Broccoli raab
Buckwheat, flour
Burdock
Canistel
Cardoon
Celeriac
Celtuce
Cherimoya
Chickpea, flour
Citrus, oil
Crabapple
Cress, garden
Cress, upland
Elderberry
Feijoa
Fennel, Florence
Filbert, Oil
Goat, meat byproducts
Goat, kidney
Goat, liver
Hickory nut
Horse, meat
Jaboticaba
Jackfruit
Kohlrabi
Kumquat
Lemongrass
Spanish Lime
Loganberry
Longan
Loquat
Lychee, dried
Mamey apple
Parsley, turnip rooted
Pawpaw
Peppermint, oil
Pummelo
Quince
Quinoa
Radish, Oriental, tops
Radish, tops
Rape greens
Rye, flour *
Salsify, roots
Salsify, tops
Sapote
Sorghum
Soursop
Spearmint
Spearmint, oil
Sugar apple
Tree Tomato
Water, dilution, source NS**
Water, tapwater-direct (drinking) **
Water-Indirect (cooking) **
Water, bottled water **
Water, commercial beverage**
Almond, oil-babyfood
Almond-babyfood
Arrowroot, flour-babyfood
Basil, fresh leaves-babyfood
Coriander, leaves-babyfood
Coriander, seed-babyfood
Corn, field, meal-babyfood
Egg, white (solids)-babyfood
Ginger-babyfood
Guar, seed-babyfood
Guava-babyfood
Honey-babyfood
Papaya-babyfood
Passionfruit, juice-babyfood
Passionfruit-babyfood
Peach, dried-babyfood
Pepper, bell-babyfood
Plum, prune, dried-babyfood
Potato, tuber, w/peel-babyfood
Sesame, oil-babyfood
Sesame, seed-babyfood
Soybean, soy milk-babyfood
Triticale, flour-babyfood
Turkey, liver-babyfood

* Rye flour translated to EPA Commodity “Rye, Whole grain”

**Water translated by EPA Office of Water

4 CHARACTERISTICS AND FORMATS OF THE FOOD COMMODITY INTAKE DATABASE

4.1 Introduction

The Food Commodity Intake Database contains several different data files related to intake of over 500 EPA-defined commodities. It is based on food consumption surveys conducted by the U.S. Department of Agriculture's Agricultural Research Service. The surveys are the 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII 1994-96) and its Supplemental Children's Survey (CSFII 1998). The database was developed for use in estimating human exposure to pesticide residues through food and beverage intake.

The general file structure and information concerning the use of the data are discussed in Sections 4.2 and 4.3. Section 4.4 contains the data file formats, which provide detailed descriptions of all of the fields included in each file. The field number, name, width, and type are given, along with a full description of the field and its applications, valid values for the field, and, if applicable, a reference to the question number on the original CSFII questionnaire. Section 4.5 contains miscellaneous notes about the data files.

4.2 Database Structure

The Commodity Intake Database is made up of 18 data files. These data files are classified into four categories as follows: (1) 8 Documentation Files, (2) 2 EPA Commodity Files, (3) 3 Commodity Intake Data Files, and (4) 5 Supporting Data Files. Records in the different data files can be linked or connected through the key fields that are described Section 4.3.2. (See section 1 "\document\section1.wpd" for a tree diagram and summary description of data files on this database.)

4.2.1 Documentation Files (files in directory "\document")

Each of the 5 documentation sections and the 3 appendices listed in the Table of Contents is contained in a separate WordPerfect file, as indicated below.

Section1.wpd	Table of contents, File names
Section2.wpd	Essential information
Section3.wpd	Methodology for translating food intakes into EPA commodity intakes
Section4.wpd	Data file characteristics and formats
Section5.wpd	Control counts
Appenda.wpd	Appendix A. EPA's use of food consumption

	data in assessing dietary risk from pesticides
Appendb.wpd	Appendix B. CSFII 1994-96, 1998 Methodology (Section 3 from CSFII 1994-96, 1998 CD-ROM documentation)
Appendc.wpd	Appendix C. CSFII 1994-96, 1998 Sampling Weights and Statistical Notes (excerpts from sections 5 and 6 of the CSFII 1994-96, 1998 CD-ROM documentation)

4.2.2 EPA Commodity Files (files in directory "\epa_comm")

The Environmental Protection Agency (EPA) provided 2 files. Both are included as WordPerfect documents:

foodvoc.wpd	EPA commodity list
csffcm.wpd	Procedures for assigning EPA cooked status, food form, and cooking method values to USDA food codes

4.2.3 Commodity intake data files (files in directory "\intake")

The commodity intake data files include 2 files that were created by translating the CSFII 1994-96, 1998 food intakes into EPA commodities, and one file of sample person data. These files, which are in ASCII delimited format, include:

comm9498.txt	Commodity intakes for CSFII 1994-96, 1998
ffcm9498.txt	Commodity intakes by cooked status, food form, and cooking method for CSFII 1994-96, 1998
smpl9498.txt	Sample person data for CSFII 1994-96, 1998

4.2.4 Supporting data files (files in directory "\support")

Five data files provide supporting documentation for the translation of the CSFII 1994-96, 1998 intake data into EPA commodities. These files, which are in ASCII delimited format, are:

fc_comm.txt	CSFII 1994-96, 1998 food codes linked to EPA commodities (grams of commodities per 100 grams of food)
fcdesc.txt	Food code descriptions
fcincl.txt	Food code include statements (extensions of food code descriptions)
fcscheme.txt	Food code outline
moddesc.txt	Descriptions of recipe modifications

4.3 General nature of the data

The data files that make up the Commodities Intake Database are in the following formats:

Directory	Format
\Document	WordPerfect
\EPA_Comm	WordPerfect
\Intake	ASCII delimited
\Support	ASCII delimited

ASCII delimited fields are separated by the caret (^) symbol and alpha-numeric/character fields are enclosed in tilde (~) marks. The majority of the data fields are numeric, but there are some alpha-numeric text fields. The numeric fields include explicit decimal points, and do not include leading zeros.

Some fields have "missing" values due to a nonspecific or absent response, or due to the lack of the necessary data for calculations. These fields are not left blank, but are filled with codes to indicate the separate categories of "refused," "don't know," and "not ascertained," or sometimes "indeterminable." The usual convention for a one column field is a '7' for "refused," '8' for "don't know," and '9' for "not ascertained." Two-column fields have values of '97,' '98,' and '99' to represent these categories of “missing” values. Continuous fields may also have codes for "missing" values.

When a blank or null value is found in a field, it means that a response or calculation for that field does not apply to a particular situation. For example, the two-day sampling weight fields will be blank for sample persons who have not provided two days of intake. In general, if a skip pattern dictates that a question should not be asked of a respondent, the corresponding field on the record will be blank. Blank or Null fields appear in the data files as two delimiters side by side (e.g., ^^).

4.3.1 Data File Characteristics

The ASCII delimited files have variable record lengths. The five ASCII delimited data files have the following record counts.

File	Total number of records

comm9498.txt	3,173,197
ffcm9498.txt	5,781,938
fcincl.txt	5,652
smpl9498.txt	21,662
fc_comm.txt	99,011
fcdesc.txt	5,845
moddesc.txt	3,902

4.3.2 List of Key Fields

Specific fields in each data file have been designated as "key" fields to (1) ensure that each record is unique, and (2) allow data in different files to be linked or joined. Following is a list of fields designated as key fields:

HHID	Household identification number
SPNUM	Sample person number
DAYCODE	Day of intake identifier
COM_CODE	EPA commodity code
CKD_STAT	EPA cooked status code
FD_FORM	EPA food form code
CKG_METH	EPA cooking method code
FOODCODE	USDA food code
MODCODE	Recipe modification code

4.3.3 Sampling weights

Note: For an in-depth discussion of weighting for the CSFII 1994-96, 1998, see Appendix C in \document\appendc.wpd.

The "smpl9498.txt" data file contains several different sampling weights. By using appropriate weights: (a) data for each survey year (1994, 1995, 1996, or 1998) can be used separately; (b) data for the 3 years of CSFII 1994-96 can be combined; or (c) data for all 4 years of the surveys can be combined. In addition, weights are present for using only one day of intake for each sample person; other weights provide for the use of only sample persons who provided 2 days of intake.

Use the annual sampling weight fields, in conjunction with the year field, when data are to be analyzed for one survey year (1994, 1995, 1996, or 1998). Annual sampling weight fields are:

WTA_DAY1 The annual day 1 sampling weight for all responding CSFII 1994-96, 1998 sample persons. This annual sampling weight is used whenever the sample of interest includes sample persons who provided the first day of intake data regardless of whether they provided the second day.

WTA_2DAY The annual 2-day sampling weight for all CSFII 1994-96, 1998 sample persons with two days of intake. This annual sampling weight is used whenever the sample of interest includes only sample persons who provided 2 days of intake data.

The 3-year sampling weight fields for CSFII 1994-96 are:

WT3_DAY1 The 3-year day 1 sampling weight for all responding CSFII 1994-96 sample persons. This 3-year sampling weight is used whenever the sample of interest includes sample persons who provided the first day of intake data regardless of whether they provided the second day.

WT3_2DAY The 3-year 2-day sampling weight for all CSFII 1994-96 sample persons with two days of intake. This 3-year sampling weight is used whenever the sample of interest includes only sample persons who provided 2 days of intake data.

The sample weight field for combining all 4 years of the CSFII 1994-96, 1998 are:

WT4_DAY1 The 4-year day 1 sampling weight for all responding CSFII 1994-96, 1998 sample persons. This 4-year sampling weight is used whenever the sample of interest includes sample persons from both surveys who provided the first day of intake data regardless of whether they provided the second day.

WT4_2DAY The 4-year 2-day sampling weight for all CSFII 1994-96, 1998 sample persons with two days of intake. This 4-year sampling weight is used whenever the sample of interest includes only sample persons from both surveys who provided 2 days of intake data.

4.4 Data File Formats

4.4.1 Introduction to the File Formats for ASCII delimited files

Seven data files on the Food Commodity Intake Database are stored in ASCII delimited formats (files in directories "\intake" and "\support"). Fields are separated by the caret (^) symbol and alpha-numeric/character fields are enclosed in tilde (~) marks. Blank fields appear as two delimiters side by side (i.e., ^^).

This section on file formats describes the contents of each of the 7 data files. The fields in each file are listed and for each field, the following information is provided: field number, field name, field width, field type, and description.

- The field number identifies the position or order in which the field can be found on the data record.
- The field name will be no longer than eight characters and will always be referred to in the file formats in uppercase letters.
- The width is the maximum number of columns allocated to the field including, where appropriate, an explicit decimal point. All alpha-numeric fields are encased in tildes (~), which are not counted in the field width.
- The type of the field is either 'N' for numeric or 'A' for alpha-numeric or character. If a numeric field has a fractional part the number of decimal places follows the N. For example, 'N2' indicates a field with two decimal places.
- The description includes allowed values and their meanings and skip patterns dictated by specific field values.

4.4.2 Formats for the food commodity intake data files

4.4.2.1 Format of the Commodity Intakes file (comm9498.txt)

Field Number	Field Name	Width /Type		Description/Application/Values
1	HHID	5	N	Household identification number. 10001 - 52999 = HHID
2	SPNUM	2	N	Sample person (SP) number. 1 - 23 = SP number

3	DAYCODE	1	N	Day 1 / Day 2 / Average indicator. Note: There is one record per EPA commodity per SP per day of intake. Where two days were reported there is also a record containing daily averages. 1 = Day 1 2 = Day 2 4 = Average of day 1 and day 2
4	COM_CODE	8	N	EPA commodity code. Note: See "foodvoc.wpd" for codes and complete descriptions of each COM_CODE.
5	COM_AMT	10	N6	Total (or average) amount of the EPA commodity consumed during that day. Amount in grams per kilogram body weight

4.4.2.2 Format for the files of commodity intakes by cooked status, food form and cooking method (ffcm9498.txt)

Field Number	Field Name	Width /Type		Description/Application/Values
1	HHID	5	N	Household identification number. 10001 - 52999 = HHID
2	SPNUM	2	N	Sample person (SP) number. 1 - 23 = SP number
3	DAYCODE	1	N	Day 1 / Day 2 / Average indicator. Note: There is one record for each EPA commodity and Cooked Status-Food Form-Cooking Method combination(COM-FFCM), per SP, per day of intake. Where two days were reported there is also a record containing daily

averages.

1 = Day 1
2 = Day 2
4 = Average of day 1 and day 2

4 COM_CODE 8 N EPA commodity code.

Note: See "foodvoc.wpd" for possible codes and complete descriptions of each COM_CODE.

5 CKD_STAT 1 N EPA cooked status identifier.

0 = Not applicable
1 = Uncooked
2 = Cooked
3 = Processed oil
6 = Frozen meal
7 = Salad
8 = Sandwich
9 = Not specified as to cooked or uncooked

6 FD_FORM 1 N EPA food form identifier.

0 = not applicable
1 = Fresh
2 = Frozen
3 = Dried
4 = Canned
5 = Cured, pickled, smoked, salted
8 = Other process forms not listed above
9 = NS as to form or multiple forms

7 CKG_METH 1 N EPA cooking method identifier.

0 = None or Not applicable
1 = Baked
2 = Boiled
3 = Fried
4 = Fried or baked
5 = Boiled or baked
8 = Not specified as to cooking method or multiple cooking methods
9 = Not specified as to whether

				further cooking occurred
8	CFFCMAMT	10	N6	Total amount consumed of the commodity and cooked status-food form-cooking method combination. Amount in grams per kilogram body weight

4.4.2.3 File Format for Sample Person Data (smpl9498.txt)

Where appropriate, the description includes references to the original question number from the CSFII questionnaires. For example "H52" indicates Household question number 52. Questionnaire symbols are:

S = Screener questionnaire
H = Household questionnaire
DA = Individual Intake Questionnaire (day 1)
DB = Individual Intake Questionnaire (day 2)

Field Number	Field Name	Width /Type		Description/Application/Values
1	HHID	5	N	Household identification number. 10001 - 52999 = HHID
2	SPNUM	2	N	Sample person (SP) number. 1 - 23 = SP number
3	REGION	1	N	Region of the United States. 1 = Northeast 2 = Midwest 3 = South 4 = West
4	URB	1	N	Urbanization; Metropolitan Statistical Area(MSA) status. 1 = MSA, central city 2 = MSA, outside central city 3 = Non-MSA
5	HHSIZE	2	N	H Household size; count of household members. 1 - 23 = Count

6 INCOME 6 N H52. During the previous calendar year, approximately how much income from all sources did you and other household members have before taxes? (Please give me your best estimate.)

Note: annual incomes have been imputed for households that could not or would not provide a response to this question. See documentation for the CSFII 1994-96, 1998 for an explanation of the methods employed. See INCREP for the original response to H52. See IMPFLAG for the method of imputation employed.

0 - 99999 = Dollars
100000 = $100,000 or more

7 INCREP 1 N H52. Type of original response to H52.

* 1 = Value of INCOME is the actual amount reported.
* 5 = No household interview
* 6 = Not a household in the previous calendar year
7 = Refused
8 = Don't know
9 = Not ascertained

* Skip INCCODE.

8 INCCODE 1 A H53. Please tell me which letter on this card best represents your combined household income before taxes for the previous calendar year.

Note: H53 is only asked of households that could not or would not answer H52.

Applies if: INCREP >= 7

A = Under $5000
B = $5,000 - $9,999
C = $10,000 - $14,999
D = $15,000 - $19,999
E = $20,000 - $24,999
F = $25,000 - $29,999
G = $30,000 - $34,999
H = $35,000 - $39,999

I = $40,000 - $44,999
J = $45,000 - $49,999
K = $50,000 - $59,999
L = $60,000 - $74,999
M = $75,000 - $99,999
N = $100,000 and over
7 = Refused
8 = Don't know
9 = Not ascertained
Blank = Not applicable

9 IMPFLAG 1 N Annual income imputation flag.

1 = Not imputed, value of INCOME is the actual amount reported.
2 = Imputed, value based on H53 (INCCODE)
3 = Imputed, value based on monthly income
4 = Imputed, value based on regression equation
5 = Imputed, based on segment level mean income

10 PCTPOV 3 N Annual income expressed as a percentage of the poverty threshold. Based on INCOME (using imputed values) and HHSIZE.

0 - 299 = Percentage of the poverty threshold
300 = 300% or more

11 AGE 2 N Age of household member in years.

Note: Age at time of day 1 intake.

0 = Under 1 year old
* 1 - 89 = Age in years
* 90 = 90 or older

* Skip AGE_M.

12 AGE_M 2 N Age of household member in months. Valid only for children 11 months old or younger.

Note: Age at time of day 1 intake.

Applies if: AGE = 0

0 = Less than one month old

1 - 11 = Months of age
Blank = Not applicable

13 SEX 1 N Sex of household member.

1 = Male
2 = Female

14 WGT_SP 3 N DA30. How much do you weigh without shoes?

1 - 995 = Pounds
997 = Refused
998 = Don't know
999 = Not ascertained

15 WGT_KG 3 N1 Weight of sample person in kilograms.

1 - 452 = Kilograms

The amount of each commodity consumed is reported in grams per kilogram body weight. The CSFII 1994-96, 1998 reports each sample persons body weight (WGT_SP) in pounds, as reported by the sample person. These body weights were converted from pounds to kilograms using the following formula:

$$\text{Weight in Kilograms} = \frac{\text{Weight in pounds}}{2.2046}$$

Note: default body weights were used for sample persons whose WGT_SP = '997', '998', or '999'. See section 4.5.2 about Unreported Body Weights.

16 REL_REF 2 N S8. What is your relationship to the reference person?

0 = Reference person
1 = Spouse
2 = Natural or adopted child; step child
3 = Grandchild
4 = Parent
5 = Brother or sister
6 = Other relative
7 = Foster child
8 = Partner; roommate; girlfriend; boyfriend
9 = Roomer or boarder
10 = Employee
11 = Guest

12 = Other unrelated

17	RACE	1	N	S9. Which of the groups on this card best describes your race? 1 = White 2 = Black 3 = Asian, Pacific Islander 4 = American Indian, Alaskan native 5 = Other
18	ORIGIN	1	N	S10. Do any of these groups (from a card) represent your national origin? 1 = Mexican, Mexican American, Chicano 2 = Puerto Rican 3 = Cuban 4 = Other Spanish / Hispanic 5 = None of the above
19	PL_STAT	1	N	Pregnant / lactating status. Note: From questions H26, H27, H29 and H31. Also, these questions were only asked of households with certain characteristics as identified at screening. 1 = Pregnant 2 = Lactating 3 = Pregnant and lactating 4 = Not pregnant or lactating 5 = Not female 10-55
20	BF_STAT	1	N	Breast-feeding status. Note: From questions H29 and H30. Also, these questions were only asked of households with children 3 years old or less identified at screening. 1 = Breast-feeding 2 = Not breast-feeding 3 = Over 3 years old
21	H2O_COOK	2	N	H18. What is the main source of the water used for cooking in your home? 1 = Community water supply 2 = Well or rain cistern

(household's)
3 = Spring (household's or public)
4 = Bottled water (purchased)
96 = Other
98 = Don't know
99 = Not ascertained

22 H2O_BEVR 2 N H19. What is the main source of the water used in your home for preparing beverages such as coffee, tea, juices, and baby formula?

1 = Community water supply
2 = Well or rain cistern
(household's)
3 = Spring (household's or public)
4 = Bottled water (purchased)
96 = Other
98 = Don't know
99 = Not ascertained

23 H2O_DRNK 2 N H20. What is the main source of plain drinking water in your home?

1 = Community water supply
2 = Well or rain cistern
(household's)
3 = Spring (household's or public)
4 = Bottled water (purchased)
96 = Other
98 = Don't know
99 = Not ascertained

24 D1_H2O_O 3 N DA15. How many fluid ounces of plain drinking water, that is, tap water or any bottled water that is not carbonated, with nothing added to it, did you drink yesterday - day 1?

* 0 = None
1 - 995 = Fluid ounces
998 = Don't know
999 = Not ascertained

* Skip D1_H2O_H - D1_H2O_A

25 D1_H2O_H 1 N DA16. How much of this plain drinking water came from your home? Would you say all, most some, or none - day 1?

Applies if: D1_H2O_O > 0

```
                              * 1 = All
                                2 = Most
                                3 = Some
                                4 = None
                                8 = Don't know
                                9 = Not ascertained

                              * Skip D1_H2O_A
```

26 D1_H2O_A 1 N DA17. What was the main source of plain drinking water that did not come from your home? Was it tap water, water from a drinking fountain, bottled water, or something else - day 1?

```
                                    Applies if: D1_H2O_H > 1

                                     1 = Tap water / drinking
                                    fountain
                                     2 = Bottled water
                                     6 = Other
                                     8 = Don't know
                                     9 = Not ascertained
                                 Blank = Not applicable
```

27 D2_H2O_O 3 N DB13. How many fluid ounces of plain drinking water, that is, tap water or any bottled water that is not carbonated, with nothing added to it, did you drink yesterday - day 2?

```
                                    Applies if: COMP_D2 = 1

                                    *       0 = None
                                      1 - 995 = Fluid ounces
                                          998 = Don't know
                                          999 = Not ascertained
                                        Blank = Not applicable

                                    * Skip D2_H2O_H - D2_H2O_A
```

28 D2_H2O_H 1 N DB14. How much of this plain drinking water came from your home? Would you say all, most some, or none - day 2?

```
                                    Applies if: D2_H2O_O > 0

                                    *     1 = All
                                          2 = Most
                                          3 = Some
                                          4 = None
```

8 = Don't know
9 = Not ascertained
Blank = Not applicable

* Skip D2_H2O_A

29 D2_H2O_A 1 N DB15. What was the main source of plain drinking water that did not come from your home? Was it tap water, water from a drinking fountain, bottled water, or something else - day 2?

Applies if: D2_H2O_H > 1

1 = Tap water/drinking fountain
2 = Bottled water
6 = Other
8 = Don't know
9 = Not ascertained
Blank = Not applicable

30 COMP_D1 1 N Is there complete Day 1 intake data for this individual?

1 = Yes

31 COMP_D2 1 N Is there complete Day 2 intake data for this individual?

1 = Yes
* 2 = No

* Skip WT3_2DAY, WTA_2DAY.

32 WT3_DAY1 8 N 3-year day 1 sample weight for CSFII 1994-96. Read section 4.3.3 on sampling weights.

Applies to day 1 records with 1994, 1995, or 1996 in YEAR field.

1 - 99999999 = Weight
Blank = Not applicable

33 WT3_2DAY 8 N 3-year 2-day sample weight for CSFII 1994-96. Read section 4.3.3 on sampling weights.

Applies to day 1 and day 2 records with 1994, 1995, or 1996 in YEAR

field and 1 COMP_D2 field.

1 - 99999999 = Weight
Blank = Not applicable

34 WTA_DAY1 8 N Annual day 1 sample weight. Read section 4.3.3 on sampling weights.

Use in conjunction with YEAR field

Applies if: COMP_D1 = 1

1 - 99999999 = Weight
Blank = Not applicable

35 WTA_2DAY 8 N Annual 2-day sample weight. Read section 4.3.3 on sampling weights.

Use in conjunction with YEAR field

Applies if: COMP_D2 = 1

1 - 99999999 = Weight
Blank = Not applicable

36 WT4_DAY1 8 N 4-year day 1 sample weight. Read section 4.3.3 on sampling weights.

Applies if: COMP_D1 = 1

1 - 99999999 = Weight
Blank = Not applicable

37 WT4_2DAY 8 N 4-year 2-day sample weight. Read section 4.3.3 on sampling weights.

Applies if: COMP_D2 = 1

1 - 99999999 = Weight
Blank = Not applicable

38 YEAR 4 N Year of the survey.

1994 = 1994 sample
1995 = 1995 sample
1996 = 1996 sample
1998 = 1998 sample

39 D1_NREC 2 N Day 1: number of food records.

0 - 99 = Number

40	D2_NREC	2	N	Day 2: number of food records. Applies if: COMP_D2 = 1 0 - 99 = Number Blank = Not applicable

4.4.3 Formats for the supporting files

4.4.3.1 Food Code-to-commodity Translation File - (fc_comm.txt)

Field Number	Field Name	Width /Type		Description/Application/Values
1	FOODCODE	8	N	Food code.
2	MODCODE	6	N	Modification code.
3	COM_CODE	8	N	EPA commodity code.
4	COM_AMT	10	N3	Total amount in grams of the EPA commodity in 100 grams of the food.

4.4.3.2 Food code descriptions (fcdesc.txt)

Field Number	Field Name	Width /Type		Description/Application/Values
1	FOODCODE	8	N	Food code.
2	DESCR	200	A	Description of food.
3	ABBRDESC	60	A	Abbreviated description of food.

4.4.3.3 Food code include statements (fcincl.txt)

Field Number	Field Name	Width /Type		Description/Application/Values
1	FOODCODE	8	N	Food code.
2	SEQUENCE	2	N	Sequence number.
3	INCLUDE	80	A	Include statement

4.4.3.4 Food Code outline (fcscheme.txt)

ASCII text file

4.4.3.5 Descriptions of recipe modifications (moddesc.txt)

Field Number	Field Name	Width /Type	Description/Application/Values
1	MODCODE	6 N	Recipe modification code number.
2	MODDESC	240 A	Recipe modification description

4.5 Miscellaneous Notes

4.5.1 Responding sample persons with no foods reported for a day

There are sample persons who completed an individual intake interview but reported consuming no foods or beverages for that day. The "smpl9498.txt" fields providing the number of foods reported for a day, D1_NREC and D2_NREC, will have a value of 0 in such cases. Such sample persons do not have records in the commodity intake ("comm9498.txt") nor the cooked status-food form-cooking method commodity intake ("ffcm9498.txt") data files for that day.

4.5.2 Unreported body weights

The amount of each commodity consumed is reported in grams per kilogram body weight. The CSFII, 1994-96 reports each sample persons body weight (WGT_SP) in pounds, which was converted into kilograms. There are some sample persons who completed an individual intake interview but failed to provide a body weight. Where a body weight was not reported, the following default weights were used:

Gender and Age	Body weight (Kg)
Children < 6 months of age	6.0 Kg
Males 6 - 11 months of age	9.4 Kg
1 year	11.8
2 years	13.6
3 years	15.7
4 years	17.8
5 years	19.8
6 years	23.0
7 years	25.1
8 years	28.2
9 years	31.1
10 years	36.4
11 years	40.3
12 years	44.2

	13 years	49.9	
	14 years	57.1	
	15 years	61.0	
	16 years	67.1	
	17 years	66.7	
	> 17 years	70.0	
Females	6-11 months of age	8.8	Kg
	1 year	10.8	
	2 years	13.0	
	3 years	14.9	
	4 years	17.0	
	5 years	19.6	
	6 years	22.1	
	7 years	24.7	
	8 years	27.9	
	9 years	31.9	
	10 years	36.1	
	11 years	41.8	
	12 years	46.4	
	13 years	50.9	
	14 years	54.8	
	15 years	55.1	
	16 years	58.1	
	17 years	59.6	
	> 17 years	60.0	

Body weights used as default weights, in instances in which body weights were not reported, were provided by EPA. The reference weights for children less than 6 months of age are derived from Table 7-1 of the Exposure Factors Handbook, U.S. Environmental Protection Agency, National Center for Environmental Assessment, Office of Research and Development, Washington, DC, (EPA/600/C-99/001), February 1999. Table 7-1 is derived from: Hamill PTV, TA Dried, CL Johnson, R. Reed, AF Riche, WE Moore. Physical Growth: National Center for Health Statistics Percentiles. Am J Cain Nut 32:607-629, 1979. The weights of children 6 months through the 17th year are from Table 7-3, Body Weights of Children, adapted from the National Center for Health Statistics (NCHS), 1987: Anthropometric reference data and prevalence of overweight, United States, 1976-1980. Data from the National Health and Nutrition Examination Survey (NHANES II), Series 11, No. 238, Hyattsville, MD: U.S. Department of Health and Human Services, Public Health Service, NCHS. DHHS Publication No. (PHS) 87-1688. The weights provided for those 18 year of age and older are those traditionally used in risk assessments by EPA's Office of Pesticide Programs, Health Effects Division for adult males and females.

4.5.3 Lower limit for reporting intakes

The commodity intake data files ("comm9498.txt" and

"ffcm9498.txt") contain the amount of each EPA commodity consumed by each sample person providing at least one day of dietary intake in the CSFII 1994-96, 1998. Commodities not consumed by a sample person are not included in the database. The smallest amount that could be included in these intake files is 0.000001 grams per kilogram body weight. When an amount of a commodity consumed from a food was less than this amount, this lower limit amount of 0.000001 grams per kilogram body weight was assigned, as requested by EPA.

4.5.4 Using this database in conjunction with the Continuing Survey of Food Intakes by Individuals and the Supplemental Children's Survey (CSFII 1994-96, 1998)

The Food Commodity Intake Database was created by translating the CSFII 1994-96, 1998 food and beverage intakes into EPA commodities. The household identification number (HHID) and sample person number (SPNUM)uniquely identify each sample person and provide linkage with the survey data. These identifiers, along with the Day code (DAYCODE), provide linkage to the survey's daily intake records.

5. Control statistics for the SMPL9498.txt data file, all records, unweighted

This file contains descriptive statistics for variables in the SMPL9498.txt file. These statistics are not population estimates. They are provided for checking purposes. The SAS MEANS procedure was used to generate this listing which includes a count of records with non-missing values (N) for each variable, the unweighted mean of all values, and the minimum, maximum, and sum of each variable. They are unweighted and were computed using all values of each variable including values such as '998' indicating a "don't know" response.

