

# AN OVERVIEW OF THE USDA'S 1994-96 CONTINUING SURVEY OF FOOD INTAKES BY INDIVIDUALS AND THE DIET AND HEALTH KNOWLEDGE SURVEY

Joseph D. Goldman, Lori G. Borrud, USDA Agricultural Research Service, Martha Berlin, Westat, Inc.  
Joseph D. Goldman, USDA ARS, 4700 River Road, Riverdale, MD 20737

**Key Words:** Dietary intake, Response rates

## 1. Introduction

The U.S. Department of Agriculture (USDA) has been conducting nationwide food surveys since the 1930's in connection with its responsibility to ensure the health and well-being of Americans through improved nutrition. USDA recently completed its 10th national survey, the Continuing Survey of Food Intakes by Individuals (CSFII) and its follow-up Diet and Health Knowledge Survey (DHKS). The CSFII provides benchmark data on the food and nutrient intakes of the general and low-income U.S. populations. Researchers and Federal and State-level decision makers use the CSFII/DHKS to monitor the nutritional adequacy of American diets (Krebs-Smith, 1995), to measure the impact of food fortification on nutrient intakes (Crane, 1995), in developing dietary guidance and related programs (Cleveland, 1997), in estimating exposure to food contaminants, in evaluating the nutritional impact of food assistance programs (Kennedy, 1996), to determine serving size standards for use in food labeling, and in assessing the demand for agricultural products. The DHKS is the first national survey to provide information on dietary attitudes and knowledge, which can be linked to food and nutrient intakes for the same individuals in the CSFII (Guthrie, 1995). The specific focus of the DHKS is obtaining individuals' attitudes and knowledge about the Dietary Guidelines for Americans and on their ability to put the Guidelines into practice.

The CSFII/DHKS 1994-96 was conducted by Westat under contract to the Agricultural Research Service (ARS), USDA. A pilot study of the data collection methods was conducted in spring 1993 and data collection for the full survey ran from January 1994 through January 1997. In each of the 3 survey years, a nationally representative sample of individuals was interviewed so that annual estimates are possible. During the three years, 16,103 persons of all ages provided food consumption data for the CSFII and 5,765 adult CSFII participants answered the DHKS questionnaire.

The CSFII sample was designed to produce estimates for each of 40 analytic domains defined in terms of sex, age, and income level (an all-income group and a low-income group, where low-income is defined as households with annual income no greater than 130 percent of the

poverty guidelines). In order to produce domain estimates of adequate precision, sampling rates varied markedly across domains. Table 1 presents the achieved sample sizes of persons providing the first day of food intake data by the analytic domains.

**Table 1. Achieved sample sizes, 1994-96**

Age	Full sample		Low-income subset	
	Male	Female	Male	Female
1-2	725	707	244	237
3-5	734	735	238	239
6-11	751	734	215	215
12-19	734	732	221	219
20-29	779	726	233	236
30-39	890	809	202	209
40-49	861	903	194	227
50-59	888	864	204	187
60-69	846	790	202	208
70+	790	723	206	230

The CSFII/DHKS employed several different questionnaires. A screener questionnaire was administered in person to identify eligible sample persons (SP's) in sampled households. Two food intake interviews (the Day 1 and Day 2 Intakes) were conducted in person with each SP (a small percentage of Day 2 Intakes were administered over the telephone), with the Day 2 Intake being administered 3-10 days after the Day 1 Intake and on a different day of the week. The 3-10 day period was selected to diminish correlation between days while keeping the period between days short enough to facilitate field operations. The household questionnaire collected socioeconomic data and was to be administered in person to a knowledgeable member of all households containing an SP. The DHKS questionnaire was to be completed with a selected SP 20 years of age or older who had completed a Day 1 Intake. The DHKS was to be completed 2 to 3 weeks after the Day 2 Intake, typically by telephone and administered by the interviewer who had administered the CSFII questionnaires in the household.

## 2. Sample Design

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 that met specified precision levels for estimates of mean Day 1 saturated fat and iron intakes. The achieved sample sizes are listed in Table 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 into 3-year sample size targets. 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 annual portions of the sample were roughly equal in size over the 40 analytic domains and nationally representative.

