United States Department of Agriculture

Agricultural Research Service

NFS Report No. 96–1

December 1997

Design and Operation: The Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey, 1994–96

United States Department of Agriculture

Agricultural Research Service

NFS Report No. 96–1

December 1997

Design and Operation: The Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey, 1994–96

Katherine S. Tippett and Yasmin S. Cypel, editors

Abstract

Tippett, Katherine S. and Yasmin 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, Nationwide Food Surveys Report No. 96–1, 240pp.

This publication is the official reference for information on the design and operation of the Continuing Survey of Food Intakes by Individuals (CSFII) and the Diet and Health Knowledge Survey (DHKS) 1994–96 which were conducted by the Agricultural Research Service, U.S. Department of Agriculture. This report describes the planning and development of the CSFII/DHKS 1994–96 and their design and operation. Included is information on the sample design, questionnaire development and data collection procedures, survey management and quality control, and food data processing. Survey response and sample weight information are also included. The appendices contain publicity materials, survey instruments, measuring guides, and followup materials.

The CSFII 1994–96, which measures the kinds and amounts of foods eaten by individuals, is USDA's 10th nationwide food consumption survey, the sixth to include the collection of individual intake data. It addresses the requirements of the National Nutrition Monitoring and Related Research Act of 1990 (Public Law 101–445) for continuous monitoring of the dietary and nutritional status of the U.S. population. The Diet and Health Knowledge Survey (DHKS) is a telephone followup to the CSFII and is designed to measure attitudes and knowledge about diet and health. The surveys are popularly referred to as the "What We Eat in America Survey."

Keywords: dietary surveys, food, food codes, food consumption, health knowledge, nutrients, nutrition attitudes, nutrition monitoring, survey methodology.

Copies of this publication may be purchased from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161; phone (703) 487–4650.

Issued December 1997

Contents

[The appendixes listed below are available in the hard copy version of this report or on the CD-ROM containing microdata for the 1994-96 CSFII/DHKS. They are not available in this pdf file.]

Acknowledgments

The Food Surveys Research Group acknowledges the leadership and work of Dr. Ellen Harris, assistant director, Beltsville Human Nutrition Research Center, ARS, in the planning and development of the survey. Thanks are due also to Dr. Kenneth D. Fisher, Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services; Mr. Steven Lutz, Economic Research Service, U.S. Department of Agriculture; and Dr. Nancy A. Mathiowetz, Joint Program in Survey Methodology, University of Maryland, for their peer reviews of this report.

Contributors

Lori G. Borrud, supervisory nutritionist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Adam Chu, survey statistician, Westat, Incorporated.

Linda E. Cleveland, nutritionist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Joseph D. Goldman, mathematical statistician, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Patricia M. Guenther, nutritionist, formerly with Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Linda A. Ingwersen, home economist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Sharon J. Mickle, nutritionist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Alvin B. Nowverl, mathematical statistician, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Betty P. Perloff, supervisory nutritionist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Rhonda S. Sebastian, home economist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Katherine E. Sykes, nutritionist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Junko A. Tamaki, nutritionist, Agricultural Research Service, U.S. Department of Agriculture, Riverdale, MD.

Abbreviations

ARS	Agricultural Research Service
BOC	Bureau of the Census
CPS	Current Population Survey
CSFII	Continuing Survey of Food Intakes by Individuals
CSMR	Center for Survey Methods Research
CSREES	Cooperative State Research, Education, and Extension Service
CSUG	Continuing Survey Users' Group
CV	Coefficients of Variation
DHKS	Diet and Health Knowledge Survey
DSD	Demographic Surveys Division
DU	Dwelling Unit
EPA	Environmental Protection Agency
ERS	Economic Research Service
FASEB	Federation of American Societies for Experimental Biology
FCS	Food and Consumer Service
FDA	Food and Drug Administration
FIB	Food Instruction Booklet
FMS	Field Management System
FSIS	Food Safety and Inspection Service
FTC	Federal Trade Commission
FTS	Forms-Tracking System
HFCS	Household Food Consumption Survey
HNIS	Human Nutrition Information Service
IBNMRR	Interagency Board for Nutrition Monitoring and Related Research
ISU	Iowa State University
MSA	Metropolitan Statistical Areas
NALS	National Adult Literacy Survey
NASS	National Agricultural Statistics Service
NCHS	National Center for Health Statistics
NCI	National Cancer Institute
NDBSR	Nutrient Data Base for Standard Reference
NFCS	Nationwide Food Consumption Survey
NFS	Not Further Specified
NHANES	National Health and Nutrition Examination Survey
NHIS	National Health Interview Survey
NHLBI	National Heart, Lung and Blood Institute
NMFS	National Marine Fisheries Service
NNMRPP	National Nutrition Monitoring and Related Research Program
NPS	National Program Staff
OMB	Office of Management and Budget
PDS	Primary Data Set

PHS	Public Health Service
PSU	Primary Sampling Unit
SP	Sample Person
TIGER	Topologically Integrated Geographic Encoding and Referencing
USDA	United States Department of Agriculture
USDHHS	United States Department of Health and Human Services
USDOC	United States Department of Commerce

Chapter 1: Introduction and Overview
By Lori G. Borrud

This report provides reference information about the design and operation of two nationwide surveys conducted by the Agricultural Research Service (ARS), U.S. Department of Agriculture (USDA) from 1994 through 1996: the Continuing Survey of Food Intakes by Individuals (CSFII) and the Diet and Health Knowledge Survey (DHKS).[1] The CSFII and DHKS 1994–96 are the most recent in a series of USDA surveys designed to measure the kinds and amounts of foods eaten by Americans and their attitudes and knowledge about diet and health. The CSFII and DHKS were planned so that information collected on dietary attitudes and knowledge (DHKS) could be linked with information collected on food choices and nutrient intakes (CSFII). The surveys are popularly known as the "What We Eat in America Survey," but will be referred to in this report as the CSFII/DHKS 1994–96.

One mission of USDA is to encourage the production and availability of a sufficient, safe, and nutritionally adequate supply of food for Americans. In support of this mission, USDA conducts surveys to monitor food use and food consumption patterns in the U.S. population. The CSFII provides detailed data that serve as benchmarks of the food and nutrient intakes of the general and low-income populations. Users of the survey data include Federal Government agencies, such as the Environmental Protection Agency, Food and Drug Administration, and Federal Trade Commission; state agencies and larger county health departments; food and agricultural industries; and university researchers. The data are used to determine the food choices Americans make and to evaluate the content and adequacy of their diets in relationship to the *Dietary Guidelines for Americans* (USDA and U.S. Department of Health and Human Services 1995) and other Federal Government statements of dietary policy, such as the year 2000 nutrition objectives (U.S. Department of Health and Human Services 1991). Health interventions, such as the National Cancer Institute's "5 a day program" (Subar 1992) and nutrition education materials, such as the *Food Guide Pyramid* (USDA 1992) are developed and targeted based on survey results. In other food- and nutrition-related program and public policy applications, the data are used to

1. The CSFII and DHKS 1994–96 were planned by the Human Nutrition Information Service (HNIS), USDA. On February 20, 1994, legislation passed by Congress moved the functions of HNIS to the Agricultural Research Service, USDA.

assess the nutritional impact of USDA's food assistance programs; to estimate exposure to pesticide residues, food additives, and contaminants; and to develop food fortification, enrichment, and food labeling policies. Other uses include assessment of the demand for agricultural products and marketing facilities and food- and nutrition-related research.

The DHKS was the first national survey of dietary attitudes and knowledge in which the results could be linked to food and nutrient intakes gathered from the same individuals in the CSFII. The DHKS provides information on people's attitudes and knowledge about the *Dietary Guidelines for Americans* and the ability to put them into practice (USDA and U.S. Department of Health and Human Services 1995). It provides information on people's perceived adequacy of their own food and nutrient intakes, their knowledge about the recommendations in the *Food Guide Pyramid* (USDA 1992), the personal importance they place on dietary guidance messages, their awareness of relationships between diet and health, and their use and understanding of food labels. This type of information is used to improve understanding of factors that affect food intake. In particular, it is used by nutrition educators in academia and the Federal Government to identify ways to implement dietary guidance effectively. Federal Government agencies that use the data include USDA's Center for Nutrition Policy and Promotion and Economic Research Service, and U.S. Department of Health and Human Services' Food and Drug Administration.

The CSFII and DHKS are major components of the National Nutrition Monitoring and Related Research Program, which consists of a set of related Federal activities intended to provide information on the dietary and nutritional status of the U.S. population (U.S. Department of Health and Human Services and USDA 1993). The CSFII addresses the requirements of the National Nutrition Monitoring and Related Research Act of 1990 (Public Law 101–445) for continuous monitoring of the dietary and nutritional status of the U.S. population. CSFII data have been an integral part of the reports to Congress on nutrition monitoring in the United States (U.S. Department of Health and Human Services and USDA 1986, 1989 and Federation of American Societies for Experimental Biology 1995).

USDA has collected national information on food intakes by individuals since 1965 when a supplement to the 1965–66 Household Food Consumption Survey was added (USDA–ARS 1972). In that survey, food intake information was collected from selected household members using a 24-hour dietary recall. Individual intake data were also collected in the Nationwide Food Consumption Surveys in 1977–78 (USDA–HNIS 1983, 1984) and in 1987–88 (USDA–HNIS 1993); information on individual intakes was collected over 3 consecutive days with a 24-hour dietary

recall and a 2-day diet record. The 1965–66 Household Food Consumption Survey and the 1977–78 and 1987–88 Nationwide Food Consumption Surveys included a household food use component also.

The Continuing Survey of Food Intakes by Individuals, initiated in 1985, was the first national USDA survey of dietary intake by individuals that was independent of a household food use component (USDA–HNIS 1985). The CSFII is intended to provide frequent information on the dietary status of individuals in the U.S. population. In 1985 and 1986, the CSFII included women 19 through 50 years of age and their children 1 through 5 years (USDA–HNIS 1985, 1987, 1988). In both years, data were obtained through 6 nonconsecutive 24-hour dietary recalls that were administered every 2 months throughout the year. The first recall was administered through an in-person interview, and the remaining recalls were conducted through a telephone interview. Data were collected for the total and low-income populations in both survey years. The CSFII 1985 also included the collection of 1 day of dietary data from men 19–50 years (USDA–HNIS 1986).

The CSFII was conducted again in 1989–91 (Tippett et al. 1995). Dietary information was collected from all members of sample households for 3 consecutive days. In 1989–91, the DHKS was conducted as a telephone followup to the CSFII. Individuals identified as main meal planners/preparers were contacted about 6 weeks after completing the CSFII to answer a series of questions about knowledge and attitudes toward diet, health, and food safety (Cypel et al. 1996).

The CSFII/DHKS 1994–96 is USDA's 10th nationwide survey, the sixth to include the collection of individual intake data (table 1). The development of the CSFII/DHKS 1994–96 included substantial research and planning, as well as extensive collaboration with other organizations within and outside the Federal sector (see chapter 2). These collaborations included interagency agreements with the U.S. Bureau of the Census and USDA's National Agricultural Statistics Service; the establishment of a Continuing Survey Users' Group; and a cooperative agreement with the University of Texas–Houston School of Public Health. These research and planning activities provided the base for the survey procedures and materials utilized in the 1994–96 CSFII/DHKS.

The CSFII/DHKS 1994–96 included a target population of noninstitutionalized individuals in all 50 states and Washington, DC, and subsampling within households. In comparison with earlier surveys, the CSFII/DHKS 1994–96 included an oversampling of the low-income population, rather than separate low-income samples; a larger sample in selected sex-age categories, specifically young

children and the elderly; and the collection of knowledge, behavior, and attitude data from one adult in the household, rather than targeting the main meal planner/preparer.

In 1992, Westat, Incorporated, a private research firm in Rockville, Maryland, was awarded the contract for the CSFII 1994–96 after competitive bidding. Westat designed the sample to meet the precision and sample yield requirements specified by USDA; designed and formatted the questionnaires; worked collaboratively with USDA to further develop the survey methods and procedures; developed the promotional materials, and, as specified by the contract, conducted a pilot study of all survey procedures, operations, and materials. USDA defined the information to be collected; monitored the contract; provided technical information, such as food codes and gram weights of standard measures of food for use in coding; and provided Survey Net, a computer-assisted system for the coding of reported food intakes. Westat collected the information; edited, coded, and keyed the data; and created the sample weights. USDA did additional editing and coding; maintained the nutrient database and applied nutrient values to the food intake data; and prepared the data for public release.

Three years of continuous data collection began in 1994. In each survey year, a nationally representative sample of individuals of all ages was asked to provide food intakes for 2 nonconsecutive days (day 1 and day 2) through the administration of in-person, 24-hour dietary recalls spaced 3–10 days apart. In each CSFII household, the DHKS was administered to one adult 20 years of age or over who had provided at least 1 day of intake data. The DHKS was administered by telephone about 2 to 3 weeks after the CSFII intake data were collected.

This report contains information on all phases of the CSFII/DHKS 1994–96 design and operation. Chapter 2 provides information on the planning and development of the surveys. Chapters 3 through 8 present detailed descriptions of the sample design, data collection procedures, response results, management and quality control procedures, data-processing procedures, and calculation of sample weights. Copies of publicity materials, food measurement guides, and the 1996 survey instruments are included in the appendices.

References

Cypel, Y.S., J.A. Tamaki, C.W. Enns, et al. 1996. Nutrition attitudes and dietary status of main meal planners/preparers, 1989–91. U.S. Department of Agriculture, Agricultural Research Service, Nationwide Food Survey Report No. 91–1.

Federation of American Societies for Experimental Biology, Life Sciences Research Office. 1995. Third report on nutrition monitoring in the United States. U.S. Government Printing Office, Washington, DC.

Subar, A.S., J. Heimendinger, S.M. Krebs-Smith, et al. 1992. 5 a day for better health: A baseline study of Americans' fruit and vegetable consumption. National Institutes of Health, National Cancer Institute. Washington, DC.

Tippett, K.S., S.J. Mickle, J.D. Goldman, et al. 1995. Food and nutrient intakes by individuals in the United States, 1 day, 1989–91. U.S. Department of Agriculture, Agricultural Research Service, Nationwide Food Survey Report No. 91–2.

U.S. Department of Agriculture. 1992. The food guide pyramid. Home and Garden Bulletin No. 252.

U.S. Department of Agriculture, Agricultural Research Service. 1972. Food and nutrient intake of individuals in the United States, spring 1965. Household Food Consumption Survey 1965–66, Report No. 11.

U.S. Department of Agriculture, Human Nutrition Information Service. 1983. Food intakes: Individuals in 48 states, year 1977–78. Nationwide Food Consumption Survey 1977–78, NFCS Report No. I–1.

U.S. Department of Agriculture, Human Nutrition Information Service. 1984. Nutrient intakes: Individuals in 48 states, year 1977–78. Nationwide Food Consumption Survey 1977–78, NFCS Report No. I–2.

U.S. Department of Agriculture, Human Nutrition Information Service. 1985. Nationwide food consumption survey, continuing survey of food intakes by individuals: Women 19–50 years and their children 1–5 years, 1 day, 1985. Nationwide Food Consumption Survey, CSFII Report No. 85–1.

U.S. Department of Agriculture, Human Nutrition Information Service. 1986. Nationwide food consumption survey, continuing survey of food intakes by individuals: Men 19–50 years, 1 day, 1985. Nationwide Food Consumption Survey, CSFII Report No. 85–3.

U.S. Department of Agriculture, Human Nutrition Information Service. 1987. Nationwide food consumption survey, continuing survey of food intakes by individuals: Women 19–50 years and their children 1–5 years, 1 day, 1986. Nationwide Food Consumption Survey, CSFII Report No. 86–1.

U.S. Department of Agriculture, Human Nutrition Information Service. 1988. Nationwide food consumption survey, continuing survey of food intakes by individuals: Women 19–50 years and their children 1–5 years, 4 days, 1986. Nationwide Food Consumption Survey, CSFII Report No. 86–3.

U.S. Department of Agriculture, Human Nutrition Information Service. 1993. Food and nutrient intakes by individuals in the United States, 1 day, 1987–88. Nationwide Food Consumption Survey 1987–88, NFCS Report No. 87–I–1.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. 1995. Nutrition and your health: Dietary guidelines for Americans, 4th ed. U.S. Department of Agriculture, Home and Garden Bulletin No. 232.

U.S. Department of Health and Human Services and U.S. Department of Agriculture. 1986. Nutrition monitoring in the U.S.—A report from the Joint Nutrition Monitoring Evaluation Committee. Public Health Service, DHHS Publication No. (PHS) 86–1255.

U.S. Department of Health and Human Services and U.S. Department of Agriculture. 1989. Nutrition monitoring in the U.S.—An update report on nutrition monitoring. Public Health Service, DHHS Publication No. (PHS) 89–1255.

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

U.S. Department of Health and Human Services, Public Health Service. 1991. Healthy people 2000: National health promotion and disease prevention objectives. DHHS Publication No. (PHS) 91–50212.

Table 1. Overview of USDA dietary intake surveys, 1965–1996

Survey	Population	Sample	Type of data collected	Dietary method
1965–66 Household Food Consumption Survey	48 states	Two separate samples (basic and low income); selected household members were asked to provide intake information	Household food use; individual intake	7 consecutive days, food list-recall; 24-hr. dietary recall
1977–78 Nationwide Food Consumption Survey (NFCS)	48 states	Two separate samples (basic and low income); all household members were asked to provide intake information	Household food use; individual intake	7 consecutive days, food list-recall; 3 consecutive days: 24-hr. dietary recall and 2-day diet record
1985–86 Continuing Survey of Food Intakes by Individuals (CSFII)	48 states	Two separate samples (basic and low income); women 19–50 yr. and their children 1–5 yr. in both years and men 19–50 yr. in 1985 only	Individual intake	Women and children: 6 nonconsecutive 24-hr. dietary recalls; day 1 in person and remaining days by telephone. Men: day 1 only
1987–88 NFCS	48 states	Two separate samples (basic and low income); all household members were asked to provide intake information	Household food use; individual intake	7 consecutive days, food list-recall; 3 consecutive days: 24-hr. dietary recall and 2-day diet record
1989–91 CSFII	48 states	Two separate samples (basic and low income); all household members were asked to provide intake information	Individual intake	3 consecutive days: 24-hr. dietary recall and 2-day diet record
1989–91 Diet and Health Knowledge Survey (DHKS)	48 states	Main meal planners/preparers with a completed day-1 intake in CSFII	Dietary knowledge, behavior, and attitudes	Telephone followup to CSFII

Table 1.—Overview of USDA dietary intake surveys, 1965–1996—Continued

Survey	Population	Sample	Type of data collected	Dietary method
1994–96 CSFII	50 states	Oversampling of the low-income population; only selected household members were asked to provide intake information	Individual intake	2 nonconsecutive 24-hr. dietary recalls
1994–96 DHKS	50 states	Adults 20 yr. and over with a completed day-1 intake in CSFII	Dietary knowledge, behavior, and attitudes	Telephone followup to CSFII

Chapter 2: Planning, Research, and Development
By Lori G. Borrud and Betty P. Perloff

The CSFII/DHKS 1994–96 was preceded by a variety of planning, research, and development activities, including an evaluation of previous survey operations, collaborative efforts, a contracting process, and a pilot study.

Evaluation of Previous Survey Operations

Planning for the CSFII/DHKS 1994–96 began in 1991 with evaluation of the management, content, and procedures of the CSFII/DHKS 1989–91 and the Nationwide Food Consumption Survey 1987–88 conducted by USDA. Every facet of the survey process from the sample design and data collection to the transmission of data to ARS was examined and redesigned, as necessary, to improve the quality of the data and the timeliness of its release. Staff working groups on survey design and planning and on questionnaire development were established to plan and design the survey and to revise the CSFII/DHKS 1991 questionnaires and other interview aids for use in the 1994–96 survey. These groups continued to meet throughout the survey planning period.

Collaborative Efforts

Collaborative efforts included the establishment of the Continuing Survey Users' Group (CSUG); an interagency agreement with the National Agricultural Statistics Service for statistical consultation; two interagency agreements with the Bureau of the Census for consultation, collaboration, and research on survey design, conduct, and management; and a cooperative research agreement with the University of Texas-Houston School of Public Health for the development of Survey Net, a computer-assisted food coding system for use with the CSFII 1994–96. A cooperative research agreement with Iowa State University included consultation on survey design.

Federal collaboration

The mission of the Food Surveys Research Group of the ARS Beltsville Human Nutrition Research Center is to "monitor and assess food consumption and related behavior of the U.S. population by conducting surveys and providing the resulting information for food- and nutrition-related programs and public policy decisions." To ensure that the survey was responsive to the program and policy needs of Federal agencies, USDA convened the Continuing Survey Users' Group**.**

Proposed changes to the CSFII/DHKS 1991 questionnaires in preparation for CSFII/DHKS 1994–96 provided the basis for the meetings of the CSUG. CSUG served as the principal source of Federal user input on the survey design (see chapter 3) and on survey questionnaires (see chapter 4).

CSUG included representatives from Federal agencies that are major users of USDA food consumption survey data or that conduct major Federal surveys. CSUG members included

- Bureau of the Census, U.S. Department of Commerce
- Cooperative State Research, Education, and Extension Service, USDA
- Economic Research Service, USDA
- Environmental Protection Agency
- Federal Trade Commission
- Food and Drug Administration, U.S. Department of Health and Human Services
- Food and Consumer Service, USDA
- Food Safety and Inspection Service, USDA
- National Cancer Institute, U.S. Department of Health and Human Services
- National Center for Health Statistics, U.S. Department of Health and Human Services
- National Heart, Lung and Blood Institute, U.S. Department of Health and Human Services
- National Marine Fisheries Service, U.S. Department of Commerce
- National Program Staff, Agricultural Research Service, USDA.

Meetings with CSUG to discuss proposed changes to the CSFII/DHKS began in late 1991. In addition to full meetings for general discussion of the questionnaires and survey design, several smaller meetings were held to discuss specific topic areas. For example, the Food and Drug Administration participated in the development of food labeling questions for the DHKS that were consistent with new regulations. The Food and Drug Administration, Environmental Protection Agency, and National Heart, Lung and Blood Institute participated in the development of new questions related to sources of water and recreational and subsistence fishing. Individual CSUG members were consulted throughout the questionnaire development process. For example, Food and Consumer Service and Economic Research Service were consulted on the development of the list of sources of foods.

Comparability among the many Federal surveys that are a part of the National Nutrition Monitoring and Related Research Program is important for enhancing the capability to establish links among the surveys. The need for such

comparability was considered during the planning and development of the CSFII/DHKS 1994–96 and was balanced against current user needs.

The recommendations of the Survey Comparability Working Group of the Interagency Board for Nutrition Monitoring and Related Research (IBNMRR) on population descriptor variables were reviewed and incorporated into the questionnaires, where appropriate. For example, race and ethnicity questions provided the categories recommended in the report from the Survey Comparability Working Group (Interagency Board for Nutrition Monitoring and Related Research 1992). Also, comparability of food and nutrient intake between the CSFII 1994–96 and the Third National Health and Nutrition Examination Survey was enhanced by the 24-hour dietary recall method developed for the CSFII 1994–96. Another example of improved comparability was a change in the target population for the CSFII/DHKS 1994–96 to include all 50 states, rather than the 48 states targeted in the past. The definition of a housing unit was changed to conform with the definition for the 1991 Current Population Survey, Bureau of the Census. The latter two changes were recommended by the Research Triangle Institute in its report on USDA and National Center for Health Statistics dietary survey sampling designs (Hubbard et al. 1991).

The National Agricultural Statistics Service provided statistical expertise and consultation in the development of the survey sample design. Review of the sample design for CSFII 1994–96 was also provided by Iowa State University (ISU).

The decision to collect 2 days of dietary intake data was based in part on concerns regarding respondent burden. The decision was also based on analyses performed by Iowa State University supporting the use of 2 days of dietary intake to estimate distributions of usual nutrient intakes (Fuller 1994). Analyses performed by Iowa State University supported the decision to separate the 2 days of intake by at least 3 days to achieve statistical independence (An and Carriquiry 1991).

Staff of the Demographic Surveys Division, Bureau of the Census, helped develop the statement of work and the independent government cost estimates needed in the contracting process. The division also provided Census Bureau listing rules, instruction manuals, and other reference materials. It helped prepare early drafts of the questionnaires and pretested the DHKS questionnaire. The Census Bureau's Center for Survey Methods Research conducted cognitive research to improve the quality of the dietary data that were to be collected in the 1994–96 survey (see chapter 4).

University of Texas-Houston School of Public Health

The cooperative research agreement that began in 1990 with the University of Texas-Houston School of Public Health yielded improved methodology for food coding and for the management of the extensive technical support system required for food coding and nutrient analysis of the data. Several new inhouse software products and database management procedures, utilizing the latest computer technology, resulted from this agreement and greatly increased the efficiency of processing the data.

Survey Net is the most visible of the software products. It is a computer-assisted food coding system operating on a network that allows immediate access by all users to central technical databases, primarily the survey food coding, nutrient, and recipe databases. The CSFII contract specified that this system be installed and used by the contractor for coding all foods reported as consumed by survey respondents. The system was also used by ARS during the survey to review and finalize coded food data.

During the initial planning phase for Survey Net, the following goals were established: (1) to reduce the time required for coding food consumption data, (2) to provide flexibility in the food coding system, while maintaining a structure comparable to past surveys, and (3) to increase oversight of food coding operations conducted by survey contractors. Features were designed to meet these goals (see chapter 7).