Variable	Label	N	Mean
HHID	Household ID	21662	26262.0
SPNUM	Sample person number	21662	1.8
REGION	Region	21662	2.6
URB	Urbanization	21662	1.9
HHSIZE	Household size	21662	3.8
INCOME	Annual income: total	21662	38624.9
INCREP	Annual income: actual report	21662	2.5
IMPFLAG	Annual income: imputation flag	21662	1.4
PCTPOV	Annual income: percent of poverty	21662	205.0
AGE	Age in years	21662	25.4
AGE_M	Age in months	1551	5.5
SEX	Sex	21662	1.5
WGT_SP	Weight of SP	21662	138.9
WGT_KG	Weight of SP in Kilograms	21662	47.2
REL_REF	Relationship to reference person	21662	1.6
RACE	Race	21662	1.5

```
     ORIGIN    Hispanic origin                          21662
    4.6
     PL_STAT   Pregnant/lactating status                21662
    4.8
     BF_STAT   Breastfeeding status                     21662
    2.7
     H2O_COOK  Source of water: cooking                 21662
    2.4
     H2O_BEVR  Source of water: beverages               21662
    3.0
     H2O_DRNK  Source of water: drinking                21662
    3.3
     D1_H2O_O  Day 1: amount of water                   21662
   29.3
     D1_H2O_H  Day 1: water from home                   17359
    1.8
     D1_H2O_A  Day 1: away from home water               6513
    1.6
     D2_H2O_O  Day 2: amount of water                   20607
   29.5
     D2_H2O_H  Day 2: water from home                   16260
    1.9
     D2_H2O_A  Day 2: away from home water               5906
    1.8
     COMP_D1   Day 1 flag                               21662
    1.0
     COMP_D2   Day 2 flag                               21662
    1.0
     WT3_DAY1  Final 3-year day 1 weight                16103
16263.9
     WT3_2DAY  Final 3-year two day weight              15303
17114.1
     WTA_DAY1  Final annual day 1 weight                21662
38123.3
     WTA_2DAY  Final annual two day weight              20607
40075.0
     WT4_DAY1  Final 4-year day 1 weight                21662
12090.2
     WT4_2DAY  Final 4-year two day weight              20607
12709.1
     YEAR      Year of survey                           21662
1995.8
     D1_NREC   Day 1: number of food records            21662
   14.4
     D2_NREC   Day 2: number of food records            20607
   14.0
     ------------------------------------------------------------
--------


     Variable  Label                                         Minimum
     Maximum
```

```
     ------------------------------------------------------------
--------------
     HHID      Household ID                                10001.0
     52852.0
     SPNUM     Sample person number                            1.0
        11.0
     REGION    Region                                          1.0
         4.0
     URB       Urbanization                                    1.0
         3.0
     HHSIZE    Household size                                  1.0
        16.0
     INCOME    Annual income: total                            0.0
    100000.0
     INCREP    Annual income: actual report                    1.0
         9.0
     IMPFLAG   Annual income: imputation flag                   1.0
         5.0
     PCTPOV    Annual income: percent of poverty               0.0
       300.0
     AGE       Age in years                                    0.0
        90.0
     AGE_M     Age in months                                   0.0
        11.0
     SEX       Sex                                             1.0
         2.0
     WGT_SP    Weight of SP                                    4.0
       999.0
     WGT_KG    Weight of SP in Kilograms                       1.8
       208.7
     REL_REF   Relationship to reference person                0.0
        12.0
     RACE      Race                                            1.0
         5.0
     ORIGIN    Hispanic origin                                 1.0
         5.0
     PL_STAT   Pregnant/lactating status                       1.0
         5.0
     BF_STAT   Breastfeeding status                            1.0
         3.0
     H2O_COOK  Source of water: cooking                        1.0
        99.0
     H2O_BEVR  Source of water: beverages                      1.0
        99.0
     H2O_DRNK  Source of water: drinking                       1.0
        99.0
     D1_H2O_O  Day 1: amount of water                          0.0
       999.0
     D1_H2O_H  Day 1: water from home                          1.0
         9.0
     D1_H2O_A  Day 1: away from home water                     1.0
         9.0
```

```
     D2_H2O_O  Day 2: amount of water                          0.0
       999.0
     D2_H2O_H  Day 2: water from home                          1.0
         9.0
     D2_H2O_A  Day 2: away from home water                     1.0
         9.0
     COMP_D1   Day 1 flag                                      1.0
         1.0
     COMP_D2   Day 2 flag                                      1.0
         2.0
     WT3_DAY1  Final 3-year day 1 weight                    1404.0
    226692.0
     WT3_2DAY  Final 3-year two day weight                  1016.0
    434881.0
     WTA_DAY1  Final annual day 1 weight                     580.0
    669591.0
     WTA_2DAY  Final annual two day weight                   507.0
   1058203.0
     WT4_DAY1  Final 4-year day 1 weight                     340.0
    226692.0
     WT4_2DAY  Final 4-year two day weight                   286.0
    434881.0
     YEAR      Year of survey                               1994.0
      1998.0
     D1_NREC   Day 1: number of food records                   0.0
        56.0
     D2_NREC   Day 2: number of food records                   0.0
        52.0
     ------------------------------------------------------------
--------------
```

```
     Variable  Label
Sum
     ------------------------------------------------------------
-
     HHID      Household ID
568887640.0
     SPNUM     Sample person number
38281.0
     REGION    Region
57357.0
     URB       Urbanization
41798.0
     HHSIZE    Household size
81549.0
     INCOME    Annual income: total
836693653.0
     INCREP    Annual income: actual report
55063.0
     IMPFLAG   Annual income: imputation flag
30647.0
     PCTPOV    Annual income: percent of poverty
4441342.0
     AGE       Age in years
550082.0
     AGE_M     Age in months
8598.0
     SEX       Sex
32337.0
     WGT_SP    Weight of SP
3008008.0
     WGT_KG    Weight of SP in Kilograms
1021680.0
     REL_REF   Relationship to reference person
35644.0
     RACE      Race
33251.0
     ORIGIN    Hispanic origin
100481.0
     PL_STAT   Pregnant/lactating status
104004.0
     BF_STAT   Breastfeeding status
58905.0
     H2O_COOK  Source of water: cooking
51221.0
     H2O_BEVR  Source of water: beverages
65725.0
     H2O_DRNK  Source of water: drinking
71879.0
     D1_H2O_O  Day 1: amount of water
634554.0
```

```
     D1_H2O_H  Day 1: water from home
31896.0
     D1_H2O_A  Day 1: away from home water
10513.0
     D2_H2O_O  Day 2: amount of water
607609.0
     D2_H2O_H  Day 2: water from home
30141.0
     D2_H2O_A  Day 2: away from home water
10739.0
     COMP_D1   Day 1 flag
21662.0
     COMP_D2   Day 2 flag
22717.0
     WT3_DAY1  Final 3-year day 1 weight
261897277.0
     WT3_2DAY  Final 3-year two day weight
261897260.0
     WTA_DAY1  Final annual day 1 weight
825826029.0
     WTA_2DAY  Final annual two day weight
825825998.0
     WT4_DAY1  Final 4-year day 1 weight
261897244.0
     WT4_2DAY  Final 4-year two day weight
261897236.0
     YEAR      Year of survey
43231966.0
     D1_NREC   Day 1: number of food records
311153.0
     D2_NREC   Day 2: number of food records
287676.0
     ------------------------------------------------------------
-
```

Commodity Intake Database
Appendix A

EPA's Use of Food Consumption Data in Assessing Dietary Risk from Pesticides (Provided by EPA)

EPA is responsible for regulating the nature and amount of pesticide residues in food under the Federal Food, Drug and Cosmetic Act which authorizes EPA to set a legal and enforceable tolerance, or an exemption from the requirement of a tolerance. EPA's Office of Pesticide Programs regulates pesticides to ensure that their use does not pose unreasonable risks to human health and that pesticide residues found in food are safe. Assessing the amount of pesticide in or on the foods that we eat - both for fresh, raw foods such as lettuce or processed foods such as frozen french fries - is a complex process that requires data from numerous sources along with an understanding of risk analysis and risk management methods.

The dietary risk posed by a pesticide in food can be expressed as a function of exposure and toxicity. The dietary exposure is derived from the amount of pesticide residue that is present in and on food (the residue) and the types and amounts of food in a person's diet (i.e., food consumption). Toxicity is expressed as a reference dose to which a person can be safely exposed over time.

Dietary exposure to pesticides in foods is estimated by considering pesticide residues in foods and the amount of food consumed. In an attempt to conserve limited resources, EPA assesses dietary exposure in a tiered approach proceeding from conservative to more refined assumptions as the risk management situation dictates. Dietary exposure estimates based on tolerance level residues (farm-gate residues) reflect a Theoretical Maximum Residue Contribution (TMRC) which overestimate actual dietary exposure. To better estimate dietary exposure, EPA developed a process by which pesticide tolerance data (40 CFR 158.240) and compliance monitoring data are refined to reflect pesticide residues in food as consumed (dinner-plate residues). The best estimate of pesticide residues in food, as consumed, is termed the Anticipated Residue (AR) estimate. When estimating ARs, EPA uses all available data. It should be noted that since data sets vary in quality, considerable scientific judgement is required to derive anticipated residue estimates.

For many years, EPA has used food consumption data collected by USDA through various large, nationally representative surveys for its dietary risk assessments. These surveys have sampled thousands of households to learn about what, and how much, people eat. The Commodity Intake Database is a result of cooperative work by EPA and USDA in taking information on the foods people

reported eating in the 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII) and the Supplemental Children's Survey (SCS) and converting "foods" to the "agricultural commodity" terms and quantities used in EPA regulation of pesticide residues in foods. "Agricultural commodity" is a term used by EPA to mean plant or animal parts consumed by humans as food; when such items are raw or unprocessed, they are referred to as "raw agricultural commodities." Many food items contain more than one commodity, as, for example, an apple pie may contain the commodities apples, flour, fat, sugar and spices. (See also Section 3 of the Documentation of the Commodity Intake Database for more information on the methods used to translate foods to agricultural commodities).

Finally, toxicity of a pesticide is brought into the risk assessment equation. EPA toxicologists review study data on possible toxicity associated with exposure to a pesticide (often animal studies) and establish *reference doses* that are estimates of the level of either one-day exposure (acute) or daily exposure over a life-span (chronic) that are believed to have no significant harmful effects on humans. Safety factors are incorporated into these reference doses to account for the differences in response that may occur when animal data are used to estimate possible human response, and also to account for the variation of response to the pesticide that may occur within the human population.

Having information on pesticide residues, the number of various commodities treated with a given pesticide, the quantity of the commodities consumed by the population in the United States, in conjunction with toxicity data expressed in a reference dose, allows EPA to complete dietary risk assessments used to determine that pesticide residues found in food are within safe limits.

3 METHODS IN THE CONTINUING SURVEY OF FOOD INTAKES BY INDIVIDUALS 1994-96, 1998

The methods used in the Supplemental Children's Survey to the 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII 1998) were identical to those used in the CSFII 1994-96. The CSFII 1998 sample design shared most basic features with the CSFII 1994-96, but differed in a few respects outlined in documentation section 3.1.2, "CSFII 1998 sample design."

3.1 Sample Design

3.1.1 CSFII/DHKS 1994-96 sample design

The primary goal of the sample design for the CSFII/DHKS 1994-96 was to obtain nationally representative samples of noninstitutionalized persons residing in households in the United States for each of 40 analytic domains defined by sex, age (10 age groups), and income level (a "low-income" group and an "all-income" group) that were aimed to meet specified precision levels for estimates of mean day-1 saturated fat and iron intakes. Excluded were persons who lived in group quarters or institutions, who resided on military installations, or who were homeless. The specific precision goals were that the coefficients of variation (CVs) for mean saturated fat and iron intakes should be 3 percent or less for each of the 20 all-income sex-age domains and 5 percent or less for each of the 20 low-income sex-age domains. These precision goals were translated into 3-year target sample sizes. In addition, the sample design specified that one day-1 intake respondent 20 years of age or older be selected for the DHKS from each household with at least one day-1 intake respondent age 20 or over. For the CSFII/DHKS 1994-96, a single sample was selected that met precision requirements by income level, in contrast to past CSFII/DHKS surveys where a separate sample of low-income persons was also chosen in addition to the basic general sample.

The sample selection process was designed by Westat, Inc., a private research firm in Rockville, MD, under contract to ARS. The sample for the CSFII/DHKS 1994-96 was derived from a Westat, Inc., master sample. This master sample, which was in existence prior to the award of the contract for the CSFII/DHKS 1994-96, is a stratified, multistage area

probability sample. The sampling frame was organized using estimates of the U.S. population in 1990 (USDC/BOC 1993). The stratification plan took into account geographic location, degree of urbanization, and socioeconomic characteristics.

At the first stage of sampling, the entire United States was divided into primary sampling units (PSU's) consisting of Metropolitan Statistical Areas (MSA's) (see section 3.6, "Glossary," below), counties, or groups of counties. Because of its size, the New York MSA was divided into three PSU's. For the same reason, the Los Angeles and Chicago MSA's were each divided into two PSU's. Apart from these, each of the other MSA's constituted a single PSU. Some counties outside MSA's were grouped to form PSU's containing at least 15,000 people. A total of 1,404 PSU's was created, and 62 PSU's were selected for use in the CSFII/DHKS 1994-96, as described below.

The 24 PSU's with the largest populations were included with certainty. The remaining (noncertainty) PSU's were then assigned to 1 of 38 strata of approximately equal size (in terms of 1990 population), and one PSU was selected from each stratum with probability proportional to the 1990 population. Stratification factors included region of the country (four census regions) (see section 3.6, "Glossary," below); whether or not the PSU was an MSA and the population size of the MSA; percentage of the population that was black or Hispanic; and per capita income. Among the noncertainty strata, 26 were MSA strata and 12 were non-MSA strata.

The second stage was the selection from each PSU of 36 area segments consisting of blocks or groups of blocks. Area segments were chosen with probability proportional to size. The CSFII/DHKS 1994-96 was designed so that data collection would be spread evenly over the 3 years of the survey and over the quarters of the year. From each sampled PSU, twelve segments were subsampled for each of the 3 years of the survey, three segments for each quarter of the year. Addresses of all dwelling units in the subsampled area segments were then listed in accordance with 1990 Census listing rules and consistent with the 1990 Census definition of a housing unit (see section 3.6, "Glossary," entry for "Dwelling unit").

In the third stage, listed dwelling units in the selected area segments were drawn into the sample from the listings. For the three years of the CSFII/DHKS 1994-96, a sample of 34,016 dwelling units in all was designated for screening. Calculation of the number of dwelling units to be screened took into account the sample sizes needed to achieve the desired levels of precision specified by ARS prior to contract award, the percentages of individuals in each sex-age group living in households at or below 130 percent of the Federal poverty guidelines (DHHS 1996), a projected figure for vacant dwelling units, and a safety factor allowing for random sampling variation. Sample households were screened to identify appropriate numbers of sample persons in specified sex-age groups.

The last sampling stage involved selection of individuals from the sampled households. As described in the first paragraph of this section (section 3.1.1), the CSFII 1994-96 was designed to obtain sample sizes for the sex-age groups that would produce estimates with equivalent coefficients of variation over the sex-age groups, both for the total

population and for the low-income population. To obtain the desired numbers of individuals, sex-age subgroups were sampled at different rates. This procedure was implemented at the screening stage of the survey. The age groups used were 1 to 2 years, 3 to 5 years, 6 to 11 years, 12 to 19 years, 20 to 29 years, 30 to 39 years, 40 to 49 years, 50 to 59 years, 60 to 69 years, and 70 years and over. The approach used to select persons for the intake interviews was to designate subsets of households within which persons meeting specified sex-age/income criteria would be included in the study. For example, for a predesignated subset of households in the dwelling unit sample, only children between the ages of 1 and 2 years and low-income males between the ages of 50 and 59 years were to be included in the sample. Sampled households were assigned to the various subsets in a random fashion to ensure the unbiased selection of sample persons for the study. In addition, all infants under 1 year of age in households that contained at least one sample person 1 year or older were included in the sample.

To facilitate the selection of sample persons in the field, each screening questionnaire carried a sampling message specifying the characteristics of the persons to be included in the sample. The proportion of households receiving a particular message was determined to satisfy the target sampling rates for the various sex-age/income domains. After completing the listing of household members, the interviewer identified which, if any, of the household members fell into the sex and age groups that had been predetermined for that household. The interviewer had no discretion as to whom to include. In the CSFII 1994-96, a total of 20,126 individuals was initially selected into the sample.

Respondents for the DHKS 1994-96 were selected from among sample persons 20 years of age and over who had completed the day-1 intake interview in the CSFII 1994-96. Only one DHKS respondent per household was selected in households with eligible participants. In households with more than one CSFII participant 20 years of age or over, one of the participants was selected randomly with probability assigned to maintain distributions of all-income and low-income individuals in the six sex-age groups age 20 years and over in the DHKS that conformed approximately to the corresponding distributions of individuals in the CSFII. In the DHKS 1994-96, a total of 7,842 individuals was selected into the sample.

For more detailed information on the CSFII/DHKS 1994-96 sample design, see Tippett and Cypel (eds.) 1997, which is included on Disk 1 in \pdffiles\dor.pdf.

3.1.2 CSFII 1998 sample design

The CSFII 1998 had its roots in the Food Quality Protection Act of 1996, which required the Secretary of Agriculture to provide the Environmental Protection Agency (EPA) with information on food consumption patterns of a statistically valid sample of infants and children. This requirement followed a report entitled Pesticides in the Diets of Infants and Children (NAS/NRC 1993) that concluded that current food consumption data for children did not provide sufficient sample sizes for adequate estimation of dietary exposure to pesticide residues. In response to the

1996 mandate, the Agricultural Research Service (ARS) of the U.S. Department of Agriculture (USDA) conducted the CSFII 1998 as a supplement to the CSFII/DHKS 1994-96. CSFII 1998 data used in conjunction with CSFII/DHKS 1994-96 data, with appropriate weights (see documentation section 5, "Sampling Weights"), meet the requirement for a larger sample of children.

The goal of the sample design for the CSFII 1998 was to obtain nationally representative samples of noninstitutionalized persons 9 years of age or younger residing in households in the United States for each of 28 analytic domains defined by sex, age (7 age groups), and income level (a "low-income" group and an "all-income" group). The age groups used were under 1 year, 1 year, 2 years, 3 years, 4 years, 5 to 6 years, and 7 to 9 years.

A complex multistage area probability sample design that incorporated the same primary and second stage sampling units developed for the CSFII/DHKS 1994-96 was used to select children for the CSFII 1998. The same 62 PSU's that were selected for the CSFII/DHKS 1994-96 were used for the CSFII 1998. The PSU's were selected with probabilities proportional to the 1990 population. From each PSU, the 24 area segments used in the last 2 years of the CSFII/DHKS 1994-96 were used for the CSFII 1998. Those 24 segments were selected because they were the segments with the most up-to-date listing information.

Dwelling units (DU's) were selected from the area segments using listing information from the CSFII 1994-96 along with quality control procedures referred to as the "missed structure" and "missed dwelling unit" procedures. In preparation for the CSFII/DHKS 1994-96, interviewers had listed over 210,000 DU’s within the 1,488 area segments included in the CSFII 1998. DU’s that had been selected for the CSFII/DHKS 1994-96 were excluded from the CSFII 1998 sample. A sample of 65,519 DU’s (i.e., an average of 44 DU’s per sample segment) was drawn for the CSFII 1998 from the existing area segment listings. An additional 2,905 DU’s were added to the sample through quality control procedures referred to as the "missed structure" and "missed DU" procedures. Thus, 68,424 DU’s were selected for the CSFII 1998.

Each sampled dwelling unit was screened to determine whether it contained children who were eligible for the survey. From the DU's with children 9 years of age or younger, a sample of eligible children was selected by a probability sampling process designed to achieve the target sample sizes. Finally, to increase the number of 3-year-old girls in the sample, a special "supplemental" sample was selected and fielded in the fourth quarter of the study. The sampling procedures described above resulted in the initial selection into the sample of 6,413 children (including 2,100 low-income children).

For more detailed information on the CSFII 1998 sample design, see "Sample Design -- Supplemental Children's Survey to the 1994-96 Continuing Survey of Food Intakes by Individuals (CSFII 1998)", which is included on Disk 1 in \pdffiles\98_samp.pdf.

3.2 Data collection

3.2.1 CSFII/DHKS 1994-96 and CSFII 1998

The CSFII 1998 methods were nearly identical to those used in the CSFII 1994-96. A few questions were omitted because they were considered inappropriate for interviews with or about children, for example, questions about alcoholic beverages and smoking. See indented sections below regarding differences between CSFII 1998 and CSFII 1994-96.

Data were collected by Westat, Inc. Prior to data collection, listers visited every sample address in person to determine by visual inspection whether that location represented a dwelling unit (see section 3.6, "Glossary," below). An introductory letter and a brochure describing the survey were mailed to each dwelling unit 1 week before the initial in-person contact by the interviewer. In all materials for respondents, the survey was referred to as the "What We Eat in America" survey rather than by the official survey name. To contact individuals in the dwelling units, interviewers made at least four visits before referring the case to a supervisor. In a number of difficult cases, contact attempts exceeded the level of effort required by the contract in order to complete the interview. In cases where a dwelling unit was determined to contain a household but the household could not be contacted after four visits, interviewers were instructed to ask two neighbors for information on the number of household members and their sexes and ages as well as on the time household members were most likely to be home.

At each dwelling unit, the interviewer attempted a screening interview to determine whether any members of the household were eligible to participate in the survey. Any household member 18 years of age or older was an acceptable respondent for the screening questionnaire (screener). However, it was recommended that interviewers attempt to conduct this portion of the survey with either the main meal planner/preparer (see section 3.6, "Glossary," below) or a person knowledgeable about household characteristics such as income because those persons were the preferred respondents for the household questionnaire, which typically followed the screener. It was not necessary for the respondent(s) completing the screener and/or household questionnaire to be sample persons (see section 3.6, "Glossary," below). If a household member (see section 3.6, "Glossary," entry for "Household") refused to complete the screener, the interviewer was instructed to ask the household member for information on the number of household members and their sexes and ages so that the number of eligible respondents could be determined. (The number of eligible respondents was important for calculating the response rates provided in documentation section 4, "Response Results.")

At the beginning of the screening interview, the interviewer reminded the respondent about the letter and brochure that had been sent and provided new ones if the respondent did not remember. During the interview, information was collected on the number of persons living in the household; the first name of the person or one of the persons who owned or rented the home (reference person); the first name of the reference person's spouse, if any; and the first name, race, ethnicity (Hispanic or non-Hispanic), date of birth, age, sex, and relationship to the reference person of any other people living in the household, including friends, relatives, roomers, boarders, employees, and

household members who were away from home at the time of the interview but who usually lived there.

One screener question asked whether the total income of all household members from all sources during the previous year was more or less than an amount specific to the household's size. That question was part of the strategy for meeting the low-income sample size goals discussed in documentation section 3.1, "Sample Design."

In the CSFII 1994-96, the screener income question was asked only when the household included individuals in sex and age groups specified in the sampling message for that dwelling unit. In the CSFII 1998, the question was asked in all households.

The maximum income level used, where necessary, during the screening process to determine the household's eligibility for inclusion in the low-income group corresponded to 130 percent of the Federal poverty guidelines (DHHS 1998), which are based on household size and income. This income level was selected because it is the same as one of the income criteria used to determine whether nonelderly households are eligible to participate in the Food Stamp Program. Not all households meeting the income criteria are eligible for food stamps; other criteria, such as asset limitations, must also be met. The CSFII 1994-96 and CSFII 1998 screened households for income level only, not for food stamp eligibility.

At households where one or more sample persons were selected, the interviewer administered the household questionnaire--a series of questions about the educational level and employment status of household members 15 years of age and older, household income, food assistance program participation, food expenditures, and some other food-related practices. During the household interview, the interviewer asked the respondent to identify the "female head of household" and the "male head of household"; this question was included for the benefit of researchers who wish to make historical comparisons involving those variables. Interviewers made up to three visits after screening to complete the household questionnaire before referring the case to a supervisor.

Interviewers' visits were scheduled in a manner designed to ensure that at least 10 percent of day-1 food intake interviews took place on each day of the week. A label specified 3 days of the week that would be acceptable for collecting day-1 food intake information from that attached to the survey materials for each household. Repeated in-person visits were made as necessary to attempt to complete day-1 intakes with sample persons on the scheduled days of the week. In some cases, when repeated visits had been made on different scheduled days and at different times, interviewers were permitted to change the day of the week in order to obtain an interview. In households with more than one sample person, if one of the sample persons was not at home when the interviewer visited, the protocol required the interviewer to make up to three additional visits in an attempt to obtain a day-1 intake for that sample person. Often the number of visits required by the contract was exceeded in order to obtain the interview. An extensive range of strategies was employed in order to convert refusals, sometimes involving efforts by two or more interviewers.

Day-1 intakes were to be collected in person. Before conducting the day-1 interview, the interviewer told the sample person that her or his participation would involve two in-person interviews (and possibly, for one sample person in the household, the DHKS interview by telephone). At the conclusion of the day-1 interview, the interviewer notified the sample person that she or he would be returning in a few days to conduct another interview.

According to the survey protocol, the day-2 interview was to be conducted 3 to 10 days after the day-1 interview but not on the same day of the week. In the CSFII 1994-96, less than 1 percent of day-2 interviews were conducted sooner than 3 days after the day-1 interview, 20 percent were conducted more than 10 days after the day-1 interview, and 1 percent were conducted on the same day of the week as the day 1 intake exactly 1 week later. In the CSFII 1998, less than 1 percent of day-2 interviews were conducted sooner than 3 days after the day-1 interview, 17 percent were conducted more than 10 days after the day-1 interview, and 2 percent were conducted on the same day of the week as the day 1 intake exactly 1 week later. Five percent of day-2 interviews in the CSFII 1994-96 and 16 percent in the CSFII 1998 were conducted by telephone, with supervisory permission. Sample persons interviewed by telephone were asked to report food quantities using the measuring guides that had been used in the day-1 interview (described below) and given to the household.

The day-1 and day-2 questionnaires were very similar. Both included a 1-day dietary recall using a multiple-pass method in order to maximize the sample person's ability to remember what she or he ate and drank [Tippett and Cypel (eds.) 1997, DeMaio et al. 1993, Guenther et al. 1995].

The 1-day recall began with the sample person being asked to report everything eaten or drunk the previous day between midnight and midnight. The interviewer did not interrupt the sample person during this initial listing of the day's intake. The sample person was invited to add any other items remembered as the interview progressed. Then, for each food and drink listed, the interviewer asked the name of the eating occasion and the time it began.

For the CSFII 1998, the introduction was revised to delete references to coffee and alcoholic beverages, and the category "alcoholic beverage break" was deleted from the card the interviewer handed the respondent as an aid in naming the eating occasion. These changes were made to both day-1 and day-2 questionnaires.

The interviewer used a Food Instruction Booklet (FIB) to probe for a complete description of every food item and the amount eaten. Under each appropriate category of food/drink listed in the FIB, there was a list of the questions (probes) the interviewer was required to ask in order to collect enough detail for the food to be coded. Probes varied with the type of food or beverage being recalled. Some examples of FIB probes are "What was the brand name?" and "Were they regular, reduced calorie, high fiber, or something else?" When appropriate, questions were asked about the use of salt ("Was salt used in cooking or preparing the [food]?") and fat ("Was any kind of fat or oil used in cooking or

preparing the [food]?") in food preparation and about additions ("Did you add anything to the [food]?"). The interviewer was directed to ask for ingredients in some categories (for example, soups; tacos, burritos, enchiladas, and fajitas; sandwiches; salads; and mixed dishes, casseroles, and stews). Interviewers were required to use the FIB to obtain a detailed description of every food item recalled by the sample person, including additions remembered as the result of questions asked in describing another food. The FIB also suggested the types of measures (weight, volume, or size) appropriate for the food.

For the CSFII 1998, the FIB was refined to reflect some changes in food products since 1996, as well as changes in food terminology. For example, food label regulation changes for milk that went into effect in January 1998 narrowed the use of the term "lowfat" from 1-percent or 2-percent milk to only 1-percent. The regulations also introduced the term "reduced-fat" for 2-percent milk. As a consequence, the term "low-fat" for milk was deleted from FIB probes, and respondents were asked to specify the percent fat in the milk they used.

Measuring guides used to aid the sample person in estimating amounts were household measuring cups (1/4 cup, 1/3 cup, ½ cup, and 1 cup) and spoons (1/4 teaspoon, ½ teaspoon, 1 teaspoon, and 1 tablespoon); a 12-inch ruler with 1/8-inch increments marked; "thickness sticks," a set of 8 small rectangular pieces of hard plastic, each 1/8 inch in thickness; a laminated card printed with concentric circles 1 inch to 6 inches in diameter, two perpendicular 6-inch rulers, pictures of a fish filet and chicken parts, and diagrams specifying the dimensions to be measured or estimated when describing and quantifying various shapes. The cups and spoons could also be used to measure the capacity of tableware. One additional measuring guide, a 2-cup measuring cup, could be used only when the sample person referred to a bowl or cup in her or his home. The sample person could then fill the bowl or cup with water to represent the amount eaten or drunk, and the interviewer could measure the volume of water by pouring it into the 2-cup measure.

After each item on the initial list of the day's intake had been described and quantified, the interviewer reviewed for the sample person all the foods listed for each eating occasion and probed for additional foods eaten before the first eating occasion listed, in between listed occasions, and after the last occasion listed. Then, for each food or drink reported, the interviewer asked where it was obtained and whether it was eaten at home or not. For foods eaten away from home, the sample person was also asked whether the food or drink had ever been in the home before it was eaten; this question was included for the benefit of researchers choosing to make historical comparisons involving the variable "food from the home supply."

Additional questions asked on both day 1 and day 2 pertained to whether the sample person's intake on the previous day had been usual or unusual and why, how much plain drinking water the sample person drank on the previous day and whether it came from home or another source, and how many hours of television or videos the sample person watched on the previous day. Further questions in the day-1 questionnaire included the type of salt usually used by the sample person and frequency of use at the table; whether the sample person was on a diet and, if so, the type

and source of the diet; whether the sample person considered herself or himself to be vegetarian; frequency of vitamin or mineral supplement use and type of supplement; use of fish oil and fiber supplements; whether the sample person ever had a blood cholesterol check; self-reported height and weight (without shoes); self-assessed health status; food allergies; physician-diagnosed medical conditions; frequency of vigorous exercise; cigarette smoking status and number of cigarettes smoked per day; and consumption (ever or never) of alcoholic beverages during the past 12 months. The day-2 interview contained an additional question on the consumption (ever or never) of 28 foods during the past 12 months.

For the CSFII 1998, questions on exercise, smoking, and consumption of alcoholic beverages were removed from the questionnaires.

Proxy interviews were conducted routinely for child sample persons under 6 years of age and any other sample persons (including adults) who could not report for themselves due to physical or mental limitations; proxy interviews were not permitted for any other reason. Proxy interviews were not considered to be an acceptable substitute for an in-person interview with adult sample persons who were difficult for the interviewer to reach or who were nonrespondents. Child sample persons 6 to 11 years of age (6 to 9 years of age in CSFII 1998) were asked to provide their own food intake data assisted by an adult household member (referred to as the assistant). The preferred proxy or assistant was the person responsible for preparing the sample person's meals. If the sample person, proxy, or assistant could not provide enough descriptive or quantitative information about the foods eaten, it was sometimes necessary to seek that information from another caregiver such as a babysitter or school cafeteria personnel. It was permissible for any number of caregivers to contribute intake data for a sample person.