A complex multistage area probability sample design was used to select persons for the intake and DHKS interviews. Excluded were persons living in group quarters or institutions, residing on military installations, and the homeless. 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 SP's within the households. The following summarizes the major features of the design:

The PSU's comprising the first-stage sampling frame were defined to be Metropolitan Statistical Areas (MSA's) or groups of contiguous counties. Because of large populations, the New York MSA was divided into three PSU's and the Chicago and Los Angeles MSA's were divided into two PSU's. Each of the remaining MSA's comprised a single PSU. Counties outside of MSA's were grouped, as necessary, to form PSU's with a minimum 1990 population of 15,000 persons, were as heterogeneous as possible, and were still small enough to permit convenient travel across the PSU by interviewers.

A total of 62 PSU's were selected. The 24 PSU's in the first-stage sampling frame 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 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 nonMSA strata.

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 three segments each, and a set of three segments per PSU was assigned to each of the 12 quarters of the 3-year survey period. By this method, interviewers were in the field in each PSU at all times of each year.

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 33,000 DU's were selected for the three years.

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 required the interviewer to recanvass the entire segment, and, in general, all DU's not previously listed were added to the sample. Segments were selected for the missed structure procedure randomly in such a way that all of the added DU's were selected with the same overall probability of selection as the rest of the sample.

To keep the interviewing workload to manageable levels within the segment, there was an upper limit of 10 missed or new DU's that would be added to the sample. When the actual numbers of missed DU's exceeded these limits, a subsample of the missed DU's was retained in the sample.

During the three years of the CSFII/DHKS, subsampling was required for 18 of the 501 segments in which the missed structure procedure was applied. A total of 692 DU's were added to the sample through the missed structure procedure.

The second procedure, referred to as the missed DU procedure, required interviewers to determine whether a structure such as an apartment building or a DU listed at a single address contained DU's that had not been listed. Any missed DU's found by this process were added to the sample. An upper limit of 4 new or missed DU's per structure was set before subsampling would take place but this limit was never exceeded. A total of 273 DU's were added to the sample through the missed DU procedure.

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. The approach used to select persons for the intake interviews was to designate subsets of households within which 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 assigned to the various subsets randomly to ensure the unbiased selection of SP's for the survey. 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. Some past USDA food surveys required the selection of all household members as sample persons, a method which sometimes placed a large respondent burden on a household. This procedure of subsampling persons within households served to allow a sample of some sex-age groups, primarily young children and older adults, larger than that found naturally in households and also served to lessen respondent burden within a household.

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 in the sample. 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. The number of sampling messages changed during the second and third years as sampling rates were modified. The proportion of households receiving a particular message was determined to satisfy the sampling rates for the various sex-age-income domains. The number and configuration of the sampling messages was a function of these sampling rates. A total of 19,830 SP's were identified through the screening procedure.

In general, the process of deriving the sampling rates, constructing the sampling messages, and allocating the messages to households may be summarized as follows: First, estimates of the number of persons in each sex-age-income domain were calculated. The initial sampling rates for each domain are then the ratio of the sample size targets to these estimated population counts. Next, domains with similar initial sampling rates and characteristics were grouped together. In the first year, 24 groups of one or more sex-age-income domains were created. The highest sampling rate within a group then became the final sampling rate for each domain in the group.

The groups were ordered by the final sampling rate and the groups were then collected cumulatively into the sampling messages. For example, in the first year the

group consisting of all-income males 1-2, all-income females 1-2, low-income males 1-2, low-income females 1-2, and low-income males 50-59 had the highest sampling rate. The group with the second highest rate consisted of only low-income males 60-69. The group with the third highest rate consisted of only low-income males 40-49.

Consequently, sample message 1 selected from households all children 1-2 as well as low-income men 50-59, sample message 2 selected all children 1-2 and low-income men 50-69, and sample message 3 selected all children 1-2 and low-income men 40-69. The last sampling message selected all members of a household. The proportion of households in the sample assigned to a message was computed using the same data that was used in computing the initial sampling rates. Households were assigned sample messages randomly in these proportions prior to the release of field assignments to interviewers.

The three-year sample size goals were 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).

Respondents for the DHKS were selected from among SP's 20 years of age and over who had completed the Day 1 intake without a proxy. 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. 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 be similar to that of the corresponding intake respondents by sex, age, and income. A total of 6,360 SP's were selected for the DHKS.