Cooperation with the University of Texas also produced a new recipe processing system for managing the recipe database and calculating the nutrient content of recipes; a program for nutritional analysis of Survey Net output; administration software for tracking and managing Survey Net files containing unknown foods and recipe modifications; new relational database formats and routines for managing 18 data files associated with the 3 central databases (over 60 megabytes of data); and a data conversion pipeline for moving updated database files from the management arena into Survey Net. In addition, the Food Intake Analysis System was developed to provide user-friendly software, comparable to Survey Net, for researchers who wish to use CSFII technical databases for their own research (see chapter 7).

Contracting Process

Results from the staff review of survey operations and suggestions from collaborating agencies were incorporated into the statement of work, which was

the basis for the request for proposals for the CSFII/DHKS 1994–96. Staff from the National Center for Health Statistics, Demographic Surveys Division, and National Agricultural Statistics Service reviewed the statement of work. USDA followed standard government contracting procedures. In September 1992, Westat was awarded the contract to conduct the CSFII/DHKS 1994–96.

The contract for the survey was fixed-price and included 175 deliverables per data collection year. It established strong management and quality control procedures and redesigned data collection methods and procedures for processing. The contract specified annual response rates for each type of interview and outlined procedures for achieving those rates. It also specified the number of individuals to be interviewed in the 40 sex-age and income subdomains for 3 years to meet precision requirements (see chapter 3).

Pilot Study

As part of the contract, a pilot study duplicating the planned survey design on a small scale was conducted from April to June 1993. CSUG members reviewed the pilot study questionnaires. The Office of Management and Budget requested that the pilot study include a test of the effects of a respondent incentive—an insulated nylon sack—on response rates. USDA had included an option for a third nonconsecutive day of food intake data collection in the pilot study; however, due to concerns about maintaining optimal response rates with the increase in respondent burden, this option was not exercised.

Questionnaires, data collection methods, field management procedures and instructional manuals, data entry and processing procedures, and electronic data delivery planned for use in the CSFII/DHKS 1994–96 were tested during the pilot study, evaluated, and revised as appropriate. Twenty-two interviewers and two senior interviewers were trained for the pilot study in a 7-day session. The training provided extensive hands-on practice with the questionnaires and interview aids through the scripted mock interviews and practice interviews (see chapter 6).

The pilot study was conducted at 10 sites across 4 regions of the country and in urban and rural areas. To capture experiences and provide feedback on procedures and materials, interviewers maintained a notebook as part of their review of completed questionnaires, mailed back a debriefing questionnaire, and attended an in-person debriefing. Taped interviews and field observations of interviews provided additional information for identifying areas for improvement. The interviews were observed by USDA and Westat staff in the respondents' homes.

Between April and June 1993, 434 sample persons were identified for the pilot study based on 623 screened households (98.4-percent response rate). Of the 434, 351 completed day-1 and day-2 individual intake interviews, for a response rate of 81 percent. Additionally, 188 household questionnaires were completed for an 87 percent response rate, and 115 DHKS questionnaires were completed for an 88 percent response rate.

The pilot study provided an excellent opportunity to further refine the quality of survey instruments and improve the efficiency of survey operations. Modifications that had been made to the CSFII 1989–91 questionnaires and data collection procedures tested in the pilot study facilitated the collection of high-quality data in the main survey. Feedback from interviewer field notebooks, mailed questionnaires, and in-person debriefings after the pilot study resulted in further revisions to the questionnaires and procedures. High response rates were achieved. Also, Survey Net was an efficient system for processing dietary intake data more quickly and precisely. Electronic delivery of survey data from the contractor to USDA each week was successful. Automated data-tracking systems, developed by Westat for use with the CSFII/DHKS, were effective in monitoring data collection and processing activities, and survey management software used in monitoring the status of data review and editing was a valuable tool. The incentive was approved by the Office of Management and Budget for use in the CSFII/DHKS 1994–96.

References

An, B., and A.L. Carriquiry. 1991. Estimation of the correlation among days for transformed nutrient intake. Iowa State University, Unpublished report prepared for the Human Nutrition Information Service, U.S. Department of Agriculture.

Fuller, W.A. 1994. Estimation in the presence of measurement error: Dietary assessment research series I. Center for Agricultural and Rural Development, Iowa State University, Staff Report No. 94–SR73.

Hubbard, M., E. deLeon-Wong, D. Hungerford, et al. 1991. Sampling designs and population descriptors of nationwide food consumption surveys and national health and nutrition examination surveys. Final report. Prepared for the Department of Health and Human Services, National Nutrition Monitoring System Steering Committee. Research Triangle Institute, Research Triangle Park, NC.

Interagency Board for Nutrition Monitoring and Related Research. 1992. Improving comparability in the national nutrition monitoring and related research program: Population descriptors. Report of the IBNMRR Survey Comparability Working Group. Photocopy.

Chapter 3: Sample Design
By Adam Chu and Joseph D. Goldman

The primary goal of the sample design for the CSFII/DHKS 1994–96 was to obtain a nationally representative sample 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 (an "all-income" group and a "low-income" group) that met specified precision levels for estimates of mean day-1 saturated fat and iron intakes.[1] The specific precision goals required the coefficients of variation (*CV*'s) for mean saturated fat and iron intakes to be 3 percent or less for each of the 20 all-income sex-age domains and to be 5 percent or less for each of the 20 low-income sex-age domains. These precision goals were translated by Westat into 3-year sample size targets (table 2). 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. The design of the 3-year sample was such that the annual portions of the sample were roughly equal in size over the 40 analytic domains, and each year was nationally representative.

A complex, multistage, area probability sample design was used to select persons for the intake and DHKS interviews. The sample design was based on a Westat master sample that existed before the contract for the CSFII/DHKS 1994–96 was awarded.[2] The design included the selection of geographical areas called primary sampling units (PSU's), area segments within the sampled PSU's, households within the selected segments, and sample persons (SP's) within the households. The major features of the design are summarized below:

- The first-stage sample was a stratified sample of 62 PSU's consisting of metropolitan statistical areas (MSA's) or groups of counties. PSU's were selected within strata of approximately equal size, with probabilities proportional to the 1990 population.

---

1. For the CSFII/DHKS 1994–96, a single sample was selected that met precision requirements by income level. This differs from past CSFII/DHKS surveys where a separate sample of low-income persons was chosen in addition to the basic sample.

2. Persons living in group quarters or institutions, residing on military installations, and the homeless were excluded.

- Thirty-six area segments (consisting of census blocks or groups of blocks) were selected from each PSU, for a total of 2,232 area segments for the 3-year survey. The 36 segments selected from each PSU were divided into 12 sets of 3 segments each, and a set of 3 segments per PSU was assigned to each of the 12 quarters of the 3-year survey period.

- Within the sampled segments, lists of dwelling units (DU's) were prepared by Westat interviewers. More than 100,000 DU's were listed for each year of the survey. A self-weighting sample was selected from each listing. Approximately 9,500 DU's were selected for the first year, approximately 11,500 were selected for the second year, and approximately 12,000 were selected for the third year. The increased numbers of DU's selected did not necessarily result in increased numbers of SP's. Sampling rates also changed throughout the survey (see "Derivation of sampling rates and sampling messages" below).

- Within the occupied DU's identified during screening, households were identified and household members eligible for the survey were selected by a probability sampling process designed to achieve the specified sample sizes for various sex-age-income domains (see table 2).

- From households containing SP's 20 years of age or older who completed the day-1 intake interview, one SP was randomly selected for the DHKS.

Selection of Primary Sampling Units

At the first stage of sampling, the entire United States was divided into PSU's consisting of MSA's, counties, or groups of counties. The sampling frame of PSU's was created from county-level data contained in the 1990 Census Public Law 94–171 (Public Law 94) and data files from the Bureau of Economic Analysis, U.S. Department of Commerce (U.S. Department of Commerce-Bureau of the Census 1991a). The Public Law 94 data file provided county-level population counts by race and Hispanic origin, while the Bureau of Economic Analysis file provided the corresponding income information.

Because of their size, the New York MSA was divided into three PSU's and the Los Angeles and Chicago MSA's were each divided into two PSU's. Each of the other MSA's constituted a single PSU. Counties outside MSA's were grouped, as necessary, to form PSU's that (1) had a minimum 1990 population of 15,000 people, (2) were as internally heterogeneous as possible, and (3) were still small enough to permit convenient travel across the PSU by interviewers. From the

more than 3,000 counties in the United States, a total of 1,404 PSU's was created, and 62 PSU's were selected for use in the CSFII/DHKS 1994–96.

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 used to select the noncertainty PSU's included the region of the country (four census regions), 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 distribution of the sampled PSU's by census region and MSA status is summarized in table 3. The nature of the PSU's does not allow for state-level estimates.

Selection of Area Segments

The second-stage sampling units were area segments, which were defined to be individual census blocks or a group of blocks. A sample of 36 area segments was randomly selected from each PSU, with probability proportional to population. The 36 segments were then divided into 12 sets of 3 segments each, and a set of 3 segments per PSU was assigned to each of the 12 quarters of the 3-year survey period. Segments were assigned to the quarters of the year in a balanced, random manner to ensure a wide spread of the segment sample within each quarter for each PSU. This balanced sampling was carried out to improve sampling precision by reducing the design effects resulting from the homogeneity of persons within segments. This method also achieved the general sample design requirement of having data collection spread evenly over the 3 years of the survey and over the quarters of the year.

As part of the sampling process, a frame of area segments for each of the 62 sample PSU's was created. This frame was constructed from the Census Bureau's 1990 Public Law 94 datatape, which contains population, housing counts, and limited geographic information for each block in the United States (U.S. Department of Commerce–Bureau of the Census 1991a). To ensure that the segments would be of sufficient size for use in sampling, small blocks were combined with adjacent blocks to form segments that had a minimum expected size of 60 DU's. After the frame was constructed, the area segments were sorted before sample selection into minority strata (based on black and Hispanic households) and geographically within minority strata. For each of the 3 years of

the study, a systematic sample of 12 area segments was selected from the sorted frame, with probabilities proportional to the number of DU's in the segment.

A national sample of segments that Westat had previously selected and listed in the selected PSU's was used to reduce sampling costs. The sample developed for the National Adult Literacy Survey (NALS) used basically the same sampling procedures required for the CSFII/DHKS 1994–96, except that high-density minority segments were selected at about twice the rate of the nonminority segments. For the NALS, the segments were deliberately made much larger than needed so that they could serve as the equivalent of a master segment sample that could be used for other studies. Fifty-six percent of the 2,232 segments required for the CSFII/DHKS sample could be drawn from the previously selected NALS segments. The remaining segments were selected to yield the desired overall probabilities of selection, while maximizing the overlap with the NALS sample. NALS listings were updated through standard quality control procedures.

Selection of Dwelling Units

The sample of DU's was selected from the sample of area segments. The procedures used to select the DU's included the creation of segment-level lists of DU's, use of special procedures for handling a few extremely large segments (chunks) in the listing process, the selection of DU's from the segment listings, and special field procedures used to verify and update the listing information.

The purpose of listing was to create a list of DU's from which a sample could be selected for interviewing. For the sample to be representative of the population of interest, it was essential that the listing be carried out accurately and systematically, so that every DU in a designated segment was included. The process of listing involved an interviewer walking or driving through every street, road, alley, or boundary in the segment and recording on forms the address and description of every DU within the boundaries of the selected segments. The maps necessary to list the segments were generated using the U.S. Census Bureau's map-producing database called TIGER (Topologically Integrated Geographic Encoding and Referencing) (U.S. Department of Commerce–Bureau of the Census 1991c). The TIGER file is a geographic database where all map features are digitized and stored along with attribute information.

Census data indicated that some of the sampled segments were very large. To reduce the listing workload in the large segments, an additional stage of sampling was introduced. In general, these segments (defined as segments with an estimated 500 or more DU's) were divided into two or more smaller chunks of

approximately equal size, and one chunk was selected for listing with probability proportional to estimated size. Of the 744 segments (including NALS segments) selected for each year of the CSFII/DHKS, 54 were chunked using these procedures in 1994, 38 were chunked in 1995, and 58 were chunked in 1996. Although the selected chunks were treated like all other segments in the subsequent stages of selection, their probabilities of selection were properly adjusted to reflect the additional stage of selection.

For the first year, a sample of about 9,500 DU's was selected from the 744 segments (or chunks) designated for the first year of data collection. It was estimated that 9,500 DU's were necessary to yield approximately one-third of the required number of SP's within each of the sex-age-income groups defined by the survey design specifications prior to data collection. This estimate took into account the percentages of individuals in each sex-age group living in households, the percentages of individuals in each sex-age group living in households at or below 130 percent of the Federal poverty guidelines (U.S. Department of Health and Human Services 1994), projected response rates, a projected rate for vacant DU's, and a safety factor allowing for random sampling variation. The number of DU's selected was increased to about 11,500 for the second year and about 12,000 for the third year because the sampling rates of individuals changed throughout the survey (see "Derivation of sampling rates and sampling messages" below). The increased numbers of DU's selected did not necessarily result in increased numbers of SP's overall. The procedure for selecting DU's for the first year follows. The same procedure was used for the second and third years. The only change was the number of DU's selected.

To select the sample, the overall national sampling rate (f) was computed by dividing 9,500 by the estimated number of DU's ($\hat{N}$) based on the DU counts obtained during listing. Specifically, $\hat{N}$ was calculated from the formula:

$$\hat{N} = \sum_{h=1}^{62} \left[\frac{1}{P_h}\right] \sum_{j=1}^{12} \left[\frac{N_{hj}^{L}}{P_{hj}}\right] \qquad [1]$$

where

P_h is the probability of selecting PSU h,
N_{hj}^{L} is the number of DU's listed in segment j in PSU h, and
P_{hj} is the within-PSU probability of selecting segment j in PSU h for the first year of the survey.

For the NALS segments, N_{hj}^{L} reflected the numbers of DU's originally listed for the NALS (not including any new or missed structures added through the "missed structure" or "missed DU" procedures described below). This is because the selection of DU's was restricted to those DU's originally listed for NALS. However, it does not mean that new construction had no chance of selection from NALS segments. New construction (and also DU's that were missed in the original NALS listing process) still had appropriate chances of selection through the missed structure and missed DU procedures. As documented below, the within-segment sampling rates used to select the DU's were designed to produce a self-weighting national sample of approximately 9,500 DU's. Note that the procedure for selecting the DU's within the NALS and non-NALS segments was slightly different because of the desire to avoid selecting those DU's previously selected for NALS in the NALS segments.

Selection of dwelling units outside of National Adult Literacy Survey segments

N_{hj}^{L} denotes the number of DU's that were listed in non-National Adult Literacy Survey (non-NALS) segment j in PSU h. The N_{hj}^{L} DU's in the segment were then subsampled with equal probabilities at a rate of

$$f_{hj}^{(w)} = \frac{f}{P_h P_{hj}} \qquad [2]$$

where

P_h is the probability of selecting the PSU and
P_{hj} is the conditional probability of selecting the segment within the PSU.

The within-segment sampling rate, $f_{hj}^{(w)}$ given by formula 2 was designed to give each DU in the segment an overall probability of selection equal to f (that is, $P_h P_{hj} f_{hj}^{(w)} = f$). The actual selection of DU's within a segment was accomplished by first creating a file of unique line numbers corresponding to the DU's listed in the segment, and then selecting the line numbers systematically using a random start and a skip interval equal to $1 / f_{hj}^{(w)}$. A systematic sampling algorithm was used to make the selections (Hansen et al. 1953).

Selection of dwelling units in National Adult Literacy Survey segments

Let N_{hj}^{L} denote the number of DU's that were originally listed for the NALS in segment j in PSU h. For NALS segments, the count N_{hj}^{L} does not include any structures or DU's that were added as a result of the missed structure or missed DU procedures. Of the N_{hj}^{L} DU's in the segment that were originally listed for NALS, the n_{hj}^{NALS} DU's sampled for the NALS were identified and excluded from the sampling process. The remaining $N_{hj}^{L} - n_{hj}^{NALS}$ DU's were then subsampled at a rate of

$$f_{hj}^{(w)} = \frac{f}{P_h P_{hj} \left[\frac{N_{hj}^{L} - n_{hj}^{NALS}}{N_{HJ}^{L}} \right]} \qquad [3]$$

where the term $N_{hj}^{L} - n_{hj}^{NALS} / N_{hj}^{L}$ in the denominator of formula 3 is the probability that a DU in the segment was not previously selected for NALS.

The selection of DU's within a NALS segment was accomplished by first creating a file of unique line numbers corresponding to the DU's listed in the segment, deleting the line numbers corresponding to the DU's previously selected for NALS, and then systematically selecting the line numbers using a random start and a skip interval equal to $1 / f_{hj}^{(w)}$. The overall sampling rate for DU's in the NALS segments is the same as that in the non-NALS segments (that is, $P_h P_{hj} (N_{hj}^{L} - n_{hj}^{NALS} / N_{hj}^{L}) / N_{hj}^{NALS} f_{hj}^{(w)} = f$).

Missed structure and dwelling unit procedures

Two separate quality control procedures were used to verify and update the listing information for all of the segments selected for the CSFII/DHKS 1994–96. Both procedures were conducted during data collection. The first of these, referred to as the missed structure procedure, was applied whenever the first DU in the

segment was selected for the CSFII/DHKS sample. Two versions of the missed structure procedure were used. The original procedure was used during the first 2 years of the survey. The procedure was modified at the start of the third year of the survey to account for the possibility that large amounts of new construction may have occurred in the NALS segments. Descriptions of both versions follow.

When a segment was designated for the missed structure procedure during the first 2 years of the survey, the interviewer recanvassed the entire segment, and all DU's not previously listed were added to the sample (except as noted below). Because the probability of selecting the first DU in a segment was equal to the within-segment sampling rate, all of the added DU's were selected at the same overall rate (that is, they had the same overall probability of selection) as the rest of the sample.

The above rule for designating the missed structure procedure segments during the first 2 years of the survey applied to the NALS and non-NALS segments. Because DU's selected for the NALS were excluded from the CSFII/DHKS sample, NALS segments that were designated for the missed structure procedure in the NALS were not designated for this procedure in the CSFII/DHKS. In effect, the updating work performed for the NALS was ignored for the CSFII/DHKS. However, no bias was introduced because the new or missed DU's still had their appropriate chances of selection for the CSFII/DHKS.

For the third year of CSFII/DHKS 1994–96, the missed structure procedure was modified so that the rules for designating the missed structure segments were different depending on whether the segment was a NALS or a non-NALS segment. For the non-NALS segments, the original rules applied; that is, a non-NALS segment was designated for the missed structure procedure if the first DU in the segment was selected for the sample. However, for the NALS segments, a modified rule was adopted. The modified rule was designed because large amounts of new construction could have occurred since the NALS segments were originally listed in 1991. Under the modified rule, an NALS segment was designated for the missed structure procedure if any of the first four DU's in the segment were selected for the CSFII. Therefore, on average, the NALS segments were designated for the missed structure procedure at four times the rate of non-NALS segments.

In those segments selected for the missed structure procedure, the interviewer prepared a list of all DU's that were not included in the original listing forms (that is, the new or missed DU's). This information was then sent to the central office, where a subsample of the new or missed DU's was selected by computer at rates

designed to yield the same overall probabilities of selection as the other DU's in the sample. Therefore, in general, all of the new or missed DU's in the non-NALS segments were added to the CSFII sample. On the other hand, only one in four of the new or missed DU's in the NALS segments were added to the sample to compensate for the fact that the NALS segments had four times as many chances of being selected for the missed structure procedure.

The second procedure, referred to as the missed DU procedure, applied to structures containing many DU's (for example, apartment buildings) and all DU's listed at a single address. If the first DU in the given structure was selected for the CSFII/DHKS sample, then the entire structure was checked to identify DU's that may have been omitted from the listing sheets. Any missed DU's found by this process were added to the sample.

To keep the interviewing workload to manageable levels within the segment, maximum limits were established for the number of missed or new DU's that could be added to the sample. These limits were 10 per segment for the missed structure procedure and 4 per structure for the missed DU procedure. When the actual numbers of missed DU's exceeded these limits, a subsample of the missed DU's was retained in the sample. In the first year of the CSFII/DHKS, subsampling was required for 4 of the 93 segments when the missed structure procedure was applied. In the second year, subsampling was required for 6 of the 126, and in the third year, subsampling was required for 8 of the 282 segments when the missed structure procedure was applied. Subsampling of the missed DU's was not required for any structures when the missed DU procedure was applied in any year. During the first year of the survey, 77 DU's were added to the sample through the missed DU procedure and 128 DU's were added through the missed structure procedure. During the second year, 100 DU's were added to the sample through the missed DU procedure and 219 DU's were added through the missed structure procedure. During the third year, 96 DU's were added to the sample through the missed DU procedure and 464 DU's were added through the missed structure procedure.

Results of the dwelling unit sampling process

A total of 32,932 DU's was selected for the 3 years of the survey. In addition, 1,084 DU's were added in the field as a result of the missed structure procedure and the missed DU procedure. Therefore, the total number of DU's included in the sample was 34,016. Of these, 33,560 completed either the full or abbreviated screener questionnaire or were vacant or non-DU's, and 4,189 were either vacant or non-DU structures. Of the 29,827 occupied DU's, 9,664 (32 percent) had

household members who were eligible for the survey. The percentage of screened households with eligible SP's decreased over 3 years, from 39 percent to 33 percent and then to 28 percent, as a result of the changes in the sampling rates of individuals. The results of the DU sampling process are summarized in table 4.

Selection of Sample Persons for Intake Interviews

The CSFII 1994–96 was designed to obtain a sample that would produce estimates with equivalent precision over the sex-age domains, for both the total population and the low-income population. To obtain the targeted numbers of individuals, different sex-age domains were sampled at different rates. The approach used to select persons for the intake interviews was to designate subsets of households where only persons meeting specified sex-age/income criteria would be included in the sample. For example, for one predesignated subset of households in the DU 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 randomly assigned to the various subsets to ensure the unbiased selection of SP's for the study. In addition, all infants under 1 year of age in households that contained at least one SP 1 year or older were included in the sample.

To facilitate the selection of SP's in the field, each screening questionnaire carried a sampling message specifying the characteristics of the persons to be included. These sampling messages were assigned at Westat's home office and the interviewers had no discretion as to whom to include. A total of 24 distinct sampling messages were employed for the first year of the CSFII/DHKS 1994–96—21 messages were employed in the first half of the second year, a slightly different set of 21 messages were employed in the second half of the second year, 13 messages were employed in the first half of the third year, and 17 messages were employed in the second half of the third year.

The proportion of households that received a particular message was determined to satisfy the target sampling rates for the various sex-age-income domains. The number and configuration of the sampling messages was a function of these sampling rates. The initial 24 messages used in the first year of CSFII/DHKS 1994–96 were derived from estimates based on a previous survey and on the pilot study experience. Once screening operations began and the results could be analyzed, the target sampling rates were adjusted to meet the sex-age-income domain goals as closely as possible. New sets of sampling messages were introduced at the beginning and midway into the final 2 years of the survey.

After completing the listing of household members, the interviewer identified which, if any, of the household members were eligible to be interviewed. A total of 19,830 SP's were identified through the screening process during the 3 years of the survey, with 6,868 in the first year, 6,576 in the second year, and 6,386 in the third year.

Derivation of sampling rates and sampling messages

Year 1. The form of the sampling messages used in the first year of CSFII/DHKS 1994–96 to select SP's was determined as follows. First, estimates of the number of persons in each sex-age-income domain were obtained from the March 1992 Current Population Survey (CPS) public use file (U.S. Department of Commerce–Bureau of the Census 1993). Second, coverage rates from the 1992 National Health Interview Survey (NHIS) (U.S. Department of Health and Human Services–National Center for Health Statistics 1994) were applied to the March 1992 CPS counts to obtain estimates of the numbers of persons who would be covered by an area probability sample. CPS estimates included adjustments to compensate for the known undercounting of certain groups of individuals and were expected to be somewhat larger than the corresponding counts obtained from the CSFII/DHKS listing operations where similar undercounting could be expected. Without the downward adjustment of the CPS estimates through application of NHIS coverage rates, the derived sampling rates might have been underestimated. Initial sampling rates were then defined for each sex-age-income group as the ratio of the sample size targets to the downward-adjusted, estimated population counts.

Third, some adjustments to the initial sampling rates were implemented. For 5 of the 20 sex-age groups, the proportion of low-income persons was high enough so that using the initial sampling rate for the total population would achieve both the all-income sample size target and the low-income sample size target. For these groups, the low-income sampling rates were adjusted by replacing them with the all-income sampling rate.

For the remaining 15 groups, different all-income and low-income sampling rates were used. The initial low-income rates were retained without adjustment. The all-income rates were adjusted by replacing them with the rates expected to obtain $n^{non\text{-}L} = n^{all} - n^{L}$ non-low-income sample persons from the all-income population, where n^{all} and n^{L} are the all-income and low-income sample size targets. Both the all-income and low-income sample size targets were expected to be met as a result of this adjustment to the initial rates and the combination of the sampling of the all-income population and a supplemental sampling of the low-income population.

Table 5 shows the sample size targets, estimated population counts, and initial and adjusted sampling rates for each sex-age-income group. Column 4 shows the CSFII/DHKS sample size targets from table 2. As stated above, the initial sampling rates are the ratio of the sample target sizes in column 4 and the population counts in column 5.