The first use of Spanish-language questionnaires in the CSFII and DHKS was in 1994-96. Interviewers who were bilingual in English and Spanish were provided with questionnaires and survey materials translated into standard Spanish and received an extra day of training in their use. The Spanish questionnaires reduced the number of language barrier cases and provided a standardized translation of the questionnaire content. They also minimized the need for interpreters, a practice that raises concerns about consistency of interpretation and interview length. If a sample person spoke neither English nor Spanish, a family member or neighbor 16 years of age or older was permitted to serve as an interpreter.

Spanish questionnaires were used in 2.8 percent of CSFII 1994-96 interviews (excluding screeners) and 4.4 percent of CSFII 1998 interviews.

The CSFII 1994-96 and CSFII 1998 used in-kind incentives. The interviewer told the screener respondent that each participating household would receive a gift. A set of measuring cups and spoons was given to the screener respondent after the screener was completed and the household was found to contain any sample person(s). An insulated nylon sack was given to each sample person prior to the collection of the intake, and at the conclusion of the day-2 interview each responding

sample person received a travel-type beverage mug as a thank-you gift for participating.

Average questionnaire administration time in the CSFII 1994-96 was about 7 minutes for the screener, 19 minutes for the household questionnaire, 32 minutes for the day-1 intake, and 29 minutes for the day-2 intake.

Average questionnaire administration time in the CSFII 1998 was about 7 minutes for the screener, 20 minutes for the household questionnaire, 32 minutes for the day-1 intake, and 30 minutes for the day-2 intake.

3.2.2 Diet and Health Knowledge Survey 1994-96

The DHKS was conducted only with respondents 20 years of age and older and so was not part of the CSFII 1998. This section is included because DHKS 1994-96 data are included in this release.

The Diet and Health Knowledge Survey was conducted as a telephone follow-up to the CSFII 1994-96. According to survey design, telephone contact was to be initiated 2 to 3 weeks after the day-2 intake. For households without telephones or with unlisted numbers not provided to interviewers, in-person interviews were the designated mode of contact.

When all sample persons in a household either had completed a day-1 intake or had been judged to be day-1 nonrespondents, the DHKS respondent was randomly selected by a computerized process from among eligible CSFII sample persons 20 years of age and over who had provided a day-1 intake. Sample persons were not eligible if their intake(s) had been completed by proxy, nor were any proxies allowed to complete the DHKS. Due to these criteria, not all households had a DHKS respondent. The interviewer scheduled an appointment for the telephone interview when the selected DHKS respondent had completed a day-2 intake. The same interviewer who administered the CSFII typically administered the DHKS. This continuity of interviewers maintained any rapport established between interviewer and respondent and was expected to have a beneficial effect on the response rate. Interviewers operating out of their own homes administered the questionnaire from a hard copy without computer assistance.

The interviewer mailed a DHKS reminder card 3 to 5 days prior to the scheduled interview. In addition to the appointment date and time, this card contained a list of response categories for selected questions in the DHKS questionnaire. During the interview, the respondent was directed to look at the set of response categories applicable to a particular question, thus reducing the need for the interviewer to repeat the response options. The card served both as an appointment reminder and as a means of improving the flow of the interview.

The first telephone contact was attempted on the scheduled day and time; if this attempt was unsuccessful, additional calls were made as needed at different times of the day and on different days of the week to reach respondents. The survey protocol required at least six telephone attempts at each number (as needed to obtain the interview), followed by four in-person visits. In a number of difficult cases, contact attempts exceeded the required level of effort in order to complete the

interview. Overall, the DHKS interview in 1994-96 took an average of 30 minutes to complete; it took longer to complete the DHKS in person (34 minutes, on average).

The telephone interview began with a request to speak to the person with whom the appointment had been made. The interviewer identified herself or himself and reminded the respondent that during the CSFII she or he had been told she or he would be recontacted later by telephone to answer a few more questions about food and nutrition issues. The DHKS respondent's name and age were verified at this time.

The gift that was provided at the end of CSFII day 2 also served as an incentive to complete the DHKS. Pretests and interviewer debriefings suggested that interest in the questionnaire content was also a motivating factor in completing the interview for some respondents.

Of all DHKS 1994-96 interviews, 84 percent were completed by telephone and 16 percent in person. The primary reasons for conducting interviews in person were that the household did not have a telephone or that limitations were posed by respondents' physical conditions (e.g., hard of hearing, feeble). Another reason was language barrier cases where an interpreter was needed.

In 1994-96, 74 percent of DHKS interviews were completed between 2 and 3 weeks after the last CSFII interview, as contractually specified. Interviews completed earlier than 2 weeks or later than 3 weeks were considered mistimed. Four percent of cases were completed earlier than 2 weeks due to reasons such as prior knowledge of extended periods of absence from the household (e.g., hospitalization, travel) and interviewer error. In 22 percent of cases, the length of time between the CSFII and the DHKS interviews was extended beyond 3 weeks because numerous contacts were required to complete the interview. These mistimings often centered on broken appointments where respondents were, for example, too busy or not at home at the scheduled time. Refusal conversion efforts also contributed to mistimings; some cases required intensive, prolonged efforts on the part of two or more interviewers to complete the interview.

In the DHKS 1994-96, a Spanish version of the questionnaire was available for use by bilingual interviewers. It served to reduce the number of language barrier cases and provided a standardized translation of the questionnaire content. The Spanish questionnaire also minimized the need for interpreters, a practice that raises concerns about consistency of interpretation and interview length. In 1994-96, 147 DHKS interviews (2.6 percent) were conducted using the Spanish questionnaire. In 1994-96, there were 61 cases (1.1 percent of DHKS interviews) where bilingual interviewers and telephones were not available or the respondent spoke a foreign language other than Spanish, interpreters were used. In these in-person interviews, the interpreters were required to be 16 years of age or older.

The content of the DHKS 1994-96 questionnaire was governed by a need for data on knowledge and attitudes about the Dietary Guidelines for Americans (USDA/DHHS 1990), food labeling issues, and dietary behaviors related to fat intake. Information from the DHKS can contribute to the

research base needed to develop food guidance materials and identify strategies for targeting nutrition education efforts. Thus, the data collected include self-perceptions of the adequacy of intake levels of nutrients and other dietary components, awareness of diet-health relationships, perceived importance of following dietary guidance for specific nutrients and other dietary components, behaviors related to fat intake and food safety, knowledge about food sources of fats and cholesterol, and self-perceptions about weight status. Also asked in the DHKS 1994-96 was a new series of questions regarding food labels. It covered use of various sections of the food label, use of specific information on the nutrient panel, frequency of using food labels when buying specified categories of food, ease of understanding food label information, and level of confidence in food label information.

3.3 Data Processing

3.3.1 Food coding and editing

The food intake data for the CSFII 1998 were coded and edited using Survey Net, the same computer-assisted food coding and data management system used with the CSFII 1994-96. Survey Net was developed cooperatively by ARS and the University of Texas-Houston Health Science Center's School of Public Health, and was tailored specifically to the questions, quality control needs, and data processing needs of the CSFII 1994-96. A general-use version of the software, the Food Intake Analysis System (FIAS), is available to researchers interested in using ARS survey food coding and nutrient databases. [For FIAS program and price information contact the University of Texas-Houston Health Science Center, School of Public Health, P.O. Box 20186, Houston, Texas 77225. Phone: (713) 500-9775. Fax: (713) 500-9329.]

Survey Net is a multilevel software system used by both the survey contractor and ARS. It operates on a computer network with multiple users accessing a set of central databases. These include (1) a food coding database containing food descriptions and food measures with their corresponding gram weights, (2) a predefined recipe database, and(3) the Survey Nutrient Database. All three databases are available with their documentation in the \TSF98 directory on Disk 2.

Westat's food coders used Survey Net to match descriptions of foods eaten by sample persons to foods listed in the food coding database. Coders entered partial or complete words or phrases from the sample person'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 sample person'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 which list 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 are included with the recipe database on the CD-ROM. Recipe modification numbers appear in the field MODCODE in record type 30 (rt30.dat).

There were three main purposes for recipe modifications: to record the specific type of fat, the type of milk, and 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, lowfat, 2-percent, 1-percent, or skim) used in their preparation. Some foods commonly modified for both 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 one can of frozen concentrate was mixed with four cans of water, instead of three cans of water.

Another aspect of the flexibility of food coding in the CSFII 1994-96 and CSFII 1998 is the use of combination codes, whose development and auxiliary use in analyses are discussed in detail in documentation section 3.3.8, "Combination codes." Combinations were often instances of one food being added to another, such as margarine to toast or gravy to potatoes. For some types of food made up of several components that are relatively easy to describe and quantify separately (such as sandwiches and salads) as well as for some mixed dishes, two or more food codes linked together in a food combination present a more precise picture of what was actually eaten by respondents than if a single food code is used.

Each food in the combination was coded separately and assigned the same combination type number (COMBTYPE) and sequence number (COMBNUM) in record type 30 fields (rt30.dat) separate from the food code. There were 11 combination types: beverage, cereal, bread/baked product, salad, sandwich, soup, frozen meal, ice cream/frozen yogurt, vegetable, fruit, and other mixture. Two-digit sequence numbers (01 and so on) linked the foods in a particular combination with each other and distinguished them from foods in other combinations. For example, a sample person might have cereal with milk in the morning and again in the afternoon. All the components of these two combinations would be assigned the combination type number for a cereal combination. The morning cereal with milk would be assigned one sequence number, and the afternoon cereal with milk would be assigned a different sequence number.

Survey Net's capabilities include a "copy foods" feature that allowed entries from a particular eating occasion, day, or sample person to be copied to a different eating occasion or day for the same person or to the food intake of another sample person in the same household. Survey Net also automatically performed gram weight checks of food quantities entered against maximum and minimum values established by ARS for each food. This weight check allowed coders to correct entry errors immediately. Coders recorded any questions regarding their food and

quantity selections in a notepad within Survey Net, which coding supervisors then reviewed and answered.

3.3.2 Processing of intakes by ARS

Westat electronically transmitted all coded intakes to ARS. All entries in each intake requiring review or resolution by ARS were highlighted in Survey Net's food summary screens. These included all "unknowns" (those foods or quantities that could not be coded by Westat coders); 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.

3.3.3 Food coding database

As mentioned previously, three databases are used in Survey Net. These include a food coding database (food descriptions, food measures, and gram weights of those measures); a recipe database; and a nutrient database.

The food coding database for CSFII 1998 contained 7,321 food codes, each bearing a complete description of the food and, if relevant, the preparation method. Each food code consists of 8 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.

Documentation section 12.1, "Food Coding Scheme," provides an outline of the major food groups and subgroups identified by the first 1 to 3 digits of the food code. Documentation section 12.2, "Food Codes and Abbreviated Descriptions," provides a list of the complete 8-digit food codes with abbreviated descriptive information about each code. Below are examples of the information found in documentation section 12.2.

CODE NUMBER	ABBREVIATED FOOD DESCRIPTION
28141010	Chicken, fried, pot, veg, dessert (froz meal, lg meat)
53105260	Cake, choc, devil's food/fudge, w/icing, homemade

More detailed food descriptive information is available on Disk 2 in the following files -- Food Description File (\tsf98\fcdb\cbdes.txt), Food Includes File (\tsf98\fcdb\cbincl.txt), Subcode Descriptions File (\tsf98\fcdb\cbsubdes.txt), and Subcode Includes File (\tsf98\fcdb\cbsubinc.txt). For example, information from the Food

Descriptions and Food Includes files is provided below for both food items listed above.

CODE NUMBER	COMPLETE FOOD DESCRIPTION
28141010	Chicken, fried, with potatoes, vegetable, dessert (frozen meal, large meat portion) (Include Banquet Extra Helping Fried Chicken Dinner; 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, NS from home recipe, from mix or bought RTE; Jack-in-the-Box Double Fudge Cake)

Sample persons varied in their knowledge of foods as well as in their ability to recall or describe foods eaten. Thus, the descriptions of foods provided by sample persons varied from very specific to very general. Also, sample persons 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 documentation section 3.3.5, "Recipe database.") In other cases, foods were reported with descriptions lacking only one detail. These foods were placed in codes providing as much detail as given and noting the one lacking detail as "not specified" in the code description, e.g., "chicken breast, fried, no coating, not specified as to skin eaten."

Identification by brand names is widespread in the food coding database. Several types of survey codes are specific to brands in the description of the code or in the weights provided. Codes may be unique to a particular brand if warranted, such as for breakfast cereals that differ in fortification levels, or they may encompass several brands of similar foods, such as cheese crackers. When appropriate, measures and their gram weight equivalents are specified by brand.

The guidelines used to decide if a new code is needed for a brand name food are the same as for other foods. A new code may be created for one or more of the following reasons: (1) no code presently exists for a food similar to the food reported, (2) the reported food contains either sizable amounts or intentionally reduced amounts of one or more nutrients, (3) the food is likely to be reported again, or (4) the form or type of food is of special interest to data users. Special effort is made to incorporate ethnic foods and foods modified to be lower in fat, sodium, or sugar.

3.3.4 Food measures and weights

Prior to the CSFII 1994-96, the food coding database's list of food measures and their corresponding weights in grams were examined for consistency by a Weights and Measures Team that included members from both ARS and the National Center for Health Statistics, 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. Cup weights for breakfast cereals and fluid ounce weights for infant formulas were updated based on new information from the manufacturers. 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 as needed were also added to the list. There are presently over 30,000 weights for measures of foods in the food coding database.

3.3.5 Recipe database

The purpose of the recipe database is to provide information for use during generation of the Survey Nutrient Database. It contains a recipe entry for each unique food code in the food coding database. These entries include ingredients and their amounts, as well as information for determining changes that may occur in nutrients during cooking. Foods that are not mixtures, e.g., whole milk, are represented as single-ingredient recipes. Ingredients are identified with codes linking them to the Primary Data Set of nutrient values (see documentation section 3.3.6.2, "Primary Data Set"). 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 sample person nor were they developed to determine the intake of specific food ingredients. A variety of popular, regional, and specialty cookbooks were consulted to aid in constructing representative recipes. Recipes for many of the commercially available mixtures were estimated from label information (Marcoe and Haytowitz 1993).

In preparation for the CSFII 1994-96, recipes for "Not Further Specified" (NFS) food codes were reviewed. These NFS codes are used when sample persons are unable to provide further detail about a food. For example, the "Milk, NFS" code is used when sample persons do not give the fat content of the milk they drank. The present recipe for "Milk, NFS" is a composite of whole milk, 2-percent milk, 1-percent milk, and skim milk in proportions that reflect milk production statistics. The "Milk, NFS" recipe is revised each year to reflect the most current production data. Recipes for other NFS codes may be based on composites, as for milk, or they may be based on the form of food most frequently consumed in the food group in question. For example, the recipe for "Bread, NFS" is white bread.

3.3.6 Survey Nutrient Database

About the Survey Nutrient Database

The Survey Nutrient Database is maintained specifically for use with nationwide food surveys (Perloff et al. 1990). It is updated once a year when a nationwide food survey is under way. 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).

The Survey Nutrient Database includes values for food energy and the following nutrients and food components: protein, total fat, saturated fatty acids, monounsaturated fatty acids, polyunsaturated fatty acids, 19 individual 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, moisture (water), selenium*, caffeine*, and theobromine*. Values for the three items with asterisks (*) were added to the database for the first time with this release.

The Survey Nutrient Database contains two files of nutrient values: (1) The Survey Nutrient Values, Set 1, which includes data for each unique survey food code from the food coding database (see documentation section 3.3.3, "Food coding database" above); and (2) the Survey Nutrient Values, Set 2, which is identical to Set 1 of the Survey Nutrient Values with the following exception: In recipes where salt is 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 sample person indicated salt was used in cooking the food, or if she or he did not know, data were selected from Set 1. If salt was not used, data were selected from Set 2.

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. These nutrient values are used to create the Survey Nutrient Database. The Primary Data Set is updated each year when a nationwide survey is being conducted. The main source of data for this version of the Primary Data Set (1998) was Release No. 11 of the USDA Nutrient Database for Standard Reference (USDA/ARS 1996), the same as used for the CSFII 1994-96. Unpublished data collected by the Nutrient Data Laboratory were also used as needed, especially for new products and for foods that recently changed. The most notable changes were to folate values as discussed below. As the survey was conducted, data for new foods were added as they were reported by sample persons,

and the final number of foods in the data set was 3,067. New values in the Primary Data Set can be identified by the "date added/modified" field [see 1998 formats document accompanying the Technical Support Files (on Disk 2 in \tsf98\formats\formats.txt)].

Selenium, caffeine, and theobromine values were added to the Primary Data Set for this release. The selenium content of plants, in particular cereal grains, is strongly influenced by the quantity of biologically available selenium in the soil in which they grow and, hence, their geographical origin (Holden et al. 1991). Values for major dietary contributors of selenium are based on laboratory analyses of food samples drawn from retail outlets according to nationwide sampling plans, in order to provide average values appropriate for national food surveys (Holden et al. 1991, Gebhardt et al. 1990).

Most of the values for major contributors of nutrients are 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 from data for similar foods (Gebhardt 1992). For each value in the Primary Data Set, a source code is present that indicates whether the value is analytical or imputed.

Folate values in this version of the Primary Data Set were updated to reflect regulations that became effective on January 1, 1998, requiring the addition of folic acid to enriched cereal grain products subject to standards of identity (DHHS/FDA 1996).These products include flour, cornmeal and grits, farina, rice, macaroni, noodles, bread, rolls, and buns. Folic acid may continue to be added (with some restrictions on amounts) to breakfast cereals, infant formulas, medical foods, food for special dietary use, and meal replacement products. For the most part, values in this data set were calculated based on enrichment levels specified in the regulations, since analytical values were not yet available. For those foods where the enrichment level is given as a range, the midpoint was used to set the value. Food items containing any of these products as ingredients, such as baked products made with enriched flour, were also updated.

The state of analytical methodology for measuring nutrients in foods has been evaluated by Beecher and Matthews (1990), and they reported that adequate methodology for folate is lacking. The current microbiological method approved by the Association of Official Analytical Chemists International applies 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. Recent research on determining the folate content of high-protein and high-carbohydrate foods indicates that additional improvements in methodology are needed (Martin et al. 1990).

Data users should note that values for carotenes are those used by ARS in arriving at the values for total vitamin A and are not solely beta-carotene. Also, the values for vitamin E (quantified as alpha-tocopherol equivalents) are based on somewhat limited data.

Recipe calculations

Entries in the recipe database identify the Primary Data Set (PDS) item(s) used to derive the Survey Nutrient Values, Set 1 and Set 2. As mentioned in the recipe database discussion, some survey food codes have a one-to-one correspondence with items in the Primary Data Set and are 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
01079	Milk, 2% Fat, with Vit A	100 grams

However, many survey food codes require 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. Perloff has described how this method is used for generating values in the Survey Nutrient Database, including how factors estimating changes in nutrients due to cooking or processing are used in the calculations (Perloff 1985). Factors for calculating moisture and fat changes are stored in each recipe. Factors for estimating losses in 18 vitamins and minerals are stored in a separate data file, the Nutrient Retention Factors File, which is accessed during the recipe calculation procedure. The presence of special codes in the recipe entries indicate when the retention factors are used. Retention factors for selenium and vitamin E are not available.

3.3.7 Multi-year databases

The nutrient intake data for the CSFII 1998 were calculated using the 1998 values from the multi-year food coding, nutrient, and recipe databases that are included only on Disk 2. Some foods changed between the CSFII 1994-96 and the CSFII 1998. For example, folic acid is now added to enriched grain products. In such cases, both the Primary Data Set and the Survey Nutrient Database contain multiple records for the different nutrient levels in the food. Multiple records also exist for some food weights and recipes. Multiple records do not exist for modified recipes.

All records in the multi-year food coding, nutrient, and recipe databases have start- and end-date fields indicating the time period

when each record was available for coding. These date fields can be used to extract a single-year version from the multi-year database.

3.3.8 Combination codes

Rationale for and development of combination codes

A notable feature available on the CSFII 1994-96 and 1998 combined data set is combination codes. Data users can find combination code data in record type 30 (rt30.dat) fields COMBNUM (positions 104-5) and COMBTYPE (positions 106-7).

Combination codes were developed for two distinct purposes. First, a greater level of specificity in coding is possible when sufficient detail about the foods that make up a combination is collected. For some foods, two or more food codes linked together in a food combination present a more precise picture of what was actually eaten by respondents than if a single food mixture code is used. Second, the use of combination codes provides insight into patterns of food consumption--what types and amounts of foods are eaten together and what types and amounts of foods are eaten as separate items. This information is helpful in answering questions about not only what people are eating, but how they are eating it and how much. For example, do adults and children consume milk differently? Do adults get more of their milk from drinking it as a beverage, or from adding it to another food, such as coffee or cereal?

Recognition of the need for a way to express food combinations through multiple food codes began with the NFCS 1977-78. For the NFCS 1977-78, three "partition codes" were developed to indicate foods that were part of a sandwich, part of a salad, or part of a frozen meal, as shown in table 3-1 on the next page. Approximately 12 percent of all foods were assigned one of these partition codes.

Table 3-1. Use of partition codes and combination codes, NFCS 1977-78 through CSFII 1996

	1977-78 NFCS#	1985-86 CSFII*	1987-88 NFCS#	1989-91 CSFII##
	Percent			
Partition code type:				
Sandwich	10.7	13.0	12.6	13.1
Salad	1.0	4.2	4.0	3.9
Frozen meal	**	**	**	NA
Mixture	NA	1.7	1.7	2.2
Soup	NA	.1	.1	.2
Beverage	NA	NA	8.3	9.3
Missing	.1	**	**	NA
Single item	88.2	81.0	73.3	71.3

	1994 CSFII	1995 CSFII	1996 CSFII
	Percent		
Combination type:			
Sandwich	13.8	13.7	14.8
Salad	5.1	5.1	5.2
Frozen meal	**	**	.0
Other mixture	5.0	5.3	5.3
Soup	.6	.6	.5
Beverage	7.4	8.3	8.1
Cereal	6.1	6.3	5.9
Baked product	7.2	7.3	7.2
Ice cream	.4	.4	.5
Vegetable	3.5	3.8	3.8
Fruit	.4	.5	.4
Single item	50.5	48.8	48.5

#Basic sample.
*Women and children, basic and low-income samples.
##Combined basic and low-income samples.
** Calculated value is <0.1%.

The number of partition codes and the utilization of these codes increased gradually through the years. In the CSFII 1985-86, partition codes were added for mixtures and for soups, and 19 percent of all foods were assigned a partition code. In the NFCS 1987-88, a partition code was added for beverages with additions (for example, coffee with cream and sugar) or with multiple ingredients (for example, "health shakes," that is, milk- or juice-based drinks with fruit, cereals, and other ingredients pureed together), and the percent of foods assigned a partition code increased to 27 percent.

The increased use of partition codes was also due to the concurrent increase in the number of nutrients in the Survey Nutrient Database, such as fiber. Greater specificity in reporting and coding of foods was necessary in order for appropriate nutrient values to be assigned. The use of partition codes allowed this information to be coded and at the same time avoided having to add unmanageable numbers of new food codes.

In the CSFII 1994-96, nearly one-half of all foods items were reported in combination. This near-doubling in the number of foods that were part of a combination is attributable in part to two major changes in the way mixture information was collected and coded. First, there was an expansion of the concept of "partition codes" with the addition of five more combination codes for cereal, baked product, ice cream, vegetable, and fruit combinations. These codes were used to code the ingredients in selected mixtures, as well as to link accompanying food items with the foods they were combined with "at the table," such as cream cheese on a bagel, margarine on a baked potato, or banana or berries on cereal. Second, the Food Instruction Booklet (FIB) was revised to standardize the collection of details about additions to foods and about mixtures, thus enabling greater specificity in food coding.

> There were no changes in combination codes between CSFII 1994-96 and CSFII 1998.

Data collection and coding of combinations in the CSFII 1994-96

The FIB is described in documentation section 3.2.1, "CSFII/DHKS 1994-96 and CSFII 1998." Under each category of food/drink in the FIB, there was a set of questions (probes) the interviewer was required to ask in order to collect enough detail for the food to be coded. For the CSFII 1994-96, major changes made to the FIB include not only more food

categories, but also more standardized probes, including probes about ingredients of foods and any additions to foods.

For the CSFII 1994-96, interviewers and coders were trained on how to record and code combinations. Following instructions in the FIB, the interviewers recorded ingredients of mixtures such as sandwiches and salads and placed brackets around them to identify them as one food item eaten. They also used brackets to link foods added together "at the table," such as the cream added to coffee and the jam spread on toast. Coders used this information to code the foods as eaten in combination. If insufficient information was available to code separately all the food items included in a salad or sandwich (for example, when detailed descriptions or amounts of ingredients were not given), the coder would attempt to find a close single-code match for the combination in the food coding database.

If enough information was available to code a combination as two or more separate food items, all food codes for that combination were assigned both a combination type number and a sequence number. The coder chose the combination type from a list of categories provided by ARS (see table 3-2 below). Each combination was assigned a sequence number which served to distinguish that particular combination from other combinations consumed by that sample person on that intake day. The combination type and sequence number are labeled as COMBTYPE and COMBNUM, respectively, on record type 30 (rt30.dat).

Table 3-2. Combination types (and type numbers)--foods with additions or foods in combination

Beverage (01)--
- Coffee/tea with: milk, cream/cream substitute; sugar/sugar substitute
- Water with: lemon; lime; fruit juice
- Infant formula with: instant baby cereal added to formula
- All milkshake/float ingredients coded separately
- All beverage/mixed drink ingredients coded separately

Cereal (02)--
- Ready-to-eat breakfast cereals with: milk; sugar/sugar substitute; fruit
- Cooked cereals such as oatmeal, cream of wheat, grits with: milk; sugar/sugar substitute; fruit; margarine/butter; gravy
- Several breakfast cereals in a mixture coded separately
- Instant baby cereal with: formula, milk, water, beverage added

Bread/baked product (03)--
- Toast, rolls, buns, bagels, biscuits, muffins, sweet breads, pancakes (including potato), waffles with: margarine/butter; jam/jelly; cheese/cream cheese; whipped cream; syrup; fruit; gravy
- Cakes, pies, brownies, cookies with: ice cream; whipped cream; fruit
- Crackers with: meat; cheese; dip; peanut butter; jam/jelly; margarine/butter
- Nacho chips/corn chips with: cheese; dip; refried beans, etc. (nacho supremes)
- Rice cakes with: peanut butter; jelly; cheese, etc.
- Tortilla with salsa

Salad (04)--
- All salad ingredients coded separately and/or additions
- Green leafy salads, pasta salads, fruit salads, potato salad, taco salad, egg salad
- Salad dressing added to salad

Sandwich (05)--
- All sandwich ingredients coded separately and/or additions
- "Filled" tacos, enchiladas and burritos
- Hamburger, hot dogs with ingredients coded separately and/or additions
- Quesadilla

--continued

Table 3-2 -- continued

Soup (06)--
  * Soup with: crackers; cheese; croutons; green spring onions
  * All soup ingredients coded separately and/or additions

Frozen meal (07)--
  * Frozen meal with: catsup, tartar sauce, margarine/butter
  * All frozen meal ingredients coded separately

Ice cream/frozen yogurt (08)--
  * Ice cream or frozen yogurt with: syrup; toppings; fruit; nuts; whipped cream; candy; cookies
  * All ingredients of a sundae coded separately

Vegetables (09)--
  * French fries with: catsup; gravy; steak sauce; vinegar; dressing
  * Potato chips with: dip
  * Potatoes with: gravy; sour cream; toppings; butter/margarine added
  * Beans, legumes with: sauce; margarine/butter
  * Vegetables (not specified as salad) with: margarine/butter; sauce; dip; dressing
  * Vegetables in a mixture coded separately

Fruit (10)--
  * Fruit with: whipped topping; sugar; milk/cream; syrup; honey
  * Fruits in a mixture (not specified as salad) coded separately

Other mixtures (99)--
  * Rice with: butter; gravy; sauce
  * Pasta/spaghetti with: butter; gravy; sauce
  * Meat, poultry, fish with: gravy; sauce; onions
  * Eggs with: catsup, salsa
  * Pizza with: grated cheese
  * Yogurt (not frozen) with: nuts, fruit, cereal, etc.
  * Foods/mixtures of foods that do not fit in other combination categories

---

Examples of analyses using combination codes

The presence of combination codes in the CSFII 1994-96 and 1998 combined data set may be useful in planning analyses, especially concerning salads, sandwiches, or foods combined "at the table," such as cereal and milk or corn chips and salsa. Examples 1 and 2 below illustrate how using combination codes can affect frequencies and mean food and nutrient intakes. Example 3 illustrates how combination codes can also provide insight into food consumption patterns.

Two-day intake data from the CSFII 1994 were used for all examples. The estimates are unweighted.

Example 1 (using combination types to measure frequency and mean intake of specific food mixtures)--To fully account for all reports of a food mixture such as a sandwich or salad, consideration must be made of the different ways that foods may have been recorded and coded. Depending on how a food was reported, it may have been coded as a single item or as multiple items linked via a combination type and sequence number. Including both ways of reporting in an analysis requires familiarity with the food coding database, but it can give a more complete picture of the consumption of that food.

For instance, suppose the research objective was to determine consumption of hamburgers and cheeseburgers. All of the hamburgers and cheeseburgers that were coded as a single item received codes in the range 275-10210 through 275-10690 in the CSFII 1994 food coding database. The number of reports, mean intakes by sex-age group, and sources of hamburgers and cheeseburgers coded as a single item are presented in tables 3-3 and 3-4.

Table 3-3. Number of reports and mean intake of hamburgers and cheeseburgers, single-code items only*, CSFII 1994 (unweighted)

Sex and age (years)	Number of individuals	Number of reports	Mean intake per report (gm)
Children < 6	1,140	150	96
Children 6-11	506	67	134
Teens 12-19	529	108	207
Women 20+	1,541	120	188
Men 20+	1,547	215	220
Total	5,263	660	---

*Includes hamburgers or cheeseburger codes in the range 275-10210 through 275-10690 regardless of whether that food was eaten in combination with another food or not.

