SURVEY OF INCOME AND PROGRAM PARTICIPATION USERS' GUIDE

(Supplement to the Technical Documentation)

Third Edition
Washington, D.C.
2001

Prepared by:

Westat
1650 Research Boulevard
Rockville, Maryland 20850

In association with:

Mathematica Policy Research, Inc.
600 Maryland Avenue, S.W., Suite 550
Washington, D.C. 20024-2512

Contract No. 50-YABC-7-66016

U.S. DEPARTMENT OF COMMERCE
ECONOMICS AND STATISTICS ADMINISTRATION
U.S. CENSUS BUREAU

Acknowledgments

The third edition of the *Survey of Income and Program Participation (SIPP) Users' Guide* was prepared for the U.S. Census Bureau by Westat. Charles T. Nelson was the Government Project Officer for the project within the Census Bureau, and Pat Doyle also provided invaluable support and guidance to the effort. Many other staff from a number of divisions within the Census Bureau shared their expertise and provided useful comments. In particular, we would like to thank Patrick Benton, John Boies, Judith Hubbard Eargle, Donald Keathly, Karen Ellen King, Gordon Lester, Stephen Mack, Mike McMahon, Thomas Palumbo, Donna Riccini, and Mahdi Sundukchi.

Chapters of the third edition were prepared by Louis Rizzo, Marianne Winglee, Alan Martinson, and Ilene France of Westat; Larry Radbill of Mathematica Policy Research, Inc.; Julie Sykes (then of Mathematica Policy Research, Inc.); and Elizabeth Sheley (Independent Consultant). Alan Martinson, Marty Franklin, Laurie Tomasino, and Carol Dominique of Westat provided editorial and production support; Julie Phillips (Independent Consultant) prepared the Index; and Ana Horton of Westat designed the cover. Garrett Moran served as the Westat Project Director.

****

Because this edition of the *Users' Guide* builds on the previous editions, we also include the following acknowledgments, which appeared in the second edition.

The first edition of the Survey of Income and Program Participation (SIPP) Users' Guide was prepared by Daniel Kasprzyk (then Office of the Director), Pat Doyle (Mathematica Policy Research, Inc.), Arnold Goldstein (Population Division), Patricia Kelly (Office of the Director), and David B. McMillen (then Office of the Director).

The second edition was prepared by the Data Access and Use Staff of the Data User Services Division. Geneva Burns coordinated the effort, assisted by Jackson Morton and J. Paul Wyatt. Andrea Meier of the Survey of Income and Program Participation Branch in the Statistical Methods Division prepared Chapter 8, "SIPP Cross-Sectional Weighting Procedures," under the direction of Rajendra P. Singh. We would like to thank our colleagues within the Census Bureau and our SIPP file users for their helpful comments.

Contents

Chapter **Page**

1 Introduction 1-1

Evolution and History of SIPP 1-1
Uses of SIPP 1-3
The Survey 1-4
Nonsampling Errors, Sampling Errors, and Weighting 1-6
SIPP Public Use Files 1-7
Comparison of SIPP with Other Surveys 1-9
Guide to This Document 1-11
Where to Go for More Information 1-13

2 SIPP Sample Design and Interview Procedures 2-1

Organizing Principles 2-1
Sample Design 2-5
Following Rules 2-9
Interview Procedures 2-16
Nonresponse 2-17

3 Survey Content 3-1

The SIPP Interview 3-1
Core Content 3-2
Topical Content 3-6

4 Data Editing and Imputation 4-1

Types of Missing Data 4-1
Goals of Imputation 4-2
Assessing the Influence of Imputed Data on Analysis 4-3
An Overview of the Process 4-3
Phase 1: Data Editing and Imputation Procedures for the Core Wave Files 4-6
Phase 2: Data Editing Procedures for the Full Panel Files 4-15
Confidentiality Procedures for the Public Use Files 4-17

5 Finding SIPP Information 5-1

Published Estimates from SIPP 5-1
SIPP Public Use Microdata Files 5-1
Sources for Obtaining SIPP Microdata 5-12
Other Sources of Information About SIPP 5-13

Chapter **Page**

6 Nonsampling Errors 6-1

Undercoverage 6-1
Nonresponse 6-1
Measurement Errors 6-2
Effects of Nonsampling Error on Survey Estimates 6-3

7 Sampling Error 7-1

Direct Variance Estimation 7-1
Using GVFs to Approximate Variance Estimates 7-4
Variance Estimation with Imputed Data 7-6

8 Using Sampling Weights on SIPP Files 8-1

What Weights Are and Why They Should Be Used 8-1
Weights Available in SIPP Files 8-3
Choosing a Weight 8-3
How Weights Are Constructed 8-4
Using Weights in the Core Wave Files 8-8
Using Weights in the Topical Module Files 8-16
Using Weights in the Full Panel File 8-16
Pooling Data from Two or Three Panels 8-19

9 The SIPP Public Use Files 9-1

Types of SIPP Data Files 9-1
Understanding the ID Variables in SIPP 9-2
Identifying Persons and Their Relationships 9-4
Working with Multiple Files 9-9
The Balance of Section II 9-9