Once the adjusted sampling rates were calculated for each sex-age-income group, the groups were ordered by the magnitude of the rates and, in some cases, combined with other groups with similar sampling rates. Where groups were combined, the highest sampling rate among the groups was assigned to each of the groups in the combination. The result was 24 distinct groups, each consisting of 1 or more of the 40 sex-age-income groups. Table 6 shows these combined sex-age-income groups and their adjusted and final sampling rates.

Table 7 shows the 24 sampling messages. The messages are cumulative. For example, message 1 indicates that all children age 1 and 2 and low-income males age 50–59 would be selected from a household assigned that message, while those persons and low-income males age 60–69 would be selected from a household assigned message 2. Additionally, all infants under 1 year of age were selected only if another person 1 year of age or older was also selected through the sampling messages. The rightmost column of table 7 shows the proportion of all DU's selected for the sample assigned each sampling message. That is, 16.63 percent of all DU's were assigned message 1 and 17.35 percent of all households were assigned message 24. The proportion of DU's assigned to sampling message *i* was calculated from the formula:

$$prop_i = \frac{(r_i - r_{i+1})}{r_1} \qquad [4]$$

where

r_i is the corresponding final sampling rate given in the last column of table 5,

r_{i+1} is the final sampling rate given in the preceding row of the table (where $r_{25} = 0$ by definition), and

$r_1 = 0.2004$ is the targeted sampling rate (corresponding to the last row of the table).

The sampled DU's within each PSU were randomly assigned to the various messages in the proportions given in the rightmost column of table 7. This was accomplished by computing $N_{DU}prop_i$ (rounded to the nearest integer) for each

message i = 1, 2, …, 24, where N_{DU} is the number of sampled DU's in the PSU, and then randomly assigning the required number of DU's to message i.

Year 2. The numbers of completed day-1 intakes obtained during the first year of the survey generally met or exceeded the designated 1-year targets, with some exceptions. Shortfalls occurred in four all-income domains and in seven low-income domains. To compensate, the sampling rates established for the first year of the survey were modified to make up for the shortfall equally in the subsequent 2 years of the survey. For example, suppose that, at the end of the first year, 60 day-1 intake interviews were obtained for a domain where the 3-year target was 207. At that rate, 180 completed interviews would be available at the end of the survey, short of the target of 207. To make up for the shortfall, the original sampling rate was increased by about 20 percent to obtain an expected 74 completed interviews in each of the next 2 years of the survey (207 – 60) / 2 = 74. Similarly, for those domains where there was an excess of completed interviews in the first year, a corresponding downward adjustment was made to the original sampling rates.

Ideally, it would have been desirable to use all of the information available at the end of the first year to make the necessary changes for the second year. Unfortunately, this was not possible because of the amount of time needed to process the survey results and to prepare interviewer materials for the first quarter of the second year of the survey. In order to proceed with the preparation of materials the sample yield results from only the first two quarters of the first year were used to design the sampling rates for the second year.

The procedures used to construct the sampling messages for the second year were analogous to those previously described for the first year. Once the initial sampling rates were calculated for each sex-age-income group, the groups were ordered by the magnitude of the rates, and, in some cases, combined with other groups with similar rates. The result was 21 distinct groups.

Year 3. The adjustments in sampling rates were successful in eliminating 7 of the 11 shortfalls in sample yields observed after the first year of data collection. However, the sample yields of several other groups decreased and at the end of the second year, there were 10 domains where the 3-year sample size targets would probably not be met if rates were not adjusted. These shortfalls occurred in four all-income domains and six low-income domains. To compensate for these shortfalls, the sampling rates established for the second year of the survey were modified to make up for the shortfall in the final year of the survey.

As was done at the end of the first year to prepare for the second year, it would have been desirable to use all of the information available at the end of the second year of the survey to make the necessary changes for the third year. This was not possible, however, and in order to proceed with the preparation of materials, the sample yield results from the first three-quarters of the second year, along with complete results from the first year, were used to design the sampling rates for the third year. A different approach than that used for the second year's adjustments was taken. The sampling rates for the third year were constructed by first projecting the sample yields for each of the sex-age-income domains through the first 2 years. These counts had to be projected because preparation for the third year had to be completed before the second year data collection was completed. It was necessary to project the sample yields for the fourth quarter of year 2 for the all-income domains and for the third and fourth quarters for the low-income domains. Once completed, the difference between the 3-year target, and the projected 2-year sample yields provided a new target for the third year and sampling rates and sample messages were designed to meet those targets. For example, for all-income males age 1–2, the actual yield for the first year was 255 and the actual yield for the first three-quarters of the second year was 204. A projection of 524 males was calculated from these actual counts and an estimate of the last quarter's yield. The 3-year target for all-income males age 1–2 was 719, so the target for the third year was calculated as (719–524=195). A yield ratio, the ratio of the number of completed day-1 intake interviews to the number of DU's where an all-income male age 1–2 was assigned for sampling, was calculated from the results of the first seven quarters. The projected number of DU's needed for sampling to meet the target for males age 1–2 was then calculated by dividing the third year target by the yield ratio. In this example, the yield ratio was 0.02687, so the number of DU's expected in the required 195 day-1 interviews was calculated as 195/0.02687=7,256.

The procedures used to construct the sampling messages for the third year were analogous to those previously described for the first year. Once the initial numbers of DU's to be sampled were calculated for each sex-age-income group, the groups were ordered by the magnitude of these numbers, and, in some cases, combined with other groups with similar requirements. The result was 13 distinct groups.

As in the second year, adjustments were made to the sampling messages midway through the third year. Unlike the second year, where the changes were made to the existing set of messages, the mid-third-year adjustments were made by recomputing the required rates to reflect actual returns through the first quarter of the third year and then reworking the sampling messages using these rates.

Classification of households to income classes

Under the procedures adopted for the CSFII/DHKS, the screener (see chapter 4) contained a question on income status (Q S14) that was asked only when necessary during screening because of the belief that asking about income during the initial contact might increase nonresponse to the survey. Therefore, if the sampling message indicated that income information was unnecessary, the question was not asked. For example, one message indicated that all persons (1 year of age or older) in these households were to be included in the sample regardless of income level. Similarly, another message selected persons 1–2 years of age and low-income males 50–59 years of age. If a household assigned this message did not include males 50–59 years of age, the sampling of SP's could proceed without collecting income data in the screener. In these cases, the income information was requested during the household interview, using an identically worded question (Q H47a) from the more detailed household questionnaire.

Occasionally, the interviewers were unable to obtain the income information necessary to select SP's for the intake interviews. In such cases, a rule based on the composition of the household was used to assign the household to one of the income groups for sampling purposes. The rule used was the following: If the household contained one or more children under 6 years of age, but no males 18 years of age or over, it was treated as low income for sampling purposes. Otherwise, the household was treated as non-low income. This rule was expected to be reasonably effective in identifying low-income households because more than 60 percent of children under 6 years of age who live in households headed by a female with related children under 6 years and no spouse present are living below Federal poverty guidelines (U.S. Department of Commerce–Bureau of the Census 1991b).

It should be noted that the sampling rule given above was adopted simply to facilitate the sampling of SP's in the field. Some households that were classified as low income by this rule may have turned out to be non-low income, and vice versa. For base weighting purposes, such households were weighted according to their income status as determined by the sampling rule and not their actual income status. However, for the purpose of determining sample yields, the response to either Q S14 of the screener questionnaire or Q H47a of the household questionnaire was generally used to establish income status. Where a response to Q S14 or Q H47a was not available, a series of five income-status imputation rules were used to determine low-income status for the purpose of determining sample yields. The five rules were applied sequentially, that is, if rule 1 could not be used, then rule 2 was used, and so on. The five income-status imputation rules were:

1. Annual income from household questionnaire items H52 or H53 was used, along with household size, to determine low-income status.

2. Monthly income from household questionnaire items H57a–H57f was totaled and used, along with household size, if rule 1 could not be applied.

3. Household questionnaire item H58 used a handcard to ask if last month's income was above or below the appropriate low-income cutoff based on household size. The result of this question was used if the previous two rules could not be applied.

4. Household questionnaire item H59 asked about food stamp use. If the answer was yes, the household was assigned to the low-income group. If the answer was no, the household was assigned to the non-low-income group. This rule was used only if the previous three rules could not be applied.

5. Finally, if none of the above four rules could be applied, the sampling rule, based on household composition was used. Under the sampling rule, households with one or more children under 6 years of age and no males 18 years or older were treated as low-income. All other households were treated as non-low-income.

It was necessary to use these rules to classify about 2 percent of all SP's with day-1 intakes as either low-income or non-low-income.

Results of SP sampling process

Table 8 summarizes the number of SP's eligible for intake interviews, the corresponding numbers completing the first intake interview, and the success of the survey process in achieving the sample size goals. As shown, the sample size goal was met or exceeded for 14 of the 20 all-income sex-age domains. For all of the remaining six all-income sex-age domains, at least 98 percent of the CSFII goals were achieved. Among the low-income domains, the sample size goals were met or exceeded for 14 of the 20 sex-age domains. For four of the remaining six low-income sex-age domains, at least 96 percent of the CSFII target was achieved. The two low-income domains with the greatest shortfalls were females 50 to 59 years of age (about 9 percent short of the goal) and males 40 to 49 years of age (about 6 percent short of the goal).

Selection of Sample Persons for the DHKS 1994–96

Respondents for the DHKS were selected from among SP's 20 years of age and over who had completed the day-1 intake interview without a proxy.[3] 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 in the field using a specially designed sampling program in each interviewer's laptop computer. Unlike the intake interviews, there were no specific numerical sample size targets for the DHKS. However, there was the requirement that the distribution of DHKS respondents by age, sex, and income be similar to that of the corresponding intake respondents. Although it was recognized that restricting the DHKS sample to only one respondent per household might distort the distribution of DHKS respondents somewhat, the random sampling procedures used to select respondents were reasonably effective in meeting the study goals. As table 9 shows, the distribution of SP's selected for the DHKS and the corresponding distribution of DHKS respondents are generally comparable to the distribution of SP's completing the day-1 intake interview. The selection was made with probability assigned to maintain distributions of all-income and low-income individuals in the 6 sex-age groups age 20 years and over in the DHKS that conformed approximately to the corresponding distributions of individuals in the CSFII. Approximately one-half of the households had more than one eligible SP for the DHKS. In all 3 years of the survey, 6,294 individuals were selected into the DHKS 1994–96 sample, 2,047 in the first year, 2,159 in the second year, and 2,088 in the third year.

3. In 1994–96, 191 SP's age 20 or older completed the day-1 intake with the assistance of a proxy.

References

Hansen, M., W. Hurwitz, and W. Madow. 1953. Sample survey methods and theory. Vol. 1. John Wiley & Sons, New York.

U.S. Department of Commerce, Bureau of the Census. 1991a. Census of population and housing, 1990: Public law 94–171 data. Machine-readable data file.

U.S. Department of Commerce, Bureau of the Census. 1991b. Poverty in the United States: 1990. Current Population Reports, Series P–60, No. 175.

U.S. Department of Commerce, Bureau of the Census. 1991c. TIGER/Line census files, 1990. Machine-readable data file.

U.S. Department of Commerce, Bureau of the Census. 1993. Current population survey, March 1992. Machine-readable data file.

U.S. Department of Health and Human Services. 1994. Poverty guidelines. Federal Register 59(28):6277.

U.S. Department of Health and Human Services, National Center for Health Statistics. 1994. 1992 national health interview survey. CD-ROM Series 10, No. 6, SETS Version 1.21. U.S. Government Printing Office, Washington, DC.

Table 2. Sample size targets, CSFII/DHKS 1994–96

Sex and age (years)	Sample size targets	
	Low income*	All income (total sample)
Male		
1–2	207	719
3–5	207	719
6–11	207	719
12–19	207	719
20–29	207	793
30–39	207	850
40–49	207	850
50–59	207	850
60–69	207	850
70 and over	207	793
Female		
1–2	207	719
3–5	207	719
6–11	207	719
12–19	207	719
20–29	207	739
30–39	207	793
40–49	207	850
50–59	207	850
60–69	207	793
70 and over	207	719
Total	4,140	15,482

* The income level used during the screening process corresponded to 130 percent of the Federal poverty guidelines (U.S. Department of Health and Human Services 1994), 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 criteria are eligible for food stamps; other criteria, such as asset limitations, must also be met. The CSFII 1994–96 screened households for income level only, not for food stamp eligibility.

Table 3. Distribution of PSU's by census region and MSA status, CSFII/DHKS 1994–96

Census region	Type of PSU			Total
	Certainty MSA	Noncertainty MSA	Non-MSA	
Northeast	6	6	1	13
Midwest	5	8	4	17
South	6	7	5	18
West	7	5	2	14
Total	24	26	12	62

Table 4. Results of the DU sampling process, CSFII 1994–96

Survey year	DU's selected from listings	DU's added in the field	Total DU's in sample	Vacant or non-DU's	Occupied DU's with eligible SP's *	Occupied DU's with no eligible SP's	Non-responding DU's**
1994	9,423	205	9,628	1,161	3,266	5,067	134
1995	11,504	319	11,823	1,337	3,379	6,954	153
1996	12,005	560	12,565	1,691	3,019	7,686	169
1994–96	32,932	1,084	34,016	4,189	9,664	19,707	456

* Eligible SP's refers to household members designated for intake interviews by the SP sampling process.
** Nonresponding DU's are those where a screener questionnaire was not completed.

Table 5. Sample size targets, estimated population counts, and initial and adjusted sampling rates, CSFII/DHKS 1994

(1) Sex	(2) Age (years)	(3) Income	(4) Sample size target (3 years)	(5) Population counts based on CPS totals and NHIS coverage (× 1,000)	(6) Initial sampling rate (× 1,000)	(7) Sample for sex-age group meets target for low-income(*)	(8) Adjusted sampling rate (× 1,000)
Male	1–2	All	719	3,612	0.1991	*	0.1991
Male	3–5	All	719	5,248	0.1370	*	0.1370
Male	6–11	All	719	10,627	0.0677		0.0651
Male	12–19	All	719	12,682	0.0567		0.0514
Male	20–29	All	793	16,189	0.0490		0.0421
Male	30–39	All	850	18,454	0.0461		0.0397
Male	40–49	All	850	14,296	0.0595		0.0496
Male	50–59	All	850	9,844	0.0863		0.0730
Male	60–69	All	850	8,844	0.0961		0.0845
Male	70+	All	793	7,559	0.1049		0.0965
Female	1–2	All	719	3,587	0.2004	*	0.2004
Female	3–5	All	719	5,273	0.1363	*	0.1363
Female	6–11	All	719	10,240	0.0702		0.0694
Female	12–19	All	719	12,530	0.0574		0.0533
Female	20–29	All	739	16,474	0.0449		0.0408
Female	30–39	All	793	20,325	0.0390		0.0348
Female	40–49	All	850	16,113	0.0528		0.0452
Female	50–59	All	850	10,927	0.0778		0.0685
Female	60–69	All	793	10,554	0.0751		0.0706
Female	70+	All	719	11,056	0.0650	*	0.0650
Male	1–2	Low	207	1,184	0.1749		0.1991
Male	3–5	Low	207	1,517	0.1364		0.1370
Male	6–11	Low	207	2,758	0.0750		0.0750
Male	12–19	Low	207	2,726	0.0759		0.0759
Male	20–29	Low	207	2,264	0.0915		0.0915
Male	30–39	Low	207	2,249	0.0920		0.0920
Male	40–49	Low	207	1,322	0.1565		0.1565
Male	50–59	Low	207	1,034	0.2002		0.2002
Male	60–69	Low	207	1,239	0.1671		0.1671
Male	70+	Low	207	1,487	0.1392		0.1392
Female	1–2	Low	207	1,118	0.1852		0.2004
Female	3–5	Low	207	1,552	0.1334		0.1363
Female	6–11	Low	207	2,863	0.0723		0.0723
Female	12–19	Low	207	2,921	0.0709		0.0709
Female	20–29	Low	207	3,446	0.0601		0.0601
Female	30–39	Low	207	3,471	0.0596		0.0596
Female	40–49	Low	207	1,884	0.1099		0.1099
Female	50–59	Low	207	1,542	0.1342		0.1342
Female	60–69	Low	207	2,251	0.0920		0.0920
Female	70+	Low	207	3,883	0.0533		0.0650

Table 6. Final sampling rates assigned to each message, CSFII/DHKS 1994

(1)	(2)	(3)	(4)	(5)	(6)
Sex	**Age (years)**	**Income**	**Adjusted sampling rate (× 1,000)**	**Sampling message number**	**Final sampling rate (× 1,000)**
Female	30–39	All	0.0348	24	0.0348
Male	30–39	All	0.0397	23	0.0397
Female	20–29	All	0.0408	22	0.0408
Male	20–29	All	0.0421	21	0.0421
Female	40–49	All	0.0452	20	0.0452
Male	40–49	All	0.0496	19	0.0496
Male	12–19	All	0.0514	18	0.0514
Female	12–19	All	0.0533	17	0.0533
Female	30–39	Low	0.0596	16	0.0601
Female	20–29	Low	0.0601	16	0.0601
Female	70+	All	0.0650	15	0.0651
Female	70+	Low	0.0650	15	0.0651
Male	6–11	All	0.0651	15	0.0651
Female	50–59	All	0.0685	14	0.0694
Female	6–11	All	0.0694	14	0.0694
Female	60–69	All	0.0706	13	0.0709
Female	12–19	Low	0.0709	13	0.0709
Female	6–11	Low	0.0723	12	0.0730
Male	50–59	All	0.0730	12	0.0730
Male	6–11	Low	0.0750	11	0.0759
Male	12–19	Low	0.0759	11	0.0759
Male	60–69	All	0.0845	10	0.0845
Male	20–29	Low	0.0915	9	0.0920
Female	60–69	Low	0.0920	9	0.0920
Male	30–39	Low	0.0920	9	0.0920
Male	70+	All	0.0965	8	0.0965
Female	40–49	Low	0.1099	7	0.1099
Female	50–59	Low	0.1342	6	0.1342
Female	3–5	All	0.1363	5	0.1370
Female	3–5	Low	0.1363	5	0.1370
Male	3–5	Low	0.1370	5	0.1370
Male	3–5	All	0.1370	5	0.1370
Male	70+	Low	0.1392	4	0.1392
Male	40–49	Low	0.1565	3	0.1565
Male	60–69	Low	0.1671	2	0.1671
Male	1–2	All	0.1991	1	0.2004
Male	1–2	Low	0.1991	1	0.2004
Male	50–59	Low	0.2002	1	0.2004
Female	1–2	Low	0.2004	1	0.2004
Female	1–2	Low	0.2004	1	0.2004

Table 7. Sampling messages by sex, income, and age, CSFII/DHKS 1994

Message (number)	Male		Female		Dwelling units assigned sampling message (proportion)
	All-income	Low-income	All-income	Low-income	
	(Age)				
1	1–2	50–59	1–2		0.1663
2	1–2	50–69	1–2		0.0527
3	1–2	40–69	1–2		0.0863
4	1–2	40+	1–2		0.0111
5	1–5	40+	1–5		0.0138
6	1–5	40+	1–5	50–59	0.1215
7	1–5	40+	1–5	40–59	0.0667
8	1–5, 70+	40–69	1–5	40–59	0.0223
9	1–5, 70+	20–69	1–5	40–69	0.0374
10	1–5, 60+	20–59	1–5	40–69	0.0429
11	1–5, 60+	6–59	1–5	40–69	0.0147
12	1–5, 50+	6–49	1–5	6–11, 40–69	0.0105
13	1–5, 50+	6–49	1–5, 60–69	6–19, 40–59	0.0073
14	1–5, 50+	6–49	1–11, 50–69	12–19, 40–49	0.0216
15	1–11, 50+	12–49	1–11, 50+	12–19, 40–49	0.0249
16	1–11, 50+	12–49	1–11, 50+	12–49	0.0339
17	1–11, 50+	12–49	1–19, 50+	20–49	0.0093
18	1–19, 50+	20–49	1–19, 50+	20–49	0.0093
19	1–19, 40+	20–39	1–19, 50+	20–49	0.0218
20	1–19, 40+	20–39	1–19, 40+	20–39	0.0155
21	1–29, 40+	30–39	1–19, 40+	20–39	0.0062
22	1–29, 40+	30–39	1–29, 40+	30–39	0.0058
23	1+		1–29, 40+	30–39	0.0245
24	1+		1+		0.1735

Table 8. Number of SP's eligible for intake interviews; number completing day 1; and corresponding sample size targets by income, sex, and age, CSFII 1994–96

Sex and age (years)	Low-income households			All households		
	Eligible SP's in low-income households	SP's completing day-1 intake	CSFII 1994–96 low-income sample size target	Eligible SP's in all-income households	SP's completing day-1 intake	CSFII 1994–96 all-income sample size target
Males						
Under 1	69	61	NA	213	187	NA
1–2	252	245	207	803	725	719
3–5	257	238	207	850	734	719
6–11	225	215	207	867	751	719
12–19	233	218	207	881	734	719
20–29	262	229	207	1,017	779	793
30–39	243	201	207	1,157	890	850
40–49	237	195	207	1,138	861	850
50–59	231	204	207	1,186	888	850
60–69	223	202	207	1,092	846	850
70 and over	221	206	207	993	790	793
Females						
Under 1	79	73	NA	222	195	NA
1–2	246	237	207	794	707	719
3–5	250	238	207	834	735	719
6–11	220	214	207	841	734	719
12–19	234	216	207	876	732	719
20–29	256	236	207	960	726	739
30–39	221	207	207	963	809	793
40–49	247	226	207	1,142	903	850
50–59	202	188	207	1,071	864	850
60–69	224	209	207	1,001	790	793
70 and over	243	230	207	917	723	719
Total, excluding children < 1	4,727	4,354	4,140	19,383	15,721	15,482
Total	4,875	4,488	4,140	19,818	16,103	15,482

NOTE: Table excludes SP's who were selected for the survey but became ineligible before completing the day-1 intake. Classification by income, sex, and age reflects imputed values. Age is that at the time of screening.

Table 9. Number of SP's completing the day-1 intake interview and number selected for completing the DHKS interview, 1994–96

Income, sex, and age (years)	SP's completing day-1 intake*		SP's selected for DHKS**		SP's completing DHKS	
	(number)	(percent)	(number)	(percent)	(number)	(percent)
Low-income						
Males						
20–39	430	17	264	15	239	15
40–59	399	16	287	16	260	16
60 and over	408	16	297	16	259	16
Females						
20–39	443	17	297	16	270	16
40–59	414	16	309	17	293	18
60 and over	439	17	352	19	323	20
Total	2,533	100	1,806	100	1,644	100
All-income						
Males						
20–39	1,669	17	983	16	874	15
40–59	1,749	18	1,120	18	1,036	18
60 and over	1,636	17	1,080	17	987	17
Females						
20–39	1,535	16	933	15	847	15
40–59	1,767	18	1,119	18	1,047	18
60 and over	1,513	15	1,059	17	974	17
Total	9,869	100	6,294	100	5,765	100

* Includes all SP's completing the day-1 intake.

** Excludes SP's who became ineligible before completing the DHKS.

Chapter 4: Questionnaire Development and Data Collection Procedures
By Patricia M. Guenther, Linda E. Cleveland, and Linda A. Ingwersen

This chapter describes the questionnaire development process and the data collection procedures for the CSFII/DHKS 1994–96. Table 10 summarizes the questionnaires and data collection procedures. For each questionnaire, the table identifies the type of data collected, the respondent, the mode of administration, the average administration time, and the number of contact attempts required (Westat 1995). Copies of the questionnaires are provided in appendix C. This chapter also describes the Food Instruction Booklet the interviewers used to collect the descriptions and amounts of foods the respondents consumed and the measurement guides the respondents used to estimate the amounts of food eaten (see appendix B).

Questionnaire Development

The questionnaires used in the CSFII 1989–91 and DHKS 1991 were the starting point for the development of the questionnaires for the 1994–96 surveys. The questionnaires were developed by staff working groups. As described in chapter 2, Federal users expressed their needs in specific content areas through the Continuing Survey Users' Group (CSUG). Draft questionnaires were tested several ways to prepare the versions used in the pilot study, and the questionnaires were further refined after the pilot study.

Continuing Survey of Food Intakes by Individuals

In the CSFII 1989–91, interviewers collected 24-hour dietary recalls from respondents and trained the respondents to keep 2-day dietary intake records. A structured grid format was used for the recall and the recordkeeping forms. The interviewers completed debriefing questionnaires and participated in focus groups periodically so that USDA staff could gain some insight into how data collection could be improved. The interviewers believed that the recordkeeping was too burdensome and difficult for respondents. Interviewers often had to collect information missing from the records when they were retrieved. USDA staff also found the level of detail on the self-administered records less than desirable for coding purposes.

In addition to respondent burden, another issue addressed in the development process was underreporting. While much of the research on underreporting focused on the food diary method (Mertz et al. 1991), underreporting has also been a concern in 24-hour recalls (Briefel et al. 1995). To address these issues, the

dietary data collection method chosen for the CSFII 1994–96 was two interviewer-administered 24-hour recalls, using a multiple-pass approach, collected 3 to 10 days apart. The 3-day minimum separation ensured that nutrient intakes on the 2 days would be statistically uncorrelated and 10 days was chosen as the maximum limit of endurance for interviewer-respondent rapport (An and Carriquiry 1991). The exclusive use of 24-hour recalls eliminated the burden of recordkeeping for the respondent, and the multiple-pass approach was expected to reduce underreporting.