### **3. Response Rates**

With the several stages of data collection and the response demands of the CSFII and DHKS, a critical concern was the achievement of high response rates. Because the rate at which screening is completed affects the response rates for all subsequent interviews, particular emphasis was placed on maintaining as high a screening

rate as possible. Towards this goal, the screener questionnaire was designed to minimize the frequency of the necessity of collecting income data at the time of screening. This was made possible by the sampling message procedure described above. Also, the use of neighbor information ("neighbors" being someone living next door, an apartment manager, a postman, or anyone who was likely to know something about the household composition) was permitted under strict protocol. Two confirming sources were required to make use of neighbor information in completing the SP selection procedure. Of the 29,371 completed screeners, 3,102 (10.6%) were completed with neighbor information. Most of these (91.6%) were for households with no eligible SP's. In practice, if neighbor information identified eligible SP's, further attempts were made to complete the screening and subsequent interviews. By using these procedures, a 3-year screening response rate of 98.5 percent was achieved.

An overall response rate is the estimate of the response rate of all possible sample units. The overall response rates for the intake interviews and the household interview may be calculated as the product of the screening rate and the participation rate, where the participation rate is the rate at which eligible SP's and households already identified through a completed screening interview complete the subsequent interviews. For the 1994-96 CSFII the Day 1, 2-day and household participation rates were 81.3 percent, 77.2 percent, and 86.0 percent respectively. This resulted in an 80.0 percent overall Day 1 response rate, a 76.1 percent overall 2-day intake response rate, and an 84.6 percent overall household response rate. The calculation of the overall DHKS response rate is more complicated because an additional factor is required reflecting the rate at which screened households containing eligible adult SP's had an adult SP complete a Day 1 interview. Nevertheless, 91.6 percent of SP's selected to participate in the DHKS did so, and the overall DHKS response rate was 73.5 percent. See Tippet and Cypel (1997) for further details on response rate calculations.

#### **4. Intake Data Collection Methodology**

Interviewer's 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 three days of the week that would be acceptable for collecting Day 1 food intake information from sample persons 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 had been made on different scheduled days and at different times, interviewers were

permitted to change the day of the week in order to obtain an interview. This goal of balancing intake interviews across the days of the week was met. The distribution ranged from 11.2 percent of Day 1 interviews taking place on Fridays to 17.1 percent of Day 1 interviews taking place on Saturdays.

Interviewers were allowed 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 sampled person who was not available on that day, no further appointment could be made. These procedures were established to prevent possible bias from SP's changing their intakes in anticipation of the interview.

Day 1 intake questionnaires were administered in person. Before conducting the Day 1 interview, the interviewer told the SP that his/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 (s)he 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 the main meal preparer because this person could possibly provide more details about food preparation than other household members could. The food preparation information could be applied to the intakes of other SP's in the household.

Several small gifts were given to sample persons to encourage participation in the survey. A set of measuring cups and spoons was given to eligible households after completion of the Screener. An insulated nylon bag with the survey logo was presented to each SP before the Day 1 interview. After completing the Day 2 interview, the interviewer gave the SP a travel-type drinking mug, also imprinted with the survey logo, as a thank-you for participating.

Proxy interviews were conducted routinely for SP's under 6 years of age and any other SP's (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 who had 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 sample person, proxy, or assistant could not provide enough descriptive or qualitative information about the foods eaten, it was sometimes necessary to seek that information from other care-givers, such as baby-sitters, day-care personnel, or school cafeteria personnel. One percent of the intakes completed over the three years required data retrieval to be complete.

The Day 1 individual intake questionnaire began with a 24-hour dietary recall, using a multiple-pass approach. The 24-hour dietary recall task is a difficult one for respondents. In preparation for the CSFII 1994-96, researchers at the U.S. Bureau of the Census reviewed the methodology for the previous survey, the CSFII 1989-91, then conducted a series of cognitive interviews to learn how respondents interpreted the questions, understood the concepts, and formulated their answers (DeMaio, 1993). The cognitive research determined that multiple cues prompted respondents to perform different cognitive tasks at different points in the interview and stimulated more reports of foods and eating occasions. In addition, the researchers found that a chronological listing of foods consumed the previous day may not be the best strategy. A revised method was developed for the CSFII 1994-96

The first pass began with the respondent being asked to report everything (s)he ate or drank the previous day between midnight and midnight using any recall strategy desired in order to recall as many foods as possible. 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 (s)he 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 (FIB). The FIB is an extension of the intake questionnaire and contains probes used by the interviewers to elicit detailed descriptions of foods and amounts. The probes are designed to capture the information needed to assign appropriate food codes and vary with the type of food and to ensure standard administration by the interviewer.

