

Contract No. 53-3198-5-20
MPR Project No. 7635-407

Authors: Pat Doyle
Thomas Fraker
Pamela Haines

FINAL REPORT

A CONCEPTUAL STUDY OF METHODOLOGIES
FOR ANALYZING DATA FROM THE 1985
CONTINUING SURVEY OF FOOD
INTAKES BY INDIVIDUALS

July 31, 1987

Prepared for:

U.S. Department of Agriculture
Food and Nutrition Service
3101 Park Center Drive
Room 212
Alexandria, Virginia 22302

Prepared by:

Mathematica Policy Research, Inc.
600 Maryland Avenue, S.W.
Suite 550
Washington, D.C. 20024
(202) 484-9220

EXECUTIVE SUMMARY

A. INTRODUCTION TO THE CSFII

The Continuing Survey of Food Intakes by Individuals (CSFII) is a longitudinal survey of women in the continental United States who are ages 19-50 years and their children ages 1-5 years. An individual participates in the survey for one year, during which time he or she is interviewed six times at two-month intervals. Demographic, income, and program participation data are gathered for the households of the survey participants. For the women and their young children, data are also gathered on food intakes during the day prior to each wave (round) of the survey. At the end of a survey year the existing panel of survey participants is dissolved and a new panel is constituted.

The initial panel of the CSFII was first interviewed in the spring of 1985. In addition to a core sample of age-eligible women and children from all income strata, the CSFII-1985 included several supplemental samples, most notably, a sample of women ages 19-50 years and their children ages 1-5 years in low-income households. Wave 1 data were obtained for 5,200 individuals in the core and low-income samples. Natural attrition, plus an intentional reduction in the size of the low-income sample by approximately 1,200 persons, reduced the number of respondents to 2,275 by the sixth wave of the survey.

Under contract to the Food and Nutrition Service, Mathematica Policy Research produced an analysis file that contains merged wave 1 data for individuals in both the core and the low-income samples of the CSFII-1985. Revised sample weights exist for the merged cases. These can

be used to weight the cases up to Census counts of the United States' population of women ages 19-50 years and their children ages 1-5 years. This analysis file has been used by FNS to generate a wide range of descriptive statistics on dietary behavior and as a basis for regression analyses of food stamp, WIC, and Child Care Food Program impacts on iron intake by young children.

B. GOALS OF THIS STUDY

This study has two goals. The first goal is to provide FNS with guidance regarding the development of future CSFII analysis files and, more specifically, the further development of analysis files for the CSFII-1985. The study's second goal is to examine the CSFII's capacity to support a wide range of analyses of dietary and related behavior. MPR's recommendations for future CSFII file development are summarized in the next section, followed by a summary of our conclusions regarding the survey's research potential.

C. RECOMMENDATIONS FOR CSFII FILE DEVELOPMENT

Due to sample attrition, each successive wave of the CSFII-1985 provides fewer observations on the individuals in the survey's core and low-income samples. Furthermore, if attrition was nonrandom with respect to individual characteristics,¹ then the respondents in each successive wave are increasingly less representative of the population cohorts that they originally were selected to represent. For these reasons, MPR recommends

¹Preliminary evidence indicates that attrition from the CSFII-1985 samples was nonrandom with respect to at least some individual characteristics.

that all cross-sectional analyses of the CSFII be conducted on wave 1 data. As noted in Table ES.1, the most immediate implication of this recommendation is that there should be no development of CSFII-1985 cross-sectional files beyond the existing wave 1 merged core and low-income file.

The construction of a single longitudinal file containing merged data for the core and low-income samples is the most effective use to which remaining resources for CSFII-1985 file development could be put.

Table ES.1 shows that MPR recommends that the unit of observation for this file be the individual who is age- and sex-eligible for the CSFII's dietary intake survey and that household characteristics be treated as attributes of the individuals. This approach avoids the knotty problem of defining and following a household over time and it has the added benefit of being consistent with the design of the wave 1 analysis file.

In the presence of sample attrition, the construction of a longitudinal file for any data set requires that tradeoffs be made between the competing objectives of a well-defined reference period of maximum length, maximum sample size, minimum discarding of valid data, and minimization of imputation of missing data for sample attriters. With these objectives in mind, MPR recommends that FNS select one of the three longitudinal file design options specified in Item 3 of Table ES.1. Further research on the tradeoffs associated with each option should be conducted so as to permit a well-informed selection to be made.

Whichever design option for the longitudinal file is selected, it will be necessary to compute revised sample weights. The second part of Table ES.1 recommends research on patterns of sample attrition, findings from which would be input into the computation of new weights. Several

TABLE ES.1

**SUMMARY OF RECOMMENDATIONS FOR FUTURE DEVELOPMENT
OF CSFII-1985 DATA FILES AND RELATED RESEARCH**

Recommendations for Development of CSFII-1985 Data Files

1. There should be no further development of cross-sectional analysis files for the core and low-income samples of the CSFII-1985. All cross-sectional analyses should be conducted on the existing wave 1 analysis file.
2. A single longitudinal file should be developed. The unit of observation in this file should be the individual who is eligible to participate in the dietary intake component of the survey. Household characteristics should be included in the file as attributes (i.e., characteristics) of the individuals.
3. In designing a longitudinal file, three alternatives for dealing with attrition of individuals from the survey between waves 1 and 6 are as follows:
 - o Restrict the file to individuals who completed all 6 interviews.
 - o Restrict the file to individuals who completed at least 4 interviews; impute responses for the missing interviews.
 - o Restrict the file to individuals who completed at least 4 interviews; include data in the file for the first interview and 3 other randomly selected interviews.