Cognitive research. USDA commissioned the Bureau of the Census to find ways to improve the 24-hour recall procedures and other questions asked of respondents. In 1992, the Census Bureau's Center for Survey Methods Research (CSMR), reviewed, revised, and tested the 24-hour dietary recall protocol used in the 1989–91 CSFII and made recommendations for improving the quality and completeness of the data collected during the individual intake interviews. After reviewing the procedures for the 1989–91 CSFII and those used for the dietary component of the Third National Health and Nutrition Examination Survey, CSMR recognized that the 24-hour dietary recall task was a difficult one for respondents compared to the task typically required of survey respondents (U.S. Department of Health and Human Services–Public Health Service 1994). The 1989–91 CSFII dietary recall procedures focused respondents' attention on time and eating occasions as cues for recalling foods eaten. However, chronological order is not necessarily the best recall strategy for everyone (Means et al. 1991) and respondents do not always think in terms of "eating occasions." Also, the 1989–91 CSFII questionnaire asked the respondent to perform a fairly complex task only once; however, research has shown that questioning respondents multiple times, using different forms of the question, can generate more recall (Means et al. 1989; Fisher and Quigley 1991; Fries et al. 1995).

CSMR used concurrent think-aloud interviews to learn about the cognitive processes and recall strategies subjects used to answer questions and to identify misperceptions about questions (DeMaio et al. 1993). Subjects were instructed to verbalize their thoughts while they answered the survey questions presented by the interviewer. This procedure was used in 17 interviews conducted at the CSMR cognitive laboratory in Maryland over 2 phases of research, which allowed revisions to be tested.

CSMR found that subjects used a variety of strategies to recall foods, such as activities, meals, and time. The indepth probing for details about the foods reported was also found to be successful in prompting the recall of additional foods. The proposed procedure allowed respondents to use their own recall

strategy in reporting foods consumed and to think about the same general question in different ways to elicit greater recall. CSMR called the revised procedure a multiple-pass approach to the 24-hour dietary recall.

The passes prompted respondents to perform different cognitive tasks, a process useful in eliciting greater recall of foods. First, the respondent was instructed to tell the interviewer everything he or she had to eat or drink yesterday, from midnight to midnight ("Tell me everything you ate or drank yesterday, from midnight to midnight. Include everything eaten at home or away--even snacks, coffee breaks, or alcoholic beverages.").[1] The first pass made respondents focus on reporting what they ate and drank, which was the main intent of the dietary intake questionnaire. It also instructed respondents to report on foods easily forgotten, such as snacks and food eaten away from home. Most important, the first pass gave the respondents the flexibility to use their own cognitive strategies in an uninterrupted recall.

After the first listing of foods, questions on the time a food was eaten and the name of the eating occasion were asked. The main purpose of the next pass was to develop a more complete listing of individual foods. Questions included: (1) "Did you have anything else on your (food item)?", or, if more appropriate, "Did you have anything else in your (food item)?" and (2) "Did you have anything else with your (grouping of foods)?" This set of questions targeted items that CSMR found to be frequently not mentioned during the first listing (for example, milk on cereal and butter on bread). Then, respondents were asked to review the previous day's intake at another time. However, its focus was on unreported eating occasions or on anything else consumed yesterday: "Did you nibble or sip on anything while preparing a meal or while waiting to eat that you haven't already told me about?" and "Did you have anything else to eat or drink yesterday?" The final pass gathered the detailed descriptions, amounts, and sources of the foods.

Pilot study. In 1993, Westat conducted a pilot study (see chapter 2). While the research conducted by CSMR provided an understanding of the respondent's interpretation of the questions asked under laboratory conditions, the pilot study offered USDA the opportunity to look at the data collection procedures and questionnaires from the interviewer's point of view under survey conditions. In preparation for the pilot study, USDA staff evaluated the CSMR recommendations.

---

1. The National Health and Nutrition Examination Survey III interview also started with a general listing of foods and beverages consumed, but respondents reported by time of eating and in chronological order.

A multiple-pass approach was incorporated into the dietary intake questionnaires, although the passes were revised. The initial listing of foods was not changed. However, the questions recommended for the detailed list were modified. Rather than ask several questions about what was eaten in, on, or with foods reported at a specific occasion, the second pass asked the detailed questions about the reported foods. The detailed list was obtained through use of the Food Instruction Booklet. The probes developed by CSMR were integrated back into the Food Instruction Booklet, mainly by probing for "additions"—items added to foods by respondents.

After the pilot study, the last pass was developed into a more extensive series of questions that asked whether the respondent reported all food and beverage intake before the first reported eating occasion (but after midnight of the previous day), between the following eating occasions, and then after the last eating occasion (but before midnight). The final review questions focused on time periods between reported eating occasions to pick up foods that may have been forgotten earlier. (See the section on the day-1 individual intake interview below for a description of the final recall procedure.)

In addition to the changes proposed for the 24-hour recall, CSMR proposed revisions to other questions in the food intake questionnaires. Question wording and response options were revised so that reporting errors resulting from the respondents' misinterpretation of questions could be minimized. Major improvements were made in the structure and content of questions about the sources of foods. Response options to this question were revised to capture information about home-grown vegetables and about fish that were not commercially purchased to address the particular needs of the Environmental Protection Agency. Handcards listing response options were revised to help respondents understand the level of detail required when reporting their responses (see appendix C). Lastly, terms used in questions were clarified; for example, "water" was changed to "plain drinking water."

Other changes were made to the content and administration of the dietary intake questionnaires prior to the pilot study. These included revising interviewer instructions to clarify the administration of the instruments, as well as the procedures for administering the 24-hour recall to children when assisted by another household member. USDA and Westat staff worked collaboratively to further develop data collection procedures, questionnaires, handcards, and portion-size measurement aids. Westat staff developed questionnaire formats that improved the ease of questionnaire administration.

USDA staff reviewed and revised the 1989–91 Food Instruction Booklet, which contained probes the interviewers used to elicit detailed descriptions of foods and amounts reported by respondents. The probes were designed to capture the information needed to assign appropriate food codes (see chapter 7) and varied with the type of food. For example, milk had a probe for fat content and bread had a probe for type of grain. CSMR staff also reviewed the Food Instruction Booklet. They recommended that the format be more standardized and that the booklet be changed from landscape to portrait style. Both suggestions were adopted.

To evaluate the multiple-pass approach to the 24-hour recall during the pilot study, USDA and Westat staff observed interviews in respondents' homes. A mailed interviewer debriefing questionnaire, a review of interviewers' logs, and an interviewer debriefing in a focus group at the end of the study provided additional insight. Interviewers and Westat and USDA staff observers concluded that the food intake questions did not overburden respondents or interviewers. The level of detail collected was judged by USDA to be sufficient for food coding and subsequent nutrient intake calculation.

Following the pilot study, Westat and USDA staff met to revise the questionnaires, the Food Instruction Booklet, and other aids for the main survey. The 24-hour recall procedure used is described below in the section on the day-1 individual intake interview. The format of the questionnaires was improved to reduce interviewer error. For the household questionnaire, the income and head-of-household questions were redesigned to reduce interviewer and respondent confusion. Employment status questions were revised to correspond to the questions used in the Current Population Survey.

The Food Instruction Booklet was also revised after the pilot study. Probes were standardized by writing the questions exactly as they were to be read when collecting the detailed food descriptions. Earlier 24-hour recall procedures required the interviewers to frame the questions needed to collect the information called for in the Food Instruction Booklet. Changes in the marketplace and interviewer suggestions resulted in additional changes.

Further refinements were made to the Food Instruction Booklet in the middle of the first year of data collection and at the beginning of the second and third years of data collection. Revisions were made to keep up with changes in the marketplace and to improve the interface with the Food Coding Database (see chapter 7). Changes included increasing the number and specificity of food categories and adding more examples of quantity measures.

Measurement guides. In 1989–91, the measurement guides used by respondents to estimate the amounts of foods eaten included a set of stainless steel measuring cups and spoons, a 6-inch ruler, and a picture of concentric circles for estimating diameters. In 1994, a 12-inch ruler replaced the 6-inch ruler. In addition, a pint measuring cup was added for use in measuring volumes of household cups, glasses, and bowls, and the thickness sticks used in the National Health and Nutrition Examination Survey III dietary intake component were added because of the known difficulty in estimating the height or thickness of foods. The full set of measurement guides used in CSFII 1994–96 is described below in the section on the day-1 individual intake interview.

Diet and Health Knowledge Survey

The DHKS questionnaire used in the 1991 survey served as the starting point for development of the questionnaire for the DHKS 1994–96. The questionnaires were developed by an inhouse working group.

USDA staff reviewed the 1991 DHKS questionnaire and proposed changes for DHKS 1994–96. Two primary criteria drove decisions about questionnaire content. First, questions were to focus on respondents' knowledge or beliefs about the *Dietary Guidelines for Americans* and their ability to put the guidelines into practice (USDA and DHHS 1990). The guidelines were a focus because they are the basis for the Federal Government's nutrition policy and education activities. Second, questions had to benefit from the unique feature of the DHKS that allows linking people's knowledge, attitudes, and behavior with their dietary intakes. In addition to these two primary criteria, content decisions were driven by research needs identified in the Ten-Year Comprehensive Plan for the Nutrition Monitoring and Related Research Program (U.S. Department of Health and Human Services and USDA 1993). The plan identified the need for research on consumer use and understanding of the nutrition information on food labels and on the predictive capabilities of knowledge, attitude, and behavior questions for assessing intakes of nutrients of public health interest. The plan also called for the development of questions framed in the context of theories of behavior change.

With these criteria in mind, proposed changes were presented to Federal users for discussion at a Continuing Survey Users' Group meeting in 1992. Written comments were requested, and the comments were considered in subsequent revisions of the questionnaire by the inhouse working group.

A number of food safety questions were proposed for deletion so the questionnaire could address new issues. Most of the food safety questions on the DHKS 1991 did not meet the criteria established for the DHKS 1994–96. Nonetheless, to

ensure that deletion of questions would not leave policy or monitoring needs unmet, agencies conducting food safety initiatives, notably the Food Safety and Inspection Service (FSIS), Environmental Protection Agency, and Food and Drug Administration, were consulted. As a result, modified forms of questions about washing and peeling vegetables and fruits prior to consumption were retained, but other food safety questions were dropped.

While the DHKS 1994–96 was being developed, the Nutrition Labeling and Education Act of 1990 was generating major changes in food labeling regulations, which were at various stages in the legislative process. The new food label was viewed as an important tool for helping consumers put the dietary guidelines into practice, and the DHKS 1994–96 provided a vehicle for collecting data to assess relationships between food label use and dietary intakes. Therefore, an interagency working group was formed by USDA to consider food labeling issues for the DHKS. This working group included staff from FDA and FSIS, the agencies responsible for food labeling regulations, as well as USDA staff involved in questionnaire development. The purpose of the working group was to identify appropriate content areas, formulate draft questions, ensure consistency with food labeling regulations, avoid unnecessary duplication in data collection, and promote comparability with other Federal nutrition surveys.

One section of the DHKS 1994–96 included questions that measure the frequency with which people use various sections of the food label, seek information on various dietary components, and use food labels to choose various types of foods. Additional questions measured respondents' level of confidence and knowledge about terms used on food labels and their attitudes toward label use. The attitude questions were framed in the context of a theory called "diffusion of innovations" (Rogers 1983). The attitudinal stage of this theory consists of five perceived attributes of innovations: (1) "relative advantage" over earlier ideas, (2) "compatibility" with existing values and needs, (3) "complexity" of use, (4) "trialability," [sic] and (5) "observability" by others. In formulating DHKS questions, the new food label was viewed as an innovation, and questions were designed to tap each of these attributes. The questions were reviewed independently by members of the inhouse working group and revised until a consensus was reached regarding the attribute measured by each question and its wording.

Another DHKS section, new in 1994, included a series of questions about dietary behaviors likely to influence fat intake. These questions were added to provide an opportunity to evaluate their predictive validity. They focused on fat intake because it is associated with major health problems in the United States, including coronary heart disease, obesity, and cancer. One of the dietary guidelines is

devoted to fat, and it has been highlighted as a current public health issue in nutrition monitoring reports (FASEB 1989, 1995; U.S. Department of Health and Human Services and USDA 1986). These questions also assess acceptance of basic food guidance concepts that have been promoted in advice on how to put the dietary guidelines into practice.

Many of the questions on fat intake were derived from research by Kristal on dimensions of behavior that correlate with fat intake (Kristal et al. 1990 a,b,c 1992). The dimensions he identified are (1) avoiding fat as a flavoring, (2) avoiding meat, (3) substituting specially manufactured low-fat foods for higher fat foods, (4) modifying commonly used foods so they are lower in fat, and (5) changing the overall cuisine. Content ideas were also derived from research assessing relationships between food intakes and fat intake (Block et al. 1985; Gorder et al. 1986; Gorbach et al. 1990; Schectman et al. 1990; Georgiou and Arquitt 1992; Krebs-Smith et al. 1992; Thompson et al. 1992).

The DHKS questionnaire was pretested by USDA staff in collaboration with survey design experts at the Census Bureau's Demographic Surveys Division to prepare the version used in the pilot study. The objectives of the pretests were to evaluate the questionnaire in actual interview situations and to revise it to improve comprehension of the questions by respondents and interviewers, resolve context issues, correct skip patterns, improve ease of administration, and meet time constraints on the length of the interview.

Four pretests were conducted. Trained interviewers administered draft questionnaires by telephone, and researchers monitored the interviews. In three of the pretests, random digit dialing was used to select sample persons (SP's), and in the fourth pretest, Census Bureau employees volunteered to be interviewed. All SP's were 20 years of age or older. The testing was conducted as an iterative process whereby weaknesses in the questionnaire were identified, corrected, and then retested. No more than six interviews were conducted using any single version of the questionnaire. This process allowed rapid recognition and revision of problems onsite by the researchers. A total of 17 interviews was conducted, 11 with women and 6 with men. In this manner, DHKS questionnaire flow, skip patterns, and question wording were refined before the pilot study.

Additional changes were made to the DHKS questionnaire based on interviewer experiences during the pilot study. The introduction was revised to improve its flow and tone. Minor wording changes were made to a few questions. Interviewer instructions were added as needed to standardize responses to frequently asked questions. A shorter alternative set of questions on attitudes about food label use was developed for people who do not use labels. A postcard listing the major sets

of response categories used in the DHKS was developed to streamline the interview by decreasing the need for interviewers to repeat response categories. The postcard, which was mailed to SP's prior to the DHKS interview, also served as a DHKS appointment reminder.

Data Collection Procedures

During the data collection process, interviewers conducted the following steps:

- Mailed introductory letters and brochures to sample addresses.
- Used maps and other information to locate sampled addresses and verify, by visual inspection, that they met the definition of a "dwelling unit."
- Checked the area around sampled addresses for dwelling units that may have been missed during the listing process (see chapter 3).
- Conducted a screening interview at each sampled dwelling unit to determine if any household members were eligible to participate.
- Administered the household questionnaire.
- Conducted two food-intake interviews with each person selected for the CSFII.
- Conducted the DHKS interview with selected persons.
- Completed a noninterview report form for each missing screener, household questionnaire, individual intake questionnaire, and DHKS questionnaire (see chapter 5).

Extensive efforts were used to gain participation in the CSFII/DHKS 1994–96. Community-level efforts included press releases and radio spot announcements. Newspaper articles related to current and past surveys were put in notebooks for the interviewers to refer to, if necessary, to gain SP cooperation. The interviewers also had photo-identification badges and a tote bag with the survey logo on it to help establish their legitimacy. In all materials prepared for SP's, the surveys were referred to as the "What We Eat in America Survey," rather than by the official name, to make it easier for SP's to remember.

Spanish questionnaires

Spanish-language questionnaires were first used in the CSFII and the DHKS in 1994. All questionnaires were translated into Spanish, as were the introductory letter, the survey brochure, the flyers, the questionnaires (but not the Food Instruction Booklet), the handcards, and the DHKS postcard.

Materials were translated by a native, Spanish-speaking Westat employee and back-translated by a subcontractor whose first language was Spanish and who had professional experience as a translator for Federal Government agencies. To check the accuracy and utility of the translation, the back-translator received a copy of the materials translated into Spanish and translated them into English without having seen the original English version. The translator checked the original English against the back-translated English for changes in meaning. When questions arose, the translator and the back-translator consulted on the most accurate translation.

Spanish questionnaires were used in 2.7 percent of CSFII 1994 interviews and 2.4 percent of the DHKS interviews. If an SP spoke neither English nor Spanish, a family member or neighbor 16 years of age or older served as an interpreter.

Introductory letter

Interviewers mailed the introductory letter and brochure (see appendix A) describing the survey to each dwelling unit (DU) 1 week before they intended to contact the household. The introductory letter included a Westat toll-free telephone number to call for more information. A total of 130 calls was received during the CSFII/DHKS 1994. Of these, 11 percent requested additional information about the survey; 15 percent wanted to verify the legitimacy of the survey; 17 percent wanted to set an appointment, reschedule an appointment, or contact a particular interviewer; and 57 percent were calling for other reasons but generally expressed an unwillingness to participate in the survey. Information about each call was recorded and passed along to the regional supervisor and the interviewer assigned to the case.

Incentives

Incentives were also used to gain SP participation. The interviewer told the screener respondent that each participating household would receive a gift. A set of high-quality measuring cups and spoons was given to the screener respondent after the screener was completed and the household was found to contain any SP's. An insulated nylon sack, bearing the survey logo, was given to each SP before the day-1 interview, and at the conclusion of the day-2 interview each responding SP

received a travel-type beverage mug, also bearing the survey logo, as a thank-you gift for participating. The gift provided at the end of day 2 also served as an incentive to complete the DHKS. The measuring cups and spoons and the nylon sacks were offered before the respondents completed their tasks to establish a social contract and help ensure participation. The incentives were not intended to be payment for cooperation.

Contact procedures

At each sample address, the interviewer determined if the structure was an occupied DU and carried out the missed DU procedure if instructed to do so (see chapter 3). Then, the interviewer attempted a screening interview to determine if any members of the household were eligible to participate in the survey. In cases where no household member could be contacted after four visits, interviewers were instructed to ask two neighbors about the number of people living in the household, the sex and age of each, and the time household members were most likely to be home.

Based on the information provided by the neighbors, the interviewer followed the SP selection instructions in the screening questionnaire. If SP's were selected, the interviewer continued his or her efforts to contact the sampled household to complete the screening questionnaire (screener) and necessary interviews.

Screening interview

Any household member 18 years of age or older was an acceptable respondent for the screener. However, it was recommended that interviewers attempt to conduct this portion of the survey with either the main meal planner/preparer or a person knowledgeable about household characteristics because they were the preferred respondents for the household questionnaire, which typically followed the screener (see table 10). It was not necessary for the respondent completing the screener or household questionnaire to be an SP.

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 the originals. 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.

Some screening respondents were asked whether the total income of all household members from all sources during 1993 was more or less than an amount specific to the household's size. The screener income question was part of the strategy used for oversampling the low-income population to meet the precision goals discussed in chapter 3, and was asked only when the household included individuals in sex and age groups specified on a computer-generated label that varied among DU's. This label and a similar label guided the interviewer to select persons to complete the individual food intake questionnaires.

Household interview

At households where at least one SP was selected, the interviewer administered the household questionnaire in person. It included a series of questions about the educational and employment status of household members 15 years and older, household income, food assistance program participation, food expenditures, and other food-related practices.

The respondent to the household questionnaire did not have to be an SP. The preferred respondent was the main meal preparer/planner for the household. When that person was unavailable, another household member who was knowledgeable about the household was asked to respond.

The interviewers had discretion, however, to complete the intake interviews before administering the household questionnaire. Interviewers might exercise this option, for example, if no qualified household questionnaire respondent was available or if an SP would have to leave the home before the household questionnaire could be completed.

Day-1 individual intake interview

Interviewers' visits were scheduled to ensure that at least 10 percent of day-1 food intake interviews took place on each day of the week. A label attached to the survey materials for each household specified 3 days of the week that would be acceptable for collecting day-1 food intake information from SP's in that household. Repeated in-person visits were made, as necessary, to attempt to complete day-1 intakes with SP's on one of the scheduled days of the week. In some cases, when repeated visits were 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 1994, 10 percent of all intake interviews were conducted

on Sunday, 17 percent on Monday, 18 percent on Tuesday, 16 percent on Wednesday, 13 percent on Thursday, 12 percent on Friday, and 14 percent on Saturday. Day of the week was a factor in the weighting—the sums of the weights for each day of the week were set equal to each other (see chapter 8).

A three-phase approach was used for nonresponse conversion. First, interviewers made at least three visits, and often more, to attempt to complete all necessary interviews. For nonresponse cases, supervisors instructed the interviewer to either approach the SP, offering advice on how to approach nonresponse cases, or to complete a noninterview report form, which listed the reason for the nonresponse and was returned to the supervisor for further review. Second, supervisors reviewed the noninterview report forms. Additional attempts to contact these SP's were made by reassigning them to the original interviewer, to another local interviewer, or to an interviewer from an adjacent area. Finally, the supervisor considered the response rates and the noninterview report forms to determine the likely benefit of additional attempts by a senior interviewer or another experienced interviewer.

Interviewers were permitted to make an appointment only for the first time an intake interview was to be administered to any household member. Multiple interviews could be administered on the appointed day, but if the interviewer needed to return to interview any sample person who was not available on that day, no further appointment could be made. It was believed that people's eating behavior might be influenced if they knew that on the following day they would be asked to report what they had eaten. The same reasoning applied to the day-2 intake interviews, so no appointments were allowed for these interviews.

Day-1 intake questionnaires were administered in person. Before conducting this interview, the interviewer told the SP that his or her participation would involve two in-person interviews (and possibly, for one SP in the household, the DHKS interview by telephone). At the conclusion of the day-1 interview, the interviewer notified the SP that he or she would be returning in a few days to conduct another interview. Whenever possible, the interviewers conducted the first day-1 intake interview with the SP who was also the main meal preparer/planner because this person could possibly provide more details about food preparation than other household members.

Proxy interviews were conducted routinely for SP's under 6 years of age and any others (including adults) who could not report for themselves due to physical or mental limitations. Proxy interviews were not permitted for any other reason. They were not considered to be an acceptable substitute for an in-person interview

with adult SP's who were difficult to reach or refused to be interviewed. Children 6 to 11 years of age were asked to describe their own food intake assisted by an adult household member (referred to as the assistant). The preferred proxy or assistant was the person responsible for preparing the SP's meals.

Interviewers often used school menus, which they obtained from newspapers, school personnel, or household members, during interviews with children to help them identify what they had eaten at school. If the SP, proxy, or assistant could not provide enough descriptive or quantitative information about the foods eaten, it was sometimes necessary to seek that information from other caregivers, such as babysitters, daycare personnel, or school cafeteria personnel. In 1994, 2 percent of the interviews required the use of the data retrieval procedures that had been developed for such instances.

The day-1 individual intake questionnaire began with a 24-hour food recall, using the multiple-pass approach. The first pass began with the respondent being asked to report everything he or she ate or drank the previous day between midnight and midnight. The interviewer did not interrupt the respondent during this initial listing of the day's intake. The respondent was invited to add any other items he or she remembered as the interview progressed.

During the second pass, for each food and drink that had been listed, the interviewer asked for the name of the eating occasion and the time it began, and for detailed food descriptions and amounts consumed. The interviewers were trained to read the questions verbatim from the questionnaire and to read the food probes verbatim from the Food Instruction Booklet. When appropriate, questions were asked about the use of salt and fat in food preparation and about additions to reported foods like coffee and bread. The interviewer was directed to ask for ingredients in some categories (for example, sandwiches; salads; mixed dishes, casseroles, and stews; soups; and tacos, burritos, enchiladas, and fajitas). Interviewers were required to use the Food Instruction Booklet to obtain details of every food item recalled by the respondent, including additions remembered as a result of questions asked in describing another food. The booklet also specified the types of measures (weight, volume, or size) appropriate for recording the amount of food consumed.

Measuring guides were used to help respondents estimate the amount of foods and beverages consumed. They included a set of four stainless steel measuring cups (1/4 cup, 1/3 cup, ½ cup, 1 cup) and four measuring spoons (1/4 teaspoon, ½ teaspoon, 1 teaspoon, 1 tablespoon); eight 1/8-inch-thick rectangular sticks for estimating the thickness of meat, poultry, and cheese; an easy-to-read 12-inch ruler for reporting dimensions in inches; and a pint measuring cup. A laminated card

with illustrations of a fish fillet and chicken parts was used to ensure adequate description of pieces. A set of concentric circles on the card helped the respondents quantify the diameter of some foods, such as pancakes.

The measuring cup was used when the respondent referred to a bowl or cup in his or her home. The respondent could then fill the bowl or cup with water to represent the amount he or she ate or drank, 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 was described and quantified, the interviewer reviewed for the respondent 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. This review was the respondent's third pass through the day. Then, for each food or drink reported, the interviewer asked where it had been obtained and whether it had been eaten at home or not.