Table 3-4. Places where hamburgers and cheeseburgers were obtained, single-code items only*, CSFII 1994 (unweighted)

Sex and age (years)	Store**	Restau-rant	Fast food	School cafeteria	Other
			Number		
Children <6	6	6	135	1	2
Children 6-11	4	3	52	8	0
Teens 12-19	0	2	96	5	5
Women 20+	3	3	111	0	3
Men 20+	1	3	208	0	3
Total	14	17	602	14	13

*Includes hamburgers or cheeseburger codes in the range 275-10210 through 275-10690 regardless of whether that food was eaten in combination with another food or not.
**Includes prepared sandwiches or sandwich ingredients purchased from stores.

It is not surprising that most of the hamburgers and cheeseburgers coded as a single item were from fast food places, because the preferred method given in the FIB for reporting standardized items such as fast food sandwiches from national chains was as a single item. Nonfast-food (or nonstandardized) hamburgers and cheeseburgers were more commonly coded as multiple food items linked with a combination code, because the FIB specified probes for the ingredients of nonstandardized sandwiches.

Using only the single-item food codes does not consider those hamburgers and cheeseburgers that were coded as multiple food items linked with a combination code. One way to expand the definition of hamburgers and cheeseburgers would be to include all sandwich combinations (COMBTYPE = 05) containing at least one code from the range 215-00100 through 215-40100 (ground beef) and one code from the range 510 ----- through 518 ----- (yeast breads and rolls). Other ingredients might also be part of these combinations. For example, this group would include a report of a sandwich with ground beef, lettuce, tomato, and ketchup on a kaiser roll.

The numbers of reports and amounts resulting from adding combinations of food items eaten as hamburgers and cheeseburgers to hamburgers and cheeseburgers coded as a single item appear in tables 3-5 and 3-6. The number of reports of hamburger and cheeseburger consumption is nearly double that shown in table 3-3, and the distribution is less dominated by the fast food sandwiches, as expected.

Table 3-5. Number of reports and mean intake of hamburgers and cheeseburgers, single-code items and combinations, CSFII 1994 (unweighted)

Sex and age (years)	Number of individuals	Number of reports	Mean intake per report (gm)
Children < 6	1,140	233	107
Children 6-11	506	160	148
Teens 12-19	529	216	207
Women 20+	1,541	272	183
Men 20+	1,547	430	220
Total	5,263	1,311	---

Table 3-6. Places where hamburgers and cheeseburgers were obtained, single-code items and combinations, CSFII 1994 (unweighted)

Sex and age (years)	Store*	Restaurant	Fast food	School cafeteria	Other
	Number				
Children <6	52	12	144	12	13
Children 6-11	50	9	60	33	8
Teens 12-19	43	6	125	29	13
Women 20+	91	27	135	5	14
Men 20+	107	30	251	5	37
Total	343	84	715	84	85

*Includes prepared sandwiches or sandwich ingredients purchased at stores.

Example 2 (using combination codes to aggregate food groups for nutrient analyses)--Similarly, assessments of the nutrient contributions of specific foods can be affected if the food was often eaten as part of a mixture that was coded in combination with other foods as well as separately. Lettuce can serve as an illustration of this type of situation. There is a series of codes in the 1994 food coding database for lettuce-based salads coded as a single item (751-43000 through 751-46000 and 751-48000). An example of these lettuce-based salads is 751-43000 (lettuce, salad with assorted vegetables including tomatoes and/or carrots, no dressing). The nutrient contribution of this food group is shown in table 3-7.

Table 3-7. Number of reports, mean intake, and nutrient contribution of lettuce-based salads, single-code salads only*, CSFII 1994 (unweighted)

Sex and age (years)	Number of reports	Mean salad intake per report (gm)	Energy (kcal)	Protein (gm)	Carbohydrate (gm)	Fat (gm)
Children <6	24	38	7	.4	1.4	.1
Children 6-11	21	57	12	.7	2.1	.3
Teens 12-19	15	85	34	1.9	3.2	1.7
Women 20+	46	115	36	1.8	4.3	1.6
Men 20+	38	140	57	3.0	5.3	2.9

*Includes the nutrients from all lettuce-based salad codes (751-43000 through 751-46000 and 751-48000) regardless of whether that food was eaten in combination with another food or not.

Using only the single-code salads has two deficiencies that can be corrected by the use of combination codes. First, it can be noted from examination of the food coding database that the lettuce-based salads coded as a single item do not include salad dressing. This is because the FIB specified probes for salad dressing in order to obtain as much information as possible about the type and amount of salad dressing eaten. Salad dressing is always linked to salad via a combination type (04, salad) and sequence number. Consequently, if only single-code salads are considered, the contribution of lettuce-based salads to total fat intake is underestimated. Second, restricting the analysis to single-code lettuce-based salads misses any salad-type combinations with lettuce coded simply as lettuce (751-13000).

When all lettuce-containing salad combinations (COMBTYPE = 04) are added to all single-code lettuce-based salads (this time incorporating any other ingredients linked to them via combination type 04 and sequence number), the nutrient contributions are considerably different, as shown in table 3-8. Not surprisingly, the contribution of lettuce-based salads to nutrient intake, most notably energy and fat, is dramatically increased when mixtures linked by combination codes are included. Mean salad intakes also increased markedly, and the number of reports of salads increased five- to fourteen-fold across the sex-age groups.

Table 3-8. Number of reports, mean intake, and nutrient contribution of lettuce-based salads, single-code and combination salads, CSFII 1994 (unweighted)

Sex and age (years)	Number of reports	Mean salad intake per report (gm)	Energy (kcal)	Protein (gm)	Carbohydrate (gm)	Fat (gm)
Children <6	144	77	77	1.5	4.1	6.3
Children 6-11	109	101	101	1.6	5.5	8.5
Teens 12-19	106	170	197	4.9	9.2	16.3
Women 20+	594	179	179	4.2	9.9	14.3
Men 20+	541	205	210	4.4	11.4	17.2

Example 3 (using combination codes to examine food consumption patterns)--The manner in which individuals consume foods, that is, separately or together with other foods, may be determined by using combination codes. It can be expected that population subgroups vary in their consumption patterns. For instance, children consume milk primarily as a single-code item whereas adults more often consume milk in combination with another food, such as coffee or cereal, as shown in table 3-9 on the next page. Although nearly one-third (32.6 percent) of all reports of milk by women were milk consumed as part of a beverage combination (such as in coffee), the largest percentage of the total quantity (in grams) that was consumed by women was provided by milk consumed as a single item (57.1 percent). Milk added to cereal made a substantial contribution to total milk consumption for all sex-age groups.

Limitations of combination codes

While combination codes may be used to identify foods eaten together, disaggregation of combinations is not sufficient to enable researchers to look at the total intake of a specific food. For example, a researcher who wished to look at the total intake of tomatoes from all sources could not arrive at that number by combining tomatoes reported separately with those that were reported as part of a combination. That method of analysis would miss tomatoes that are included as ingredients in many single-code mixtures such as 283-10220 (chili beef soup) and 581-30010 (lasagna with meat and/or poultry).

Table 3-9. Milk consumption by combination type, CSFII 1994 (unweighted)

Sex and age (years)	Combination type			
	Single item	Beverage combination	Cereal combination	Other combinations
	Number of reports			
Children <12	3,383	200	1,665	15
Teens 12-19	471	27	283	3
Women 20+	788	739	719	18
Men 20+	959	641	764	24
All ages	5,601	1,607	3,431	60
	Percent of all milk reports			
Children <12	64.3	3.8	31.6	0.3
Teens 12-19	60.1	3.4	36.1	0.4
Women 20+	34.8	32.6	31.8	0.8
Men 20+	40.2	26.8	32.0	1.0
All ages	52.4	15.0	32.1	0.6
	Percent of total quantity consumed			
Children <12	70.5	4.0	25.3	0.2
Teens 12-19	67.1	2.5	30.0	0.4
Women 20+	57.1	11.1	31.2	0.6
Men 20+	61.1	8.4	29.6	0.9
All ages	65.6	6.1	27.9	0.5

3.4 Quality Control

At every step during the development and execution of the CSFII/DHKS 1994-96 and CSFII 1998, quality control has been one of ARS' primary concerns. During the process of CSFII/DHKS 1994-96 questionnaire development, ARS solicited input from the Continuing Survey Users Group, which is made up of representatives from 13 Federal agencies, as well as other Federal users. The CSFII intake questionnaire underwent cognitive testing by the Census Bureau's Center for Survey Methods Research, and the "multiple-pass" approach used for the first time in 1994 was developed to optimize the completeness of intake data collected [Tippett and Cypel(eds.) 1997, DeMaio et el. 1993, Guenther et al. 1995]. The DHKS questionnaire was revised and expanded with input from members of the Continuing Survey Users Group and an in-house DHKS working group. It

was then pretested for comprehensibility and flow by ARS in collaboration with the Census Bureau's Demographic Surveys Division. ARS staff and the contractor revised the Food Instruction Booklet used in conjunction with the intake section of the CSFII questionnaire, expanding the booklet to standardize probing by interviewers and ensure the collection of adequate detail for food coding.

All field supervisors, interviewers, and coders attended extensive training sessions. All sessions were scripted for consistency and were monitored by ARS staff. Bilingual interviewers attended an additional day of training in the use of Spanish language questionnaires. Practice interviews were reviewed by supervisors, and telephone retraining was conducted when necessary. Detailed instruction manuals were provided to supervisors, interviewers, and coders.

Electronic communications permitted close tracking by the contractor and ARS of assigned cases in the field, their completion status, and documents in various stages of processing. Electronic delivery of survey data facilitated the timely resolution of such issues as errors in sample person selection or clarification by the interviewer of data received by the home office.

Survey Net, a computer-assisted food coding system (see documentation section 3.3.1, "Food coding and editing," for additional information on Survey Net) developed under a cooperative agreement between ARS and the University of Texas-Houston Health Science Center, School of Public Health, provided efficiency and accuracy in on-line coding of foods and editing. Edit checks were built into the system to reduce data entry of erroneously high or low food amounts and to catch some of the most common reporting, recording, and coding errors.

A pilot study duplicating the planned survey design on a small scale was conducted from April to June of 1993. The pilot study tested the questionnaires, data collection methods, field management procedures, data entry and processing procedures, and survey management software slated for implementation in the CSFII 1994-96. This experience provided an excellent opportunity to further refine the quality of survey instruments and improve the efficiency of survey operations.

As a result of the pilot study, interviewer training was lengthened to 7 full days to allow more thorough coverage of survey procedures. Modifications that had been made to the questionnaires and data collection procedures were judged to be effective in reducing respondent burden and facilitating the collection of high-quality data. Interviewer field notebooks and debriefing after the pilot study provided feedback resulting in further revision of the questionnaires. Survey management software programs used by the contractor and ARS were found to be effective tools for monitoring survey activities and improving the efficiency of survey operations.

ARS data processing activities were reviewed by a panel of outside experts in November 1994. The panel's primary recommendation was that ARS scale back its exhaustive review of the data by prioritizing tasks and streamlining the mechanics of data processing. Quality control

procedures described in this section reflect ARS' implementation of the panel's recommendations.

Achieving acceptable response rates in the CSFII/DHKS 1994-96 was a priority for ARS. By contract, Westat, Inc., was required to meet specified response rate requirements for each questionnaire (screener, household, individual intake, and DHKS).

Many steps were taken to monitor interviewer performance. These included partial reinterviews of 10 percent of each interviewer's cases to validate contact of households, audiotaping of at least one intake interview and two DHKS interviews per interviewer per year, and in-person observations.

Interviewers were instructed to edit their own work as soon as possible after the interview to identify and correct errors in recording and to permit (with supervisory permission) retrieval of any missing information from the respondents. Completed questionnaires were reviewed within 2 days of receipt at the Westat central office to determine whether they met ARS minimum criteria. If not, callbacks were made to obtain missing information. Reviews sometimes led to telephone mini-retrainings of interviewers. Field staff memos and a quarterly newsletter provided all interviewers with answers to questions raised during training and in the field, as well as feedback on problem areas detected in data review by the contractor and ARS.

Food coders were required to pass a certification test developed by ARS before they were allowed to code survey data. Initially, 100 percent of each food coder's work was verified by blind double-coding with resolution of any differences. At the supervisor's discretion, this adjudication process was applied to less of the coder's work; 10 percent of the food coder's work continued to be verified routinely. Problems in the food coding process were discussed at biweekly food coding meetings. ARS monitored coder performance by occasionally observing food coders at work, by periodically attending coder meetings and refresher trainings, and by comparing information recorded on the questionnaire to coded entries.

Accuracy of nonfood data entry was verified by routine 100 percent independent double entry with resolution of differences by coding supervisors. Nonfood data were edited for reasonableness, logic, and consistency; supervisors resolved discrepancies.

ARS verified the accuracy of weekly data delivery by checking each hard-copy document received against an electronic list of documents. At least 10 percent of all food intake questionnaires were reviewed for accuracy in coding and data entry. In addition, all foods and food amounts that could not be coded by the contractor (i.e., "unknowns") were reviewed and coded by ARS food coding staff. Other food codes and amounts flagged by the contractor as questionable were reviewed for accuracy. All recipe modifications (see documentation section 3.4.1, "Food coding and editing") done by Westat were reviewed by ARS coding staff.

A series of reviews was conducted on food intake data. Values of food variables falling outside reasonable parameters were flagged, checked against information recorded on the questionnaire, and corrected if in error. ARS reviewed audiotaped intake interviews for proper interviewing techniques. Any problems in interviewer or coder performance detected by ARS were brought to the attention of the contractor.

ARS review of nonfood data was also extensive, encompassing over 30 specific edit checks for reasonableness, consistency, and logic. Values falling outside of reasonable or expected parameters were checked against information recorded on the questionnaire and corrected if in error.

All screeners from eligible households were reviewed to confirm that proper sampling procedures had been followed. Sampling errors were immediately brought to the contractor's attention.

All household questionnaires and DHKS questionnaires were reviewed to ensure that proper interviewing and coding procedures had been followed. Any interviewer or coder problems were summarized in periodic reports to Westat. Also, audiotaped DHKS interviews were reviewed by ARS, and general feedback was provided to the contractor.

The accuracy of the Survey Nutrient Database was also a priority for ARS. Numerous quality control checks were performed on various components of the Survey Nutrient Database, such as nutrient values for new or updated codes in the Primary Data Set, the recipe file, and the file of weights for household measures. Final nutrient values in the 1994-96 Survey Nutrient Database were confirmed by a series of comparisons to earlier Survey Nutrient Databases, with subsequent review of values falling outside of reasonable parameters. After food codes were aggregated by type of food, averages of nutrients from those foods were subjected to many of the same rigorous outlier checks conducted for Primary Data Set codes.

Every nutrient intake value (daily total) from each responding sample person's intake was tested for reasonableness against parameters for individuals of that age and sex. In addition to detecting errors in coding of foods or amounts, this provided an additional quality check of the nutrient database.

3.5 Glossary

Age--Calculated from date of birth, if given. Otherwise, age as given by respondent. For responding sample persons (see "Responding sample person"), this is the age as of the day-1 intake; for others, this is the age on the day of screening.

Alpha-tocopherol equivalent--See "Vitamin E."

Assistant--Person who assisted in the dietary recall for a sample person age 6 to 9 years in CSFII 1998 and age 6 to 11 years in CSFII 1994-96.

Black--See "Race."

Breast-fed child--A child 3 years of age or younger at the time of the household interview who was identified by the household respondent as being breast fed currently. Breast- fed sample persons were included in the weighting process, and the survey data set includes information on breast-fed children as discussed in section 7.6.2, "Breast-fed Children."

Calcium conversion factor--A factor that expresses the amount of calcium in 100 grams of a given milk product (that is, any food code beginning with "1") as a proportion of the amount of calcium in 100 grams of fluid whole cow's milk. For example, the calcium conversion factor for cheddar cheese was calculated by dividing the amount of calcium in 100 grams of cheddar cheese (721 milligrams) by the amount of calcium in 100 grams of fluid whole cow's milk (119.4 milligrams), resulting in a calcium conversion factor of 6.04. Used in calculation of calcium equivalent as described below.

Calcium equivalent--The amount, expressed in grams, of whole fluid cow's milk that has the same quantity of calcium as the reported milk product. Derived by multiplying the amount of the milk product eaten, expressed in grams, by the calcium conversion factor (see "Calcium conversion factor" above.) For example, the calcium equivalent of 2 ounces (57 grams) of cheddar cheese is calculated by multiplying 57 grams x 6.04 (the calcium conversion factor for cheddar cheese) = 344 grams. Thus, the amount of calcium in 57 grams of cheddar cheese is equal to the amount of calcium in 344 grams of whole fluid milk. Intakes of total milk and milk products may be compared among population groups using calcium equivalents to take into account the different calcium densities of milk products subgroups (for example, fluid milk and cheese) that may be used in varying proportions by the population groups. The calcium equivalent is present on record type 30 (rt30.dat) in the field CALEQ.
Carotenes--Beta-carotene and other provitamin-A carotenoids. See "Vitamin A."

Central city--See "Urbanization."

Combination--Foods combined together and consumed as a unit that were coded using two or more food codes; identified by the record type 30 (rt30.dat) fields COMBNUM and COMBTYPE. For more discussion of combinations, see sections 3.3.1, "Food coding and editing," and 3.3.8, "Combination codes."

Dietary fiber--Total dietary fiber including both the insoluble fraction (cellulose, hemicellulose, and lignin) and the soluble fraction (for example, gums in cereal grains and pectin in fruits and vegetables).

Dietary intake--See "Food intake."

Dwelling unit--House, apartment, room, or group of rooms occupied as separate living quarters, when the occupants do not live and eat with any other person in the structure and when there is direct access from the outside or through a common area or hall. Synonymous with "housing unit" as described in the definition of "households" for the 1990 Census (Baugher and Lamison-White 1996).

Eating occasion--Any report of eating or drinking by a sample person. Each change in time of eating reported on the questionnaire was considered to be a separate eating occasion.

Educational level--For each household member 15 years of age or older, the household respondent was asked to name the highest grade of formal schooling completed, starting with "kindergarten or less" and continuing in 1-grade or 1-year increments to "5 or more years of college." Formal schooling does not include trade or vocational schooling, company training, or tutoring, unless credit is given which would be accepted at a regular school or college. High school equivalency (GED) was considered equal to completing grade 12.

Employment status--For each household member 15 years of age or older, the household respondent was asked whether the person worked during the week preceding the interview and, if so, how many hours. "Work" includes any full-time or part-time activity for which money, goods, or services were received. Employment includes active duty in the armed forces. An individual was also "employed" if he or she had a job but was not actually at work that week. Full-time status equals 35 hours or more worked during the week; part-time status equals 1 to 34 hours. See

discussion of the field EMP_STAT in section 9.3, "Additional Documentation on Calculated Variables."

Ethnic origin--The screener respondent reported whether or not each household member was of Mexican/Mexican-American/Chicano, Puerto Rican, Cuban, or other Spanish or Hispanic origin.

Exercise--Sample persons 12 years of age or older were asked "How often do you exercise vigorously enough to work up a sweat?"

Female head of household--Person indicated as such by the household respondent. (Included for purposes of historical comparison.)

Folate--Total folate content; includes naturally occurring folate and added folic acid. Folate values have been updated to reflect the regulation requiring enriched grain products to include added folic acid beginning January 1998.

Food intake--All beverages (except plain water with nothing in it) and foods ingested. Does not include inedible parts of foods (such as bones, rinds, and seeds); uneaten portions of food; or vitamin, mineral, or other supplements.

Health status--Self-appraised.

Height--Self-reported.

Home food supply--Foods and beverages ingested at home (including food obtained away from home and carried home to be eaten) and food items carried from home and eaten elsewhere, such as those in picnics and packed lunches. (Included for purposes of historical comparison.) See the file formats for record type 30 (rt30.dat) fields EATHOME and EVERHOME.

Household--All persons who regularly share a house, an apartment, a room, or a group of rooms used as separate living quarters. Household membership is based on the place where a person usually lives or sleeps for 6 or more months per year and where the person is free to return at any time. Includes persons temporarily absent, such as those who were in the hospital or traveling; students who live away from the sampled dwelling unit in dormitories or sorority or fraternity housing while attending school, who are scheduled to return to the household at the end of the term, and who use the sampled dwelling unit as their permanent address; domestic or other employees who usually live and sleep at the sampled dwelling unit; boarders or roomers who usually live and sleep at the sampled dwelling unit; and persons temporarily visiting the dwelling unit who have no usual place of residence elsewhere, such as a visitor who is house hunting. Excludes former household members who live in institutions, nursing homes, convents, etc.; persons working abroad; and members of the armed forces stationed elsewhere. Excludes students who live in an off-campus dwelling unit while attending school, persons who take their meals in the household but usually lodge or sleep elsewhere, domestic or other employees who live in an adjacent but separate dwelling unit, and persons temporarily visiting the household who have a usual place of residence elsewhere to which they are free to

return at any time. Excludes noninstitutional group quarters of nine or more unrelated persons living and eating together.

Household income--Household respondent's estimate of the total income from all sources, before taxes, of all household members for the calendar year prior to the interview. Includes income of roomers/boarders. Excludes income of live-in employees. See discussion of the field INCOME in section 9.3, "Additional Documentation on Calculated Variables."

Household member--See "Household."

Household respondent--Person who answered the household questionnaire, usually either the main meal planner/preparer or a person knowledgeable about household characteristics such as income; not necessarily a sample person.

Household size--Number of individuals in a household.

Income--Both monthly and annual household income were collected. See section 9.3, "Additional Documentation on Calculated Variables," for a discussion of income and for information on imputed incomes.

Key field--A frequently-used field (variable) included in all record types (data files). See section 7.4.1 for a list of the key fields.

Lactating female--A female household member 10 to 55 years of age identified by the household respondent as currently breast-feeding a child 3 years of age or less.

Main meal planner/preparer--Person who usually plans and/or prepares the household's meals or does the major food shopping. This person was the preferred household respondent, proxy, and assistant.

Male head of household--Person indicated as such by the household respondent. (Included for purposes of historical comparison.)

Metropolitan Statistical Area--A geographic area consisting of a large population nucleus together with adjacent communities that have a high degree of economic and social integration with that nucleus; defined by the Federal Office of Management and Budget for use in the presentation of statistics by agencies of the Federal government (USDC/BOC and APDU 1993).

Midwest--See "Region."

Niacin--Nicotinic acid and nicotinamide present in foods. Does not include potential niacin that could be converted from dietary tryptophan, a niacin precursor, in the body.

Nonmetropolitan areas--See "Urbanization."

Nonrespondent--Sample person who did not complete an interview.

Northeast-- See "Region."

Nutrient intake--Nutrient content of all foods and beverages (except plain water with nothing in it) ingested. Excludes vitamin, mineral, and other supplements.

One-day dietary recall--A recall of beverages and foods ingested during the day preceding the interview--the 24 hours from 12:00 a.m. (midnight) to 11:59 p.m.

Percentage of poverty level--Household income for the previous calendar year expressed as a percentage of the Federal poverty thresholds (Baugher and Lamison-White 1996) adjusted for inflation. See discussion of the field PCTPOV in section 9.3, "Additional Documentation on Calculated Variables."

Poverty level--See "Percentage of poverty level."

Pregnant female--Female household member 10 to 55 years of age identified by the household respondent as currently pregnant.

Proxy--Knowledgeable adult who completed the dietary recall for children under 6 years of age and other sample persons unable to report for themselves due to physical or mental limitations or because of illness. Proxy interviews were not substituted for in-person interviews with adult sample persons who were difficult for the interviewer to reach or who were nonrespondents.

Race--The screener respondent reported the race of each household member as white, black, Asian/Pacific islander, American Indian/Alaskan native, or some other race.

Recommended Dietary (or Energy) Allowances (RDA or REA)-- Levels of nutrient (or energy) intake considered by the Food and Nutrition Board of the National Academy of Sciences to be adequate to meet the known nutritional needs of practically all healthy individuals (NRC/FNB 1989). In a population group whose usual intake approximates or exceeds the RDA, the likelihood of deficiency is small (NRC/FNB 1989).

Region--An area of the United States as defined by the U.S. Department of Commerce for the 1990 Census of Population. The four census regions and their States are as follows:

(1) Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont;

(2) Midwest: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin;

(3) South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia;

(4) West: Alaska, Arizona, California, Colorado, Hawaii, Idaho,

Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming.

Responding sample person--Household member who was selected to participate in the individual intake component of the survey and who provided at least 1 day of dietary intake data.

Retinol equivalents--See "Vitamin A."

Sample person--Household member selected to participate in the individual intake component of the survey.

Sampling weights--Weights required in analysis to compensate for variable probabilities of selection, differential nonresponse rates, and possible deficiencies in the sampling frame. See section 5, "SAMPLING WEIGHTS."

Screening respondent--Household member 18 years of age or older who answered the screening questionnaire (screener).

Source of food--The place where each food or beverage (or most of the ingredients of a mixed item) was obtained, for example, from a store, restaurant, vending machine, or Meals on Wheels; as a mail order purchase; or as a gift from someone else. This information was provided by the sample person, proxy, or assistant.

South--See "Region."

Suburban areas--See "Urbanization."

Supplements--Vitamins and minerals ingested in a form other than in food or beverage. Not included in food and nutrient intake data.

Urbanization--Based on Metropolitan Statistical Areas (MSA's) defined by the Federal Office of Management and Budget (OMB) using information and recommendations provided by the U.S. Bureau of the Census. The three levels of urbanization are as follows:

(1) MSA, central city: All OMB-designated central cities, as defined by their corporate city limits, located in 1990 MSA's. These are primarily the urban cores of the MSA's. Although some MSA's contain no central city, most MSA's contain one or more.

(2) MSA, outside central city: The remaining counties or county equivalents located in MSA's.

(3) Non-MSA: All counties or county equivalents that were located outside of 1990 MSA's.

Vitamin A--Vitamin A activity derived from both preformed vitamin A (retinol) and provitamin A carotenoids. Values are expressed as international units (IU) and as micrograms of retinol equivalents (RE). One IU equals 0.3 micrograms of retinol, 0.6 micrograms of beta-carotene, or 1.2 micrograms of other carotenoids having vitamin A

activity. One RE equals 1 microgram of retinol, 6 micrograms of beta-carotene, or 12 micrograms of other provitamin A carotenoids.

Vitamin E--Vitamin E activity derived from alpha-, beta-, and gamma-tocopherol and alpha-tocotrienol. Values are expressed as milligrams of alpha-tocopherol equivalents. One alpha-tocopherol equivalent equals 1 milligram of alpha-tocopherol, 2 milligrams of beta-tocopherol, 10 milligram of gamma-tocopherol, or 3.3 milligrams of alpha-tocotrienol.

Weight--Self-reported.

Weighting factors--See "Sampling weights."

West--See "Region."

White--See "Race."

3.6 References

Baugher, E. and L. Lamison-White. 1996. Poverty in the United States: 1995. U.S. Bureau of the Census, Current Population Reports, Series P60-194 (see appendix A).

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.

DeMaio, T.M., S. Ciochetto, and W.L. Davis. 1993. Research on the Continuing Survey of Food Intakes by Individuals. In: 1993 Proceedings of the Section on Survey Research Methods, papers presented at American Statistical Association, San Francisco, August 8-12, 1993, and American Association for Public Opinion Research, St. Charles, IL, May 20-23, 1993, vol. II, pp. 1021-1026. American Statistical Association, Alexandria, VA.

DHHS (U.S. Department of Health and Human Services). 1998. 1998 Poverty Guidelines. Federal Register 63(36):9235-9238.

DHHS/FDA (U.S. Department of Health and Human Services, Food and Drug Administration). 1996. Food standards: Amendment of standards of identity for enriched grain products to require addition of folic acid. Final rule. Code of Federal Regulations, Title 21, Parts 136, 137, and 139. Federal Register 61(44--March 5, 1996):8781-8797.

Federal Register. October 22, 1990. Public Law 101-445. National Nutrition Monitoring and Related Research Act of 1990.

Gebhardt, S.E. 1992. Imputing nutrient values. In: 17th National Nutrient Data Bank Conference Proceedings, Baltimore, MD, June 7-10, 1992, pp. 143-153. International Life Sciences Institute Press, Washington, DC.

Gebhardt, S.E., J.M. Holden, and C.S. Davis. 1990. Preliminary report: A nationwide study of the selenium content of U.S. foods. In: M.R. Stewart, ed., Proceedings of the 15th National Nutrient Databank Conference, Blacksburg, VA, June 3-6, 1990, pp. 43-52. The CBORD Group, Inc., Ithaca, NY.

Guenther, P.M., T.J. DeMaio, L.A. Ingwersen, et al. 1995. The multiple-pass approach for the 24-hour recall in the Continuing Survey of Food Intakes by Individuals (CSFII) 1994-96. International Conference on Dietary Assessment Methods, Boston, January 23, 1995.

Holden, J.M., S. Gebhardt, C.S. Davis, and D.G. Lurie. 1991. A nationwide study of the selenium contents and variability in white bread. Journal of Food Composition and Analyses 4:183-195.

Marcoe, K.K., and D.B. Haytowitz. 1993. Estimating nutrient values of mixed dishes from label information. Food Technology 47(4):69-75.

Martin, J.I., W.O. Landen, Jr., A.M. Soliman, et al. 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: 16th National Nutrient Databank Conference Proceedings, San Francisco, June 17-19, 1991, pp. 129-132. The CBORD Group, Ithaca, NY.

NAS/NRC (National Academy of Sciences, National Research Council). 1993. Pesticides in the diets of infants and children. National Academy Press, Washington, DC.

NRC/FNB (National Research Council, Food and Nutrition Board). 1989. Recommended Dietary Allowances, 10th ed. National Academy Press, Washington, DC.

Perloff, B.P. 1985. Recipe calculations for NFCS database. In: S. Murphy and D. Rauchwarter, eds., Proceedings of Tenth National Nutrient Data Bank Conference, San Francisco, July 22-24, 1985, pp. 11-21. National Technical Information Service, Springfield, VA.

Perloff, B.P., R.L. Rizek, D.B. Haytowitz, et al. 1990. Dietary intake methodology II: USDA's nutrient data base 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.

Tippett, K.S., and Y.S. Cypel (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, Continuing Survey of Food Intakes by Individuals 1994-96, Nationwide Food Surveys Rep. No. 96-1. (Included on Disk 1 in \pdffiles\dor.pdf.)

USDA/ARS (U.S. Department of Agriculture, Agricultural Research Service). 1996. USDA Nutrient Database for Standard Reference, Release 11. ARS Nutrient Data Laboratory Home Page <http://www.nal.usda.gov/fnic/foodcomp>.

USDA/DHHS (U.S. Departments of Agriculture and Health and Human Services). 1990. Nutrition and your health, Dietary Guidelines for Americans. Home and Garden Bulletin No. 232, 3rd ed.