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 FIB to

obtain details of every food item recalled by the respondent, including additions remembered as the result of questions asked in describing another food. The FIB 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. The guides included a set of stainless steel measuring cups (1/4 cup, 1/3 cup, 1/2 cup, and 1 cup) and four measuring spoons (1/4 teaspoon, 1/2 teaspoon, 1 teaspoon, and 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, glass or cup used in her/his home. The respondent could then fill the eating utensil with water to represent the amount (s)he 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 had been described and quantified, the interviewer reviewed for the respondent all the foods listed for each 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.

Interviewers were trained to review and edit the intake questionnaires as soon as possible after leaving the respondent's homes. Legibility, accuracy, and completeness were to be checked using a standard list. Explanations were added if necessary.

The Day 2 interview was to be conducted 3 to 10 days after the Day 1 interview but not on the same day of the week. While the majority of the interviews were conducted within the 3-10 day window, 0.4 percent of Day 2 interviews were conducted sooner than 3 days after the Day 1 interview, 2.5 percent were conducted on the same day of the week, and 20.5 percent were conducted more than 10 days after the Day 1 interview.

## 5. Data

The coding of the intake data is a complex process that is described in another paper from this session (LaComb,

1997). The result of the coding process is a food code that identifies each food reported by each SP and an amount, in grams, of each food. The survey nutrient database contains the amounts of nutrients and other dietary components in 100 grams of each food. Thus, the data outcome of the survey is a data set that includes the quantified daily intake of foods and nutrients for each SP. These data are available to the public on CD-ROM in a package that includes the data collected from the household, intake, and DHKS interviews as well as the specific information about each food reported and daily totals of foods and nutrients for each SP. The package also includes sample weights (Chu, 1997), technical support files such as the nutrient database, and documentation. The 1994 and 1995 data have been made available in annual releases. By March 1998, a package containing all three years of the CSFII/DHKS 1994-96 data will be available to the public on CD-ROM.

## 6. Acknowledgments

We would like to thank Graham Kalton and Howard Riddick for their useful comments and suggestions.

## 7. References

Chu, A., and J. Goldman. 1997. Weighting Procedures for USDA's Continuing Survey of Food Intakes by Individuals 1994-96. 1997 Proceedings of the Section on Survey Research Methods, American Statistical Association.

Cleveland, L.E., D.A. Cook, S.M. Krebs-Smith, and J. Friday. 1997. Method for Assessing Food Intakes in Terms of Servings Based on Food Guidance. *American Journal of Clinical Nutrition*, 65(suppl):1254S-1263S.

Crane, N.T., D.B. Wilson, D.A. Cook, C.J. Lewis, E.A. Yetley, and J.I. Rader. 1995. Evaluating Food

Fortification Options: General Principles Revisited with Folic Acid. *American Journal of Public Health*, 85(5):660-666.

DeMaio, T.J., S. Ciochetto, W.L. Davis. 1993. Research on the Continuing Survey of Food Intakes by Individuals. 1993 Proceedings of the Section on Survey Research Methods, Vol.II, American Statistical Association, pp. 1021-1026.

Guthrie, J.F., J.J. Fox, L.E. Cleveland, and S. Welsh. 1995. Who Uses Nutrition Labeling, and What Effects Does Label Use Have on Diet Quality. *Journal of Nutrition Education*, 27(4):163-172.

Kennedy, E., J. Ohls, S. Carlson, and K. Fleming. 1996. Measuring the Dietary Quality of Americans' Food Consumption: The Healthy Eating Index. Nutrition and Food Security in the Food Stamp Program. U.S. Dept. of Agriculture, Food and Consumer Services. January 1996, pp. 75-90.

Krebs-Smith, S.M., D.A. Cook, A.F. Subar, L. Cleveland, and J.E. Friday. 1995. U.S. Adults' Fruit and Vegetable Intakes, 1989 to 1991: A Revised Baseline for the Healthy People 2000 Objectives. *American Journal of Public Health*, 85(12):1623-1629.

LaComb, R., A. Green, and L. Ingwersen. 1997. Survey Net: Coding and Management of CSFII Food Intake Data. 1997 Proceedings of the Section on Survey Research Methods, American Statistical Association.

Tippett, K.S., and Y.S. Cypel (eds.). 1997. Design and Operation: The Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey, 1994-96. U.S. Department of Agriculture, Agricultural Research Service, Nationwide Food Surveys Report No. 91-1.