Additional questions asked on day 1 and day 2 were whether the respondent's intake on the previous day had been usual or unusual and why; how much plain drinking water the respondent drank on the previous day and whether it came from his or her home or another source; and, how many hours of television or videos the respondent watched the previous day. Additional questions on the day-1 questionnaire included the type of salt usually used by the respondent and frequency of use at the table; whether the respondent was on a diet and, if so, the type and source of the diet; whether the respondent considered himself or herself to be a vegetarian; frequency of vitamin or mineral supplement use and type of supplement; use of fish oil and fiber supplements; whether the respondent ever had his or her blood cholesterol checked; 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 specific foods during the past 12 months.

Interviewers were trained to review and edit the intake questionnaires as soon as they left the respondents' homes. Legibility, accuracy, and completeness were checked using a standard list. Explanations were added, if necessary.

Day-2 individual intake interview

The day-2 interview was conducted 3 to 10 days after the day-1 interview, but not the same day of the week. Less than 1 percent of day-2 interviews were conducted sooner than 3 days after the day-1 interview, 2 percent were conducted the same day of the week, and 24 percent were conducted more than 10 days after the day-1 interview.

Supervisory permission was given to conduct the interview by telephone when interviewers were unable to complete the questionnaire in person. Three percent of day-2 interviews were conducted by telephone in 1994. Respondents interviewed by telephone were asked to refer to the measuring cups and spoons given to the household when reporting food quantities on day 1.

Diet and Health Knowledge interview

When all SP's in a household either had completed a day-1 intake or been judged to be day-1 nonrespondents, the DHKS sample person was randomly selected by a computerized process from among eligible CSFII sample persons 20 years of age and over who had completed a day-1 intake interview. An SP was not eligible if the intake interview was completed by proxy, nor were any proxies allowed to complete the DHKS. Therefore, not all households had a DHKS respondent.

Telephone contact to conduct the DHKS interview was initiated 2 to 3 weeks after the DHKS respondent completed the day-2 intake interview or was determined to be a day-2 nonrespondent. The purpose of the time delay was to minimize any influence that reporting food intake might have on responses to attitude questions. Interviewers attempted to schedule an appointment for the DHKS interview at the completion of the day-2 intake interview or as soon as the individual was judged to be a day-2 nonrespondent. For households without telephones or with unlisted numbers not previously provided to interviewers, interviews were conducted in person. Typically, the interviewer who administered the CSFII also administered the DHKS.

Three to 5 days before the scheduled interview date, the interviewer mailed a colorful DHKS reminder card. This card contained the appointment date and time and sets of response categories used in the DHKS questionnaire. During the interview, the respondent was directed to look at the set of response categories applicable to the particular question being asked in order to reduce the need for repetitious reciting of response options. Therefore, the card served to improve 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 various times of the day and on various days of the week to reach respondents. At least six telephone attempts were required for each telephone number, followed by four in-person visits to obtain the interview. In a number of difficult cases, contact attempts exceeded the required level of effort to complete the interview.

Eighty-four percent of 1994 DHKS interviews were completed by telephone. Interviewers were required to get supervisor approval before conducting an interview in person. The primary reason for in-person interviews was that the household did not have a telephone. Other major reasons included physical limitations, such as a hearing difficulty and language barriers.

In 1994, 73 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. About 5 percent of interviews were completed earlier than 2 weeks for reasons such as prior knowledge of extended periods of absence (for example, hospitalization or travel). For 23 percent, more than 3 weeks had elapsed. These mistimings were often caused by broken appointments when respondents were, for example, too busy or not at home at the scheduled time. Refusal conversion efforts also contributed to mistimings by lenghtening the time between the day-2 intake interview and the DHKS interview.

The interview began with a request to speak to the person with whom the appointment had been made. The interviewer identified himself or herself and reminded the respondent that during the CSFII, he or she had been told about later contact to answer a few more questions about food and nutrition issues. The DHKS respondent's name and age were verified to make sure the correct person was being interviewed.

The DHKS questionnaire covered knowledge, behavior, and attitudes about diet and health issues. Topics included (1) knowledge of the recommended number of servings from five major food groups, (2) self-assessment of one's own diet (overall and for specific dietary components), (3) perceived importance of dietary guidance concepts, (4) awareness of relationships between diet and health, (5) knowledge about food sources of fat, (6) food intake behaviors, (7) knowledge and attitudes about food label use, (8) behaviors related to food label use, and (9) food-handling practices.

References

An, B., and A.L. Carriquiry. 1991. Estimation of the correlation among days for transformed nutrient intake. Iowa State University, Unpublished report prepared for the Human Nutrition Information Service, U.S. Department of Agriculture.

Block, G., C.M. Dresser, A.M. Hartman, and M.D. Carroll. 1985. Nutrient sources in the American diet: Quantitative data from the NHANES II survey. II. Macronutrients and fats. American Journal of Epidemiology 122:27–40.

Briefel, R.R., M.A. McDowell, K. Alaimo, et al. 1995. Total energy intake of the U.S. population: The third national health and nutrition examination survey, 1988–91. American Journal of Clinical Nutrition 62(suppl.):1072S–1080S.

DeMaio, T.J., 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. Vol. 2, pp. 1021–1026. American Statistical Association, Alexandria, VA.

Federation of American Societies for Experimental Biology. 1989. Nutrition monitoring in the United States—An update report on nutrition monitoring. DHHS Publication No. (PHS) 89–1255. U.S. Government Printing Office, Washington, DC.

Federation of American Societies for Experimental Biology. 1995. Third report on nutrition monitoring in the United States. U.S. Government Printing Office, Washington, DC.

Fisher, R.P., and K.L. Quigley. 1991. Applying cognitive theory in public health investigations: Enhancing food recall with the cognitive interview. *In* J.M. Tanur, ed., Questions about questions. Inquiries into the cognitive bases of surveys, pp. 154–169. Russell Sage Foundation, New York.

Fries, E., P. Green, and D.J. Bowen. 1995. What did I eat yesterday? Determinants of accuracy in 24-hour food memories. Applied Cognitive Psychology 9:143–155.

Georgiou, C.C., and A.B. Arquitt. 1992. Different food sources of fat for young women who consumed lower-fat diets and those who consumed higher-fat diets. Journal of the American Dietetic Association 92:358–360.

Gorbach, S.L., A. Morrill-LaBrode, M.N. Woods, et al. 1990. Changes in food patterns during a low-fat dietary intervention in women. Journal of the American Dietetic Association 90:802–809.

Gorder, D.D., T.A. Dolecek, G.G. Coleman, et al. 1986. Dietary intake in the multiple risk factor intervention trial (MRFIT): Nutrient and food group changes over 6 years. Journal of the American Dietetic Association 86:744–751.

Krebs-Smith, S.M., F.J. Cronin, D.B. Haytowitz, and D.A. Cook. 1992. Food sources of energy, macronutrients, cholesterol, and fiber in diets of women. Journal of the American Dietetic Association 92:168–174.

Kristal, A.R., B.F. Abrams, M.D. Thornquist, et al. 1990a. Development and validation of a food use checklist for evaluation of community nutrition interventions. American Journal of Public Health 80:1318–1322.

Kristal, A.R., A.L. Shattuck, and H.J. Henry. 1990b. Patterns of dietary behavior associated with selecting diets low in fat: Reliability and validity of a behavioral approach to dietary assessment. Journal of the American Dietetic Association. 90:214–220.

Kristal, A.R., A.L. Shattuck, H.J. Henry, and A.S. Fowler. 1990c. Rapid assessment of dietary intake of fat, fiber, and saturated fat: Validity of an instrument suitable for community intervention research and nutritional surveillance. American Journal of Health Promotion 4:288–295.

Kristal, A.R., E. White, A.L. Shattuck, et al. 1992. Long-term maintenance of a low-fat diet: Durability of fat-related dietary habits in the Women's Health Trial. Journal of the American Dietetic Association 92:553–559.

Means, B., A. Nigam, M. Zarrow, et al. 1989. Autobiographical memory for health-related events. U.S. Department of Health and Human Services, National Center for Health Statistics, Vital and Health Statistics, Series 6, No. 2.

Means, B., G.E. Swan, J.B. Jobe, and J.L. Esposito. 1991. An alternative approach to obtaining personal history data. *In* P.P. Biemer, R.M. Groves, L.E. Lyberg, eds., et al., Measurement errors in surveys, pp. 167–183. John Wiley & Sons, New York.

Mertz, W., J.C. Tsui, J.T. Judd, et al. 1991. What are people really eating? The relation between energy intake derived from estimated diet records and intake determined to maintain body weight. American Journal of Clinical Nutrition 54:291–295.

Rogers, E.M. 1983. Attributes of innovations and their rate of adoption. *In* Diffusion of innovations, 3rd ed., pp. 210–240. The Free Press, Macmillan Publishing Co., New York.

Schectman, G., P. McKinney, J. Pleuss, and R.G. Hoffman. 1990. Dietary intake of Americans reporting adherence to a low cholesterol diet (NHANES II). American Journal of Public Health 80:698–703.

Thompson, F.E., M.F. Sowers, E.A. Frongillo, and B.J. Parpia. 1992. Sources of fiber and fat in diets of U.S. women aged 19 to 50: Implications for nutrition education and policy. American Journal of Public Health 82:695–702.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. 1990. Nutrition and your health: Dietary guidelines for Americans, 3rd ed. U.S. Department of Agriculture, Home and Garden Bulletin No. 232.

U.S. Department of Health and Human Services, Public Health Service. 1994. Plan and operation of the third national health and nutrition examination survey, 1988–94. National Center for Health Statistics. Vital Health Stat. 1(32).

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

U.S. Department of Health and Human Services and U.S. Department of Agriculture. 1986. Nutrition Monitoring in the United States—A report from the Joint Nutrition Monitoring Evaluation Committee. Public Health Service, DHHS Publication No. (PHS) 86–1255.

Westat, Inc. 1993. Interview Manual. Continuing survey of food intakes by individuals and the diet and health knowledge survey, 1994–96. Unpublished document.

Westat, Inc. 1995. The continuing survey of food intakes by individuals/diet and health knowledge survey, 1994–96: 1994 survey operations report. Prepared for the Agricultural Research Service, U.S. Department of Agriculture.

Table 10. —Summary of questionnaires and data collection procedures for the 1994–96 CSFII/DHKS*

Questionnaire	Type of data collected	Respondent	Mode of administration	Average administration time (minutes)	Contact attempts required (number)
Screener	List of household members and each one's age, race, and ethnicity	Household member 18 yr. or older	In person	8	Two visits to sampled dwelling unit, then two more with neighbor
Household	Educational and employment status of household members 15 yr. and older, household income, food assistance program participation, food expenditures, and other food-related practices	Main food preparer/meal planner for the household or any adult knowledgeable about household characteristics, especially income. Did not have to be a sample person	In person	20	Three visits after screening
Day-1 intake	24-hr. dietary recall, time and name of eating occasions, source of foods, water intake, vegetarianism, supplement use, height and weight, allergies, smoking, exercise frequency, type of diet, health status, and consumption of alcoholic beverages	Sample person. Adult proxy for children under 6 yr. and persons mentally or physically incapable of self-report; adult assistance required for children 6-11 yr.	In person	33	Three visits; additional visits with supervisor approval
Day-2 intake	24-hr. dietary recall and a food list question that asked whether specific foods were consumed in the past year	Sample person or adult proxy	In person, telephone only if approved by supervisor in advance (3% in 1994)	30	Three visits, then telephone with supervisor approval

Table 10. —Summary of questionnaires and data collection procedures for the 1994–96 CSFII/DHKS*—Continued

Questionnaire	Type of data collected	Respondent	Mode of administration	Average administration time (minutes)	Contact attempts required (number)
Diet and Health Knowledge Survey	Dietary knowledge, behavior, and attitudes such as perceived adequacy of food and nutrient intake, perceived importance of dietary guidance food intake behaviors, awareness of diet-health relationships, and use and understanding of food labels	Sample person 20 yr. or older who completed day-1 interview; selected by an automated, randomized procedure	Telephone, in person only if approved by supervisor in advance (16% in 1994)	Telephone interview, 30; in-person interview, 35; overall, 31	Six telephone, then, four in-person visits with supervisor approval

* Adapted from Westat, Inc. (1993).

Chapter 5: Response Rates
By Joseph D. Goldman and Alvin B. Nowverl

As with any survey, not all individuals drawn into the sample participated. Survey response rates provide a measure of this undercoverage. This chapter provides reasons for nonresponse, response rate calculations, and tables showing overall response rates for CSFII/DHKS 1994–96.

ARS reports overall response rates when referring to final survey results. An overall response rate estimates the proportion of all eligible individuals or households that responded to a particular interview. "All eligible individuals or households" is the sum of those who actually responded to an interview (participants) and an estimate of those who were eligible to participate but did not do so (eligible nonparticipants). In general, ARS assumes that nonparticipants in various phases of the survey were eligible in the same proportions as participants.

Overall response rates can often be stated as the product of participation rates. Participation rates reflect a rate of participation relative to a particular interview. For example, the household interview participation rate is the proportion of screened households containing sample persons (SP's) who provided a household interview. This is not an estimate. The number of screened households and the number of participating households are known. An overall household response rate needs to account for all households in the sample that contain SP's. This requires estimating the number of interview-eligible households that were not screened.

The first, and often the only, interview asked of a household is the screening interview. The screening interview determines whether there are household members who are eligible for the intake interviews. Household members who are eligible are called sample persons (SP's). If there are no SP's in a household, the household's participation concludes with the screening interview. If one or more SP's are identified in the household, the household questionnaire is administered during the household interview and the day-1 food intake questionnaires are administered to each SP during the day-1 interviews. At a later time, day-2 interviews are conducted with each SP who participated in a day-1 interview. The DHKS questionnaire is administered to selected adult day-1 participants (no more than one per household) during the DHKS interview.

There were several reasons for nonresponse (nonparticipation). Refusals accounted for more than 50 percent of nonresponse for each type of interview. Inability to contact the household or SP was the next largest category of

nonresponse. Other reasons for nonresponse include the unavailability of members of a household for the field period, household members who moved and could not be located, language problems for which a suitable translator could not be found, and households in buildings or communities where access was limited.

The overall response rate formulas are provided below, along with derivations and assumptions made about the eligibility of nonparticipants. Table 11 provides a summary of the response results for the CSFII/DHKS 1994–96.

Screening

The screening rate, R_S, is defined as the proportion of nonvacant households in the sample that were screened. That is,

$$R_S = \frac{H_S}{H_S + H_{NS}}, \qquad [1]$$

where

H_S is the number of screened households and
H_{NS} is the number of unscreened households.

The number of addresses either vacant or ineligible for screening, H_v, is not relevant to this calculation. It is assumed that all such addresses were identified, that is, that all of the households represented by H_{NS} were eligible to be screened. Because screening is the first step of the interview process, the calculation of the screening rate is a simpler calculation than that of the other overall response rates. The screening rate is a factor in the computation of each of the other overall rates.

Day 1

The overall day-1 response rate, R_{D1}, may be written as the product of the screening rate, R_S, and the day-1 participation rate, R_{PART1}.

$$R_{D1} = R_S * R_{PART1}. \qquad [2]$$

The day-1 participation rate, R_{PART1}, is defined as the proportion of screened SP's eligible for day 1 who provided complete day-1 intake interviews, that is,

$$R_{PART1} = \frac{N_{D1}}{N_{SP\text{-}S(1)}},$$

where

N_{D1} is the number of SP's participating in day 1, and
$N_{SP\text{-}S(1)}$ is the number of screened SP's eligible for day 1.

An overall day-1 response rate estimates the proportion of all eligible SP's who responded to the day-1 interview. Equation 2 was derived from this definition as follows: The total number of SP's eligible for day 1, N_{SP1}, is

$$N_{SP1} = N_{SP\text{-}S(1)} + N_{SP\text{-}NS(1)} = H_S * \frac{N_{SP\text{-}S(1)}}{H_S} + H_{NS} * \frac{N_{SP\text{-}NS(1)}}{H_{NS}},$$

where

$N_{SP\text{-}S(1)}$ is the number of screened SP's eligible for day 1 and
$N_{SP\text{-}NS(1)}$ is the number of day-1-eligible SP's in the unscreened households.

It is assumed that the proportion of day-1-eligible SP's is the same among the unscreened households as it is among the screened households—

$$\frac{N_{SP\text{-}S(1)}}{H_S} = \frac{N_{SP\text{-}NS(1)}}{H_{NS}}.$$

N_{SP1} may then be estimated by N^*_{SP1}, where

$$N^*_{SP1} = H_S * \frac{N_{SP\text{-}S(1)}}{H_S} + H_{NS} * \frac{N_{SP\text{-}S(1)}}{H_S}$$

$$= (H_S + H_{NS}) * \frac{N_{SP\text{-}S(1)}}{H_S} = \frac{H_S + H_{NS}}{H_S} * N_{SP\text{-}S(1)} = N_{SP\text{-}S(1)} + \frac{H_{NS}}{H_S} * N_{SP\text{-}S(1)}.$$

Therefore, the overall day-1 response rate, R_{D1}, the ratio of day-1-participating SP's to the estimated total number of day-1-eligible SP's, is $R_{D1} = \frac{N_{D1}}{N^{*}_{SP1}} = \frac{N_{D1}}{N_{SP-S(1)} + \frac{H_{NS}}{H_S} * N_{SP-S(1)}}$.

Equivalently,

$$R_{D1} = \frac{N_{D1}}{N^{*}_{SP1}} = \frac{N_{D1}}{\frac{H_S + H_{NS}}{H_S} * N_{SP-S(1)}} = \frac{H_S}{H_S + H_{NS}} * \frac{N_{D1}}{N_{SP-S(1)}} = R_S * R_{PART1}$$

which shows R_{D1} expressed as the product of the screening rate and the day-1 participation rate, as stated by equation 2.

2 Days

The overall 2-day response rate, R_{D2}, may be written as the product of the screening rate, R_S, and the 2-day participation rate, R_{PART2}.

$$R_{D2} = R_S * R_{PART2} \qquad [3]$$

The 2-day participation rate, R_{PART2}, is defined as the proportion of screened SP's eligible for day 2 who provided complete day-2 intake interviews.

$$R_{PART2} = \frac{N_{D2}}{N_{SP-S(2)}},$$

where

N_{D2} is the number of SP's participating in day 2 and
$N_{SP-S(2)}$ is the number of screened SP's eligible for day 2.

Note that a discussion of participation in day 2 is synonymous with a discussion of 2-day participation. If an SP participated in day 2, the SP must also have participated in day 1. Similarly, an SP eligible for day 2 must have been eligible for day 1.

An overall 2-day response rate estimates the proportion of all eligible SP's who responded to the day-2 interview. The derivation of equation 3 from this definition is analogous to the derivation of the overall day-1 response rate in equation 2.

Household Interview

The overall household interview response rate, R_H, may be written as the product of the screening rate, R_S, and the household interview participation rate, R_{PARTH}

$$R_H = R_S * R_{PARTH}. \quad [4]$$

The household interview participation rate, R_{PARTH}, is defined as the proportion of screened households from which a complete household interview was obtained.

$$R_{PARTH} = \frac{N_H}{N_{H-S}},$$

where

N_H is the number of households with complete interviews and
N_{H-S} is the number of households found through screening and eligible for a household interview.

An overall household interview response rate estimates the proportion of all eligible households providing complete household interviews. The derivation of equation 4 from this definition is analogous to the derivation of the overall day-1 response rate in equation 2.

Diet and Health Knowledge Survey

The calculation of the overall DHKS response rate is more complicated than the calculation of the other overall rates because it must incorporate a modified (household level) day-1 participation rate for adults. Nevertheless, the definition can be stated simply as the product of the screening rate and two participation rates,

$$R_{DHKS} = R_S * R_{20} * R_{PARTDHK} \quad [5]$$

where

$R_{PARTDHK}$ is the DHKS participation rate and
R_{20} may be called the adult participation rate.

The DHKS participation rate, $R_{PARTDHK}$, is defined as the proportion of screened households from which a DHKS interview was attempted that actually provided complete DHKS interviews, that is, households with an SP age 20 or older who completed day 1 in person (not by proxy) and a DHKS interview. This may be written as

$$R_{PARTDHK} = \frac{N_{PARTDHK}}{N_{DHKELIG-A}}$$

where

$N_{PARTDHK}$ is the number of households with complete DHKS interviews and
$N_{DHKELIG-A}$ is the number of households from which a DHKS interview was attempted.

Because the overall rate must account for screened households with adult SP's who did not complete day 1, a second participation rate needs to be defined. This adult participation rate, R_{20}, is the proportion of screened households with one or more SP's age 20 or older that had at least one of these adult SP's complete day 1 (either in person or through a proxy).

$$R_{20} = \frac{N_{20-A}}{N_{20-A} + N_{20-NI}}$$

where

N_{20-A} is the number of screened households containing at least one adult SP and from which at least one of these adult SP's completed day 1 and
N_{20-NI} is the number of screened households containing at least one adult SP and from which none of the adult SP's completed day 1.

An overall DHKS response rate estimates the proportion of all DHKS-eligible SP's who responded to the DHKS interview. Equation 5 was derived from this definition as follows: The total number of households with DHKS-eligible respondents, $N_{DHKELIG}$, may be written as

$$N_{DHKELIG} = N_{DHKELIG-A} + N_{DHKELIG-NI} + N_{DHKELIG-NS}$$

where

$N_{DHKELIG-A}$ is the number of households for which at least one DHKS respondent was identified and a DHKS interview was attempted, that is, at least one adult SP completed a day-1 intake in person (not by proxy);
$N_{DHKELIG-NI}$ is the number of screened households with at least one DHKS-eligible SP, but where no day-1 intakes were completed by an age-eligible SP; and
$N_{DHKELIG-NS}$ is the number of unscreened households with at least one DHKS-eligible SP.

Of these three components, $N_{DHKELIG-A}$ is known and $N_{DHKELIG-NI}$ and $N_{DHKELIG-NS}$ must be estimated.

To estimate $N_{DHKELIG-NI}$, first let N_{20-NI} be the number of households with at least one age-eligible SP, but with no complete day-1 intakes for any of these SP's. Also, let N_{20-A} be the number of households with at least one age-eligible SP and at least one complete day-1 intake among these adult SP's. It is assumed that, had the screened, nonparticipating adults completed day-1 interviews, they would have done so in-person in the same proportion that was actually observed among the households with participating adults; that is,

$$\frac{N_{20-A}}{N_{DHKELIG-A}} = \frac{N_{20-NI}}{N_{DHKELIG-NI}}$$

$N_{DHKELIG-NI}$ is then estimated by $N^*_{DHKELIG-NI}$ where

$$N^*_{DHKELIG-NI} = \frac{N_{20-NI}}{N_{20-A}} * N_{DHKELIG-A}.$$

It is also assumed that DHKS-eligible households would have been found among the unscreened households in the same proportion as was actually observed among the screened households; that is,

$$\frac{N_{DHKELIG-A} + N_{DHKELIG-NI}}{H_S} = \frac{N_{DHKELIG-NS}}{H_{NS}}.$$

$N_{DHKELIG-NI}$ is then estimated by $N^*_{DHKELIG-NS}$ where

$$N^*_{DHKELIG-NS} = \frac{H_{NS}}{H_S} * (N_{DHKELIG-A} + N^*_{DHKELIG-NI}).$$

$N_{DHKELIG}$, the total number of households eligible for the DHKS is then estimated by $N^*_{DHKELIG}$ where

$$N^*_{DHKELIG} = N_{DHKELIG-A} + N^*_{DHKELIG-NI} + N^*_{DHKELIG-NS}$$

$$= N_{DHKELIG-A} + \left(\frac{N_{20-NI}}{N_{20-A}} * N_{DHKELIG-A}\right) + \left[\frac{H_{NS}}{H_S} * (N_{DHKELIG-A} + N^*_{DHKELIG-NI})\right]$$

$$= N_{DHKELIG-A} + \left(\frac{N_{20-NI}}{N_{20-A}} * N_{DHKELIG-A}\right) + \left[\frac{H_{NS}}{H_S} * \frac{(N_{20-A} + N_{20-NI})}{N_{20-A}} * N_{DHKELIG-A}\right]$$

The overall DHKS response rate, R_{DHKS}, is

$$R_{DHKS} = \frac{N_{PARTDHK}}{N^{*}_{DHKELIG}}$$

$$= \frac{N_{PARTDHK}}{N_{DHKELIG-A} + \left(\frac{N_{20-NI}}{N_{20-A}} * N_{DHKELIG-A}\right) + \left[\frac{H_{NS}}{H_S} * \frac{(N_{20-A} + N_{20-NI})}{N_{20-A}} * N_{DHKELIG-A}\right]}$$

where

$N_{PARTDHK}$ is the number of complete DHKS interviews.

R_{DHKS} may also be written as

$$R_{DHKS} = \frac{N_{PARTDHK}}{N_{DHKELIG-A} + N^{*}_{DHKELIG-NI} + \left[\frac{H_{NS}}{H_S} * (N_{DHKELIG-A} + N^{*}_{DHKELIG-NI})\right]}$$

$$= \frac{N_{PARTDHK}}{\left(1 + \frac{H_{NS}}{H_S}\right) * (N_{DHKELIG-A} + N^{*}_{DHKELIG-NI})}$$

$$= \frac{N_{PARTDHK}}{\left(1 + \frac{H_{NS}}{H_S}\right) * \left[N_{DHKELIG-A} + \left(\frac{N_{20-NI}}{N_{20-A}} * N_{DHKELIG-A}\right)\right]}$$

$$= \frac{N_{PARTDHK}}{\left(1 + \frac{H_{NS}}{H_S}\right) * \left(1 + \frac{N_{20-NI}}{N_{20-A}}\right) * N_{DHKELIG-A}}$$

$$= \frac{H_S}{(H_S + H_{NS})} * \frac{N_{20-A}}{(N_{20-A} + N_{20-NI})} * \frac{N_{PARTDHK}}{N_{DHKELIG-A}}.$$

Therefore, R_{DHKS} may be expressed as the product of the screening rate and two participation rates,

$$R_{DHKS} = R_S * R_{20} * R_{PARTDHK}, \quad [5]$$

where

R_S is the screening rate,

$R_{PARTDHK} = \frac{N_{PARTDHK}}{N_{DHKELIG-A}}$, is the DHKS participation rate, and

$R_{20} = \frac{N_{20-A}}{N_{20-A} + N_{20-NI}}$ is the adult participation rate.