USDC/BOC (U.S. Department of Commerce, Bureau of the Census). 1993. Current Population Survey, March 1992. Conducted for the Bureau of Labor Statistics. Machine readable data file.

USDC/BOC and APDU (U.S. Department of Commerce, Bureau of the Census, and Association of Public Data Users). 1993. A guide to State and local census geography. U.S. Department of Commerce, Bureau of the Census, 1990 CPH-I-18.

[Note: This Appendix includes information from the documentation section, Chapters 5 and 6, of the following CD-ROM: U.S. Department of Agriculture, Agricultural Research Service. 2000. Continuing Survey of Food Intakes by Individuals 1994-96, 1998. NTIS Accession No. PB2000-500027.]

5. SAMPLING WEIGHTS

5.1 Introduction to Sampling Weights Discussion

In general, the analysis of data from surveys having complex designs requires the use of sampling weights to compensate for variable probabilities of selection, differential nonresponse rates, and possible deficiencies in the sampling frame. The CSFII/DHKS 1994-96 data set release contained sets of sampling weights appropriate for use in the analysis of the annual data sets as well as sampling weights appropriate for the analysis of the 3 years combined. In addition to the sampling weights provided with the 1994-96 (3-year) release, this combined CSFII 1994-96, 1998 (4-year) release provides sampling weights appropriate for use in the analysis of the 4-year data set and sampling weights for use in the analysis of the CSFII 1998 data separately. Table 5-1 provides counts of children in the combined 1994-96, 1998 data set. Tables 5-2 and 5-3 summarize the sampling weight sets.

Guidance in the choice of appropriate sampling weights and in the application of the reporting guidelines followed by the USDA in the preparation of general statistical reports is provided in Section 5.2 below. Sections 5.3 through 5.5 document the construction of the weights. Section 5.6 discusses variance estimation procedures appropriate with the analysis of data from this data set.

Sampling weights appropriate for the analysis of the combined CSFII 1994-96 data at the household level were made available in the spring of 1999. Those 3-year household weights have been included with this release. Section 5.7 provides the documentation that accompanied the original release of the household weights.

Although the Diet and Health Knowledge Survey (DHKS) was not administered for CSFII 1998, the DHKS data records and sampling weights from 1994-96 have been included with this release on record type 50. Furthermore, sampling weights designed for the analysis of household-level data from both the 3-year and 4-year data sets have been included with this release.

Table 5-1. Number of children providing intakes in CSFII 1994-96 and CSFII 1998

	1994-96	1998	Total
Under 1 year old	376	1,175	1,151
1 year old	711	373	1,084
2 years old	705	402	1,107
3 years old	492	1,344	1,836
4 years old	511	1,348	1,859
5 years old	475	409	884
6 years old	256	343	599
7 years old	233	71	304
8 years old	236	53	289
9 years old	258	41	299
0 - 9 years old	4,253	5,559	9,812

Table 5-2. Final sampling weights* provided with the combined CSFII 1994-96, 1998 data release

	Annual#	1994-96 (3-year)	1994-96, 1998 (4-year)	Available on record types
One day of intake	WTA_DAY1	WT3_DAY1	WT4_DAY1	RT20, RT25, RT30, RT35, RT40
Two days of intake	WTA_2DAY	WT3_2DAY	WT4_2DAY	RT20, RT25, RT30, RT35, RT40
Household level	Not provided	WT3_HH	WT4_HH	RT15
DHKS@	WTA_DHK	WT3_DHK	Not provided	RT50
DHKS with two days of intake@	WTA_DHK2	WT3_DHK2	Not provided	RT50

* Columns 3, 4, and 5 give the names of the sampling weights where weights are available. Jackknife replicate weights for variance estimation are also provided for each of these sets of sampling weights (see section 5.6.2, "Estimation of Sampling Errors").

\# These weights are appropriate for separate analysis of years 1994, 1995, 1996, or 1998.

@ DHKS sampling weights are only applicable for 1994, 1995, and 1996. The DHKS was not administered for the CSFII 1998.

Table 5-3. Summary of sampling weights included in the combined CSFII 1994-96, 1998 release

Ages 0-19

		Sample size	Sum of weights	CV*	VIF#
Day 1:	1994	2,298	76,641,610	64.83	1.42
	1995	1,981	77,498,715	69.98	1.49
	1996	1,952	78,316,471	59.52	1.35
	1998	5,559	40,134,208	209.16	5.37
	4-year	11,790	77,485,571	111.23	2.24
	3-year	6,231	77,485,604	62.06	1.39
2-day:	1994	2,223	76,641,600	76.07	1.58
	1995	1,904	77,498,713	80.44	1.65
	1996	1,853	78,316,485	73.44	1.54
	1998	5,304	40,134,206	213.58	5.56
	4-year	11,284	77,485,611	122.57	2.50
	3-year	5,980	77,485,635	74.29	1.55

Ages 20+

		Sample size	Sum of weights	CV*	VIF#
Day 1:	1994	3,291	182,865,657	58.47	1.34
	1995	3,345	184,451,592	69.42	1.48
	1996	3,236	185,917,776	51.72	1.27
	3-year	9,872	184,411,673	59.68	1.36
2-day:	1994	3,088	182,865,609	70.04	1.49
	1995	3,168	184,451,679	86.34	1.75
	1996	3,067	185,917,706	63.91	1.41
	3-year	9,323	184,411,625	73.78	1.54

* CV is the population coefficient of variation for the sampling weights (standard deviation / mean) expressed as a percentage

\# The variance inflation factor, VIF = 1 + (CV / 100)**2

-- continued

Table 5-3. Summary of sampling weights included in the combined CSFII 1994-96, 1998 release -- continued

All ages

		Sample size	Sum of weights	CV*	VIF#
Day 1:	1994	5,589	259,507,267	65.80	1.43

	1995	5,326	261,950,307	72.14	1.52
	1996	5,188	264,234,247	56.76	1.32
	1998	5,559	40,134,208	209.16	5.37
	4-year	21,662	261,897,244	91.40	1.84
	3-year	16,103	261,897,277	64.05	1.41
2-day:	1994	5,311	259,507,209	77.59	1.60
	1995	5,072	261,950,392	87.73	1.77
	1996	4,920	264,234,191	69.11	1.48
	1998	5,304	40,134,206	213.58	5.56
	4-year	20,607	261,897,236	104.52	2.09
	3-year	15,303	261,897,260	77.74	1.60

Household

	Sample size	Sum of weights	CV*	VIF#
4-year	12,364	98,574,787	85.67	1.73
3-year	8,067	98,574,761	45.88	1.21

* CV is the population coefficient of variation for the sampling weights (standard deviation / mean) expressed as a percentage

\# The variance inflation factor, VIF = 1 + (CV / 100)**2

5.2 Guidance for Sampling Weights and Reporting

5.2.1 Sampling weight guidance

As noted above, it is generally necessary to use sampling weights in the analysis of data from surveys having complex designs. This data release contains a variety of sets of sampling weights designed to be used in various situations. The choice of which sampling weight to use was straightforward with the CSFII/DHKS 1994-96 release. Day 1 weights are used whenever day 1 intakes are analyzed and generally whenever analyzing CSFII data at the person level. The 2-day weights need to be used when a subset of the sample is used that is restricted to 2-day respondents. The 3-year weights are generally used if all 3 years of data are being analyzed. The annual weights are generally used if the individual years are analyzed separately. However, results do not tend to change very much if the annual and 3-year sampling weights are used interchangeably because sampling procedures and the target population were the same in each of 1994, 1995, and 1996.

With the CSFII 1998 the situation changes somewhat. Because only children 9 years old or younger were targeted in 1998 and relatively few of those children were in the age group 7-9 years, the weights constructed for use with the CSFII 1998 and the combined CSFII 1994-96, 1998 sample have several features that should be noted. Among these features are:

1) The CSFII 1998 weights are more variable than the other annual weights. This is mainly due to the unequal distribution of ages in the 1998 sample as seen in Table 5-1. It should be noted that the weights for a subset of the CSFII 1998 sample that is more equally distributed across ages, such as children 1-5 years or children 7-9 years, are considerably less variable.

2) The combined CSFII 1994-96, 1998 weights are more variable than the CSFII 1994-96 weights for children 0-9 years old. This is due to the more variable CSFII 1998 weights and to the difference in distribution of ages between the two samples.

3) For convenience, there are sampling weights for adults 20 years and older in the set of 4-year weights. These are exactly the same weights found in the 3-year weight set. Adults were not sampled in 1998.

4) Although no data was collected for persons 10-19 years in the CSFII 1998, the 4-year weights for these persons are slightly different than the 3-year weights. This is because the final calibration process for the 4-year weights was done for persons 6-19 years as a group. The calibration adjustments necessary for the 4-year weights for 6-19 year olds differed from the adjustments necessary for the 3-year weights due to the inclusion of children 6-9 years from CSFII 1998.

It will be the USDA's convention to use the 4-year combined CSFII 1994-96, 1998 weights whenever a statistical presentation uses data from the CSFII 1994-96, 1998 data set and displays statistics for children 9 years and under. For statistical presentation of data for persons 10-19, years USDA also recommends the usage of the 4-year combined weights for the reasons explained in item (4) above.

Furthermore, the USDA recommends caution in analyzing the CSFII 1998 by itself. Unlike the annual samples of CSFII 1994-96, the CSFII 1998 is a supplemental sample,

designed to be merged with the CSFII 1994-96 in order to increase the overall sample size of children of certain ages. The CSFII 1998 sampling weights provide some calibration of the CSFII 1998 sample to the population of 0 to 9 year olds but the fact that there are proportionately fewer children 7 to 9 years in the sample than children of other ages might affect analyses of groups that include both children 7 years or older and younger children.

5.2.2 Reporting guidance

It is the USDA's convention to follow guidelines derived from a report of the Life Sciences Research Office (FASEB/LSRO 1995) in identifying or flagging estimates of means, percentages, and percentiles presented in general reports that might be less statistically reliable than other estimates due to small cell size or high relative variability. The guidelines for determining such estimates take into account the complex sample design of a survey and the procedures used to weight the data by specifying the use of a broadly calculated design effect. The design effect is a measure of the variability introduced into an estimate by these factors.

Each estimate has a unique design effect. A "broadly calculated" design effect might be an average of design effects among related statistics or population groups. For the convenience of having a single measure of this type of variability, it is the USDA's convention to use a variance inflation factor (VIF) in this role in the presentation of general statistical tables. A VIF is solely a function of the sampling weights. Variance inflation factors for the CSFII 1994-96 and CSFII 1998 sampling weight sets are presented in table 5-3 above.

Prior to the release of data from the CSFII 1998, the USDA has used by convention a single VIF, derived from the weights of individuals of all ages, in the presentation of statistics from USDA survey data. This convention will be changed for the analysis of data from the combined CSFII 1994-96, 1998 sample. Whenever a statistical presentation is based on data for persons under 20 years of age from both CSFII 1994-96 and CSFII 1998, a VIF based on the 4-year weights on persons 0-19 years will be used in applying the reporting guidelines. If statistics for adults are also provided, a VIF based on the weights of persons 20 years and older will be used. If statistics for persons from both groups are presented, for example, a table showing statistics for various age groups including an all-ages group, the VIF for persons 0-19 years will be used. The VIFs that would be used in such reports are:

Day 1, 0-19:	2.24
Day 1, 20+:	1.36
Day 1, all:	2.24
2-day, 0-19:	2.50
2-day, 20+:	1.54
2-day, all:	2.50

The reporting guidelines generally followed are:

1) An estimated mean is flagged when it is based on a cell size of less than 30 times the average design effect (VIF) or when its coefficient of variation (cv) is equal to or greater than 30 percent. The cv is the ratio of the estimated standard error of the mean to the estimated mean, expressed as a percentage. Note that the cv statistic refereed to here is relative to the estimate of the mean, hence the use in the numerator of the standard error rather than the standard deviation as used in the calculation of the (population) coefficient of variation shown in Table 5-3.

2) An estimated proportion (percent) that falls above 25 percent and below 75 percent is flagged when it is based on a cell size of less than 30 times the average design effect (VIF) or when the cv is equal to or greater than 30 percent.

   An estimated proportion of 25 percent or lower or 75 percent or higher is flagged when the smaller of np and n(1-p) is less than 8 times the average design effect, where "n" is the cell size on which the estimate is based and "p" is the proportion expressed as a fraction.

3) Estimated percentiles are flagged according to rules that parallel the cell size rules applied to proportions (guideline 2). Estimated percentiles inside the 25 to 75 range are flagged when the cell size is less than 30 times the average design effect. Estimates of the 25 and lower percentiles are flagged when the cell size is less than 8 times the average design effect divided by p, where p is the level of the percentile expressed as a fraction. Estimates of the 75 and higher percentiles are flagged when the cell size is less than 8 times the average design effect divided by 1 - p.

5.3 CSFII 1998 (Annual) Sampling Weights

5.3.1 CSFII 1998 weighting design

The approach used in weighting the CSFII 1998 data followed the approach used in weighting the 1994, 1995, and 1996 person-level data. These annual data sets were weighted separately in the following steps. First, a base weight equal to the reciprocal of the probability of selection was assigned to each sample person. The base weights were then adjusted for nonresponse within weighting classes defined by variables that were determined to be correlated with response rates. Finally, the nonresponse-adjusted weights were ratio adjusted to population estimates from the March Current Population Survey (CPS) of the appropriate year (USDC/BOC 1994, 1995, 1996, 1998) to compensate for random variation in the observed sample counts and possible undercoverage of certain groups in the area sample frame. Two sets of weights were constructed for the CSFII 1998: a set for sample persons who completed the day-1 interview and a set for sample persons who provided 2 days of intake.

5.3.2 Base weights

The base weight associated with a sample person is the reciprocal of the overall probability of including that person in the survey. For the CSFII 1998, sample persons were selected through a complex multistage sample design involving the selection of primary sampling units (PSUs), area segments within PSUs, households within segments, and finally persons (sample persons) within households. Consequently, the following components were required to calculate the overall probabilities of selection:

1. The probability of selecting the PSU.

2. The probability of selecting the segment within the PSU.

3. The probability of selecting the household within the segment.

4. The probability of selecting an eligible sample person from within the household.

For any sample person, the product of these four factors is the probability of being selected for the CSFII.

5.3.3 CSFII 1998 nonresponse adjustments

Not all sample persons completed an intake interview. To compensate for this, the following procedures were used to adjust the sample person base weights. First the weights were adjusted for screening nonresponse. These adjustments were made within classes created by grouping segments by census region, MSA status, minority status (percent of the population that was black or Hispanic), and quarter of the year of field operations. Within each class, the base weight of each eligible sample person was increased by a factor corresponding to the screener nonresponse rate within the class.

These screener nonresponse-adjusted weights were then adjusted again to account for person nonresponse. A different set of weighting classes was used for this adjustment. These classes were defined by income level, age, sex, census region, MSA status, quarter of the year of field operations, and minority status of the segment. The result of this step was a set of nonresponse-adjusted base weights for responding sample persons.

5.3.4 CSFII 1998 population adjustments

Lastly, the nonresponse-adjusted weights were calibrated using an iterative process called "raking ratio weighting" so that the sum of the final weights equaled the corresponding 1998 March CPS population totals (USDC/BOC 1998) within cells defined by the following variables:

1. Sex
2. Age group (0-2, 3-5, 6-9)
3. Home ownership
4. Season of intake (winter, spring, summer, fall)
5. Day of week of intake
6. Census region
7. MSA status (metropolitan/nonmetropolitan)
8. Household income as percentage of poverty level (using the appropriate poverty thresholds)
9. Household received food stamps in past 12 months
10. Number of persons in the household 18 and older
11. Presence in household of children under 6 years
12. Presence in household of children 6 to 17 years
13. Presence of female head of household 40 years or younger and no one in the household under 18 years
14. Employment status (for children this was the status of the female head, or if there was no female head, the male head of household)
15. Race (black or nonblack)
16. Ethnic origin (Hispanic or non-Hispanic)

Table 5-4 shows the adjustments necessary for calibration for the weighting class age 0-5 years and Table 5-6 shows the same for the weighting class age 6-9 years. Column 1 provides the number of children with the various characteristics. Column 2 provides the weighted percentages of the persons within the weighting class in each of the categories using the nonresponse-adjusted sampling weights. Column 3 shows the target percentage from the CPS, which is also the weighted percentage for the sample using the final, calibrated weights.

Table 5-4. Children 5 years and younger: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments, and population targets, day 1, CSFII 1998

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Total	5,051	100.0	100.0
Age/sex			
Male			
0-2	950	39.5	25.1
3-5	1,573	21.9	26.1
Female			
0-2	1,000	18.3	23.8
3-5	1,528	20.3	25.0
Home ownership			
Home owned	2,828	60.1	58.3
Home not owned	2,223	39.9	41.7
Season of intake			
Winter	1,166	25.7	25.0
Spring	1,240	27.1	25.0
Summer	1,667	29.6	25.0
Fall	978	17.6	25.0
Day of week of intake			
Sunday	961	18.9	14.3
Monday	786	15.8	14.3
Tuesday	771	15.3	14.3
Wednesday	638	12.7	14.3
Thursday	580	11.2	14.3
Friday	787	15.6	14.3
Saturday	528	10.5	14.3
Census region			
Northeast	906	17.9	17.3
Midwest	1,115	22.4	23.9
South	1,661	32.8	34.3
West	1,369	27.0	24.5
MSA status			
MSA (metropolitan)	4,134	80.6	81.7
Non-MSA	917	19.4	18.3

--continued

Table 5-4. Continued.

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Household income as percentage of poverty level			
0-75%	895	14.9	14.6
76-130%	924	15.3	13.4
131-300%	1,656	34.9	34.8
Over 300%	1,576	35.0	37.3
Household received food stamps in past 12 months			
Yes	1,073	17.4	17.8
No	3,978	82.6	82.2
Presence in household of persons 18 and older			
Exactly 1	655	11.7	13.5
Exactly 2	3,656	74.5	74.0
Other than 1 or 2	740	13.8	12.5
Presence in household of children 6-17			
Children 6-17	2,326	44.8	44.5
No children 6-17	2,725	55.2	55.5
Employment status of female head of household (or male head if there is no female head)			
Have job	2,644	53.4	58.4
Do not have job	2,407	46.6	41.6

--continued

Table 5-4. Continued.

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Race			
Black	749	13.6	15.8
Non-black	4,302	86.4	84.2
Ethnic origin			
Hispanic	901	16.1	17.7
Non-Hispanic	4,150	83.9	82.3

* Calculated using 1998 Current Population Survey data (USDC/ BOC 1998) except for the variables "season of intake" and "day of week of intake." Since the goal of the CSFII was to estimate behavior on an average day, each day of the week received an equal value of 14.3 percent, and each season received a value of 25 percent.

Table 5-5. Persons 6 to 9: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments, and population targets, day 1, CSFII 1998

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Total	508	100.0	100.0
Sex			
Male	279	56.8	50.9
Female	229	43.2	49.1
Home ownership			
Home owned	304	66.2	65.2
Home not owned	204	33.8	34.8
Season of intake			
Winter	134	23.5	25.0
Spring	126	25.9	25.0
Summer	156	31.1	25.0
Fall	92	19.5	25.0
Day of week of intake			
Weekend (Fri - Sun)	228	45.4	42.9
Weekday (Mon - Thr)	280	54.6	57.1
Census region			
Northeast	77	12.6	18.6
Midwest	110	23.6	23.0
South	187	33.6	34.4
West	134	30.2	24.0
MSA status			
MSA (metropolitan)	411	82.1	80.7
Non-MSA	97	17.9	19.3

--continued

Table 5-5. Continued.

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Household income as percentage of poverty level			
0-75%	101	16.1	13.8
76-130%	85	11.5	12.1
131-300%	149	33.0	35.5
Over 300%	173	39.4	38.5
Household received food stamps in past 12 months			
Yes	99	17.8	16.2
No	409	82.2	83.8
Presence in household of persons 18 and older			
Exactly 1	75	12.6	18.0
Exactly 2	365	74.6	69.7
Other than 1 or 2	68	12.8	12.3
Presence in household of children under 6			
Children under 6	257	45.5	41.4
No children under 6	251	54.5	58.6
Employment status of female head of household (or male head if there is no female			
Have job	278	54.3	63.7
Do not have job	230	45.7	36.3

--continued

Table 5-5. Continued.

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Race			
Black	77	14.7	16.4
Non-black	431	85.3	83.6
Ethnic origin			
Hispanic	81	12.7	15.3
Non-Hispanic	427	87.3	84.7

* Calculated using 1998 Current Population Survey data (USDC/ BOC 1998) except for the variables "season of intake" and "day of week of intake." Since the goal of the CSFII was to estimate behavior on an average day, each day of the week received an equal value of 14.3 percent, and each season received a value of 25 percent.

5.4 CSFII 1994-96, 1998 (4-Year) Combined Person-Level Sampling Weights

5.4.1 Introduction to person-level sampling weights discussion

Although the CSFII 1998 was a nationally representative sample of children 9 years of age and younger, it was primarily intended to serve as a supplement to the sample of children in the CSFII 1994-96. A composite estimation approach was used to combine the CSFII 1994-96 and CSFII 1998 samples. Under this approach, the combined estimator xcomp, is considered to be a linear combination of the corresponding CSFII 1994-96 and CSFII 1998 estimates, i.e.,

```
    xcomp = a * x[94-96] + (1 - a) * x[1998],
```

where a is a constant between 0 and 1.

Assuming that x[94-96] and x[1998] are both unbiased estimates, the composite estimate, x[4-year], will also be unbiased for any value of a. The approximately optimal value of a, i.e., the value that minimizes the variance of x[4-year], is a function of the effective sample sizes of the CSFII 1994-96 and the CSFII 1998:

```
    a = eff[94-96]   / (eff[94-96] + eff[1998])
```

where eff[94-96] = n[94-96] / (1 + cv[94-96:w]**2),

n[94-96] = the actual CSFII 1994-96 sample size,

cv[94-96:w]**2 = the square of the coefficient of variation (expressed as a percentage) of the CSFII weights, and

eff[1998] is similarly defined with the CSFII 1998 sample size and weights.

The factors a and (1 - a) are known as compositing factors and were computed by sex and age group for the person-level weights.

5.4.2 Day 1 person-level weights

The nonresponse-adjusted day 1 CSFII weights described in section 5.3.3 were recalibrated to the corresponding 1994-96 CPS population totals. This was done so that the CSFII weights would be consistent with the previously computed CSFII 1994-96 (3-year) weights. The recalibration of the CSFII weights was done separately for (1) children age 5 years or younger and (2) children 6-9 years of age. The procedures used for calibration were exactly the same as those described in section 5.3.4 except that the 1994-96 CPS totals were used as control totals.

Next, the compositing factors a and (1 - a) were computed using the CSFII 1994-96 weights and the recalibrated CSFII weights by sex / age groups. Table 5-6 shows the day 1 compositing factors.

Next, the CSFII 1998 sample was combined with the CSFII 1994-96 sample by applying the appropriate CSFII compositing factor (1 - a) to each recalibrated CSFII 1998 day 1 weight and by applying the appropriate CSFII compositing factor a to each CSFII 1994-96 day 1 weight. This was done for all children age 9 years or younger in the combined sample.

Finally, the penultimate combined weights described in the above paragraph were calibrated one final time to the March 1994-96 CPS totals along the dimensions used in the original calibration of the CSFII 1994-96 day 1 weights. This final calibration process was done separately for children 5 years of age and younger and for persons 6 - 19 years of age.

Table 5-6. Compositing factors for children age 9 and under completing the CSFII day 1 Intake

Sex	Age	1994-96 sample size	1998 sample size	Total sample size	1994-96 compositing factor (a)	1998 compositing factor (1-a)
Male	< 1	184	576	760	0.22	0.78
	1	362	174	536	0.70	0.30
	2	353	200	553	0.67	0.33
	3	251	687	938	0.28	0.72
	4	244	670	914	0.27	0.73
	5	246	216	462	0.54	0.46
	6	125	184	309	0.45	0.55
	7-9	383	95	478	0.83	0.17
Female	< 1	192	599	791	0.22	0.78
	1	349	199	548	0.67	0.33
	2	352	202	554	0.67	0.33
	3	241	657	898	0.32	0.68
	4	267	679	946	0.30	0.70
	5	229	192	421	0.59	0.41
	6	131	159	290	0.50	0.50
	7-9	344	70	414	0.86	0.14
Total		4,253	5,559	9,812		

5.4.3 Two-day person-level weights

The procedure followed in constructing the day 1 combined weights was followed in constructing the combined two-day weights. The two-day CSFII 1998 weights were recalibrated to the 1994-96 CPS population totals, compositing factors were computed based on both the recalibrated CSFII 1998 two-day weights and the CSFII 1994-96 two-day weights by sex and age groups, penultimate combined weights were created by applying the appropriate compositing factors to the appropriate weights, and a final raking procedure was used to calibrate the penultimate weights.

5.5 CSFII 1994-96, 1998 (4-Year) Combined Household-Level Sampling Weights

5.5.1 Introduction to household-level sampling weights discussion

To permit calculation of household-level estimates for items collected in the household interview (e.g., amount and sources of income, general information about food shopping practices, the amounts spent on food, source of cooking and drinking water, food stamp eligibility), a set of household weights for analysis of the combined CSFII 1998 and CSFII/DHKS 1994-96 data sets were computed. The procedures followed in constructing these household-level weights were similar to those used in constructing the CSFII 1998 and combined CSFII 1994-96, 1998 person-level weights. First, household-level weights were constructed for the CSFII 1998 by adjusting a base weight for nonresponse and then calibrating the nonresponse adjusted weights to population totals. Secondly, a compositing approach was used to combine the CSFII 1998 and the CSFII 1994-96. The construction of the CSFII 1994-96 household weights as documented for the release of that sampling weight set is included in this section as section 5.7.1.

5.5.2 CSFII 1998 Household Base Weights

The first step was to assign a base weight to each responding CSFII 1998 household that is equal to the reciprocal of the probability of retaining the household for the household interview. For the CSFII 1998 (and also for the CSFII 1994-96), only those households with eligible SPs were eligible for the household interview. Thus, the probability of including a household in the study was equal to the probability that any of its members was selected for the intake interviews. Under the procedures used to select persons for the CSFII 1998, the probability of selecting a household for the household interview is equal to maximum probability of selection of the SPs in the household. Hence, the base weight for the I-th sampled household was computed from the formula:

$$w = \min \{w[1], w[2], \ldots, w[n]\},$$

where w[1], w[2], ..., w[n] are the corresponding base weights of the SPs in the household. In general, the household base weights varied by quarter, as well as within quarter depending on the composition of the household. In particular, households with children under 1 or 3-4 years of age had considerably smaller weights (larger probabilities of selection) than households where the only children were between 7 and 9 years of age.

5.5.3 Household-level nonresponse Adjustments

The procedures used for nonresponse adjustments followed those used in constructing the CSFII 1994-96 household weights and were essentially as follows. First, the base weights were adjusted for screening nonresponse within classes defined by Census region, MSA status, minority status (percent of the population that was black or Hispanic), and quarter of field operations. Within each class, the base weight of each eligible sample person was increased by a factor corresponding to the screener nonresponse rate within the class.

Next, the screener-adjusted weights were adjusted to account for household nonresponse. The weighting classes used for this adjustment were defined by income level, Census region, MSA status, and minority status of the segment. Note that for the purpose of weighting, those households that contained at least one sample person who completed at least one intake interview were considered to be "respondents" regardless of whether a household interview was completed.

5.5.4 Household-level population adjustments

Lastly, the nonresponse-adjusted weights were calibrated using the same iterative process called "raking ratio weighting" used in calibrating the person-level weights so that the sum of the final weights equaled the corresponding 1994-96 March CPS population totals (USDC/BOC 1994, 1995, 1996). Since the CSFII 1998 was restricted to households with children 9 years of age or younger (i.e., households without children 9 years or younger had no chance of selection for the CSFII 1998), the totals were only for households with children 9 years of age or younger. Cells defined by the following variables were used:

1. Home ownership and age of the head of household
2. Season of household interview (winter, spring, summer, fall)
3. Day of week of household interview
4. Census region
5. MSA status (metropolitan/nonmetropolitan)
6. Household income as percentage of poverty level (using the appropriate poverty thresholds)
7. Household received food stamps in past 12 months
8. Presence in household of persons 18 and older
9. Presence in household of children under 6 years
10. Presence in household of children 6 to 17 years
11. Presence of female head of household 40 years or younger and no one in the household under 18 years
12. Employment status of the head of household
13. Race (black or nonblack) of the head of household
14. Ethnic origin (Hispanic or non-Hispanic) of the head of household
15. Household size

5.5.5 Combined CSFII 1994-96, 1998 household samples

The same compositing approach used to combine the person-level samples was used in combining the household samples. Compositing factors were computed using the CSFII 1994-96 household weights and the CSFII 1998 household weights by income / household composition groups. Table 5-7 shows the household compositing factors. Note that the choice of the household composition grouping "households with children 7-9 years of age only" followed from the design of the CSFII 1998, which selected a proportionately small group of 7-9 year olds, resulting in some large CSFII 1998 household weights for such households. Using this group for compositing purposes reduced the impact of these large weights when the samples were combined.

Next, the CSFII 1998 sample was combined with the CSFII 1994-96 sample by applying the appropriate CSFII 1998 compositing factor (1 - a) to each CSFII 1998 household weight and by applying the appropriate CSFII 1994-96 compositing factor a to each CSFII 1994-96 household weight.

Finally, these penultimate combined weights were calibrated one final time to the March 1994-96 CPS totals along the dimensions specified above. Unlike the calibration of the CSFII 1998-only household sample, this time the population totals represented all U.S. households. The same cells listed in section 5.5.4 were used.

Table 5-7. Compositing factors for CSFII households with children 9 or younger

Income group	HH comp.	1994-96 sample size#	1998 sample size	Total sample size	1994-96 compositing factor (a)	1998 compositing factor (1-a)
>= 130% poverty	Children 7-9 only	299	50	349	0.88	0.12
	Others	1,796	2,813	4,609	0.51	0.49
< 130% poverty	Children 7-9 only	99	12	111	0.90	0.10
	Others	787	1,422	2,209	0.48	0.52
Total		2,981	4,297	7,278		

Households with at least one SP who completed the day 1 Intake.