Recommendations for Related Research

4. The CSFII's rules regarding the following of children who move out of their original households are more restrictive than those for women; consequently, the CSFII-1985 longitudinal sample of children may not be representative of the cohort of young children of women who were ages 19-50 in the spring of 1985. Research on the implications of this for the inclusion of young children in the longitudinal analysis file is recommended.

TABLE ES.1 (continued)

Recommendations for Related Research (continued)

5. To inform the selection among the three alternative longitudinal file designs for dealing with sample attrition, research is recommended on the implications of each alternative for: sample size, loss of valid data, volume of required imputation, and feasibility of computing change in household income.
6. Research on whether sample attrition between waves 1 and 6 was correlated with individual or household characteristics is recommended prior to the computation of revised sample weights for the longitudinal file.
7. The CSFII methodology for collecting data on the change in household income between successive interviews raises questions about the quality of the resultant data. Research on the quality of those data is recommended prior to the construction of income measures for the longitudinal analysis file.
8. Patterns over time in a case's data on household composition, WIC participation, and food stamp participation should be examined. Longitudinal editing should be used to correct gross wave-to-wave inconsistencies.

other areas of research that would also provide critical information for the development of a longitudinal data file are also identified in the table.

D. THE CSFII'S RESEARCH POTENTIAL

The CSFII's greatest strength with respect to its capacity to support analyses of dietary and related behavior is the up to six observations on 24-hour dietary intakes that it provides for individual participants in the survey. This is twice as many daily observations as provided by the Nationwide Food Consumption Survey (NFCS). The survey's greatest weakness is its highly restrictive sample. It is not possible to generalize findings from dietary analyses of the CSFII's core and low-income samples beyond the subpopulations of women and their young children from which they were drawn.¹

The CSFII can support a wide range of cross-sectional and longitudinal analyses of dietary and related behavior. Table ES.2 identifies several general analytic methods that could be applied to CSFII data. It also identifies many specific cross-sectional and longitudinal analyses that could be conducted on the basis of CSFII data, as well as several that could not. In neither respect is the table exhaustive. To avoid excessive redundancy with the table, we highlight here only four of the more interesting analytic methods and specific analyses.

¹For wave 1 only, the CSFII-1985 obtained dietary intake data for a sample of 658 men ages 19-50 years who were residing with women ages 19-50 years. Due to the small sample size and the absence of repeated observations on dietary intakes, the research potential of this sample is not explored in this report.

1. Measurement of Usual Diet

The measurement of usual diet on the basis of short-term (e.g., one day, three days, or six days) intake data is subject to considerable error. Some types of measurements are known to be biased. For example, simple estimates of the distribution of the usual daily intake of a nutrient for a subpopulation are biased if computed on the basis of short-term data. This means that derived estimates of the percent of the subpopulation that attains (or fails to attain) the RDA for that nutrient are also biased. However, the National Research Council (1986) proposes a technique that essentially purges short-term data of day-to-day and season-to-season fluctuations in an individual's daily dietary intake. The adjusted data will support the computation of unbiased estimates. The CSFII design, with repeated observations over days of the week and seasons of the year is well-suited to the application of this technique.

2. Estimation of Program Impacts

Regarding the dietary impacts of food and nutrition programs, the CSFII is best suited to analyses of the WIC Program. However, CSFII data will support analyses of the impacts of other food and nutrition programs as well. This report develops models of the impacts of five USDA programs on dietary intakes by women and their young children and on the food expenditures of their households.¹ The dietary intake model incorporates both the direct effect of participation in a program and the spillover effect of participation by some other household member. Given the CSFII's

¹The five programs are WIC, the Food Stamp Program, the School Breakfast Program, the National School Lunch Program, and the Child Care Food Program.

samples of women and young children, for some programs (e.g., food stamps and WIC) both effects may be present, whereas for the school-based programs only spillover effects can be present.

3. Dietary Intakes and Duration Since Receipt of Food Stamps

The CSFII is the first large-scale dietary survey to provide information on the date of last receipt of food stamps by survey respondents or their households. NFCS files providing similar information will not be available for approximately 18 months. The principal motivation for obtaining this information in the 1987 NFCS (it was not obtained in the 1977-78 NFCS) is to permit analyses of associations between the timing of individual dietary intakes and household food expenditures and timing of the receipt of food stamps. The analysis of the first of these two associations could be conducted now on the basis of the wave 1 analysis file for the CSFII-1985. Due to its two-month reference period for food expenditure information, the CSFII cannot support an analysis of associations between the timing of household food expenditures and the timing of the receipt to food stamps.

4. A "Fixed-Effects" Model for Longitudinal Analyses of Dietary Data

Because of the multiple observations on each sample case, a longitudinal data file provides an opportunity to obtain more reliable statistical estimates than can be obtained on the basis of a cross-sectional file with an equal or somewhat larger number of cases. However, the choice of an appropriate model for analyzing longitudinal data is critical and the available models tend to be complicated. A model that appears to be appropriate for a variety of longitudinal analyses of dietary

intakes is the "fixed-effects" model. For example, this model could be used to explain the deviation of each CSFII respondent's dietary intake in each wave of the survey from her average intake in all waves as a function of the deviation of her program participation status (and income, household size, etc.) in the same wave from her average participation status (and income, household size, etc.) in all waves.