This participation rate, R_{20}, is the proportion of screened households with one or more SP's age 20 or older that had at least one of these adult SP's complete day 1 (either in person or through a proxy).

Table 11.—CSFII/DHKS Response Rates for 1994–96

Category	1994–96	1994	1995	1996
Total dwelling units selected	34,016	9,628	11,823	12,565
Dwelling units vacant or determined not to be households	4,189	1,161	1,337	1,691
Occupied dwelling units ($H_S + H_{NS}$)	29,827	8,467	10,486	10,874
Screened households (H_S)	29,371	8,333	10,333	10,705
Screening rate (R_S)	**98.5%**	**98.4%**	**98.5%**	**98.4%**
Total SP's identified through screening [1]	19,830	6,868	6,576	6,386
Screened SP's eligible for day 1 ($N_{SP\text{-}S(1)}$)	19,818	6,864	6,570	6,384
SP's completing day 1 (N_{D1})	16,103	5,589	5,326	5,188
Day-1 participation rate (R_{PART1})	81.3%	81.4%	81.1%	81.3%
Overall day-1 response rate (R_{D1})	**80.0%**	**80.1%**	**79.9%**	**80.0%**
Screened SP's eligible for day 2 ($N_{SP\text{-}S(2)}$)	19,813	6,863	6,567	6,383
SP's completing day 2 (N_{D2})	15,303	5,311	5,072	4,920
Two-day participation rate (R_{PART2})	77.2%	77.4%	77.2%	77.1%
Overall 2-day response rate (R_{D2})	**76.1%**	**76.2%**	**76.1%**	**75.9%**
Screened households eligible for household interview ($N_{H\text{-}S}$)	9,658	3,264	3,376	3,018
Households completing household interview (N_H)	8,302	2,813	2,892	2,597
Household interview participation rate (R_{PARTH})	86.0%	86.2%	85.7%	86.1%
Overall household interview response rate (R_H)	**84.6%**	**84.8%**	**84.4%**	**84.7%**
Households with adult SP's completing day 1 and eligible for DHKS ($N_{DHKELIG\text{-}A}$)	6,294	2,047	2,159	2,088
DHKS participants ($N_{PARTDHK}$)	5,765	1,879	1,966	1,920,
DHKS participation rate ($R_{PARTDHK}$)	**91.6%**	**91.8%**	**91.1%**	**92.0%**
Screened households with at least one adult SP ($N_{20\text{-}A} + N_{20\text{-}NI}$)	7,803	2,527	2,700	2,576
Screened households with at least one adult SP who completed day 1 ($N_{20\text{-}A}$)	6,360	2,073	2,188	2,099
Adult participation rate (R_{20})	81.5%	82.0%	81.0%	81.5%
Overall DHKS response rate (R_{DHKS})	**73.5%**	**74.1%**	**72.7%**	**73.8%**

[1] In several cases, SP's were identified through screening but left the population of interest before they had a chance to participate. This happened when SP's moved out of the United States, were institutionalized, or died.

Chapter 6: Management and Quality Control
By Sharon J. Mickle, Katherine E. Sykes, Rhonda S. Sebastian, and Junko A. Tamaki

The implementation of strong management and quality control at every step during the development and execution of the CSFII/DHKS 1994–96 was a primary objective. This chapter provides an overview of the quality control features incorporated into the operations of the survey.

Management Structure and Reporting Requirements

ARS staff oversaw the management of all CSFII/DHKS 1994–96 operations. Staff included nutritionists, survey methodologists, statisticians, food specialists, and the contract officer's representative or project manager. The staff evaluated all survey methods and revised procedures as appropriate, monitored the survey contract, reviewed the data provided by the contractor, and prepared the data for public release.

Westat was required to assemble a group of key personnel and maintain their involvement throughout the course of the survey. The responsibilities of these personnel were to develop the sample design and survey materials, including questionnaires, manuals, and promotional materials; hire and train interviewers to collect data; conduct and monitor field operations; and process and transmit the data to ARS. All plans and materials were reviewed and approved by ARS.

The survey contract called for frequent communication between ARS and Westat staff. Regular meetings began shortly after the contract was awarded. These included an orientation meeting for the introduction of staff, review of the roles of both parties, review of survey procedures and requirements, identification of communication channels, and discussion of the transfer procedures for survey materials. At quarterly meetings, ARS and Westat staff discussed data collection progress, data-processing status, and activities for the next quarter. Additional meetings were held as needed and covered specific issues such as data processing efficiency, revisions to the Food Instruction Booklet, and development of survey weights.

ARS monitored survey progress through regular reports. The contract required monthly progress reports, daily field monitoring and data-processing status reports, weekly tracking reports of field status and data processing, annual survey operation reports, and a final operations report covering survey development and 3 years of data collection.

The monthly reports provided a summary of significant survey activities during the calendar month and status of those activities. The activities included hiring and training personnel, materials' development, data collection, data preparations, and weekly delivery of data. Sample yield information, which was the basis for response rate calculations, was also included in the monthly report. Detailed tables provided counts of sampled dwelling units (DU's), numbers of occupied DU's, eligible and ineligible households, numbers of sampled persons identified, and numbers of sampled persons participating. The yield information was provided for the current quarter as well as cumulatively.

The daily field-monitoring and forms-processing reports provided a snapshot of data collection and processing activities. The current status of contact with each sampled DU was captured. This information allowed supervisors to monitor the productivity of the interviewers. The status of questionnaire processing allowed monitoring of coder productivity. Weekly tracking reports provided information on the field status and processing status for each questionnaire type, coder error rates, cumulative questionnaire response rates by primary sampling unit and by interviewer, and running totals of intake interviews conducted by day of week. The annual survey operations report included a discussion of the execution of the sample design; summaries of the data collection activities, such as data-tracking systems, field and coding staff performance, and interviewer debriefings; response results; and weighting and variance estimation procedures. The inclusion of daily and weekly reports was a significant change in survey monitoring, allowing ARS staff the opportunity to assess the status of survey operations on a daily basis.

Materials Development

ARS and Westat staff worked collaboratively in preparing materials for the pilot study, including the questionnaires, measurement aids, and Food Instruction Booklet (see chapter 4); publicity materials; and instructional manuals and training programs for field supervisors, interviewers, and coders. ARS and Westat staff used the results of the pilot study to revise the survey questionnaires and related materials for the main survey. Instructional manuals and training programs for field supervisors, interviewers, and coders were also revised to reflect information obtained from the pilot study.

Interviewers used publicity materials to describe the survey to respondents, establish the legitimacy of the survey, gain respondent cooperation, and answer respondent questions. The publicity materials included a brochure, flyers, and a factbook. Along with the introductory letter described in chapter 4, a colorful brochure (see appendix A) was mailed to households 1 week in advance of the

interviewer's initial contact. The brochure included general information about the survey, uses of the survey data, and results from an earlier USDA food consumption survey. The flyers were designed to further encourage the participation of adults and teens (see appendix A). The factbook contained news articles reporting past survey findings, press releases announcing the survey, and factsheets listing uses of the data. A photo identification badge was used by interviewers to establish their legitimacy.

A trainer's guide containing scripted lectures, role plays, and exercises was developed by Westat to train field supervisors, senior interviewers, and interviewers. It was further refined during supervisor/senior interviewer training. An interviewer manual provided question-by-question instructions for administering survey questionnaires, including responses for anticipated questions from respondents. The manual assisted interviewers during training and in the field by serving as a key source of reference information for interview procedures. Any information provided to interviewers during training was included in the manual. Interviewers were also given detailed instructions for entering and transmitting data to the home office about interviewer production and field results (see "Field Management" below).

Manuals were developed for food intake and nonintake coding. The food intake coding manual included instructions for coding the food intake and diet and health-related questions from the day-1 and day-2 intake questionnaires. The nonintake manual provided instructions for coding the screener, household, and DHKS questionnaires. In addition, coders were given guidelines for handling infrequent or unusual responses. The content of the manuals and guidelines are discussed later in this chapter.

Training

For data collection, Westat recruited and trained 5 regional supervisors, 5 senior interviewers, and 90 interviewers (10 of whom were bilingual). Senior interviewers were available to travel to all locations within a region to convert nonresponse cases and observe interviewers at work. Westat also recruited and trained 14 coders. All field and coding staff received in-person training monitored by ARS.

Supervisor and senior interviewer training

Supervisors and senior interviewers attended an 8-day training session. The first 6 days covered topics that would be presented in interviewer training. The last 2 days were devoted to supervisory responsibilities, including assigning work to interviewers, monitoring field progress, conducting validation interviews and interviewer observations, and reviewing taped interviews. Supervisors were also taught a standardized procedure designed to address nonresponse conversion. ARS staff met with the contractor to evaluate the training at the end of each day. Minor changes were made to the materials to reflect what had been learned from supervisor training prior to the interviewer training.

Interviewer training

Field interviewers were trained in a 7-day session (see the abbreviated interviewer training agenda at the end of the chapter). Bilingual interviewers received an additional day of training on the use of Spanish language questionnaires and materials.

Prior to training, interviewers were required to complete a home-study assignment based on field procedures in the interviewer manual and the Food Instruction Booklet. These exercises were submitted to their supervisors for review and feedback on the first day of training.

The training was conducted by Westat in five groups or communities with each of the five supervisors serving as the principal instructor for the interviewers in their region. This arrangement maximized trainee participation and provided ample opportunity for supervisory staff to observe and evaluate trainee performance. To ensure standardization of information presented to interviewers, each supervisor used the trainer's guide, which included all lecture scripts, role plays, and written exercises.

Training scripts and exercises were designed to provide increasingly complex situations that the interviewers were likely to encounter. Throughout training, extensive hands-on experience was provided with the questionnaires, the Food Instruction Booklet, and the measuring guides. Trainees were given the opportunity to interview respondents who were recruited for the training session.

After returning home, interviewers were required to complete a practice interview with a neighbor or family member and a scripted telephone interview with their supervisor or senior interviewer. After supervisory review of the practice

interview and successful completion of the mock interview, supervisors authorized the interviewers to begin work.

Ongoing training included the use of periodic quizzes. These quizzes were developed and administered by supervisors over the telephone. The quizzes were used to assess the interviewer's understanding of complex areas of the questionnaires that were common sources of error. The supervisor provided retraining on the questions that the interviewer answered incorrectly.

Ninety interviewers were trained during the first session in January 1994. An additional 44 interviewers were trained later in 1994, 1995, and 1996 to replace those who quit or were released. During the course of the survey, 15 to 20 percent of the interviewers terminated their employment. The contractor was required to submit to ARS quarterly reports that contained the exit interviews of all interviewers who left. The exit interviews were reviewed by ARS as a means of monitoring changes in the field staff.

At the end of each of the first 2 survey years, the field staff was brought together to receive 1-day refresher training. This training focused on refinements to the materials for the next survey year, as well as on areas of survey administration needing improvement. Interviewers were collectively debriefed following retraining.

Food coder training

Food coders began their training by attending 5 days of field supervisor training, which included an introduction to the survey, training on the day-1 and day-2 intake questionnaires, and sessions on the nonintake documents (screener, household, and DHKS questionnaires). In addition, coders were required to complete the interviewer home-study assignment before attending coder-specific training. Coder training continued with a 9-day training session that focused on Survey Net, the computer-assisted food coding system cooperatively developed by ARS with the University of Texas-Houston School of Public Health (see chapter 7). The coders were also trained to complete the three-stage quality review of intake questionnaires (see the "Data Processing" section below) and to code the nonfood questions on the intake questionnaires.

The materials used in coder training included general instructions for reviewing and coding food intake questions. These instructions included procedures for the three-stage quality review of intakes and for Survey Net coding. The coders were also given ARS food-coding guidelines and numerous coder aids, such as laminated copies of combination code descriptions and eating-occasion codes.

During Survey Net training, lectures, demonstrations, a tutorial, and exercises were used to introduce material. The training approach was designed to provide extensive hands-on practice. Coders were trained to code food intake data using the various features of Survey Net, including the food term search to identify proper food codes, recipe modification feature, and a copy feature for copying the same food within or across intakes.

Food coders were subsequently trained to manually code the diet and health-related questions on the intake questionnaire. Codebooks provided the file layouts, variable names and values, special notes to coders, and logics used in machine editing. Manual coding included correcting interviewer errors, marking fields to be skipped, entering codes for "don't know" and "refused" responses, and translating marginal notes into codes.

Before completion of training, coders were required to use Survey Net to code a final test set of three intakes provided by ARS. Coders had to achieve less than a 5 percent error rate on each intake for certification to code intake questionnaires. With experience, food coders achieved error rates that averaged to 0.4 percent, well below the rate allowed at initial certification. Error rates varied little among coders.

Food coders periodically received refresher training on Survey Net and the intake review. Survey Net retraining included searching for foods, entering quantities, and modifying recipes. Intake review retraining emphasized strategies for improving the efficiency and effectiveness of the review.

Nonintake coder training

Four of the 14 food coders were initially trained in a 4-day session to code and machine edit the nonintake documents (screener, household, and DHKS questionnaires). Four new nonintake coders were later trained so that food coders could exclusively code intakes.

The nonintake training materials included general instructions for reviewing and coding nonintake questions for each nonintake document and training exercise. The manual contained an introduction to nonintake data collection, the data flow, the general edit, and coding conventions. Training exercises were scripted to incorporate examples of many of the data problems the coders could encounter. ARS staff monitored all training sessions and provided feedback on the quality of coders' work throughout the survey.

Field Management

Strict management controls were required for data collection. Supervisors conducted weekly telephone conferences with interviewers to discuss field performance, using interviewer production and response rate information obtained through Westat's automated Field Management System (FMS). Supervisors and field interviewers used FMS-loaded laptop computers for quick e-mail communication and transmission of FMS information about each sample DU and each sample person (SP). FMS data were transmitted to the home office once a week. The FMS also allowed supervisors to produce specialized reports to determine whether survey procedures were being followed.

Field management efforts also included a systematic three-phase approach to handling nonresponse conversion. In phase 1, interviewers made the required number of attempts to complete the interviews. In phase 2, supervisors determined whether the interviewer should make further attempts to complete the interviews or transfer the case to another local interviewer who handled unsuccessful interviews. In phase 3, supervisors assessed whether further action was needed by a senior interviewer to complete the interview.

Several procedures were used to monitor interviewers on the quality of their performance and the data they collected. These included taped interviews, in-person observations, validation of interviews, and the three-stage quality review. Each interviewer audiotaped three interviews during the survey year--one intake and two DHKS interviews. (Also, 12 household interviews were taped.) The supervisor selected the cases to be taped, and interviews secured respondent permission in advance of taping. Tapes were mailed to supervisors who evaluated the dynamics of the interviewer-respondent interaction and the administration of the questionnaires. Interviewers received feedback on their performance and retraining if problems were found.

Both ARS and Westat conducted in-person observations of interviewers at work. During the first year, the objective was to evaluate interviewer performance and assess the survey instruments in actual field conditions. During the following 2 years, observations focused on the performance of new interviewers.

Validation interviews were conducted to verify that an interview had been done at the assigned address according to survey procedures. The original interview process was not replicated but simply confirmed. Supervisors and senior interviewers validated at least 10 percent of all selected DU's. DU's included households with sample persons, ineligible households, and units reported as vacant. Most of the validations were conducted by telephone. Those DU's

without a telephone, or where there were no reported numbers, were validated in person.

Several methods were used to facilitate communication between the home office and field staff, and provide retraining, if appropriate. E-mail was used to address issues related to data clarification and SP selection errors, provide information regarding respondent inquiries or concerns from the survey toll-free line, or respond to requests for refusal conversion letters. Field staff memos provided additional information on questions raised from training and interviewer diaries. An interviewer newsletter issued three times each year, called "Food for Thought," answered questions on different situations encountered in the field and on procedures, questionnaire items, and the Food Instruction Booklet. Interviewer anecdotes were also included.

All interviewers were collectively debriefed and retrained each year. Interviewers completed a questionnaire at the end of each year about their experiences with the survey materials and procedures. Results from the questionnaire provided the basis for an in-person debriefing, held in conjunction with refresher training. The questionnaire results and summarized discussions from the debriefing were included in the annual survey operations report.

Data Processing

Data collected in the field were to be processed and delivered to ARS within 30 days of receipt at Westat's home office. This requirement allowed for early identification and resolution of any problems in data collection or processing.

Data processing began when completed questionnaires arrived at Westat's home office and were registered in an automated forms-tracking system. The forms-tracking system was developed to monitor the progress of all survey documents through every step of data processing by recording when and by whom each task was completed.

Nonintake data processing

The flow of nonintake data and the steps involved in processing are presented in figure 1. At receipt, nonintake documents were given a general edit and a check for acceptability based on meeting minimum criteria. The minimum criteria were established by ARS to ensure a minimum level of acceptable data from each questionnaire (see minimum criteria at the end of the chapter). The document was then either entered into the forms-tracking system if it met the minimum criteria or

held for problem resolution if it did not meet the minimum criteria. Failure to pass the general edit or meet the minimum criteria generated procedures to retrieve missing data from the interviewer, the original respondent, or another knowledgeable respondent within the household.

Documents were then prepared for data entry. Nonintake coders assigned numeric codes for "don't know," "refused," and other handwritten responses, zero-filled dollar amount fields, and corrected any interviewer errors. Ten percent of the manual coding of nonintake documents was verified by having a second coder visually review the coding and identify and correct any errors.

Nonintake data were keyed with programs developed by Westat using a data entry system of hardware and software. Data entry was 100 percent verified through independent entry by a different operator. A Westat computer program called COED was used to machine edit the data. This process verified that keyed data were within the acceptable ranges for a particular variable and that skip patterns had been followed. Acceptable error rates were specified by ARS to be no more than 3 percent of data points coded. Actual error rates achieved were less than 1 percent on average across all nonintake documents.

Intake data processing

Processing of the day-1 and day-2 intake questionnaires involved quality review, data entry, verification, adjudication, and machine editing. The flow of intake data and steps involved in processing are outlined in figure 2.

The quality review of intakes consisted of three stages: the minimum criteria review, a food description and amount review, and a review of the intake for the most common interviewer errors. An automated system was designed to facilitate the review and to provide interviewers with timely feedback on the quality of the intake data. Information from each stage of the review was entered into the computer and a summary page was generated for each intake questionnaire.

During the first stage, coders were required to verify that each intake questionnaire had met the minimum criteria for acceptability within 2 days of a questionnaire's receipt by the contractor. If an intake questionnaire failed to meet the minimum criteria, data retrieval was conducted to obtain the information from the interviewer or the respondent. In 1994 and 1995, approximately 230 intake questionnaires did not meet minimum criteria when they were reviewed at Westat. By using data retrieval procedures, all but 17 of these intakes met the minimum criteria.

During the second stage of review, a comparison was made between each reported food description and the relevant food probes listed in the Food Instruction Booklet. The coder noted any foods where required probes were not asked by the interviewer.

The third stage of the review consisted of a check for general interviewer errors, such as an eating occasion left blank or illegible handwriting. The results of this general edit, along with those from the other two stages were entered into the computer and printed summaries were generated and forwarded to the field supervisors for feedback to interviewers.

Next, the intake documents were coded using Survey Net, which provided computer-assisted access to USDA's Food Coding Database and its component parts, including food and gram-weight descriptions and the recipe file (see chapter 7). ARS updated Survey Net periodically during the study to add system enhancements and to incorporate new food information into the database.

Ten percent of intakes were verified by having a second coder independently re-enter the intake information into Survey Net. The two entries were compared using adjudication reports developed by Westat. In adjudication, the supervisory staff reviewed a report comparing the two sets of Survey Net entries, corrected entries if necessary, and tabulated the number of coder errors. Coders were then given feedback and instruction, if necessary, on any problem areas found.

Data Transmission

Westat transmitted data in electronic and paper form to USDA on a weekly basis. For each delivery, ARS verified receipt of paper documents against an electronic list. The weekly receipt and review of data facilitated early detection and resolution of any problems in the delivery itself or in the quality of the data.

Final Processing

Review and editing

After the data were received, ARS staff took a series of steps in preparation for public release. First, the food intake data coded in Survey Net underwent final processing and review. All entries in each intake questionnaire requiring review or resolution by ARS were highlighted in Survey Net's food summary screens. Entries requiring review or resolution included all unknowns (those foods or food quantities that could not be coded by Westat coders), newly created recipe modifications, and notepad entries that consisted of questions from the coders and explanations of coding decisions by the coding supervisors.

The food and nutrient intake data from the intake questionnaires, along with the data from the screener, household, and DHKS questionnaires were also reviewed for reasonableness, completeness, and logic. For example, extremely large quantities of foods were verified, intakes with five or fewer food items reported were reviewed for completeness, and illogical entries such as a 12:30 a.m. school lunch were resolved.

In addition to these steps, ARS visually reviewed a random sample of 10 percent of intakes for accuracy in coding and data entry and listened to audiotaped interviews to monitor interviewer performance. Feedback was provided to Westat on all reviewed documents on a continual basis.

Benefits of the review process

The review process served several purposes. It gave feedback to Westat on the conduct of the survey and the performance of the field and coding staff. It contributed to the development of data review guidelines, as well as the updating of procedures related to food code selection and modification of the Food Instruction Booklet, questionnaires, and other survey materials, as needed during the course of the survey. Information obtained through the review process will be used in planning and developing methodological improvements for future surveys.

External evaluation of review procedures

A formal review of inhouse data processing for the CSFII/DHKS 1994–96 was held in November 1994. The goal of the session was to evaluate the data-processing and review system and identify any adjustments that would facilitate release of high-quality survey data in a more timely manner.

The review panel consisted of nutritionists, an epidemiologist, and survey specialists representing the Bureau of the Census (two reviewers); the National Heart, Lung and Blood Institute; the University of Maryland; and the ARS National Program Staff. The panelists were asked to focus their review on the efficiency of data processing and preparation, use of limited staff, the establishment of acceptable error levels in the data, procedures for contractor feedback, and use of automation in data processing. Recommendations from the review were incorporated by ARS to assure efficient and effective data review and editing, while ensuring that the quality of the data was maintained.