5.6 Variance Estimation

5.6.1 Variance estimation fields

As described in Section 3, "Sample Design," Westat's 62 primary sampling unit (PSU) master sample was employed for both the CSFII/DHKS 1994-96 and the CSFII 1998. This sample of PSUs contains 24 PSUs selected with certainty. The remaining 38 PSUs were selected with probability proportional to size from 38 strata, 1 PSU per stratum. Area segments were then selected from each of the 62 PSUs also with probability proportional to size. The area segments were randomly allocated to the annual samples, across quarters of the year, so that the 62 PSUs were fielded at all times throughout each year. The following approach was used to create a framework of 2 sampling units per stratum to facilitate variance estimation procedures. First, 19 variance estimation strata were formed from the 38 noncertainty PSUs by pairing adjacent PSUs in the sampling frame. Each PSU within a variance estimation stratum defines what is referred to as a variance estimation unit. Next, within each of the 24 certainty PSUs, one-half of the segments were assigned to one variance estimation unit and the remaining one-half to another. Because each certainty PSU is considered to be a separate variance estimation stratum, a total of 43 variance estimation strata (each containing 2 variance estimation units) was formed by this process. See section 7.4.2, "Sampling weights and variance estimation fields," for details of identifying these variance estimation fields in the data set. This framework applies to all weighted samples, annual or combined, of the CSFII 1994-96 and CSFII 1998.

5.6.2 Estimation of Sampling Errors

Linearization method

Estimation of sampling errors may be conducted with a Taylor series linearization method using the final sampling weights described in the above sections along with the variance estimation strata and variance estimation units described in section 5.6.1. Software packages such as SUDAAN and Stata can be used to obtain estimates using the linearization method.

Jackknife replicate method

Alternatively, sampling errors may be estimated using the jackknife technique described here. The construction of jackknife replicate weights makes use of the variance estimation stratum/variance estimation unit structure described above in section 5.5.1. As an illustration of how a jackknife variance estimator can be calculated, let y denote a weighted survey estimate (for example, total fat intake) calculated using the full-sample weights. Let y(j) be the corresponding weighted estimate calculated using the j-th set of replicate weights (j = 1, 2, ..., 43). The estimated variance of y is then given by the formula

$$Var(y) = SUM\ (y(j) - y)**2,$$

where the summation extends over the 43 sets of jackknife replicate weights. Forty-three replicates were created by applying this process to each of the 43 variance estimation strata.

A jackknife replicate is created by dropping out one of the two variance estimation units from a variance estimation stratum and doubling the initial probability weights of the individuals in the other variance estimation unit in that stratum. The entire weighting process as described in the previous sections was repeated for each replicate. Individuals who were not in the current replicate subsample were assigned a corresponding replicate weight of zero. In this way, series of replicate weights were generated for each sample person or household. Together with the final, full-sample weights, these replicate weights were designed for the calculation of sampling errors.

Using a replication method to calculate sampling errors of survey-based estimates makes complicated variance estimation formulas unnecessary. The jackknife replication method used here is also designed to reflect the stratification and clustering used in the CSFII/DHKS sample design and to capture the effects of the raking ratio adjustments mentioned in the sections above.

Replicate weights are provided for use with each of the sets of sampling weights listed in Table 5-2. There are seven files altogether, found in the \jacknife directory on disk 2:

jkw4yrcs.dat	Day 1 and two-day weights for the combined CSFII 1994-96, 1998 (4-year) sample
jkwanncs.dat	Day 1 and two-day weights for annual samples (1994, 1995, 1996, 1998)
jkw3yrcs.dat	Day 1 and two-day weights for the CSFII 1994-96 combined (3-year) sample
jkw4yrhh.dat	Household weights for the combined CSFII 1994-96, 1998 (4-year) sample
jkw3yrhh.dat	Household weights for the CSFII 1994-96 combined (3-year) sample
jkwanndh.dat	DHKS and two-day DHKS weights for the annual samples (1994, 1995, 1996)
jkw3yrdh.dat	DHKS and two-day DHKS weights for the DHKS 1994-96 combined (3-year) sample

Corresponding file formats are provided in section 9.4 and SAS programs for reading the data files are provided in section 10.4.

The annual and 4-year person-level files each contain one record per CSFII respondent (21,662 records total, 5,559 from CSFII 1998). The 3-year person-level file contains one record per CSFII respondent from 1994-96 (16,103 records). The 2-day weight fields are blank for respondents providing only one day of intake. The DHKS files each contain one record per DHKS respondent in 1994-96 (5,765 records). The 2-day DHKS weight fields are blank for DHKS respondents who did not provide a second day of intake. The 4-year household-level file contains one record per CSFII household (12,364 records total, 4,297 from the CSFII 1998). The 3-year household-level file contains one record per CSFII household from 1994-96 (8,067 records). A field indicating the survey year, the full-sample sampling weights, and the variance-estimation stratum and unit are included in each file.

The replicate weighting process described above was designed and implemented by Westat, Inc., who have also created a variance estimation program, WesVarPC, which runs on computers using the Windows operating system and is available to the public at no charge. A commercial version, WesVar Complex Samples, is also available from SPSS. Information about both programs may be found at Westat's home page at <http://www.westat.com>. Note that in WesVarPC terminology, the JK2 method was used in constructing these replicate weights.

5.7 CSFII 1994-96 (3-Year) Household Sampling Weights: Original Documentation

These weights permit household level estimates using the fields that are present on household record type 15. The data contained in the record type 15 fields include household participation in programs such as WIC and Food Stamps, income and food-related expenditures, and food sufficiency. The 3-year weights, calibrated to 3-year averages of population characteristics, are intended to be used with the 3-year CSFII data set. They may be used with the annual subsets, however, as long as it is understood that the annual subsets were not calibrated to annual population characteristics. If annual totals are being estimated, the weights should each be multiplied by 3 to scale the weights appropriately. Such scaling is not necessary for the estimation of means or percentages.

5.7.1 How the 3-year household sampling weights were constructed

In general, the analysis of data from surveys having complex designs requires the use of sample weights to compensate for variable probabilities of selection, differential nonresponse rates, and possible deficiencies in the sampling frame. For the 1994-96 CSFII/DHKS, the overall probabilities of selecting sample persons were designed to vary by sex, age, and income level to meet precision goals specified by ARS. For this reason, the probability of selecting a household into the sample is directly related to the composition of the household at the time of screening. The construction of household sample weights was performed by ARS using the design developed by Westat, Inc.

Weighting design

The 3-year CSFII households were weighted in the following steps. First, a base weight equal to the reciprocal of the probability of selection was assigned to each household. The base weights were then adjusted for nonresponse at two levels within weighting classes defined by variables that were determined to be correlated with response rates. The first was a screener-level adjustment using 57 classes defined by combinations of region, quarter, MSA status, and minority status of the segment. The second was a household level adjustment using 8 classes defined by combinations of region, MSA status, minority status of segment and household income as a percentage of poverty. Finally, to compensate for random variation in the observed sample counts and possible undercoverage of some groups, the nonresponse-adjusted weights were ratio adjusted to the average population estimates from the March Current Population Surveys for 1994, 1995, and 1996.

Base weights

The base weight associated with a household is the reciprocal of the overall probability of including that household in the survey. For each year of the CSFII/DHKS, sample households were selected through a complex multistage sample design involving the selection of primary sampling units (PSUs), area segments within PSUs, and households within segments. The eligibility of households for the CSFII was determined by household income level and the sex and age of its members at the time of screening. The product of steps 1, 2, and 3 below determines the probability of selection for eligible households. Since segments were allocated for selection over the 3 years of the survey, a factor of 3 is included in probability of selecting area segments. As with the individual weights, the reciprocal of this probability is the household base weight.

1. The probability of selecting the PSU.

2. The probability of selecting the segment within the PSU.

3. The probability of selecting the household within the segment.

CSFII nonresponse adjustments

Not all households completed the household interview but all households had a member to provide a Day-1 intake. Those households that did not provide a household interview are included in the nonresponse adjustment as participating households. This was done because household eligibility and participation were determined by the presence and participation of a specific household member. There were 41 households where a Day-1 intake was completed but the household questionnaire was not. In these cases most of the household information is missing or was imputed on record type 15. Otherwise, to compensate for nonresponse, the following procedures were used to adjust the household base weights.

First the weight was adjusted for screening nonresponse. This adjustment was made within classes created by grouping households by census region (see 1994-96 CD-ROM documentation "Region" in section 14, "Glossary"), MSA status (see 1994-96 CD-ROM documentation "Metropolitan Statistical Area" in section 14, "Glossary"), minority status of area segment (high or low minority) and quarter of field operations. Within each class, the base weight of each eligible household was increased by a factor equal to the inverse of the screening rate within the class. This adjustment is the same screener adjustment made in constructing the individual sample weights.

The screener nonresponse-adjusted weight was then adjusted again to account for household nonresponse. A different set of weighting classes was used for this adjustment. A CHAID analysis was performed by ARS to determine the groupings for the household level nonresponse adjustments. The new classes were defined by income level, census region, MSA status, and minority status of the segment. Only those households which had eligible sample persons but did not complete any day 1 intakes were considered nonresponding. As in the screener nonresponse adjusted weight, this adjustment is equal to the inverse of the household response rate within the classes. The result of this step was a set of nonresponse-adjusted base weights for responding households. The nonresponse-adjusted base weight (WT_H_ADJ) is included in the weight file.

Post-stratification and population adjustments

Finally, the nonresponse-adjusted weights were calibrated using an iterative process called "raking ratio weighting" to produce final weights that sum to the average of population totals over the 3-year period of the CSFII/DHKS. The totals are from the March (1994, 1995, and 1996) Current Population Surveys (CPS). The cells used to define the totals were generally the same as those used for the individual weight reflecting household totals. Day and quarter of intake were replaced by day and quarter of the household interview. Household size was added. Race and ethnic origin variables are based on characteristics of the female head of household when present; otherwise, the male head of household.

1. Home ownership and age of the head of household
2. Season of household interview (winter, spring, summer, fall)
3. Day of week of household interview
4. Census region
5. MSA status (metropolitan/nonmetropolitan)
6. Household income as percentage of poverty level (using the appropriate poverty thresholds)
7. Household received food stamps in past 12 months
8. Presence in household of persons 18 and older
9. Presence in household of children under 6 years
10. Presence in household of children 6 to 17 years
11. Presence of female head of household 40 years or younger and no one in the household under 18 years
12. Employment status of the head of household
13. Race (black or nonblack) of the head of household
14. Ethnic origin (Hispanic or non-Hispanic) of the head of household
15. Household size

To illustrate the adjustments, table 5-8 shows, by weighting variable, the 1994-96 CSFII unweighted sample sizes, the weighted percentage distributions following nonresponse adjustments (but before calibration to population targets), and the population targets for all responding households.

Table 5-8. Unweighted household sample sizes, weighted percentage distributions following nonresponse adjustments, and population targets, CSFII 1994-96

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Total	8,067	100.0	100.0
Home ownership/age			
Home owned			
20-39	1,813	22.7	19.3
40-59	1,909	24.0	25.9
60 and older	1,598	19.8	19.6
Home not owned			
20-39	1,623	19.9	20.7
40-59	660	8.0	8.9
60 and older	464	5.6	5.6
Season of interview			
Winter	1,943	24.3	25.0
Spring	2,122	26.2	25.0
Summer	1,988	24.7	25.0
Fall	2,014	24.8	25.0
Day of week of interview			
Sunday	952	11.8	14.3
Monday	1,348	16.6	14.3
Tuesday	1,246	15.5	14.3
Wednesday	1,226	15.3	14.3
Thursday	968	12.1	14.3
Friday	919	11.2	14.3
Saturday	1,408	17.5	14.3
Census region			
Northeast	1,499	19.3	19.9
Midwest	1,958	24.1	23.9
South	2,866	34.7	35.1
West	1,744	21.9	21.1
MSA status			
MSA (metropolitan)	6,092	76.2	78.8
Non-MSA	1,975	23.8	21.2

-- continued

Table 5-8. Continued.

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Household income as percentage of poverty level			
0-75%	888	10.3	8.4
76-130%	1,156	13.2	10.9
131-300%	2,665	33.8	31.8
Over 300%	3,358	42.8	48.9
Household received food stamps in past 12 months			
Yes	1,011	11.7	9.0
No	7,056	88.3	91.0
Presence in household of persons 18 and older			
Exactly 1	2,019	24.5	31.3
Exactly 2	4,832	60.2	54.2
Other than 1 or 2	1,216	15.2	14.5
Presence in household of children under 6 and 6-17			
Children under 6			
Children 6-17	1,128	13.9	8.5
No children 6-17	1,329	16.6	9.5
No children under 6			
Children 6-17	1,380	17.0	19.9
No children 6-17	4,230	52.5	62.1
Presence of female head of household 40 or younger and no one in household under 18			
Yes	449	5.7	9.7
No	7,618	94.3	90.3
Employment status			
Have job	4,355	54.6	57.7
Do not have job	3,712	45.4	42.3

-- continued

Table 5-8. Continued.

Variable	Sample size	Nonresponse adjustment	Population targets*
	Number	Percent	
Race			
Black	993	12.1	11.6
Non-black	7,074	87.1	88.4
Ethnic origin			
Hispanic	755	9.0	8.0
Non-Hispanic	7,312	91.0	92.0
Household size			
1 Member	1,464	17.9	24.8
2 Members	2,429	30.3	32.3
3 or more members	4,174	51.8	42.9

* Calculated using 1994-96 Current Population Survey data except for the variables "season of interview" and "day of week of household interview." Since the goal of the CSFII was to estimate behavior on an average day, each day of the week received an equal value of 14.3 percent, and each season received a value of 25 percent.

5.7.2 Use of the household sampling weights

The household sample in the 1994-96 CSFII consists of all households where at least one sample person was selected and provided a Day-1 intake. This is true regardless of whether a household questionnaire was completed. The use of the weights should be restricted to household information only (record type 15). No connections to sample persons or their intakes should be assumed in using the household weights.

Summary of final household weights

Table 5-9 summarizes the set of final household weights. The table shows the sample size; the sum of the weights; the coefficient of variation of the weights (CV), defined as the ratio of the standard deviation of the weights to the mean of the weights expressed as a percentage; and the variance inflation factor (VIF), defined as 1 + (CV/100)**2. This last statistic, which is equivalent to the ratio of the mean of the squared weights to the square of the mean of the weights, represents the anticipated proportional increase in the variance of survey estimates resulting from the variation in the weights. For example, it is anticipated that the variance of a household estimate will be 1.2 times what it would have been had all the weights been equal. The VIF can be used in the role of the "broadly calculated average design effect" specified in reporting guidelines adopted by ARS (FASEB/LSRO 1995).

Table 5-9. Summary of final household sample weights

Sample size	Sum of weights	CV	VIF = 1+(CV/100)**2
8,067	98,574,761	45.88%	1.21

Variance Estimation Fields

As described in CSFII/DHKS 1994-96 documentation (USDA 1998) section 3.2.1, "Sample design," Westat's 62 primary sampling unit (PSU) master sample was employed for CSFII/DHKS 1994-96. This sample of PSUs contains 24 PSUs selected with certainty. The remaining 38 PSUs were selected with probability proportional to size from 38 strata, 1 PSU per stratum. Thirty-six area segments were then selected from each of the 62 PSUs also with probability proportional to size. The thirty-six area segments were randomly allocated to the annual samples, twelve per year and three per quarter, so that the 62 PSUs were fielded at all times throughout the three years.

The following approach was used to create a framework of 2 sampling units per stratum to facilitate variance estimation procedures. First, 19 variance estimation strata were formed from the 38 noncertainty PSUs by pairing adjacent PSUs in the sampling frame. Each PSU within a variance estimation stratum defines what is referred to as a variance estimation unit. Next, within each of the 24 certainty PSUs, one-half of the segments were assigned to one variance estimation unit and the remaining one-half to another. Because each certainty PSU is considered to be a separate variance estimation stratum, a total of 43 variance estimation strata (each containing 2 variance estimation units) was formed by this process. See CSFII/DHKS 1994-96 documentation (USDA 1998) section 7.4.2, "Sampling weights and variance estimation fields," for details on identifying these variance estimation fields in the data set.

Estimation of Sampling Errors - Linearization method

Estimation of sampling errors may be conducted with a Taylor series linearization method using the final sample weights described in CSFII/DHKS 1994-96 documentation (USDA 1998) sections 5.1.2 and 5.1.3 along with the variance estimation strata and variance estimation units described in section 5.1.4. Software packages such as SUDAAN and Stata can be used to obtain estimates using the linearization method.

Estimation of Sampling Errors - Jackknife replicate method

Alternatively, sampling errors may be estimated using the jackknife technique described here. The construction of jackknife replicate weights makes use of the variance estimation stratum/variance estimation unit structure described in CSFII/DHKS 1994-96 documentation (USDA 1998) section 5.1.4. To illustrate how a jackknife variance estimator can be calculated, let y denote a weighted survey estimate (for example, total fat intake) calculated using the full-sample weights. Let y(j) be the corresponding weighted estimate calculated using the j-th set of replicate weights (j = 1, 2, .,43). The estimated variance of y is then given by the formula

$$Var(y) = SUM\ (y(j) - y)^{**}2,$$

where the summation extends over the 43 sets of jackknife replicate weights. Forty-three replicates were created by applying this process to each of the 43 variance estimation strata.

A jackknife replicate is created by dropping out one of the two variance estimation units from a variance estimation stratum and doubling the initial probability weights of the households in the other variance estimation unit in that stratum. The entire weighting process as described in the previous sections of this document was repeated for each replicate. Households not in the current replicate subsample were assigned a corresponding replicate weight of zero. In this way, a series of replicate weights was generated for each household. Together with the final, full-sample weights, these replicate weights were designed for the calculation of sampling errors.

Using a replication method to calculate sampling errors of survey-based estimates makes complicated variance estimation formulas unnecessary. The jackknife replication method used here is also designed to reflect the stratification and clustering used in the CSFII/DHKS sample design and to capture the effects of the raking ratio adjustments mentioned in CSFII/DHKS 1994-96 documentation section 5.1.2.5.

The replicate weighting process described above was designed and implemented by Westat, Inc., who has also created a variance estimation program, WesVarPC, which runs on computers using the Windows operating system and is available to the public at no charge. The software can be downloaded from Westat's home page at <http://www.westat.com>. In WesVarPC terminology, the JK2 method was used in constructing these replicate weights.

Programs and examples of output

Note: The following programs were written to accompany the release of the CSFII 1994-96 household sampling weights, not this 1994-96, 1998 release. The main purpose of two of the programs was to merge the final household sampling weights into an existing file derived from household record type 15. That merging process is not necessary with this release because both the 3-year and 4-year final household sampling weights have been included on record type 15. Also note that the 3-year household jackknife replicate weight file, jkw3yrhh.dat, has a different format than the file read in by the following program. Appropriate input programs for both the survey data files and the replicate weight files from this release may be found in section 10. Nevertheless, these input programs and programming examples from the original documentation may be useful to users of household level data.

The following are three SAS programs used to prepare data files and three examples of using the household data and sampling weights. Program 1 is a SAS program that reads the ASCII household weight file (hhwgt.dat) and creates a SAS system file. Program 2 adds the household sampling weight to an existing CSFII household-level SAS system file. Program 3 is a modified version of Program 2 that prepares a data file used as input by two of the examples.

Example 1 is a simple SAS program that produces weighted percentages of selected household variables. Examples 2 and 3 demonstrate the use of SUDAAN and WesVarPC for the estimation of standard errors of percentages. Example 2 is a SUDAAN program; Example 3 describes the preparation procedure necessary for using WesVarPC with CSFII data. The latter two examples examine household income as a percentage of poverty level (POVCAT) and the adequacy of the food supply of the household (FOODDESC). Both variables are modified by program 3 to create two-category variables. Levels 1 and 2 of FOODDESC have been combined to identify those households where "enough food eaten" was reported. Levels 3 and 4 have been combined to identify households where "not enough food eaten" was reported. Households with other values of FOODDESC are not represented in the analysis. Levels 2 and 3 of POVCAT have been combined for those households that have income over 130% of poverty. Level 1 represents those households that are below 131% of poverty (see documentation section 3.5, "Glossary").

```
*******************************************************
*                                                     *
* program1.sas                                        *
*                                                     *
* This SAS program reads the entire household weight  *
* file and creates the SAS system file, HHWGT,        *
* containing the same data.                           *
*                                                     *
* These programs assume that the directory \data9496  *
* holds all CSFII SAS files as well as the downloaded *
* ascii file containing the household sampling        *
* weights.  The LIBNAME and FILENAME statements       *
* should be modified as appropriate.                  *
*                                                     *
*******************************************************;

options ls = 78 ps = 55;

libname dir9496 'c:\data9496';
filename hhwgt 'c:\data9496\hhwgt.dat';

data dir9496.hhwgt (compress = 'yes');
  infile hhwgt lrecl = 386;
  input hhid 1-5 wt3_hh 8-15 wt_h_adj 16-23
        (r3_hh_01-r3_hh_43) (43 * 8.)
        hh_bwt 368-375 wt_h_scr 376-383 varstrat 384-385
        varunit 386;

label hhid     = "Household ID"
      wt3_hh   = "Full-sample household weight"
      wt_h_adj = "Non-response adjusted base weight"
      r3_hh_01 = "Replicate household weight - 1"
      r3_hh_02 = "Replicate household weight - 2"
      r3_hh_03 = "Replicate household weight - 3"
      r3_hh_04 = "Replicate household weight - 4"
      r3_hh_05 = "Replicate household weight - 5"
      r3_hh_06 = "Replicate household weight - 6"
      r3_hh_07 = "Replicate household weight - 7"
      r3_hh_08 = "Replicate household weight - 8"
      r3_hh_09 = "Replicate household weight - 9"
      r3_hh_10 = "Replicate household weight - 10"
      r3_hh_11 = "Replicate household weight - 11"
      r3_hh_12 = "Replicate household weight - 12"
      r3_hh_13 = "Replicate household weight - 13"
      r3_hh_14 = "Replicate household weight - 14"
      r3_hh_15 = "Replicate household weight - 15"
      r3_hh_16 = "Replicate household weight - 16"
      r3_hh_17 = "Replicate household weight - 17"
      r3_hh_18 = "Replicate household weight - 18"
      r3_hh_19 = "Replicate household weight - 19"
      r3_hh_20 = "Replicate household weight - 20"
      r3_hh_21 = "Replicate household weight - 21"
      r3_hh_22 = "Replicate household weight - 22"
```

```
       r3_hh_23 = "Replicate household weight - 23"
       r3_hh_24 = "Replicate household weight - 24"
       r3_hh_25 = "Replicate household weight - 25"
       r3_hh_26 = "Replicate household weight - 26"
       r3_hh_27 = "Replicate household weight - 27"
       r3_hh_28 = "Replicate household weight - 28"
       r3_hh_29 = "Replicate household weight - 29"
       r3_hh_30 = "Replicate household weight - 30"
       r3_hh_31 = "Replicate household weight - 31"
       r3_hh_32 = "Replicate household weight - 32"
       r3_hh_33 = "Replicate household weight - 33"
       r3_hh_34 = "Replicate household weight - 34"
       r3_hh_35 = "Replicate household weight - 35"
       r3_hh_36 = "Replicate household weight - 36"
       r3_hh_37 = "Replicate household weight - 37"
       r3_hh_38 = "Replicate household weight - 38"
       r3_hh_39 = "Replicate household weight - 39"
       r3_hh_40 = "Replicate household weight - 40"
       r3_hh_41 = "Replicate household weight - 41"
       r3_hh_42 = "Replicate household weight - 42"
       r3_hh_43 = "Replicate household weight - 43"
       hh_bwt   = "Household base weight"
       wt_h_scr = "Screener adjusted household base weight"
       varstrat = "Variance strata"
       varunit  = "Variance estimation unit"
                ;

proc means;

run;

************* End of Program 1 ************************;
```

```
*******************************************************
*                                                     *
* program2.sas                                        *
*                                                     *
* This SAS program adds the household sampling        *
* weight, WT3_HH, to an existing household-level SAS  *
* file such as the file created from record type 15   *
* by the READ15.SAS program on the 1994-96 CD-ROM.    *
* The file created by PROGRAM1.SAS supplies the       *
* sampling weight.                                    *
*                                                     *
* These programs assume that the directory \data9496  *
* holds all CSFII SAS files.  The LIBNAME statement   *
* should be modified as appropriate.                  *
*                                                     *
*******************************************************;

options ls = 78 ps = 55;

libname dir9496 'c:\data9496';

***********************************************
*                                             *
* Delete or modify the KEEP option in the     *
* following statement to add the replicate    *
* sampling weights to the RT15 file.  The     *
* replicate weights are required if software  *
* such as WESVAR, utilizing a replication     *
* method, is used for variance estimation     *
* (see example 3).                            *
*                                             *
***********************************************;

data dir9496.rt15 (compress = 'yes');
  merge dir9496.rt15
        dir9496.hhwgt (keep = hhid wt3_hh);
  by hhid;

proc means;

run;

************* End of Program 2 ************************;
```

```
************************************************************
*                                                          *
* program3.sas                                             *
*                                                          *
* This SAS program prepares an input file for the two      *
* variance estimation programs, examples 2 and 3. It is    *
* used for three purposes.  The first purpose is to create *
* a PC SAS file in an older than current version, version  *
* 6.04, that both PC-based SUDAAN and Wesvar can read      *
* directly. Secondly, the replicate weights are collected  *
* from the file created by PROGRAM1.SAS.  Thirdly, the     *
* variables used for analysis in example programs are      *
* created.  Only the variables needed for the examples are *
* retained.                                                *
*                                                          *
************************************************************;

options ls = 78 ps = 55;

libname dir9496 'c:\data9496';
libname dir2 v604 'c:\data9496';

data dir2.pgm3 (keep = hhid wt3_hh varstrat varunit
                       r3_hh_01-r3_hh_43 under131 enough);
  merge dir9496.rt15 (keep = hhid povcat fooddesc wt3_hh
                             varstrat varunit)
        dir9496.hhwgt (keep = hhid r3_hh_01- r3_hh_43);
  by hhid;

if fooddesc in(1, 2)  then
  enough = 1;
else if fooddesc in(3, 4)  then
  enough = 2;

if (povcat eq 1)  then
  under131 = 1;
else
  under131 = 2;

label under131 = 'Income status'
      enough   = 'Enough to eat'
               ;

proc means;

run;

************* End of Program 3 ************************;
```

```
*******************************************************
*                                                     *
* example1.sas                                        *
*                                                     *
* This SAS program produces weighted frequencies of   *
* several household level variables.  The input file  *
* is the file created by PROGRAM2.SAS                 *
*                                                     *
* These programs assume that the directory \data9496  *
* holds all CSFII SAS files.  The LIBNAME statement   *
* should be modified as appropriate.                  *
*                                                     *
*******************************************************;

options ls = 78 ps = 60;
options nodate nonumber nocenter;

libname dir9496 'c:\data9496';

proc freq data = dir9496.rt15;
  tables povcat fooddesc fs_rcv12 urb region;
  weight wt3_hh;
  format povcat povcat. fooddesc fooddesc.
         fs_rcv12 yn789f.  urb urb. region region.;
  title 'Example 1: Weighted frequencies of household '
        'level data, 1994-96 CSFII';

run;

************* End of Example 1 program ***************;
```

Example 1: Weighted frequencies of household level data,
1994-96 CSFII

Annual income: % of poverty category

POVCAT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0 - 130%	19520256	19.8	19520256	19.8
131 - 350%	39468942	40.0	58989198	59.8
Over 350%	39585563	40.2	98574761	100.0

Description of food eaten in HH

FOODDESC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Enough - 1	75210037	76.3	75210037	76.3
Enough - 2	20344690	20.6	95554727	96.9
Sometimes not enough	1803550	1.8	97358277	98.8
Often not enough	336251	0.3	97694528	99.1
Not ascertained	880233	0.9	98574761	100.0

Food stamps: in last 12 months

FS_RCV12	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Yes	8693044	8.8	8693044	8.8
No	88663791	89.9	97356835	98.8
Refused	237994	0.2	97594829	99.0
Don't know	102999	0.1	97697828	99.1
Not ascertained	876933	0.9	98574761	100.0

Urbanization

URB	Frequency	Percent	Cumulative Frequency	Cumulative Percent
MSA, central city	31977978	32.4	31977978	32.4
MSA, not central city	45717307	46.4	77695285	78.8
Non-MSA	20879476	21.2	98574761	100.0

Region

REGION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Northeast	19586188	19.9	19586188	19.9
Midwest	23591612	23.9	43177800	43.8
South	34604507	35.1	77782307	78.9
West	20792454	21.1	98574761	100.0

********** End of Example 1 output listing *********;


```
/*****************************************************************
*                                                                *
* Example2.prc                                                   *
*                                                                *
* This SUDAAN program provides an example of computing the       *
* standard error of estimates from the CSFII 1994- 96.  SUDAAN   *
* is a program containing procedures designed to be used to      *
* analyze data from complex sample surveys such as the CSFII.    *
*                                                                *
* This program was written to be used by the stand- alone        *
* version of SUDAAN.  The input file is the SAS system file      *
```

```
* created by PROGRAM2.SAS which created a version 6.04 PC SAS   *
* system file.  This program provides the basic statements      *
* needed to inform SUDAAN of the CSFII sample design            *
* information needed for the estimates.                         *
*                                                               *
* The SUDAAN procedure used here is PROC CROSSTAB. The          *
* procedure call specifies a "with replacement" design          *
* (design = wr).  A NEST statement is used to define the        *
* required design parameters, VARSTRAT, the variance-estimation *
* stratum, and VARUNIT, the variance-estimation unit which is   *
* used as a primary sampling unit or PSU.                       *
*                                                               *
* Notes: The data directory must be set to the directory        *
*        containing the input file. Also, a LEVEL.DBS may be    *
*        placed in that directory to supply variable formats.   *
*                                                               *
****************************************************************/

proc crosstab data = pgm3 filetype = sas design = wr;
  nest varstrat varunit;
  weight wt3_hh;
  subgroup under131 enough;
  levels   2        2;
  tables under131 * enough;
  print nsum rowper colper serow secol /style = nchs ;

************* End of Example 2 program ***************;
```

Example 2 output listing

Research Triangle Institute
The CROSSTAB Procedure

by: Income status, Enough to eat.