Figure 1. Data flow and processing of nonintake data

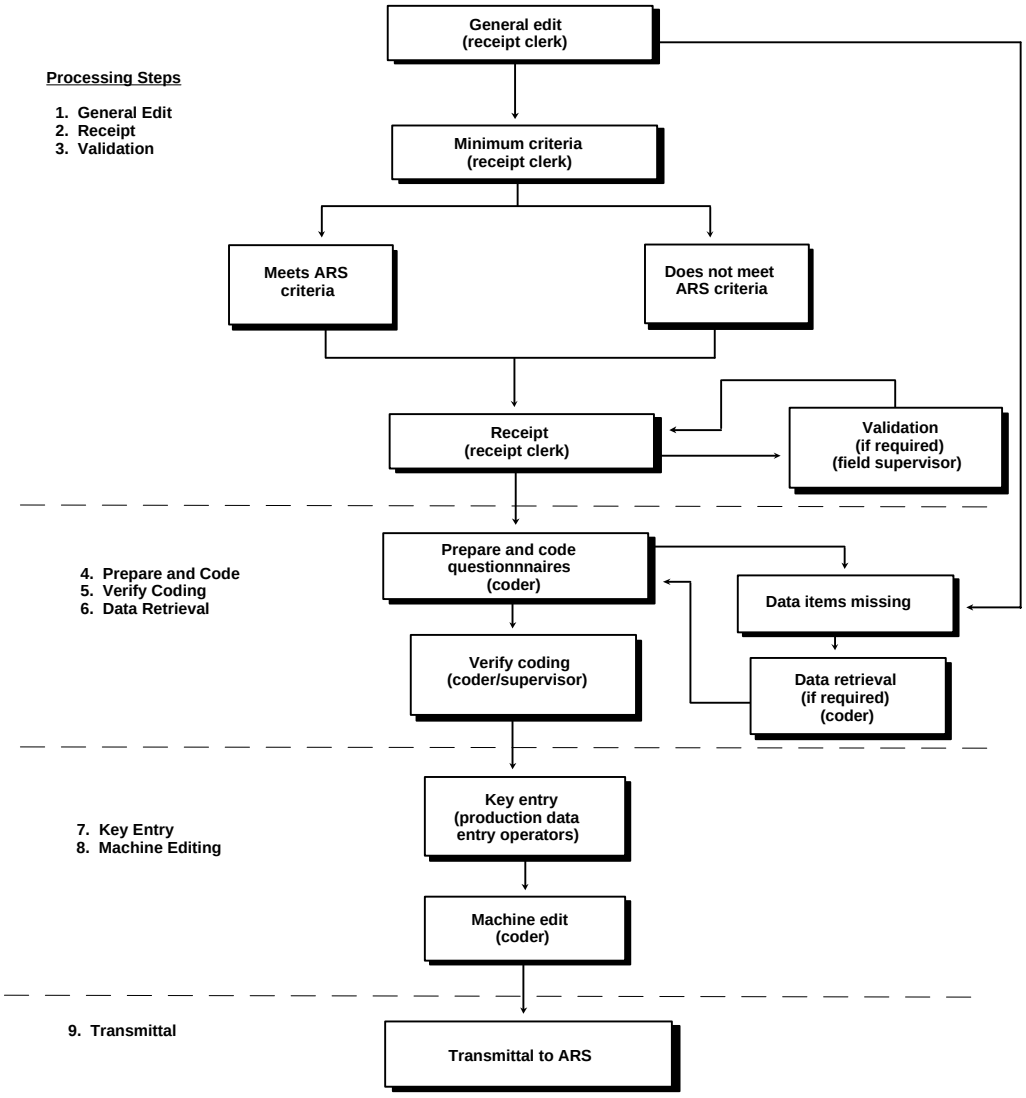
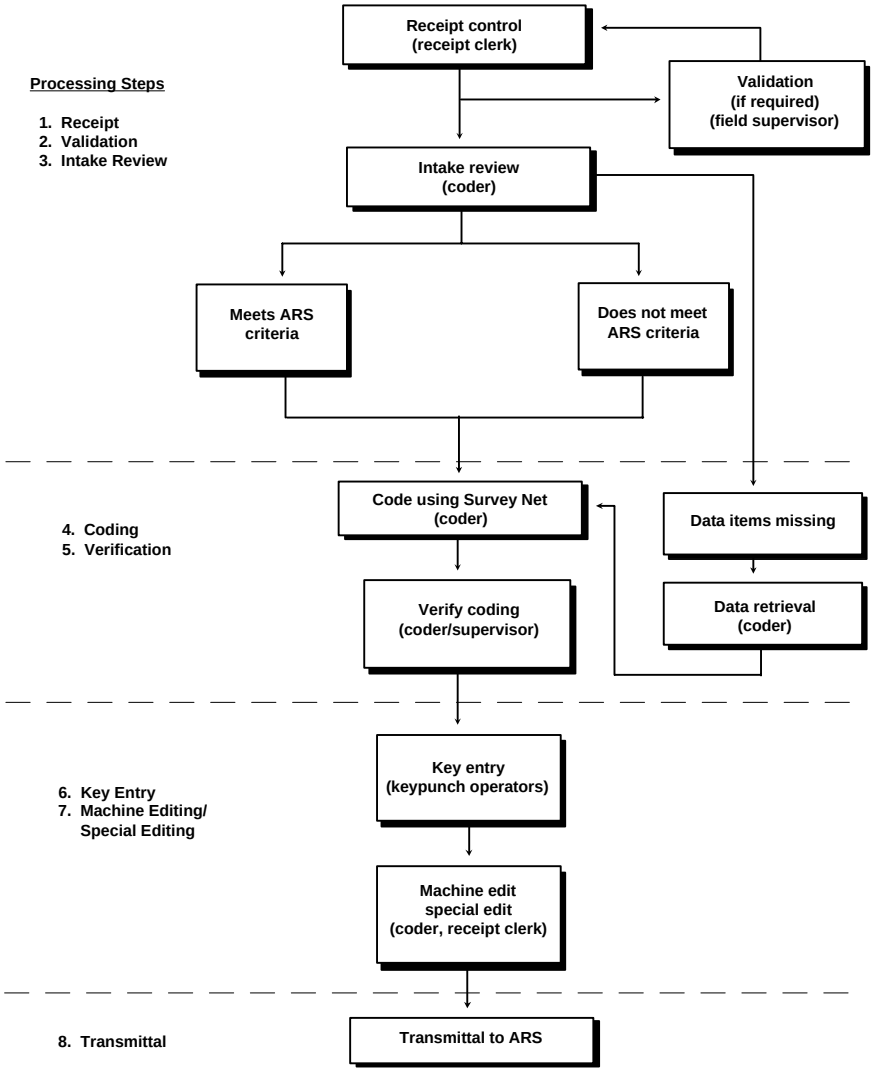


Figure 2. Data flow and processing of intake data



Abbreviated Interviewer Training Agenda

Pretraining

General Interviewing Techniques. Any interviewer new to Westat will be required to go through this session.

Day 1

1. **Introduction to What We Eat in America.** Introduction and welcome from ARS and Westat staff, including basic explanation of study.

2. **Introduction to Training**. Introduction to training staff and trainees. Brief overview of the week's agenda.

3. **Receiving Assignments and Beginning the Interview.** Overview of materials, listing process, and sample design. Also includes introduction at the door, answering questions, and using publicity materials.

4. **Overview of the Interviewing Process**. Introduction to the different questionnaires that will be used. Discussion of the basic contacting rules and timings for conducting the various types of questionnaires.

5. **Introduction to the Household Folder and Administering the Screener.** Using the household folder, including description of sections, recording screener results, and recording sample-person identifying information. Conducting the screener, including identifying the respondent, determining household members, and selecting sample persons.

6. **Quality Control of Listing**. Discussion of the missed structure and missed dwelling unit procedure including documentation.

7. **Demonstration of the Individual Intake Questionnaire**. Overview of the process for conducting the intake.

8. **Introduction to the Day-1 Intake, Food Instruction Booklet, and Measuring Guides**. Introduction to all necessary materials, introduction to the respondent and the flow of the intake, including the quick list.

Day 2

9. **Day-1 Intake Scripts.** A variety of scripts using the day-1 intake. These scripts will build in complexity and focus on bringing out various specifications (such as recording combinations).

Day 3

10. **Day-1 Intake Script and Missing-Meal Data Retrieval.** Script focusing on retrieval, including a 30-minute session for missing-meal data retrieval.

11. **Role Plays.** Trainees will complete two role plays of the day-1 intake. One of these will be scripted and the other will be a real interview of their partner. Trainees will be required to edit these questionnaires and turn them in at the end of the day.

12. **Overview of Paid Respondent Practice**. Description of process for the afternoon, including grouping trainees.

13. **Paid Respondent Practice**. Trainees will interview paid respondents using the day-1 intake questionnaire. This session will include a variety of respondents (that is, different ages, and so forth) to expose the trainees to a wide range of problems. Trainees will be required to edit these questionnaires and turn them in at the end of the day.

Day 4

14. **Answers to Questions from the Previous Day (including comments from ARS observation)**

15. **Recording Contact Attempts**. Overview of the process of entering results into the field management system, including discussion of the result codes, and practice making simple entries. As an exercise, trainees will be required to enter the status of each role play and paid respondent interview from the previous day.

16. **Day-2 Intake Scripts**. Script using the day-2 intake.

17. **Wrap-Up Exercise for the Intake**. Exercise to pull together all of the concepts and specifications relating to the intake and related materials. This session includes time for the trainees to complete and review the exercise.

18. **The Household Questionnaire**. Conducting the household questionnaire, including identifying the respondent.

19. **Editing the Individual Intake**. Exercise to practice recognizing problems with the intake.

Day 5

20. **Administering the Screener, Part 2**. Discussion of the more complex procedures and specifications for the screener, including the use of neighbor information. Introduction to different types (result codes) of completed screeners.

21. **The Household Questionnaire, Part 2.** Scheduling considerations for conducting the household questionnaire, answering respondent questions, and determining specifications for the questionnaire.

22. **Administering the DHKS**. Selecting the respondent and scheduling the interview, answering respondent questions, telephone procedures, and determining specifications for the questionnaire.

23. **Putting It All Together**. Practice doing interviews in simulated real-life setting. Working in small groups, trainees will go in front of the room to play role of interviewer (using all materials, such as measuring guides).

Day 6

24. **Recording Contact Attempts, Contact Problems and Procedures, and Avoiding Refusals**. A variety of exercises focusing on contact problems and procedures using the household folder and field management system. These exercises will build in complexity and focus on bringing out various procedures (both documentation of situation and handling the respondent). Will also include backup procedures and using e-mail.

25. **Transmitting Field Management System Data to Your Supervisor and Role Plays**. During role plays, trainees will be pulled to learn procedures for transmitting data, including a practice. When not in this session, trainees will complete a scripted role play that will include each questionnaire type. Trainees will be required to edit these questionnaires and turn them in at the end of the day.

Day 7

26. **Market Checks**. Discussion of procedures for conducting a market check and completing the form.

27. **Data Uses**. ARS presentation to further explain use of the data.

28. **Final Review of the Questionnaire - Review of Paid Respondent and Role Play Experiences**. Answer questions that still need clarification and review experience from paid respondent and role play interviews. Answer any outstanding questions from the floor.

29. **Quality Control and Administrative Procedures**. Discussion of all procedures, including receiving feedback from the supervisors on data quality (using previous presentation material as an example).

30. **Explanation of Practice Interviews**. Description of procedures for completing additional practice interviews and of the process for supervisor evaluation of them.

Criteria for Acceptable Questionnaires

Minimum criteria are defined for acceptability of completed questionnaires for determining response rates and for inclusion in the CSFII data set for analysis. These criteria do not address every facet of document review nor do they exempt the need for each appropriate question to be answered. All collected data will be processed; however, only acceptable data will be considered for analysis. The criteria for each of the survey questionnaires are listed below. Question numbers and column labels in parenthesis refer to the question or column on the CSFII 1991 survey questionnaires. The question number or column label may be different for CSFII 1994–96.

For acceptable screener information

1. Results of screening attempts must be documented. Date and time of day for each attempt must be recorded on the questionnaire.

2. Sample descriptive data must be recorded--segment identification and housing unit identifiers.

For participating households

3. Income screener information (if necessary) must be recorded.

4. Information on household composition (if necessary) must be recorded.

For an acceptable household questionnaire

1. Sample descriptive data must be recorded--segment identification and housing unit identifiers.

2. Date of interview and day of week of interview must be recorded on the document.

3. Household composition grid (columns A, B, C, and D) must be completed. These columns represent the first name or other suitable designation of the individual, the relationship to head of household, the age (or date of birth), and the sex of each household member.

4. The question relating to the number of people in the household (question 8) must be answered.

For an acceptable individual intake questionnaire

1. Sample descriptive data must be recorded--segment identification and housing unit identifiers.

2. First name or other suitable designation of the individual and person identification number must be recorded on the intake document.

3. Age in years or date of birth must be recorded.

4. Date of intake and day of week of interview must be recorded on the document.

5. No missing meals are indicated.

6. A food name (question 4) is reported, but a food description (question 5) is missing for 25 percent or fewer foods each day.

7. A food name (question 4) and a food description (question 5) are reported, but a food amount (question 6) is missing for 15 percent or fewer foods each day.

For an acceptable Diet Health Knowledge Survey questionnaire

1. Sample descriptive data must be recorded--segment identification and housing unit identifiers.

2. First name of individual or other suitable designation must be recorded on the questionnaire.

3. Results of contact attempts must be documented. Date and time of day for each attempt, including the interview date and time, must be documented on the questionnaire.

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

Survey Net

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

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

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

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

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

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

Food Coding and Editing Process

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

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

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

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

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

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

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

Technical Support Files

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

---

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

Food Coding Database

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

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

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

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

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

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

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

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

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

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

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

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

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

Recipe Database

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

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

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

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

Survey Nutrient Database

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

Chapter 8: Probabilities of Selection and Calculation of Sampling Weights
By Adam Chu, Alvin B. Nowverl, and Joseph D. Goldman

CSFII/DHKS Weighting Design

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 example, undercoverage of certain population groups). For the CSFII/DHKS 1994–96, the overall probabilities of selection were designed to vary by sex, age, and income level to meet precision goals specified by ARS (see chapter 3).

To compensate for nonresponse and noncoverage, the data were weighted in the following steps: (1) a base weight equal to the reciprocal of the probability of selection was assigned to each sample person (SP); (2) the base weights were then adjusted for nonresponse within weighting classes, defined by variables that were determined to be correlated with response rates; (3) the nonresponse-adjusted weights were ratio adjusted to population estimates from the U.S. Bureau of the Census, Current Population Survey, to compensate for random variation in the observed sample counts and possible undercoverage of certain groups in the area sample frame (U.S. Department of Commerce 1994, 1995, 1996).

The following four sets of weights were calculated: (1) a set for SP's who completed the day-1 interview, (2) a set for SP's who provided 2 days of intake, (3) a set for all DHKS respondents, and (4) a set for DHKS respondents who provided 2 days of intake. Westat, Incorporated developed the overall weighting design and performed the calculations needed for the day-1 set of weights. Using their design and base weights, ARS made all the necessary computations for the 2-day sets of weights. In general, sample weights are required for analysis of the survey data.

Base weights

The base weight associated with an SP was the reciprocal of the overall probability of including that person in the survey. The base weights inflated the sample to the population if there was no nonresponse or noncoverage in the survey. For the CSFII/DHKS, nonresponse could occur at different stages of data collection, for example, (1) before the enumeration of SP's in the household, (2) after household enumeration and the selection of SP's but before the completion of the day-1 intake interview, (3) after the day-1 intake interview but before the completion of day-2 intake interview, and (4) after the day-1 intake interview but before

completion of the DHKS interview. In addition, nonresponse to the household questionnaire could occur after the selection of SP's in eligible households. Noncoverage arose when some members of the survey population had no chance of being selected. With the CSFII/DHKS, noncoverage could occur from incomplete listings of dwelling units (DU's) in selected segments or incomplete listings of persons within DU's.

For the CSFII/DHKS, SP's were selected through a complex, multistage sample design involving the selection of primary sampling units (PSU's), area segments within PSU's, DU's within segments, and finally SP's within households (see chapter 3). Some of the sampled segments were so large that an additional stage of sampling was introduced to reduce the amount of listing required. In general, these segments were divided into two or more chunks of approximately equal size. One chunk was selected with probability proportional to size. Of the 744 segments selected for the CSFII/DHKS, 54 were divided into chunks.

Within each segment, there existed the possibility of finding DU's that were missed in the listing process, and Westat's missed-structure and missed-DU procedure was used for this purpose. In segments where there were more than 10 missed DU's found, a random subsample of 10 of the DU's was retained in the study. For this reason, a factor was added to calculation of the base weight to compensate for these DU's.

To calculate the overall probabilities of selection, the following components were required:

1. the probability of selecting the PSU,

2. the probability of selecting the segment (or chunk) within the PSU,

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

4. the probability that a dwelling unit identified through the missed structure procedure was retained for the sample, and

5. the probability of selecting an eligible SP from within the household.

For most SP's, the product of these five factors was the probability of being selected for the CSFII. These factors applied to SP's 1 year of age or older. For infants under 1 year of age, the weights were derived differently. Because infants were included in the CSFII/DHKS sample whenever another eligible SP 1 year old

or older was selected from the household, the probability of selecting an infant was the same as the probability of retaining that household for the CSFII/DHKS. Calculating the probability of selection for a DHKS respondent required a sixth factor, the probability of selecting an eligible CSFII respondent for the DHKS.

Imputing classification variables required for weighting

The assignment of base weights required that sex, age, and income status be known for all SP's. For a small number of cases (all nonrespondents), one or more of the required variables were not available in the screener questionnaire. For these cases, values of the missing variables were imputed by the methods described below.

In 1994, there were 6 cases out of 6,868 persons sampled where sex was missing.[1] For each of these SP's, a value was imputed by generating a uniform random number between 0 and 1 and setting the sex equal to male if the random number was less than 0.5. Otherwise, the sex was set to female.

For the 83 cases where age was missing, a broad age range (for example, "under 18" or "over 65") was often available in the abbreviated enumeration table ("neighbor information") of the screener questionnaire. This information was used to impute age by the following "hot-deck" procedure. First, a listing of the 6,868 SP's, sorted by message number and randomly within message number, was prepared. In the sorted listing, each case that had a missing value for age was located. For each of these cases, the next listed case was also identified that was assigned the same message number and that satisfied the same age range recorded in the abbreviated enumeration table. The age recorded for the latter cases (referred to as "donor records") was then assigned to the corresponding record with the missing age.

Income level was imputed for 12 SP's. To be consistent with the procedures developed for selecting SP's when income information was not obtained during screening (see chapter 3), the imputation was accomplished by inspecting the information recorded by the interviewer in the household folder to determine if the household included children under 6 years of age and if the household had no males over 18 years. If the information indicated this to be the case, the household and its members were coded as low income for weighting purposes. Otherwise, the household was imputed as non-low income for weighting purposes.

Development of nonresponse adjustments

Unit nonresponse (that is, whole questionnaire nonresponse) occurred when an eligible SP failed to respond to the survey for any reason. Separate adjustments were made to compensate for nonresponse in the screener interview, the day-1 intake interview, and the DHKS interview. For a given period of adjustment, the general approach was to divide the sample into a number of homogenous weighting classes, where nonresponse-adjusted weights were calculated by multiplying the base weights by the corresponding inverse of the weighted response rate for the class.

To adjust the SP base weights, the initial base weights were calculated by adjusting the reciprocal probabilities of selection for screening nonresponse. These adjustments were made within classes created by grouping segments by census region, MSA status, minority status of the segment (percent of the population that was black or Hispanic), and quarter of field operations. Within each class, the base weight of each eligible SP was increased by a factor corresponding to the screener nonresponse rate within the class.

These initial base 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 field operations, and minority status of the segment. The result of this step was a set of nonresponse-adjusted base weights for responding SP's.

Nonresponse adjustments for the DHKS weights involved a modification to the CSFII nonresponse-adjusted weights which accounted for the fact that only day-1-responding adult SP's who responded without the assistance of a proxy could be interviewed for the DHKS. If there were two or more such respondents, one was randomly selected for the DHKS. Once that adjustment was made, the resulting modified initial weights were adjusted for DHKS nonresponse within classes defined by age, household size, annual income as a percentage of poverty level, race, and whether the SP's were on a special diet, usually shopped for food, or usually prepared meals. Such information was available because SP's selected for the DHKS had already participated in the CSFII.

Population adjustments

In addition to compensating for unequal selection probabilities and nonresponse, another important function of weighting is to adjust for sampling variability and possible undercoverage in the sampling frame. Therefore, the final step in the

weighting process was to calibrate the nonresponse-adjusted weights so the sum of the final weights equaled the corresponding March Current Population Survey totals within cells defined by the following variables (U.S. Department of Commerce 1994, 1995, 1996):

1. sex,
2. age group (seven categories based on intake interview),
3. home ownership (owned versus not owned),
4. season of intake (winter, spring, summer, fall),
5. day of week of day-1 intake,
6. day of week of day-2 intake (2-day weights only),
7. census region (four regions),
8. MSA status (metropolitan versus nonmetropolitan),
9. household income as a percentage of poverty level (1994 poverty thresholds),
10. household received food stamps in the past 12 months,
11. presence in household of persons 18 and older,
12. presence in household of children under 6 years,
13. presence in household of children 6 to 17 years,
14. presence of female head of household 40 years or younger and absence of anyone under 18 years,
15. employment status (for children this was the status of the female head, or if there was no female head, the male head of household),
16. race (black versus nonblack), and
17. ethnic origin (Hispanic versus non-Hispanic).

The calibration was implemented by an iterative process known as "raking ratio weighting." This process was carried out separately for each of the following four subsets: (1) males 20 years and older, (2) females 20 years and older, (3) children 5 years and younger, and (4) persons 6 to 19 years. The variables differed slightly for the four subsets as appropriate. The same procedure and sets of variables were used for both the CSFII and the DHKS weights. Before the process was carried out for the 2-day CSFII and DHKS subsets, the nonresponse-adjusted base weights described earlier were ratio adjusted within weighting classes to reflect nonresponse between day 1 and day 2. The day of the week that the day-2 intake was collected was added to the calibrating process. Tables 12 through 15 show, by weighting variable, the CSFII 1994 unweighted sample sizes, the weighted percentage distributions following nonresponse adjustments (but before calibration to population targets), and the population targets for day-1 and 2-day respondents. Tables 16 and 17 provide the same information for the DHKS 1994.

Summary of Final Weights

Table 18 summarizes the four sets of final weights for CSFII/DHKS 1994. The table shows the sample size, the sum of the weights, the coefficient of variation of the weights (*CV*), and $1 + (CV/100)^2$. 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 was anticipated that the variance of a day-1 estimate would be 1.43 times what it would have been had all the weights been equal. The final CSFII/DHKS 1994–96 data set contains annual sampling weights for each of the 3 years of the survey and a set of 3-year sampling weights for use in analyses using 3-year combined data.

Variance Estimation Fields

As described in chapter 3, Westat's 62 PSU master sample was employed for CSFII/DHKS 1994–96. This sample of PSU's contains 24 PSU's selected with certainty. The remaining 38 PSU's were selected with probability proportional to size from 38 strata, or 1 PSU per stratum. Thirty-six area segments (12 for each of the 3 years of the CSFII/DHKS) were then selected from each of the 62 PSU's also with probability proportional to size.

---

1. In general, data provided are for 1994--the first year of the 3-year survey. Data for 1995 and 1996 are similar.

A framework was created of two sampling units per stratum to facilitate variance estimation procedures. First, 19 variance estimation strata were formed from the 38 noncertainty PSU's by pairing adjacent PSU's in the sampling frame. Each PSU within a variance estimation stratum defined what was referred to as a variance estimation unit. Next, within each of the 24 certainty PSU's, one-half of the segments were assigned to one variance estimation unit and the remaining one-half to another. Because each certainty PSU was considered to be a separate variance estimation stratum, a total of 43 variance estimation strata (each containing two variance estimation units) was formed by this process.

Replicate Weights

The final sampling weights, along with the variance estimation strata and estimation units described in "Variance Estimation Fields," may be used with a Taylor series linearization method to estimate sampling errors. Software packages such as SUDAAN (Shah et al. 1993) and PC CARP (Fuller et al. 1989) can be used to obtain estimates using the linearization method.

Sampling errors could also be estimated using the jackknife technique. Separate sets of jackknife replicate weights were constructed for the day-1 and 2-day CSFII/DHKS and 2-day DHKS (based on respondents who provided 2 days of intake). A prescribed number of subsamples called "jackknife replicates" were generated from the full sample.

The construction of jackknife replicates made use of the variance estimation stratum/variance estimation unit structure described above. A jackknife replicate was created by eliminating one of the two variance estimation units from a variance estimation stratum and doubling the initial base weights of the individuals in the other variance estimation unit. The person nonresponse adjustment was made to adjust for the loss of those persons in the variance estimation unit that was dropped. The calibration process was then repeated. When an individual was not in a replicate subsample, he or she was assigned a corresponding replicate weight of zero. In this way, a series of replicate weights was generated for each SP. Together with the final, full-sample weights, these replicates were designed for the calculation of sampling errors. Forty-three replicates were created by applying this process to each of the 43 variance estimation strata. Each replicate produces its own set of replicate weights for the entire sample.

One advantage of using such a replication method to calculate sampling errors of survey-based estimates is that it precludes the need for complicated variance estimation formulas like those produced by linearization (McCarthy 1966). It

should also be noted that the jackknife replicates were designed to reflect not only the stratification and clustering used in the CSFII/DHKS sample design, but also the weighting adjustment process. No publicly available software using the linearization method captures the effects of the raking ratio adjustment as well as jackknife replication.

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. The estimated variance of y is then given by the formula:

$$\text{Var}(y) = \sum (y(j) - y)^2$$

where the summation extends over all 43 sets of jackknife replicate weights.

This replicate weighting process has been designed and implemented by Westat Incorporated. Westat 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 (Westat 1996). WesVarPC computes weighted survey estimates and their associated sampling errors and provides confidence intervals for sample-based estimates. WesVarPC also performs modified chi-square tests of independence in weighted two-way tables using estimated design effects. These modified chi-square methods include the method proposed by Fellegi (1980) and two methods described by Rao and Scott (1981, 1984).

Factors Influencing Statistical Inference

Several factors affect the ability to make valid inferences from the data for the CSFII/DHKS. Two of these factors, survey weights and variance estimation, were alluded to in previous sections of this chapter. The third factor, the population of inference, is described next.

In general, the population of inference for any given study year of the CSFII/DHKS consisted of noninstitutionalized persons residing in the United States. Excluded were institutionalized individuals, such as those in prisons, juvenile facilities, and nursing homes. Also excluded were persons living in group quarters (for example, rooming houses), persons residing on military installations, and homeless persons.

Although the above definition is conceptually straightforward, special rules were required in a few unusual circumstances to establish eligibility for weighting and analytic purposes. These rules included the following:

- SP's who became institutionalized, died, or moved out of the country before completion of a particular interview were considered to be ineligible for that interview.

- SP's identified during screening who moved within the United States before the first intake interview and who could not be successfully traced and interviewed were considered to be eligible nonrespondents.

- An attempt was made to contact by telephone SP's who moved after completing the day-1 intake interview. If the day-2 intake interview was not obtained, the SP was considered to be an eligible nonrespondent for the combined day-1 and day-2 intake interviews, but a respondent for the day-1 intake interview only.

- SP's selected for the DHKS who later died, became institutionalized, or moved out of the country before completing the survey were treated as ineligible for the DHKS.

References

Fellegi, I. 1980. Approximate test of independence and goodness of fit based on stratified multistage samples. Journal of the American Statistical Association 71:665–670.

Fuller, W. P., W. Kennedy, D. Schnell, et al. 1989. PC CARP. Statistical Laboratory, Iowa State University, Ames. Photocopy.

McCarthy, P. 1966. Replication: An approach to the analysis of data from complex surveys. U.S. Department of Health, Education, and Welfare, Vital and Health Statistics, Series 2, No. 14.

Rao, J., and A. Scott. 1981. The analysis of categorical data from complex sample surveys: Chi-squared test for goodness of fit and independence in two way tables. Journal of the American Statistical Association 76:221–230.

Rao, J., and A. Scott. 1984. On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. Annals of Statistics 12:46–60.

Shah, B., R. Folsom, L. LaVange, et al. 1993. Statistical methods and mathematical algorithms used in SUDAAN. Research Triangle Institute, Research Triangle Park, NC.

U.S. Department of Commerce, Bureau of the Census. 1994. Current population survey, March 1994. U.S. Department of Commerce. Machine readable data file.

U.S. Department of Commerce, Bureau of the Census. 1995. Current population survey, March 1995. U.S. Department of Commerce. Machine readable data file.

U.S. Department of Commerce, Bureau of the Census. 1996. Current population survey, March 1996. U.S. Department of Commerce. Machine readable data file.

Westat, Inc. 1996. A user's guide to WesVarPC®. Westat, Inc., Rockville, MD.

Table 12. Males 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Home ownership/age					
Home owned					
20–39	305	286	27.6	27.6	26.4
40–59	434	417	26.8	27.2	25.9
60 and older	429	399	17.9	17.9	16.9
Home not owned					
20–39	274	253	19.0	18.8	20.3
40–59	136	126	6.2	6.0	7.5
60 and older	71	66	2.5	2.5	3.0
Season of intake					
Winter	373	351	21.2	21.3	25.0
Spring	420	391	25.7	25.4	25.0
Summer	438	409	28.0	28.0	25.0
Fall	418	396	25.1	25.3	25.0
Day of week of day-1 intake					
Sunday	284	268	17.1	17.1	14.3
Monday	269	253	15.3	15.3	14.3
Tuesday	234	218	14.3	14.1	14.3
Wednesday	194	176	11.8	11.5	14.3
Thursday	165	157	10.5	10.8	14.3
Friday	292	274	17.8	17.8	14.3
Saturday	211	201	13.3	13.4	14.3
Day of week of day-2 intake					
Sunday	---	249	---	15.8	14.3
Monday	---	292	---	19.1	14.3
Tuesday	---	288	---	18.8	14.3
Wednesday	---	224	---	14.0	14.3
Thursday	---	222	---	14.2	14.3
Friday	---	164	---	10.8	14.3
Saturday	---	108	---	7.2	14.3

Table 12.—Males 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Household income as a percentage of poverty level					
0–75%	156	143	5.6	5.3	5.0
76–130%	228	215	8.6	7.8	7.9
131–300%	540	514	33.1	34.4	31.0
Over 300%	725	675	52.7	52.5	56.1
Household received food stamps in past 12 months					
Yes	139	133	5.3	5.3	6.4
No	1,510	1,414	94.7	94.7	93.6
Presence of persons 18 and older in household					
Exactly 1	203	191	11.4	11.4	11.5
Exactly 2	1,034	974	63.5	63.7	59.7
More than 2	412	382	25.2	24.9	28.8
Presence of child under 6 and 6–17 in household					
Child under 6					
Child 6–17	140	135	9.8	10.1	9.0
No child 6–17	128	121	9.3	9.4	9.9
No child under 6					
Child 6–17	323	303	21.9	22.0	20.1
No child 6-17	1,058	988	59.0	58.6	61.1
Presence of female head of household 40 or younger and no one under 18					
Yes	104	98	8.3	8.4	7.3
No	1,545	1,449	91.7	91.6	92.7

Table 12.—Males 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Employment status					
Have job	1,069	1,007	74.7	75.1	69.8
Do not have job	580	540	25.3	24.9	30.2
Census region					
Northeast	305	281	18.9	18.6	20.0
Midwest	390	378	22.8	23.5	23.3
South	579	546	36.3	36.4	34.5
West	375	342	22.0	21.5	22.2
MSA status					
MSA (metropolitan)	1,195	1,115	77.3	77.4	78.8
Non-MSA	454	432	22.7	22.6	21.2
Race					
Black	172	164	10.7	10.9	10.5
Nonblack	1,477	1,383	89.3	89.1	89.5
Ethnic origin					
Hispanic	152	139	8.3	8.1	9.3
Non-Hispanic	1,497	1,408	91.7	91.9	90.7
TOTAL	1,649	1,547	100.0	100.0	100.0

* Calculated using 1994 Current Population Survey data, 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 13. Females 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Home ownership/age					
Home owned					
20–39	275	263	24.1	24.3	23.9
40–59	448	426	26.4	26.4	24.9
60 and older	392	364	19.4	19.5	19.3
Home not owned					
20–39	291	271	19.9	19.7	19.6
40–59	143	133	5.8	5.8	7.2
60 and older	93	84	4.4	4.3	5.1
Season of intake					
Winter	395	368	21.2	23.0	25.0
Spring	399	371	25.7	24.1	25.0
Summer	433	414	28.0	28.1	25.0
Fall	415	388	25.1	24.9	25.0
Day of week of day-1 intake					
Sunday	301	283	17.1	17.9	14.3
Monday	266	243	15.3	15.6	14.3
Tuesday	238	218	14.3	13.9	14.3
Wednesday	193	182	11.8	11.6	14.3
Thursday	163	156	10.5	10.9	14.3
Friday	280	269	17.8	17.7	14.3
Saturday	201	190	13.3	12.4	14.3
Day of week of day-2 intake					
Sunday	---	295	---	19.4	14.3
Monday	---	289	---	18.8	14.3
Tuesday	---	241	---	16.1	14.3
Wednesday	---	244	---	15.1	14.3
Thursday	---	205	---	13.1	14.3
Friday	---	164	---	11.1	14.3
Saturday	---	103	---	6.5	14.3

Table 13. —Females 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Household income as a percentage of poverty level					
0–75%	197	186	8.4	8.3	8.6
76–130%	217	200	8.9	8.8	10.8
131–300%	549	519	34.1	35.6	31.8
Over 300%	679	636	47.6	47.4	48.8
Household received food stamps in past 12 months					
Yes	185	171	8.3	8.1	10.4
No	1,457	1,370	91.7	91.9	89.6
Presence of persons 18 and older in household					
Exactly 1	368	338	18.9	18.6	20.8
Exactly 2	923	878	58.8	59.4	56.0
More than 2	351	325	22.3	22.0	23.2
Presence of child under 6 and 6–17 in household					
Child under 6					
Child 6–17	156	151	10.9	11.3	10.1
No child 6–17	141	130	9.7	9.5	11.0
No child under 6					
Child 6–17	338	318	22.8	22.7	21.7
No child 6-17	1,007	942	56.6	56.5	57.3
Presence of female head of household 40 or younger and no one under 18					
Yes	166	155	13.1	12.8	10.7
No	1,476	1,386	86.9	87.2	89.3

Table 13. —Females 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Employment status					
Have job	838	799	57.7	58.1	55.5
Do not have job	804	742	42.3	41.9	44.5
Census region					
Northeast	306	276	18.6	18.3	20.7
Midwest	422	405	24.8	25.0	23.5
South	581	551	36.6	37.1	34.9
West	333	309	20.0	19.7	20.9
MSA status					
MSA (metropolitan)	1,215	1,131	77.0	76.7	78.7
Non-MSA	427	410	23.0	23.3	21.3
Race					
Black	214	201	13.8	13.9	12.0
Nonblack	1,428	1,340	86.2	86.1	88.0
Ethnic origin					
Hispanic	150	139	8.5	8.4	8.4
Non-Hispanic	1,492	1,402	91.5	91.6	91.6
TOTAL	1,642	1,541	100.0	100.0	100.0

* Calculated using 1994 Current Population Survey data, 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 14. Children 5 years and younger: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Age/Sex					
Male					
0–2	311	299	22.1	21.7	25.5
3–5	299	293	27.7	27.9	25.6
Female					
0–2	314	300	22.5	22.3	24.4
3–5	303	296	27.7	28.1	24.5
Home ownership					
Home owned	687	670	57.0	56.9	53.6
Home not owned	540	518	43.0	43.1	46.4
Season of intake					
Winter	309	299	21.2	25.1	25.0
Spring	298	287	25.7	23.5	25.0
Summer	317	309	28.0	26.0	25.0
Fall	303	293	25.1	25.4	25.0
Day of week of day-1 intake					
Sunday	205	193	17.1	16.1	14.3
Monday	179	170	15.3	14.7	14.3
Tuesday	211	207	14.3	17.3	14.3
Wednesday	147	144	11.8	12.2	14.3
Thursday	139	137	10.5	11.6	14.3
Friday	196	193	17.8	16.1	14.3
Saturday	150	144	13.3	12.0	14.3
Day of week of day-2 intake					
Sunday	---	201	---	18.0	14.3
Monday	---	247	---	21.0	14.3
Tuesday	---	194	---	16.0	14.3
Wednesday	---	177	---	14.6	14.3
Thursday	---	194	---	16.1	14.3
Friday	---	95	---	7.9	14.3
Saturday	---	80	---	6.4	14.3

Table 14.—Children 5 years and younger: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Household income as a percentage of poverty level					
0–75%	255	245	19.8	19.9	18.5
76–130%	175	164	13.8	13.0	13.9
131–300%	435	419	35.8	36.8	34.8
Over 300%	362	360	30.7	30.4	32.9
Household received food stamps in past 12 months					
Yes	334	314	25.4	25.1	25.0
No	893	874	74.6	74.9	75.0
Presence of persons 18 and older in household					
Exactly 1	183	173	14.4	14.3	14.7
Exactly 2	877	853	71.6	71.6	71.7
More than 2	167	162	14.0	14.1	13.6
Presence of child 6–17 in household					
Child 6–17	546	536	45.7	46.3	44.8
No child 6–17	681	652	54.3	53.7	55.2
Employment status of female head of household (or male head if there is no female head)					
Have job	674	651	55.4	55.0	53.5
Do not have job	553	537	44.6	45.0	46.5

Table 14.—Children 5 years and younger: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Census region					
Northeast	208	205	17.2	17.4	18.8
Midwest	298	287	24.3	24.2	23.8
South	403	389	32.5	32.3	33.6
West	318	307	26.0	26.1	23.9
MSA status					
MSA (metropolitan)	942	913	77.7	77.8	81.2
Non-MSA	285	275	22.3	22.2	18.8
Race					
Black	179	165	14.4	14.0	16.5
Nonblack	1,048	1,023	85.6	86.1	83.5
Ethnic origin					
Hispanic	197	189	16.3	16.3	15.2
Non-Hispanic	1,030	999	83.7	83.7	84.8
TOTAL	1,227	1,188	100.0	100.0	100.0

* Calculated using 1994 Current Population Survey data, 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 15. Persons 6 to 19 years: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Age/Sex					
Male					
6–11	254	252	21.3	21.5	22.7
12–19	286	268	30.4	30.1	28.3
Female					
3–5	271	261	27.6	27.8	27.3
6–11	260	254	20.6	20.5	21.6
Home ownership					
Home owned	722	704	71.2	71.6	66.2
Home not owned	349	331	28.8	28.4	33.8
Season of intake					
Winter	292	278	21.2	26.4	25.0
Spring	251	244	25.7	22.5	25.0
Summer	281	271	28.0	28.3	25.0
Fall	247	242	25.1	22.8	25.0
Day of week of day-1 intake					
Sunday	160	153	17.1	14.4	14.3
Monday	175	162	15.3	15.2	14.3
Tuesday	166	163	14.3	15.5	14.3
Wednesday	161	159	11.8	14.5	14.3
Thursday	114	109	10.5	11.8	14.3
Friday	171	167	17.8	16.0	14.3
Saturday	124	122	13.3	12.8	14.3
Day of week of day-2 intake					
Sunday	---	209	---	20.4	14.3
Monday	---	211	---	21.2	14.3
Tuesday	---	163	---	15.6	14.3
Wednesday	---	119	---	11.6	14.3
Thursday	---	133	---	12.9	14.3
Friday	---	122	---	11.4	14.3
Saturday	---	78	---	6.9	14.3

Table 15.—Persons 6 to 19 years: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Household income as a percentage of poverty level					
0–75%	173	164	12.8	12.8	14.1
76–130%	124	119	10.0	9.4	12.0
131–300%	407	399	38.9	39.6	35.0
Over 300%	367	353	38.3	38.2	38.9
Household received food stamps in past 12 months					
Yes	184	171	14.1	13.6	17.3
No	887	864	85.9	86.4	82.7
Presence of persons 18 and older in household					
Exactly 1	139	135	11.2	11.4	15.5
Exactly 2	663	649	60.5	61.1	58.7
More than 2	269	251	28.3	27.6	25.8
Presence of child under 6 and 6–17 in household					
Child under 6	303	289	26.3	26.1	27.1
No child under 6	768	746	73.7	73.6	72.9
Child 6–17	60	58	6.1	6.2	6.9
No child 6–17	1,011	977	93.9	93.8	93.1
Employment status of female head of household (or male if there is no female head)					
Have job	703	689	67.4	68.4	64.3
Do not have job	368	346	32.6	31.6	35.7

Table 15.—Persons 6 to 19 years: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, CSFII 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Census region					
Northeast	150	145	14.0	14.1	18.5
Midwest	285	279	25.6	25.9	24.1
South	361	348	35.6	35.8	34.8
West	275	263	24.8	24.3	22.6
MSA status					
MSA (metropolitan)	782	751	75.5	75.2	77.4
Non-MSA	289	284	24.5	24.8	22.6
Race					
Black	142	137	14.7	14.9	15.8
Nonblack	929	898	85.3	85.1	84.2
Ethnic origin					
Hispanic	159	152	13.5	13.5	12. 9
Non-Hispanic	912	883	86.5	86.5	87.1
TOTAL	1,071	1,035	100.0	100.0	100.0

* Calculated using 1994 Current Population Survey data, 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 16. Males 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, DHKS 1994

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Home ownership/age					
Home owned					
20–39	159	156	27.4	27.3	26.4
40–59	236	236	26.8	27.1	25.9
60 and older	224	222	17.0	17.0	16.9
Home not owned					
20–39	145	140	20.0	19.7	20.3
40–59	91	89	6.2	6.2	7.5
60 and older	47	46	2.6	2.6	3.0
Season of intake					
Winter	194	194	21.1	21.4	25.0
Spring	235	231	27.5	27.4	25.0
Summer	246	242	28.0	27.9	25.0
Fall	227	222	23.4	23.3	25.0
Day of week of day-1 intake					
Sunday	149	147	16.7	16.7	14.3
Monday	139	138	14.4	14.5	14.3
Tuesday	135	132	14.9	14.7	14.3
Wednesday	113	109	13.1	12.9	14.3
Thursday	86	85	10.2	10.2	14.3
Friday	158	157	17.1	17.2	14.3
Saturday	122	121	13.6	13.8	14.3
Day of week of day-2 intake					
Sunday	---	138	---	15.4	14.3
Monday	---	174	---	21.6	14.3
Tuesday	---	165	---	17.6	14.3
Wednesday	---	137	---	14.7	14.3
Thursday	---	116	---	12.9	14.3
Friday	---	95	---	10.8	14.3
Saturday	---	64	---	7.0	14.3

Table 16.—Males 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, DHKS 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Household income as a percentage of poverty level					
0–75%	84	81	5.9	5.7	5.0
76–130%	137	134	8.1	8.0	7.9
131–300%	288	287	32.7	33.1	31.0
Over 300%	393	387	53.3	53.2	56.1
Household received food stamps in past 12 months					
Yes	87	86	5.8	5.9	6.4
No	815	803	94.2	94.1	93.6
Presence of persons 18 and older in household					
Exactly 1	179	178	12.0	12.2	11.5
Exactly 2	565	557	65.9	65.9	59.7
More than 2	158	154	22.1	21.9	28.8
Presence of child under 6 and 6–17 in household					
Child under 6					
Child 6–17	70	70	8.9	9.0	9.0
No child 6–17	75	73	10.4	10.3	9.9
No child under 6					
Child 6–17	168	167	21.3	21.5	20.1
No child 6–17	589	579	59.4	59.2	61.1
Presence of female head of household 40 or younger and no one under 18					
Yes	54	52	9.4	9.2	7.3
No	848	837	90.6	90.8	92.7

Table 16.—Males 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, DHKS 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Employment status					
Have job	588	579	75.4	75.3	69.8
Do not have job	314	310	24.6	24.7	30.2
Census region					
Northeast	162	158	16.7	16.5	20.0
Midwest	218	216	23.2	23.4	23.3
South	317	314	36.8	37.1	34.5
West	205	201	23.2	23.1	22.2
MSA status					
MSA (metropolitan)	639	628	76.9	76.7	78.8
Non-MSA	263	261	23.1	23.3	21.2
Race					
Black	102	100	12.5	12.2	10.5
Nonblack	800	789	87.5	87.8	89.5
Ethnic origin					
Hispanic	80	78	8.1	8.1	9.3
Non-Hispanic	822	811	91.9	91.9	90.7
TOTAL	902	889	100.0	100.0	100.0

* Calculated using 1994 Current Population Survey data, 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 17. Females 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, DHKS 1994

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Home ownership/age					
Home owned					
20–39	138	136	23.2	23.4	23.9
40–59	261	254	26.9	27.1	24.9
60 and older	237	229	19.6	19.5	19.3
Home not owned					
20–39	180	172	21.1	20.8	19.6
40–59	95	92	5.3	5.3	7.2
60 and older	66	64	4.0	4.0	5.1
Season of intake					
Winter	239	234	23.4	23.5	25.0
Spring	231	221	22.6	22.4	25.0
Summer	255	252	27.3	27.8	25.0
Fall	252	240	26.6	26.3	25.0
Day of week of day-1 intake					
Sunday	181	175	17.8	17.6	14.3
Monday	155	150	15.2	15.2	14.3
Tuesday	143	137	13.3	13.3	14.3
Wednesday	112	108	11.3	11.4	14.3
Thursday	96	94	11.5	11.6	14.3
Friday	170	166	18.8	17.8	14.3
Saturday	120	117	12.0	12.1	14.3
Day of week of day-2 intake					
Sunday	---	192	---	21.5	14.3
Monday	---	164	---	16.9	14.3
Tuesday	---	158	---	17.2	14.3
Wednesday	---	148	---	13.8	14.3
Thursday	---	122	---	13.7	14.3
Friday	---	93	---	10.2	14.3
Saturday	---	70	---	6.8	14.3

Table 17.—Females 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, DHKS 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Household income as a percentage of poverty level					
0–75%	144	140	8.6	8.4	8.6
76–130%	142	138	8.7	8.7	10.8
131–300%	318	310	35.2	35.4	31.8
Over 300%	373	359	47.6	47.5	48.8
Household received food stamps in past 12 months					
Yes	129	127	8.6	8.7	10.4
No	848	820	91.4	91.3	89.6
Presence of persons 18 and older in household					
Exactly 1	336	324	18.9	18.7	20.8
Exactly 2	491	476	59.1	59.0	56.0
More than 2	150	147	22.0	22.3	23.2
Presence of child under 6 and 6–17 in household					
Child under 6					
Child 6–17	86	85	10.5	10.7	10.1
No child 6–17	73	71	9.2	9.1	11.0
No child under 6					
Child 6–17	199	196	23.5	23.8	21.7
No child 6-17	619	595	56.9	56.5	57.3
Presence of female head of household 40 or younger and no one under 18					
Yes	96	91	13.3	13.0	10.7
No	881	856	86.7	87.0	89.3

Table 17.—Females 20 years and older: Unweighted sample sizes, weighted percentage distributions following nonresponse adjustments for day-1 and 2-day totals, and population targets, DHKS 1994—Continued

Variable	Sample size		Nonresponse adjustment		Population targets*
	Day 1	2-day	Day 1	2-day	
	----number----		----------percent----------		
Employment status					
Have job	493	479	57.5	57.6	55.5
Do not have job	484	468	42.5	42.4	44.5
Census region					
Northeast	201	192	20.8	20.8	20.7
Midwest	286	280	27.5	27.7	23.5
South	328	317	33.7	33.3	34.9
West	162	158	17.9	18.3	20.9
MSA status					
MSA (metropolitan)	722	697	75.9	75.7	78.7
Non-MSA	255	250	24.1	24.3	21.3
Race					
Black	130	127	12.5	12.4	12.0
Nonblack	847	820	87.5	87.6	88.0
Ethnic origin					
Hispanic	82	81	7.9	8.0	8.4
Non-Hispanic	895	866	92.1	92.0	91.6
TOTAL	977	947	100.0	100.0	100.0

* Calculated using 1994 Current Population Survey data, 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 18. Summary of final sampling weights, CSFII/DHKS 1994

Weight	Sample size	Sum of weights	*CV* (percent)	Average design effect $1+(CV/100)^2$ (number)
	----------number----------			
Day 1	5,589	259,507,267	65.80	1.43
2-day	5,311	259,507,209	77.59	1.60
DHKS	1,879	182,865,634	85.92	1.74
2-day DHKS	1,836	182,865,638	95.24	1.91

Glossary

Assistant. A knowledgeable person who assisted in the dietary intake interview for a sample person 6 to 11 years of age.

Day–1 and day–2 intake questionnaires. Questionnaires containing a 24-hour dietary recall and nonintake questions (for example, self-reported height), which were completed with each sample person. The dietary recall questions elicited a report of all foods and beverages the sample person had consumed the day before the interview (the 24 hours from midnight to midnight).

Diet and Health Knowledge Questionnaire. This questionnaire was completed with a selected sample person 20 years of age or older who had completed a day-1 intake without the assistance of a proxy. The DHKS was to be completed 2 to 3 weeks after the day-2 interview. It included information about the sample person's attitudes, knowledge, and behaviors concerning diet and health status. Topics that were addressed included the importance of dietary guidance, awareness of diet-health relationships, food labeling, and behavior related to food safety and fat 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.

Eating occasion. Any report of eating or drinking (except plain drinking water) by a sample person.

Food Instruction Booklet. A booklet used by the interviewers in association with the dietary intake questionnaires to collect the descriptions and amounts of foods consumed by the respondents. The booklet contains a series of standardized probes specific to the various types of foods that respondents might report. It also contains recording conventions, standard abbreviations, an index for locating foods, and instructions for using the measurement aids when estimating food quantities.

Food intake. All beverages (except plain drinking water) and foods ingested. Did not include inedible parts of foods (such as bones, rinds, and seeds), uneaten portions of food, or vitamin or mineral supplements.

Food Intake Analysis System. A computer-assisted food coding and nutrient analysis system that utilizes CSFII food coding, recipe, and nutrient databases. The software is available from the University of Texas-Houston School of Public Health (see chapter 7).

Handcards. Cards that list response options. They were shown to respondents when asking sensitive questions, such as income, or questions with a long list of response categories.

Household. All persons who regularly share a dwelling unit that is defined as 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. A household includes persons temporarily absent, such as those who were in the hospital or traveling; students who live away from the 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 dwelling unit as their permanent address; domestic or other employees who usually live and sleep at the dwelling unit; boarders or roomers who usually live and sleep at the 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.

A household excludes former household members who live in institutions, nursing homes, convents, and so forth; 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 member. See Household.

Household questionnaire. The questionnaire that collected socioeconomic data and was administered in person to a knowledgeable household respondent 18 years or older, not necessarily a sample person. It included questions on household income, food shopping practices, employment status, participation in government food programs (for example, Women, Infants, and Children; school lunch; food stamp), and food sufficiency.

Household respondent. A 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.

Measuring guides. A set of tools used in conducting the intake interview to help respondents estimate the quantities of foods and beverages consumed. Included were a set of four measuring cups (1/4 cup, 1/3 cup, 1/2 cup, 1 cup), four measuring spoons (1/4 teaspoon, 1/2 teaspoon, 1 teaspoon, 1 tablespoon), eight 1/8-inch-thick sticks, a ruler, a laminated card containing illustrations of fish and chicken parts, a set of 6 concentric circles (1 inch to 6 inches in diameter), 2 perpendicular 6-inch rulers, and a 1-pint measuring cup.

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 the nucleus; defined by the Office of Management and Budget for use in the presentation of statistics by the Federal Government.

Multiple-pass approach. A 24-hour recall procedure used in the dietary intake interview designed to help elicit more complete reporting of foods by respondents. The strategy consisted of three sets of questions that were asked in different forms. The first set of questions began by asking the respondents to provide a list of foods consumed during the previous day. The second set involved the collection of detailed food descriptions, and the third set contained a check for completeness of what had been reported.

Nonrespondent. A sample person who did not complete an interview.

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

Proxy. A knowledgeable adult who completed the dietary intake interview for children under 6 years of age and other sample persons unable to report for themselves due to physical or mental limitations or 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.

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

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

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

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

West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming.

Respondent. A 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.

Sample person. A household member selected to participate in the individual intake component of the survey. Sample persons who provided at least 1 day of dietary intake were considered respondents (see Respondent).

Screener questionnaire. A questionnaire that was administered to identify eligible sample persons in the sample households.

Screening respondent. A household member 18 years of age or older who answered the screening questionnaire.

Survey Net. The computer-assisted food coding and nutrient analysis system developed by ARS and used for processing data from the CSFII 1994–96.

24-hour dietary recall. A recall of beverages (excludes plain drinking water) and foods ingested during the day preceding the interview (the 24 hours from midnight to midnight).