Income status Enough to eat	Sample Size	Row Percent	Col Percent	SE Row Percent	SE Col Percent
Total					
Total	8007	100.00	100.00	0.00	0.00
Enough	7791	97.81	100.00	0.19	0.00
Not enough	216	2.19	100.00	0.19	0.00
Below 131% of poverty					
Total	2083	100.00	19.90	0.00	0.96
Enough	1908	91.80	18.68	0.71	0.93
Not enough	175	8.20	74.54	0.71	3.19
131%+					
Total	5924	100.00	80.10	0.00	0.96
Enough	5883	99.30	81.32	0.11	0.93
Not enough	41	0.70	25.46	0.11	3.19

********** End of Example 2 output listing *********;

```
*****************************************************************
*                                                               *
* Example3                                                      *
*                                                               *
* Example 3 used the WesVarPC software to estimate percentages  *
* and their standard errors.  The file created by PROGRAM2.SAS  *
* provided the input.  During the preparation step, the SAS     *
* file PGM3.SSD was imported, the analysis variables, full      *
* sample weight and replicate weights identified, and the       *
* replication method JK2 selected.                              *
*                                                               *
* The output shown below was produced by a table request of     *
* under131 * enough, asking for percentages of the sum of the   *
* weights, and with other specifications as shown below.        *
*                                                               *
*****************************************************************/

Example 3 output listing

 PC WESVAR VERSION NUMBER:                    2.12
 TIME THE JOB EXECUTED:                       08:19:23 02/19/99
 INPUT DATASET NAME:                         C:\data9496\Pgm3.var
 OUTPUT LISTING:                             C:\data9496\pgm3.LST

 OPTION NOSUMMARY IS:                         OFF
 OPTION FUNCTION LOG IS:                      OFF
 OPTION ALIGNMENT IS:                         OFF
 OPTION EXPORT IS:                            OFF
 VARIANCE ESTIMATION METHOD:                  JK2
 FINITE POPULATION CORRECTION FACTOR:         1.00000
 VALUE OF ALPHA (CONFIDENCE INTERVAL %):      0.05000 (95.00000 %)
 DEGREES OF FREEDOM:                          INFINITE
 t VALUE:                                     1.960

 OPTION COMPLETE IS:                          ON
 FULL SAMPLE WEIGHT:                          WT3_HH
 REPLICATE WEIGHTS:                           R3_HH_01...R3_HH_43
 ANALYSIS VARIABLES:                          None Specified.
 COMPUTE STATISTIC:                           None Specified.
 TABLE REQUESTS:                              UNDER131*ENOUGH

 FACTOR(S):                                   1.00

 NUMBER OF REPLICATES:                        43
 NUMBER OF OBSERVATIONS READ:                 8067
 WEIGHTED NUMBER OF OBSERVATIONS READ:        98574761.000
```

TABLE REQUEST : UNDER131 * ENOUGH

UNDER131	ENOUGH	EST_TYPE	ESTIMATE	STDERROR	N
Below 131%	Enough	PERCENT	18.27	0.16	1908
Below 131%	Not enough	PERCENT	1.63	0.12	175
Below 131%	MARGINAL	PERCENT	19.90	0.11	2083
131%+	Enough	PERCENT	79.54	0.13	5883
131%+	Not enough	PERCENT	0.56	0.10	41
131%+	MARGINAL	PERCENT	80.10	0.11	5924
MARGINAL	Enough	PERCENT	97.81	0.16	7791
MARGINAL	Not enough	PERCENT	2.19	0.16	216
MARGINAL	MARGINAL	PERCENT	100.00	0.00	8007
Below 131%	Enough	COLPCT	18.68	0.14	1908
Below 131%	Not enough	COLPCT	74.54	3.39	175
Below 131%	MARGINAL	COLPCT	19.90	0.11	2083
131%+	Enough	COLPCT	81.32	0.14	5883
131%+	Not enough	COLPCT	25.46	3.39	41
131%+	MARGINAL	COLPCT	80.10	0.11	5924
MARGINAL	Enough	COLPCT	100.00	0.00	7791
MARGINAL	Not enough	COLPCT	100.00	0.00	216
MARGINAL	MARGINAL	COLPCT	100.00	0.00	8007
Below 131%	Enough	ROWPCT	91.80	0.58	1908
Below 131%	Not enough	ROWPCT	8.20	0.58	175
Below 131%	MARGINAL	ROWPCT	100.00	0.00	2083
131%+	Enough	ROWPCT	99.30	0.12	5883
131%+	Not enough	ROWPCT	0.70	0.12	41
131%+	MARGINAL	ROWPCT	100.00	0.00	5924
MARGINAL	Enough	ROWPCT	97.81	0.16	7791
MARGINAL	Not enough	ROWPCT	2.19	0.16	216
MARGINAL	MARGINAL	ROWPCT	100.00	0.00	8007

********** End of Example 3 output listing *********;

5.8 References

FASEB/LSRO (Federation of American Societies for Experimental Biology, Life Sciences Research Office). 1995. Joint policy on variance estimation and statistical standards on NHANES III and CSFII reports ... (Appendix III). In: Third Report on Nutrition Monitoring in the United States. Prepared for the Interagency Board for Nutrition Monitoring and Related Research. USDA Publication.

SAS Institute, Inc. 1990. SAS language: Reference, version 6 first edition. SAS Institute, Inc., Cary, NC.

Shah, BV, Barnwell, BG, and Bieler, GS. 1997. SUDAAN User's Manual, Release 7.5. Research Triangle Park, NC: Research Triangle Institute.

USDA (U.S. Department of Agriculture, Agricultural Research Service). 1998. 1994-96 Continuing Survey of Food Intakes by Individuals and 1994-96 Diet and Health Knowledge Survey. CD-ROM. Available from National Technical Information Service, Springfield, VA. (NTIS Accession No. PB98-500457)

USDC/BOC (U.S. Department of Commerce, Bureau of the Census). 1994. Current Population Survey, March 1994. Machine-readable data file.

USDC/BOC (U.S. Department of Commerce, Bureau of the Census). 1995. Current Population Survey, March 1995. Machine-readable data file.

USDC/BOC (U.S. Department of Commerce, Bureau of the Census). 1996. Current Population Survey, March 1996. Machine-readable data file.

USDC/BOC (U.S. Department of Commerce, Bureau of the Census). 1998. Current Population Survey, March 1998. Machine-readable data file.

6.3 Statistical Notes

6.3.1 Statistical software

Because of the complex sample design of the CSFII, ARS recommends that data users calculate standard errors and coefficients of variation for descriptive and related statistics using software that takes the sample design and weighting into account. The PSUs used in the design can be paired as in a stratified sample where the Taylor Series expansion method can be used. This will allow software such as SUDAAN or Stata to be used when studying population subgroups. The fields VARSTRAT and VARUNIT are located in positions 11-12 and 13 in the data file, respectively. These fields represent the nesting fields STRATUM and PSU used for Taylor Series expansion estimation of standard errors. Also, see section 7.4.2, "Sampling weights and variance estimation fields." Replicate weights, as provided by the jackknife replication method, can also be used as described in section 5.6.2, "Estimation of Sampling Errors." See section 5, "SAMPLING WEIGHTS," for more information on weighting procedures.

6.3.2 Guidelines for the use of sampling weights

Weights should always be used when calculating descriptive statistics. This is because descriptive statistics are meant to provide summary information about the entire population under study, not just the sample. Included under the heading of descriptive statistics are measures of central tendency, such as means and medians, as well as measures of variability, such as variances.

Most statistical software packages allow the user to compute weighted descriptive statistics although they may not estimate variances properly. If in doubt, the analyst is advised to consult a survey statistician.

ATTACHMENT 2

The EPA Food Commodity Vocabulary

(Source: CSFII Documentation)

U.S. Environmental Protection Agency

THE EPA FOOD COMMODITY VOCABULARY

EPA's list of food commodities, that is, agricultural food items from either plant or animal sources in raw or processed forms, includes items based on the following rationale. A food commodity is included by EPA if it is cited in the Residue Chemistry Guidelines, Table 1, entitled: Raw Agricultural and Processed Commodities and Feedstuffs Derived from Crops (Guideline prepared by the Office of Prevention, Pesticides, and Toxic Substances; available electronically from the EPA Public Access webpage. To access the guideline, go to: http://www.epa.gov/OPPTS_Harmonized/860_Residue_Chemistry_Test_Guidelines/Series. Then, select 860.1000).

Additionally, a food commodity is included in the following list if it currently has a tolerance. Other food commodities are included if foods known to contain them have been reported to be consumed in the 1994-96 CSFII, or in prior food consumption surveys, or if they are known to be ingredients in commercially available baby food. Some few other food commodities may exist in the list below as they were revealed to be consumed by humans subsequent to research undertaken by EPA staff in the preparation of numerous internal reports on use of agricultural animals or crops.

THE EPA FOOD COMMODITY VOCABULARY

Master List June 15, 2000

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95000010	Acerola	weight of fruit; excluding seed	
18000020	Alfalfa, seed	weight of dry seed	Alfalfa sprouts are the human food item.
14000030	Almond	weight of nutmeat	
14000031	Almond- babyfood	same as almond	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
14000040	Almond, oil	weight of oil	
14000041	Almond, oil- babyfood	weight of oil	
04010050	Amaranth, leafy	weight of leaf	Includes Tampala, Chinese spinach, lambsquarter, and pokeweed (pokeberry).
95000060	Amaranth, grain	dry weight of grain; include Amaranth flour	
11000070	Apple, fruit with peel	weight of apple; including peel, excluding core and stem	
11000080	Apple, peeled fruit	weight of apple; excluding peel, core and stem	
11000081	Apple, peeled fruit-babyfood	weight of apple; excluding peel, core and stem	
11000090	Apple, dried	dry weight; excluding peel, core, stem	
11000091	Apple, dried - babyfood	dry weight; excluding peel, core, stem	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
11000100	Apple, juice	weight of juice at single strength (or standard dilution)	
11000101	Apple, juice - babyfood	weight of juice at single strength (or standard dilution)	
11000110	Apple, sauce	weight of applesauce	
11000111	Apple, sauce - babyfood	weight of applesauce	
12000120	Apricot	weight of pulp, with or without peel; excluding pit	
12000121	Apricot- babyfood	weight of pulp, with or without peel; excluding pit	
12000130	Apricot, dried	dry weight of pulp, with or without peel; excluding pit	
12000140	Apricot, juice	weight of juice at single strength (or standard dilution)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
12000141	Apricot, juice- babyfood	weight of juice at single strength (or standard dilution)	
01030150	Arrowroot, flour	dry weight of flour	
01030151	Arrowroot, flour- babyfood	dry weight of flour	
95000160	Artichoke, globe	edible portion of flowerhead	
01030170	Artichoke, Jerusalem	edible portion of tuber	
04010180	Arugula	weight of leaves	
95000190	Asparagus	weight of edible portion of spears/stems	
95000200	Avocado	weight of pulp; excluding skin and pit	
09020210	Balsam pear	weight of whole fruit	Includes Balsam apple, Chinese cucumber, and Bittermelon.
95000220	Bamboo, shoots	weight of shoots	Bamboo shoots are the human food item.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95000230	Banana	weight of pulp; excluding peel; juice	
95000231	Banana- babyfood	weight of pulp; excluding peel; juice	
95000240	Banana, dried	dry weight of dried pulp; excluding peel (include weight of fruit from chips)	
95000241	Banana, dried- babyfood	dry weight of dried pulp; excluding peel (include weight of fruit from chips)	
15000250	Barley, pearled barley	dry weight of barley	
15000251	Barley, pearled barley - babyfood	dry weight of barley	
15000260	Barley, flour	dry weight of flour	Includes malt and whole barley.
15000261	Barley, flour- babyfood	dry weight of flour	Includes malt and whole barley.
15000270	Barley, bran	dry weight of bran	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
19010280	Basil, fresh leaves	weight of leaves and stems	
19010281	Basil, fresh leaves - babyfood	weight of leaves and stems	
19010290	Basil, dried leaves	dry weight of leaves and stems	
19010291	Basil, dried leaves- babyfood	dry weight of leaves and stems	
06030300	Bean, black, seed	dry weight of bean	Includes black turtle bean, Bayo, and brown bean.
06020310	Bean, broad, succulent	weight of bean and pod	Also called fava bean.
06030320	Bean, broad, seed	dry weight of bean	Also called fava bean.
06020330	Bean, cowpea, succulent	weight of bean; excluding pod	Includes cowpea, crowder pea, blackeye pea, and southern pea.
06030340	Bean, cowpea, seed	dry weight of bean	Includes cowpea, crowder pea, blackeye pea, and southern pea.
06030350	Bean, great northern, seed	dry weight of bean	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
06030360	Bean, kidney, seed	dry weight of bean	
06020370	Bean, lima, succulent	weight of bean; excluding pod	
06030380	Bean, lima, seed	dry weight of bean	
06030390	Bean, mung, seed	dry weight of bean	Bean sprouts are the human food item.
06030400	Bean, navy, seed	dry weight of bean	Includes pea bean.
06030410	Bean, pink, seed	dry weight of bean	
06030420	Bean, pinto, seed	dry weight of bean	Include calico and red Mexican bean.
06010430	Bean, snap, succulent	weight of bean and pod	Includes green bean, runner bean, and wax bean.
06010431	Bean, snap, succulent-babyfood	weight of bean and pod	Includes green bean, runner bean, and wax bean.
21000440	Beef, meat	weight of meat; excluding the weight of bone, and all nutrient fat	Consider veal as equivalent to beef. Include bison and buffalo.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
21000441	Beef, meat- babyfood	weight of meat; excluding the weight of bone, and all nutrient fat	
21000450	Beef, meat, dried	weight of dried meat; excluding bones, and trimmable fat	
21000460	Beef, meat byproducts	weight of meat; excluding bone; may contain some fat	Includes brain, heart, lung, sweetbread, tail, tripe, tongue, and head.
21000461	Beef, meat byproducts- babyfood	weight of meat; excluding bone; may contain some fat	Includes brain, heart, lung, sweetbread, tail, tripe and tongue.
21000470	Beef, fat	weight of nutrient fat only; includes nutrient fat from beef meat	
21000471	Beef,fat- babyfood	weight of nutrient fat only; includes nutrient fat from beef meat	
21000480	Beef, kidney	weight of organ including nutrient fat	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
21000490	Beef, liver	weight of organ including nutrient fat	
21000491	Beef, liver- babyfood	weight of organ including nutrient fat	
01010500	Beet, garden, roots	weight of roots; juice	
01010501	Beet, garden, roots-babyfood	weight of roots	
02000510	Beet, garden, tops	weight of leaves	Include pumpkin leaves.
01010520	Beet, sugar	dry weight of sugar (sucrose)	Sucrose is a disaccharide obtained from sugar cane and sugar beet.
01010521	Beet, sugar- babyfood	dry weight of sugar (sucrose)	Sucrose is a disaccharide obtained from sugar cane and sugar beet.
01010530	Beet, sugar, molasses	weight of molasses	
01010531	Beet, sugar, molasses - babyfood	weight of molasses	
95000540	Belgium endive	weight of leaves	Also called Witloof chicory.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
13010550	Blackberry	weight of berry	Includes Marionberry, Olallieberry, and Youngberry.
13010560	Blackberry, juice	weight of juice at single strength (or standard dilution)	Includes Marionberry, Olallieberry, and Youngberry.
13010561	Blackberry, juice - babyfood	weight of juice at single strength (or standard dilution)	Includes Marionberry, Olallieberry, and Youngberry.
13020570	Blueberry	weight of berry	
13020571	Blueberry- babyfood	weight of berry	
13010580	Boysenberry	weight of berry	
14000590	Brazil nut	weight of nutmeat	
95000600	Breadfruit	weight of pulp; excluding peel	
05010610	Broccoli	weight of flower heads and adjoining stems	
05010611	Broccoli- babyfood	weight of flower heads and adjoining stems	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
05010620	Broccoli, Chinese	weight of flower buds, adjoining stems and leaves	
05020630	Broccoli raab	weight of flower buds, adjoining stems and leaves	
05010640	Brussels sprouts	weight of leaf sprouts	
15000650	Buckwheat	dry weight of groats; whole groat flour	
15000660	Buckwheat, flour	dry weight of flour	
01010670	Burdock	weight of roots	
14000680	Butternut	weight of nutmeat	
05010690	Cabbage	weight of leaves	
05020700	Cabbage, Chinese, bok choy	weight of leaves and stems	
05010710	Cabbage, Chinese, napa	weight of leaves	
05010720	Cabbage, Chinese, mustard	weight of leaves	
95000730	Cactus	weight of pulp; excluding peel; juice	Includes Prickly pear, Cactus pads and flowers, Nopales, and Aloe vera.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95000740	Canistel	weight of pulp; excluding peel	Also called Eggfruit.
09010750	Cantaloupe	weight of pulp; excluding seeds and outer rind	Includes wintermelon.
04020760	Cardoon	weight of leaf stalks	
95000770	Carob	dry weight of bean; flour	
01010780	Carrot	weight of roots, with or without peel, excluding tops	
01010781	Carrot- babyfood	weight of roots, with or without peel, excluding tops	
01010790	Carrot, juice	weight of juice at single strength (or standard dilution)	
09010800	Casaba	weight of pulp, excluding seeds and rind	
14000810	Cashew	weight of nutmeat	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
01030820	Cassava	weight of root; dry tapioca	Includes tapioca.
01030821	Cassava- babyfood	weight of root; dry tapioca	Includes tapioca.
05010830	Cauliflower	weight of flower heads and adjoining stems	Includes Broccoflower and green cauliflower.
01010840	Celeriac	weight of tuberous root	
04020850	Celery	weight of leaf stalk	
04020851	Celery- babyfood	weight of leaf stalk	
04020860	Celery, juice	weight of juice as single strength (or standard dilution)	
04020870	Celtuce	weight of stalks and leaves	
09020880	Chayote, fruit	weight of fruit	Also called Christophine, and Mirliton.
95000890	Cherimoya	weight of fruit; excluding peel	Also called custard apple.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
12000900	Cherry	weight of fruit; including skin; excluding pit and stem	Includes sweet cherry and sour or tart cherry.
12000901	Cherry- babyfood	weight of fruit; including skin; excluding pit and stem	Includes sweet cherry and sour or tart cherry.
12000910	Cherry, juice	weight of juice at single strength (or standard dilution)	Includes sweet cherry and sour or tart cherry.
12000911	Cherry, juice- babyfood	weight of juice at single strength (or standard dilution)	Includes Sweet cherry and Sour or tart cherry.
14000920	Chestnut	weight of nutmeat	
40000930	Chicken, meat	weight of flesh; excluding the weight of bone, total nutrient fat, and skin	
40000931	Chicken, meat- babyfood	weight of flesh; excluding the weight of bone, total nutrient fat, and skin	
40000940	Chicken, liver	weight of organ; including nutrient fat	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
40000950	Chicken, meat byproducts	weight of meat; excluding bone; may include some trimmable fat and/or skin	Includes giblets (heart), necks, gizzard, feet, and tail.
40000951	Chicken, meat byproducts-babyfood	weight of meat; excluding bone; may include some trimmable fat and/or skin	Includes giblets (heart), necks, gizzard, feet, and tail.
40000960	Chicken, fat	weight of nutrient fat only; includes weight of nutrient fat from chicken meat and skin	
40000961	Chicken, fat- babyfood	weight of nutrient fat only; includes weight of nutrient fat from chicken meat and skin	
40000970	Chicken, skin	weight of skin only (0 grams nutrient fat)	
40000971	Chicken, skin- babyfood	weight of skin only (0 grams nutrient fat)	
06030980	Chickpea, seed	dry weight of bean	Also called garbanzo bean.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
06030981	Chickpea, seed - babyfood	dry weight of bean	Also called garbanzo bean.
06030990	Chickpea, flour	dry weight of flour	Also called garbanzo bean.
01011000	Chicory, roots	weight of roots	
02001010	Chicory, tops	weight of leaves	
09021020	Chinese waxgourd	weight of flesh; including or excluding peel	Includes Togan and wintermelon.
19011030	Chive	weight of leaves	
04011040	Chrysanthemum, garland	edible leaves and stems	Includes Chrysanthemum, edible leaved.
19021050	Cinnamon	dry weight of spice: stick or ground powder	
19021051	Cinnamon- babyfood	dry weight of spice: stick or ground powder	
10001060	Citrus citron	weight of pulp; excluding peel	
10001070	Citrus hybrids	weight of pulp; excluding seeds and peel	Includes tangelo, Tangor, Chironja, and Calamondin.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
10001080	Citrus, oil	weight of oil	
95001090	Cocoa bean, chocolate	weight of chocolate; cocoa butter	
95001100	Cocoa bean, powder	weight of powder	
95001110	Coconut, meat	weight of meat; excluding milk and shell	
95001111	Coconut- meat, babyfood	weight of meat; excluding milk and shell	
95001120	Coconut, dried	dry weight of meat; excluding milk and shell	
95001130	Coconut, milk	weight of milk only; excluding meat and shell	
95001140	Coconut, oil	weight of oil	
95001141	Coconut, oil- babyfood	weight of oil	
95001150	Coffee, roasted bean	dry weight of bean	
95001160	Coffee, instant	dry weight of powder or granules	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
05021170	Collards	weight of leaves	
19021180	Coriander, leaves	weight of leaves	Includes Chinese parsley and cilantro leaf.
19021181	Coriander, leaves - babyfood	weight of leaves	Includes Chinese parsley and cilantro leaf.
19021190	Coriander, seed	weight of seed	Includes Chinese parsley and cilantro seed.
19021191	Coriander, seed - babyfood	weight of seed	Includes Chinese parsley and cilantro seed.
15001200	Corn, field, flour	dry weight of whole grain flour; masa harina	
15001201	Corn, field, flour-babyfood	dry weight of whole grain flour; masa harina	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
15001210	Corn, field, meal	dry weight of whole grain or degermed meal; dry weight of corn ingredient from corn or cornmeal based chips or snacks; dry weight of hominy	Includes hominy.
15001211	Corn, field, meal-babyfood	dry weight of whole grain or degermed meal; dry weight of corn ingredient from corn or cornmeal based chips or snacks; dry weight of hominy.	Includes hominy.
15001220	Corn, field, bran	dry weight of bran	
15001230	Corn, field, starch	dry weight of corn starch	
15001231	Corn, field, starch-babyfood	dry weight of corn starch	
15001240	Corn, field, syrup	weight of syrup	
15001241	Corn, field, syrup-babyfood	weight of syrup	
15001250	Corn, field, oil	weight of oil	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
15001251	Corn, field, oil- babyfood	weight of oil	
15001260	Corn, pop	weight of kernels; excluding cob and husk	
15001270	Corn, sweet	weight of kernels; excluding cob and husk	
15001271	Corn, sweet- babyfood	weight of kernels; excluding cob and husk	
95001280	Cottonseed, oil	weight of oil	
95001281	Cottonseed, oil - babyfood	weight of oil	
11001290	Crabapple	weight of pulp; excluding core and stem; including or excluding peel	
95001300	Cranberry	weight of berry	
95001301	Cranberry- babyfood	weight of berry	
95001310	Cranberry, dried	dry weight of berry	
95001320	Cranberry, juice	weight of juice at single strength (or standard dilution)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95001321	Cranberry, juice- babyfood	weight of juice at single strength (or standard dilution)	
04011330	Cress, garden	weight of leaves	
04011340	Cress, upland	weight of leaves	Includes Yellow rocket and Winter cress.
09021350	Cucumber	weight of flesh and seeds; including or excluding peel	
13021360	Currant	weight of berry	
13021370	Currant, dried	dry weight of berry	
04011380	Dandelion, leaves	weight of leaves	
01031390	Dasheen, corm	weight of the corm	Includes taro.
02001400	Dasheen, leaves	weight of the leaves	Includes taro.
95001410	Date	weight of fruit, excluding pit	
13011420	Dewberry	weight of berry	
19021430	Dill, seed	dry weight of seed	
19011440	Dill	weight of leaves	Also called dillweed.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
70001450	Egg, whole	weight of white and yolk; excluding shell	
70001451	Egg, whole - babyfood	weight of white and yolk; excluding shell	
70001460	Egg, white	weight of egg white	
70001461	Egg, white (solids)- babyfood	rehydrated weight of egg white solids	
70001470	Egg, yolk	weight of egg yolk	
70001471	Egg, yolk- babyfood	weight of egg yolk	
08001480	Eggplant	weight of whole vegetable; including seeds, with or without peel	
13021490	Elderberry	weight of berry	
04011500	Endive	weight of leaves	Includes escarole.
95001510	Feijoa	weight of pulp, excluding peel	Also called pineapple guava.
04021520	Fennel, Florence	weight of leaves	Includes Italian and sweet fennel.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95001530	Fig	weight of fruit	
95001540	Fig, dried	dry weight of fruit	
14001550	Filbert	weight of nutmeat	Also called hazelnut.
14001560	Filbert, oil	weight of oil	Also called hazelnut.
80001570	Fish- freshwater finfish	weight of edible portion; excluding head, tail, scales, fins, viscera, inedible bones and skin	**See Appendix A ("The Fish List") for categorization of various fish species into appropriate FC, ie, for species that are specifically fresh water fish, salt water fish, crustacean-shellfish, and mollusc-shellfish.**
80001580	Fish- freshwater finfish, farm raised	weight of edible portion; excluding head, tail, scales, fins, viscera, inedible bones and skin	**See Appendix A.**

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
80001590	Fish- saltwater finfish, tuna	weight of edible portion; excluding head, tail, scales, fins, viscera, inedible bones and skin	**See Appendix A.**
80001600	Fish- saltwater finfish, other	weight of edible portion; excluding head, tail, scales, fins, viscera, inedible bones and skin	**See Appendix A.**
80001610	Fish- shellfish, crustacean	weight of edible portion; excluding shell, gills, and viscera	**See Appendix A.**
80001620	Fish- shellfish, mollusc	weight of edible portion; excluding shells	**See Appendix A.**
95001630	Flaxseed, oil	weight of oil; nutrient fat from flax seeds	Also called Solin.
03001640	Garlic	weight of bulb; excluding skin (outer scales)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
03001650	Garlic, dried	dry weight of bulb; excluding skin (outer scales)	
03001651	Garlic, dried- babyfood	dry weight of bulb; excluding skin (outer scales)	
01031660	Ginger	weight of roots; excluding peel	
01031661	Ginger - babyfood	weight of roots, excluding peel	
01031670	Ginger, dried	dry weight of roots; excluding peel	
01031680	Ginseng, dried	dry weight of roots	
23001690	Goat, meat	weight of meat; excluding weight of bone, and all nutrient fat and skin	
23001700	Goat, meat byproducts	weight of meat; excluding bone; may include some trimmable fat and/or skin	Includes brain, heart, and head.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
23001710	Goat, fat	weight of nutrient fat only; include nutrient fat from goat meat	
23001720	Goat, kidney	weight of organ including nutrient fat	
23001730	Goat, liver	weight of organ including nutrient fat	
13021740	Gooseberry	weight of berry	
95001750	Grape	weight of grape, with skin, and with or without seeds	Includes Muscadine.
95001760	Grape, juice	weight of juice at single strength (or standard dilution)	
95001761	Grape, juice- babyfood	weight of juice at single strength (or standard dilution)	
95001770	Grape, leaves	weight of leaves	
95001780	Grape, raisin	dry weight of raisin	Includes Zante currant.
95001790	Grape, wine and sherry	weight of wine or sherry	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
10001800	Grapefruit	weight of pulp; excluding seeds and rind	
10001810	Grapefruit, juice	weight of juice at single strength (or standard dilution)	
06031820	Guar, seed	weight of bean	
06031821	Guar, seed - babyfood	weight of bean	
95001830	Guava	weight of pulp; excluding peel; juice	
95001831	Guava- babyfood	weight of pulp; excluding peel; juice	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
19011840	Herbs, other	weight of leaves and stems	See Appendix B ("The Herb List"). Note that some herbs are not included in this list as they are unique FCs and are cited in this vocabulary separately. These latter include: basil, chive, coriander (cilantro), leaf, dill (dillweed), fennel, Florence (Italian and sweet), lemongrass, marjoram (oregano), parsley and savory (winter and summer).

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
19011841	Herbs, other- babyfood	weight of leaves and stems	**See Appendix B ("The Herb List"). Note that some herbs are not included in this list as they are unique FCs and are cited in this vocabulary separately. These latter include: basil, chive, coriander (cilantro), leaf, dill (dillweed), fennel, Florence (Italian and sweet), lemongrass, marjoram (oregano), parsley, and savory (summer and winter).**
14001850	Hickory nut	weight of nutmeat	
95001860	Honey	weight of honey	
95001861	Honey- babyfood	weight of honey	
09011870	Honeydew melon	weight of pulp; excluding seeds and rind	
95001880	Hop	weight of dried hops	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
24001890	Horse, meat	weight of meat; excluding bone, trimmable fat or skin	
01011900	Horseradish	weight of root, ground	
13021910	Huckleberry	weight of berry	
95001920	Jaboticaba	weight of whole fruit	
95001930	Jackfruit	weight of pulp; excluding peel	
05021940	Kale	weight of leaves	Includes mizuna.
95001950	Kiwifruit	weight of pulp; excluding peel	
05011960	Kohlrabi	weight of leaves and stems	
10001970	Kumquat	weight of fruit; including peel; excluding seeds	
03001980	Leek	weight of whole plant; including leaves and bulb	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
10001990	Lemon	weight of pulp; excluding seeds and peel	
10002000	Lemon, juice	weight of juice at single strength (or standard dilution)	
10002001	Lemon, juice- babyfood	weight of juice at single strength (or standard dilution)	
10002010	Lemon, peel	weight of peel only	
19012020	Lemongrass	weight of leaves	
06032030	Lentil	dry weight of edible seed	
04012040	Lettuce, head	weight of leaves; juice	
04012050	Lettuce, leaf	weight of leaves	Includes romaine.
10002060	Lime	weight of pulp; excluding seeds and peel	
10002070	Lime, juice	weight of juice at single strength (or standard dilution)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
10002071	Lime, juice- babyfood	weight of juice at single strength (or standard dilution)	
13012080	Loganberry	weight of berry	
95002090	Longan	weight of pulp; excluding seeds and rind	
11002100	Loquat	weight of pulp; excluding seeds and skin	
95002110	Lychee	weight of pulp; excluding peel	
95002120	Lychee, dried	weight of dried pulp	
14002130	Macadamia nut	weight of nutmeat	
95002140	Mamey apple	weight of pulp; excluding peel, core and stem	Also called Mamey.
95002150	Mango	weight of pulp; excluding peel	
95002151	Mango- babyfood	weight of pulp; excluding peel	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95002160	Mango, dried	weight of dried pulp	
95002170	Mango, juice	weight of juice at single strength (or standard dilution)	
95002171	Mango, juice - babyfood	weight of juice at single strength (or standard dilution)	
95002180	Maple, sugar	dry weight of sugar	
95002190	Maple syrup	weight of syrup	
19012200	Marjoram	weight of leaves and stems	Includes oregano.
19012201	Marjoram - babyfood	weight of leaves and stems	Includes oregano.
28002210	Meat, game	weight of meat; excluding bone, trimmable fat and skin	Includes armadillo, bear, beaver, caribou, deer, elk, frog, groundhog, moose, snake, opossum, raccoon, squirrel, turtle.
27002220	Milk, fat	weight of nutrient fat only	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
27002221	Milk, fat - baby food/infant formula	weight of nutrient fat only	
27012230	Milk, nonfat solids	Remaining weight of milk after subtracting milk fat and moisture content of milk	Includes casein; also, includes lactose added to items that are neither commercial baby foods nor infant formulas.
27012231	Milk, nonfat solids- baby food/infant formula	Remaining weight of milk after subtracting milk fat and moisture content of milk	Includes casein; also, note that lactose added to commercial baby foods or infant formulas is not included here, but is represented under EPA Food Commodity Code #27032251 (Milk, sugar [lactose]-babyfood/infant formula).
27022240	Milk, water	Moisture content of milk	
27022241	Milk, water-babyfood/infant formula	Moisture content of milk	
27032251	Milk, sugar (lactose)-baby food/infant formula	dry weight of lactose-report only when consumed from baby food/infant formula	Includes lactose added to items that are either commercial baby foods or infant formulas.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
15002260	Millet, grain	dry weight of grain	
95002270	Mulberry	weight of fruit	
95002280	Mushroom	weight of caps or caps and stems	
05022290	Mustard greens	weight of leaves and stems	
12002300	Nectarine	weight of pulp; including peel; excluding pit and stem	
15002310	Oat, bran	dry weight of bran	
15002320	Oat, flour	dry weight of flour	
15002321	Oat, flour- babyfood	dry weight of flour	
15002330	Oat, groats/rolled oats	dry weight	
15002331	Oat, groats/rolled oats-babyfood	dry weight	
95002340	Okra	weight of pods; including seeds	
95002350	Olive	weight of fruit; excluding pit	
95002360	Olive, oil	weight of oil	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
03002370	Onion, dry bulb	weight of bulb; excluding outer skin	
03002371	Onion, dry bulb- babyfood	weight of bulb; excluding outer skin	
03002380	Onion, dry bulb, dried	dry weight of bulb; excluding outer skin	
03002381	Onion, dry bulb, dried- babyfood	dry weight of bulb; excluding outer skin	
03002390	Onion, green	weight of bulb or bulb and leaves	
10002400	Orange	weight of pulp; excluding seeds and peel	
10002410	Orange, juice	weight of juice at single strength (or standard dilution)	
10002411	Orange, juice- babyfood	weight of juice at single strength (or standard dilution)	
10002420	Orange, peel	weight of peel only	
95002430	Palm heart, leaves	weight of stem and leaves	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95002440	Palm, oil	weight of oil	
95002441	Palm, oil - babyfood	weight of oil	
95002450	Papaya	weight of pulp; excluding peel and seeds	
95002451	Papaya- babyfood	as above	
95002460	Papaya, dried	weight of dried pulp	
95002470	Papaya, juice	weight of juice at single strength (or standard dilution)	
04012480	Parsley, leaves	weight of leaves and stems	
19012490	Parsley, dried leaves	dried weight of leaves and stems	
19012491	Parsley, dried leaves - babyfood	dried weight of leaves and stems	
01012500	Parsley, turnip rooted	weight of roots and leaves	
01012510	Parsnip	weight of roots with or without peel	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
01012511	Parsnip - babyfood	weight of roots with or without peel	
95002520	Passionfruit	weight of pulp; excluding seeds and peel	
95002521	Passionfruit- babyfood	weight of pulp; excluding seeds and peel	
95002530	Passionfruit, juice	weight of juice at single strength (or standard dilution)	
95002531	Passionfruit, juice-babyfood	weight of juice at single strength (or standard dilution)	
95002540	Pawpaw	weight of pulp; excluding peel	
06022550	Pea, succulent	weight of peas	
06022551	Pea, succulent- babyfood	weight of peas	
06032560	Pea, dry	dry weight of pea	
06032561	Pea, dry- babyfood	dry weight of pea	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
06012570	Pea, edible podded	weight of pea and pod	
06032580	Pea, pigeon, seed	dry weight of pea	
06022590	Pea, pigeon, succulent	weight of pea	In U.S. presently, usually found as canned pigeon peas.
12002600	Peach	weight of pulp, with or without peel; excluding pit	
12002601	Peach- babyfood	weight of pulp, with or without peel; excluding pit	
12002610	Peach, dried	weight of dried pulp, with or without peel; excluding pit	
12002611	Peach, dried- babyfood	weight of dried pulp, with or without peel; excluding pit	
12002620	Peach, juice	weight of juice at single strength (or standard dilution)	
12002621	Peach, juice- babyfood	weight of juice at single strength (or standard dilution)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95002630	Peanut	weight of nutmeat; excluding shell	
95002640	Peanut, butter	weight of ground peanuts; excluding shell	
95002650	Peanut, oil	weight of oil	
12002660	Pear	weight of pulp, with or without peel; excluding core and stem	Include Oriental pear.
12002661	Pear- babyfood	weight of pulp, with or without peel; excluding core and stem	
12002670	Pear, dried	weight of dried pulp, with or without peel	
12002680	Pear, juice	weight of juice at single strength (or standard dilution)	
12002681	Pear, juice- babyfood	weight of juice at single strength (or standard dilution)	
14002690	Pecan	weight of nutmeat	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
08002700	Pepper, bell	weight of flesh; excluding seeds and stem	Includes sweet pepper, cooking pepper, pimento, and banana pepper.
08002701	Pepper, bell- babyfood	weight of flesh; excluding seeds and stem	Includes sweet pepper, cooking pepper, pimento, and banana pepper.
08002710	Pepper, bell, dried	dry weight of flesh only	Includes sweet pepper, cooking pepper, pimento, and banana pepper.
08002711	Pepper, bell, dried-babyfood	dry weight of flesh only	Includes sweet pepper, cooking pepper, pimento, and banana pepper.
08002720	Pepper, non-bell	weight of flesh, with or without seeds; excluding stem	
08002721	Pepper, non-bell, -babyfood	weight of flesh, with or without seeds; excluding stem	
08002730	Pepper, non-bell, dried	dry weight of flesh or flesh and seeds	
19022740	Pepper, black and white	dry weight of pepper	
19022741	Pepper, black and white-babyfood	dry weight of pepper	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95002750	Peppermint	weight of leaves and stems	
95002760	Peppermint, oil	weight of oil	
95002770	Persimmon	weight of entire fruit, pulp and peel	
95002780	Pine nut	weight of nutmeat	Also called pignolia.
95002790	Pineapple	weight of pulp; excluding leaves and outer peel	
95002791	Pineapple- babyfood	weight of pulp; excluding leaves and outer peel	
95002800	Pineapple, dried	weight of dried pulp only	
95002810	Pineapple, juice	weight of juice at single strength (or standard dilution)	
95002811	Pineapple, juice- babyfood	weight of juice at single strength (or standard dilution)	
14002820	Pistachio	weight of nutmeat	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95002830	Plantain	weight of pulp; excluding skin	
95002840	Plantain, dried	weight of dried pulp only	
12002850	Plum	weight of pulp with peel; excluding pit	
12002851	Plum- babyfood	weight of pulp with peel; excluding pit	
12002860	Plum, prune, fresh	weight of plum, with peel; excluding pit	
12002861	Plum, prune, fresh-babyfood	weight of plum, with peel; excluding pit	
12002870	Plum, prune, dried	weight of dried flesh, with or without peel; excluding pit	
12002871	Plum, prune, dried-babyfood	weight of dried flesh, with or without peel; excluding pit	
12002880	Plum, prune, juice	weight of juice at single strength (or standard dilution)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
12002881	Plum, prune, juice-babyfood	weight of juice at single strength (or standard dilution)	
95002890	Pomegranate	weight of pulp; excluding peel and seeds	Seeds are usually not consumed.
25002900	Pork, meat	weight of meat; excluding weight of bone, and all nutrient fat and skin	
25002901	Pork, meat- babyfood	weight of meat; excluding weight of bone, and all nutrient fat and skin	
25002910	Pork, skin	dry weight of skin including nutrient fat	e.g., pork rind snacks.
25002920	Pork, meat byproducts	weight of meat; excluding bone; may include some trimmable fat and skin	Includes ears, jowl, chitterlings, stomach (maw), fatback, and feet.
25002921	Pork, meat byproducts-babyfood	weight of meat; excluding bone; may include some trimmable fat and skin	Includes ears, jowl, chitterlings, stomach (maw), fatback, and feet.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
25002930	Pork, fat	weight of nutrient fat only; includes weight of nutrient fat from meat	
25002931	Pork, fat- babyfood	weight of nutrient fat only; includes weight of nutrient fat from meat	
25002940	Pork, kidney	weight of organ including nutrient fat	
25002950	Pork, liver	weight of organ including nutrient fat	
01012960	Potato, chips	weight of potato from chip or stick, with or without peel	
01012970	Potato, dry (granules/ flakes)	dry weight of granules or flakes	
01012971	Potato, dry (granules/ flakes)- babyfood	dry weight of granules or flakes	
01012980	Potato, flour	dry weight of flour or potato starch	
01012981	Potato, flour - babyfood	dry weight of flour or potato starch	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
01012990	Potato, tuber, w/peel	weight of tuber; including peel	
01012991	Potato, tuber, w/peel-babyfood	weight of tuber; including peel	
01013000	Potato, tuber, w/o peel	weight of tuber; excluding peel	
01013001	Potato, tuber, w/o peel-babyfood	weight of tuber; excluding peel	
60003010	Poultry, other, meat	weight of meat; excluding weight of bone, and all nutrient fat, and skin	Includes dove, duck, emu, goose, guinea hen, ostrich, partridge, pheasant, pigeon, quail, squab, wild duck.
60003020	Poultry, other, liver	weight of organ, including nutrient fat	Includes dove, duck, emu, goose, guinea hen, ostrich, partridge, pheasant, pigeon, quail, squab, wild duck.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
60003030	Poultry, other, meat byproducts	weight of meat; excluding bone; may contain some trimmable fat and/or skin	Includes dove, duck, emu, goose, guinea hen, ostrich, partridge, pheasant, pigeon, quail, squab, wild duck. Byproducts include (as for chicken) giblets, gizzard, neck, feet, and tail, as applicable to species of poultry.
60003040	Poultry, other, fat	weight of nutrient fat only; includes weight of nutrient fat from meat and skin	Includes dove, duck, emu, goose, guinea hen, ostrich, partridge, pheasant, pigeon, quail, squab, wild duck.
60003050	Poultry, other, skin	weight of nutrient fat only; includes weight of nutrient fat from meat and skin	Includes dove, duck, emu, goose, guinea hen, ostrich, partridge, pheasant, pigeon, quail, squab, wild duck.
95003060	Psyllium, seed	dry weight of psyllium husks	
10003070	Pummelo	weight of pulp; excluding seeds and peel	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
09023080	Pumpkin	weight of pulp; excluding seeds and rind	
09023090	Pumpkin, seed	weight of dried seed only	
11003100	Quince	weight of pulp; excluding seeds and peel	
95003110	Quinoa, grain	weight of grain or Quinoa flour	
29003120	Rabbit, meat	weight of meat; excluding bone, trimmable fat and skin	
04013130	Radicchio	weight of leaves	
01013140	Radish, roots	weight of roots	
02003150	Radish, tops	weight of leaves	
01013160	Radish, Oriental, roots	weight of roots	
02003170	Radish, Oriental, tops	weight of leaves	
05023180	Rape greens	weight of leaves and stems	
95003190	Rapeseed, oil	weight of oil	Includes canola oil.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95003191	Rapeseed, oil - babyfood	weight of oil	Includes canola oil.
14013200	Raspberry	weight of berry	
14013201	Raspberry- babyfood	weight of berry	
14013210	Raspberry, juice	weight of juice at single strength (or standard dilution)	
14013211	Raspberry, juice - babyfood	weight of juice at single strength (or standard dilution)	
04023220	Rhubarb	weight of stalks; excluding leaves	
15003230	Rice, white	dry weight of grain	
15003231	Rice, white- babyfood	dry weight of grain	
15003240	Rice, brown	dry weight of grain or brown rice flour	
15003241	Rice, brown- babyfood	dry weight of grain or brown rice flour	
15003250	Rice, flour	dry weight of flour	
15003251	Rice, flour- babyfood	dry weight of flour	
15003260	Rice, bran	dry weight of bran	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
15003261	Rice, bran- babyfood	dry weight of bran	
01013270	Rutabaga	weight of roots; excluding tops	
15003280	Rye, grain	dry weight of grain; weight of whole grain flour	
15003290	Rye, flour	dry weight of flour	
95003300	Safflower, oil	weight of oil	
95003301	Safflower, oil - babyfood	weight of oil	
01013310	Salsify, roots	weight of roots	Also called oyster plant.
02003320	Salsify, tops	weight of leaves	Also called oyster plant.
95003330	Sapote, Mamey	weight of pulp; excluding peel	Includes black sapote, Mamey sapote, and white sapote.
19013340	Savory	weight of leaves and flower buds	Includes summer and winter savory.
95003350	Seaweed	weight of vegetation (wet and dry)	Includes algae, Irish moss, kelp, spirulina, agar, laver, and wakame.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95003351	Seaweed - babyfood	weight of vegetation (wet and dry)	Includes algae, Irish moss, kelp, spirulina, agar, laver, and wakame.
95003360	Sesame, seed	dry weight of seed	
95003361	Sesame, seed- babyfood	dry weight of seed	
95003370	Sesame, oil	weight of oil	
95003371	Sesame, oil- babyfood	weight of oil	
03003380	Shallot	weight of bulb; excluding skin	
26003390	Sheep, meat	weight of meat; excluding weight of bone, total nutrient fat and skin	
26003391	Sheep, meat- babyfood	weight of meat; excluding weight of bone, total nutrient fat and skin	
26003400	Sheep, meat byproducts	weight of meat; excluding bones; may include some trimmable fat and/or skin	Includes brain and tongue.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
26003410	Sheep, fat	weight of nutrient fat only; includes nutrient fat from meat	
26003411	Sheep, fat- babyfood	weight of nutrient fat only; includes nutrient fat from meat	
26003420	Sheep, kidney	weight of organ, including nutrient fat	
26003430	Sheep, liver	weight of organ, including nutrient fat	
15003440	Sorghum, grain	dry weight of grain	
15003450	Sorghum, syrup	weight of syrup	
95003460	Soursop	weight of pulp; excluding peel	
06003470	Soybean, seed	dry weight of seed or bean	
06003480	Soybean, flour	dry weight of flour, soy meal, soy protein concentrate and isolate	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
06003481	Soybean, flour- babyfood	dry weight of flour, soy meal, soy protein concentrate and isolate	
06003490	Soybean, soy milk	total weight of milk	
06003491	Soybean, soy milk- babyfood or infant formula	total weight of milk	
06003500	Soybean, oil	weight of oil	Includes lecithin.
06003501	Soybean, oil- babyfood	weight of oil	Includes lecithin.
95003510	Spanish lime	weight of pulp; excluding seeds and pulp	Includes Genip.
95003520	Spearmint	weight of leaves and stems	
95003530	Spearmint, oil	weight of oil	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
19023540	Spices, other	dry weight	**See Appendix C ("The Spice List"). Note that this list does not include some spices that are unique FCs and are listed separately in this Commodity Vocabulary. These latter spices include: cinnamon, dill (seed), coriander (cilantro)-seed, and black and white pepper.**
19023541	Spices, other- babyfood	dry weight	**See Appendix C.**
04013550	Spinach	weight of leaves; juice	
04013551	Spinach- babyfood	weight of leaves; juice	
09023560	Squash, summer	weight of flesh, seeds and peel	Includes crookneck squash, kampyo, scallop squash, straightneck squash, vegetable marrow, and zucchini.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
09023561	Squash, summer- babyfood	weight of flesh, seeds and peel	Includes crookneck squash, kampyo, scallop squash, straightneck squash, vegetable marrow, and zucchini.
09023570	Squash, winter	weight of flesh; excluding seeds and peel	Includes butternut squash, hubbard squash, calabaza, acorn squash, and spaghetti squash.
09023571	Squash, winter- babyfood	weight of flesh; excluding seeds and peel	
95003580	Starfruit	weight of fruit, including seeds and peel	Also called Carambola.
95003590	Strawberry	weight of berry; excluding leaf cap	
95003591	Strawberry- babyfood	weight of berry, excluding leaf cap	
95003600	Strawberry, juice	weight of juice at single strength (or standard dilution)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95003601	Strawberry, juice - babyfood	weight of juice at single strength (or standard dilution)	
95003610	Sugar apple	weight of pulp; excluding peel	Include atemoya.
95003620	Sugarcane, sugar	dry weight of cane sugar (sucrose)	Sucrose is a disaccharide obtained from sugar cane and sugar beet.
95003621	Sugarcane, sugar- babyfood	dry weight of cane sugar (sucrose)	Sucrose is a disaccharide obtained from sugar cane and sugar beet.
95003630	Sugarcane, molasses	weight of molasses	Includes blackstrap molasses.
95003631	Sugarcane, molasses - babyfood	weight of molasses	
95003640	Sunflower, seed	dry weight of seeds	
95003650	Sunflower, oil	weight of oil	
95003651	Sunflower, oil - babyfood	weight of oil	
01033660	Sweet potato	weight of roots, with or without peel; juice	
01033661	Sweet potato- babyfood	weight of root; juice	Likely without peel.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
04023670	Swiss chard	weight of leaves and stalks	
95003680	Tamarind	weight of pulp, including seeds; excluding peel; juice	
10003690	Tangerine	weight of pulp; excluding seeds and peel	Include mandarin.
10003700	Tangerine, juice	weight of juice at single strength (or standard dilution)	Include mandarin.
01013710	Tanier, corm	weight of corm	Also called Cocoyam.
95003720	Tea, dried	dry weight of tea leaves	
95003730	Tea, instant	dry weight of powder	
08003740	Tomatillo	weight of fruit; excluding outer husks	
08003750	Tomato	weight of pulp, seeds and skin; tomatoes without skin but with seed (example canned whole tomatoes)	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
08003751	Tomato- babyfood	weight of pulp, seeds and skin; tomatoes without skin but with seed (example canned whole tomatoes)	
08003760	Tomato, paste	weight of concentrated tomato pulp from food described as paste; excluding seeds and skin	
08003761	Tomato, paste- babyfood	weight of concentrated tomato pulp (from food described as paste); excluding seeds and skin	
08003770	Tomato, puree	weight of concentrated tomato pulp (from food described as puree or sauce); excluding seeds and skin	
08003771	Tomato, puree- babyfood	weight of concentrated tomato pulp (from food described as puree or sauce); excluding seeds and skin	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
08003780	Tomato, dried	weight of dried tomato (may include skin and/or seeds)	
08003781	Tomato, dried - babyfood	weight of dried tomato (may include skin and/or seeds)	
08003790	Tomato, juice	weight of juice at single strength (or standard dilution)	
95003800	Tomato, Tree	weight of the tree tomato, excluding peel	Also called Tamarillo.
15003810	Triticale, flour	dry weight of flour	
15003811	Triticale, flour- babyfood	dry weight of flour	
50003820	Turkey, meat	weight of meat; excluding bone, all nutrient fat, and skin	
50003821	Turkey, meat- babyfood	weight of meat; excluding bone, all nutrient fat, and skin	
50003830	Turkey, liver	weight of organ, including nutrient fat	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
50003831	Turkey, liver- babyfood	weight of organ, including nutrient fat	
50003840	Turkey, meat byproducts	weight of meat; excluding bone (may contain some fat and/or skin)	Includes gizzard, heart, neck, and tail.
50003841	Turkey, meat byproducts-babyfood	weight of meat; excluding bone (may contain some fat and/or skin)	Includes gizzard, heart, neck, and tail.
50003850	Turkey, fat	weight of nutrient fat only; includes nutrient fat from meat and skin	
50003851	Turkey, fat- babyfood	weight of nutrient fat only; includes nutrient fat from meat ans skin	
50003860	Turkey, skin	weight of skin only (0 grams nutrient fat)	
50003861	Turkey, skin - babyfood	weight of skin only (0 grams nutrient fat)	
19023870	Turmeric	weight of roots	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
01013880	Turnip, roots	weight of roots	
02003890	Turnip, tops	weight of leaves and stems	
95003900	Vinegar	Weight of vinegar made from apple, grape or rice juice in recipe	
14003910	Walnut	weight of nutmeat	
86003920	Water, dilution, source NS	water to dilute or reconstitute a juice, beverage, or soup	This FC is of interest to EPA, but will be acquired by a separate mechanism.
86003930	Water, tapwater- direct (drinking)		This FC is of interest to EPA, but will be acquired by a separate mechanism.
86003940	Water- indirect (cooking)		This FC is of interest to EPA, but will be acquired by a separate mechanism.
86003950	Water, bottled water		This FC is of interest to EPA, but will be acquired by a separate mechanism.
86003960	Water, commercial beverage		This FC is of interest to EPA, but will be acquired by a separate mechanism.

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
95003970	Water chestnut	weight of tuber	
95003980	Watercress	weight of leaves and stems; juice	
09013990	Watermelon	weight of pulp and rind; excluding seeds	To include weight of pickled watermelon rind.
09014000	Watermelon, juice	weight of juice at single strength (or standard dilution)	
15004010	Wheat, grain	dry weight of grain	Include whole kernel, cracked wheat, whole grain flour, bulgur, couscous.
15004011	Wheat, grain - babyfood	dry weight of grain	Include whole kernel, cracked wheat, whole grain flour, bulgur, couscous
15004020	Wheat, flour	dry weight of flour	
15004021	Wheat, flour- babyfood	dry weight of flour	
15004030	Wheat, germ	dry weight of germ	
15004040	Wheat, bran	dry weight of bran	
15004050	Wild rice	dry weight of grain	

EPA Food Commodity (FC) Code	Food Commodity (FC)	Weight Basis of Food Commodity (FC): amount=edible portion as consumed, except where otherwise noted (in gm/kg body weight/day)	Comments
01034060	Yam, true	weight of roots	
01034070	Yam bean	weight of roots	Also called jicama.

APPENDIX A
THE FISH LIST

FISH-SALTWATER Finfish, Tuna

Tuna

FISH-SALTWATER Finfish, Other

Alewife
Anchovy
Barracuda
Bass
Bluefish
Bonita
Butterfish
Cisco
Cod
Croaker
Dolphinfish
Drum
Flatfish
Flounder
Grouper
Haddock
Halibut
Herring
Kingfish
Mackerel
Menhaden
Monkfish
Mullet
Perch
Pollock
Pompano
Porgy
Rockfish
Roe (herring, sea urchin)
Roughy
Sablefish
Sardine
Sea Trout
Scup
Shad
Shark
Skate
Smelt
Snapper
Sole
Spot
Swordfish
Tilefish
Tomcod
Whitefish
Whiting

Eel

APPENDIX A
THE FISH LIST, Continued

FISH-SHELLFISH, Crustacean

Crab
Crayfish and Crawfish
Cuttlefish
Lobster
Shrimp and Prawns

FISH-SHELLFISH, Mollusc

Abalone
Clam
Conch
Mussel
Octopus
Oyster
Scallop
Snail
Squid

FISH-FRESHWATER, Finfish

Carp
Catfish (not farm raised)
Caviar (Sturgeon)
Pike
Salmon (Chinook, Chum, Coho, Sockeye)(not farm raised)
Tilapia (not farm raised)
Sturgeon
Trout (not farm raised)

FISH-FRESHWATER, Finfish-Farm Raised

Bluegill
Catfish, Channel
Trout (Rainbow, Brook, Lake)
Tilapia
Salmon (Chinook, Chum, Coho, Sockeye)

APPENDIX B
THE HERB LIST

Angelica
Balm
Borage
Burnet
Camomile
Catnip
Chervil, dried leaves
Chinese chive
Clary
Costmary
Curry, dried leaves
Horehound
Hyssop
Lavender
Lovage, leaves
Marigold
Nasturtium
Pennyroyal
Rosemary
Rue
Sage
Sweet bay
Tansy
Tarragon
Thyme
Wintergreen
Woodruff
Wormwood

Note that <u>some herbs</u> are <u>not</u> included in this list as they <u>are unique FCs and are cited in this vocabulary separately.</u> These latter include: basil, chives, coriander (cilantro), leaf, dill (dillweed), fennel, Florence (Italian and sweet), lemongrass, marjoram (oregano), parsley, and savory (summer and winter).

APPENDIX C
THE SPICE LIST

Allspice
Anise, seed
Anise, star
Annatto, seed
Caper, buds
Caraway
Caraway, black
Cardamom
Cassia, buds
Celery, seed
Clove, buds
Cumin
Fennel, common
Fennel, Florence, seed
Fenugreek, seed
Grains of Paradise
Juniper Berry
Lovage, seed
Mace
Mustard, seed
Nutmeg
Poppy, seed
Saffron
Vanilla

Note that this list does not include some spices that are unique FCs and are listed separately in this Commodity Vocabulary. These latter spices include: cinnamon; dill, seed; coriander (cilantro), seed; black and white pepper.

APPENDIX D

SELECTED REFERENCES FOR THE EPA FOOD COMMODITY VOCABULARY

Arthey, D. and P.R. Ashust. 1996. Fruit processing. Blackie Academic and professional. NY. 248 pp.

Bebee, C.N. 1990. The Protection of Tropical and Subtropical Fruits. 1979-April 1990. USDA National Agricultural Library. BLA No. 97.

Chaisson, C., B. Petersen, S. White, and A. Clayton. 1984. Tolerance Assessment System: Crops to Food Map. US EPA. HED Toxicology branch. 58 pp.

Codex Alimentarius. 1993. Pesticides Residues in Food. Section 2. Codex Classification of Foods and Animal Feeds. FAO/WHO. Rome, Italy. Vol. 2. 218 pp.

Considine, D.M. and G.D. Considine. 1982. Foods and Food Production Encyclopedia. Van Nostrand Reinhold, Co. NY. 2305 pp.

Ensminger, M., J. Oldfield, and W.W. Heineman. 1990. Feed and Nutrition Digest. Second Ed. Regu Publication. Clovis, CA. 795 pp.

Farrell, K.T. 1990. Spices, Condiments, and Seasonings. Second Ed. AVI Book. VanNostrand Reinhold Co. NY. 244 pp.

Heaton, D.D. 1997. A Produce Reference Guide to Fruits and Vegetables From Around the World. Nature's Harvest. Food Product Press. NY. 415 pp.

Hoseney, R.C. 1994. Principles of Cereal Science and Technology. Second Ed. American Association of Cereal Chemists. St. Paul, MN. 378 pp.

Janick, J. Editor. 1996. Progress in New Crops. American Society of Horticulture Science Press. Alexandria, VA. 600 pp.

Larkcom, J. 1991. Oriental Vegetables. Kodansha, Int. NY. 232 pp.

Magness, J.R., G. M. Markle, and C.C. Compton. 1971. Food and Feed Crops of the United States. USDA IR-4. Rutgers, NJ. IR Bulletin No. 1. Bulletin 828.

Markle, G.M., J.J. Baron, and B.A. Schneider. 1998. Food and Feed Crops of the United States. Second Ed. Meister Publications. Willoughby, OH. 517 pp.

Morton, J.F. 1987. Fruits of Warm Climates. First Ed. Media Inc., Greensboro, NC. 505 pp.

Pennington, J.A.T. 1997. Bowes and Church's Food Values of Portions Commonly Used. Seventeenth Ed. Lippincot-Raven Publishers. Philadelphia, PA. 481 pp.

APPENDIX D (continued)

Randolph, S. and M. Synder. 1993. The Seafood List. US FDA . Washington, DC 68 pp. .

Raymond, E. 1999. The Packer 1999 Produce and Availability Guide. Vance Publishing Co. Lenexa, KS. 552 pp..

Schneider, Bernard A. 1992. Comparison of USDA Human Nutrition Information Service (HNIS) Legume Vegetables Vocabulary With OPP Commodity Vocabularies. US EPA. September 4.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part I: A - CN, Entries 1 - 206. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part II: CO - E, Entries 207 - 283. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part III: F - G, Entries 284 - 379. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part IV: H - K, Entries 380 - 419. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part V: L - M, Entries 420 - 508. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part VIA: O - Pears, Entries 509 - 554. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part VIB: Peas - Pumpkins, Entries 555 - 621. US EPA.

Schneider, B.A. 1993. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part VII: Q - R, Entries 622 - 663. US EPA.

Schneider, B.A. 1994. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part VIII: S, Entries 664 - 784. US EPA.

Schneider, B.A. 1994. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part IX: T, Entries 785 - 833. US EPA.

Schneider, B.A. 1994. Response to Commodities, Subcommodities, and Commodity Groups Not Yet Included In The New DRES Vocabulary. Part X: U - Z, Entries 834 - 920. US EPA.

Schneider, B.A. 1994. Response to Public Comments on Pesticide Tolerance Crop Grouping Regulation: Proposed Rule (40CFR §180). Part II: Evaluation of Supporting Data Concerning Commodity Common Names and Proposed Additions to Several Crop Groups. CB# 12719. DP Barcode D196080. US EPA. OPPTS, HED, CBTS. June 22. 15 pp.

Schneider, B.A. 1996. Uses of Brans of Wheat, Oats, Barley, and Rye As Foods. D220744. US EPA. OPP HED Feb. 14. 32 pp.

Schneider, B.A., Yuen-Shaung Ng, H. Jamerson, C. Colledge, C. Lang, I. Hornstein, R. Cromwell. 1999. EPA Food and Feed Tolerance Vocabulary. US EPA. OPP. HED.

APPENDIX D (continued)

Schneider, Elizabeth. 1985. Uncommon Fruits and Vegetables: A Common Sense Guide. Harper and Row Publishers, Inc. NY. 546 pp.

Smith, D.S., J.N. Cash, W.K. Nip, and Y.H. Hui. 1997. Processing Vegetables Science and Technology. Technomics Publ. Co. 434 pp.

U.S. EPA. 1993. Pesticide Tolerances: Portion of Food Commodities to Be Analyzed for Pesticide Residues. Proposed Rule. OPP-3000243. Federal Register Vol. 58. No. 187. Pp. 50888-50893. September 23.

U.S. EPA. 1993. Table 1. Raw Agricultural and Processed Commodities and Feedstuffs Derived from Crops. Subdivision G: Residue Chemistry Guidelines. OPPTS860.1000