10 Using the Core Wave Files 10-1

Using the Technical Documentation of the Core Wave Files 10-2
Relationship of the Core Wave Data Files to the SIPP Survey Instrument 10-4
Structure of the Core Wave Files 10-6
Identifying Persons 10-6
Identifying Households 10-9
Identifying Families 10-11
Other Variables Describing Household and Family Composition 10-15
More About Using the SIPP ID Variables: Identifying Movers 10-20
Identifying Program Units 10-26
Income Topcoding in the 1996 Panel 10-29

Chapter **Page**

10 Using the Core Wave Files (Cont.)

Topcoding Prior to the 1996 Panel 10-35
Using Allocation (Imputation) Flags 10-36
Using Weights 10-37
Identifying States 10-38
Identifying Metropolitan Areas 10-39

11 Using Topical Module Files 11-1

Using the Technical Documentation of the Topical Module Files 11-2
Relationship of the Topical Module Data Files to the Survey Instrument 11-6
Structure of the Topical Module Files 11-7
Reference Periods and Samples 11-8
Using a Person's Monthly Interview Status Variables 11-9
Comparison of Variables in the Topical Module and Core Wave Files 11-11
Identifying People 11-13
Identifying Families 11-16
Other Variables Describing Household and Family Composition 11-19
More About Using the SIPP ID Variables: Identifying Movers 11-21
Topcoding 11-27
Using Allocation (Imputation) Flags 11-28
Using Weights 11-28
Identifying States 11-29
Identifying Metropolitan Areas 11-29

12 Using the 1990–1993 Full Panel Longitudinal Research Files 12-1

Using the Technical Documentation of the 1990–1993 Longitudinal Research Files 12-2
Relationship of the Longitudinal Research Data Files to the SIPP Survey Instrument 12-5
Structure of the Longitudinal Research Files 12-6
How to Align Data by Calendar Month 12-7
Using the Monthly Interview Status (PP-MIS) Variables 12-9
Identifying Persons 12-13
Identifying Households 12-15
Identifying Families 12-16
Variables Describing Household and Family Composition 12-19
Using Family-Level Income Variables 12-23
More About Using the SIPP ID Variables: Identifying Movers 12-23
Identifying Program Units 12-28
Using the Unearned Income Variables 12-30

Chapter **Page**

12 Using the 1990–1993 Full Panel Longitudinal Research Files (Cont.)

Income Topcoding 12-31
Using Allocation (Imputation) Flags 12-37
Using Weights 12-37
Identifying States 12-38
Identifying Metropolitan Areas 12-38

13 Linking Core Wave, Topical Module, and Longitudinal Research Files 13-1

Procedures for Linking Files 13-2
Nonmatches When Merging Files 13-15

Appendix

A SIPP Users' Guide Variable Crosswalk: 1993 to 1996 ..... A-1

By 1993 Variable Name A-2
By 1996 Variable Name A-10
By 1993 File Position A-17
By 1996 File Position A-25

B SIPP Topcoding Specifications B-1

Earnings B-1
Year of Birth (TBYEAR) B-4
Age (TAGE) B-4
Age at Receipt of Social Security Disability Benefits (TAGESS) B-5
Age Respondent Started Job or Business (TSJDATE, TEJDATE, TSBDATE, TEBDATE) B-5

C Computing the SIPP Sample Weights C-1

Wave 1 Weights C-1
Wave 2+ Weights C-12
Calendar Year and Panel Weights C-17

D Acronyms D-1

E Glossary E-1

References R-1

Index Index-1

Tables

Table		Page
1-1	Comparison of SIPP, CPS, and PSID	1-10
2-1	Summary of the 1984–1996 SIPP Panels	2-2
2-2	1996 Panel: Rotation Groups, Waves (W), and Reference Months	2-4
2-3	Household Membership	2-7
2-4	Composition of the 1990 Panel	2-8
2-5	Household Noninterview and Sample Loss Rates: 1990–1996 Panels	2-19
3-1	Types of Income Recorded in SIPP	3-5
3-2	Topical Modules Grouped Thematically	3-7
5-1	Publications in the P-70 Series	5-2
5-2	Structure of the Person-Month Format Core Wave Files	5-5
5-3	Topical Modules, by Panel and Wave	5-6
5-4	Topical Modules, by Subject	5-10
5-5	Structure of Topical Module Microdata File	5-11
5-6	Telephone Numbers for Information About Specific Aspects of SIPP	5-16
7-1	Variance Stratum Code and Variance Unit Code in SIPP Files, 1990–1993	7-3
8-1	Weighted and Unweighted Point-in-Time Estimates of Percentages Based on Core Wave 1 of the 1990 SIPP Panel for January 1990	8-2
8-2	Weight Variables in SIPP Files for the 1996 and 1990–1993 Panels	8-3
8-3	Final Person Weights for Four Reference Months and One Interview Month in Wave 1 of the 1991 Panel	8-10
8-4	Household, Reference Month, and Interview Month Weights for Members of a Household for a Given Month in Wave 1 of the 1990 Panel	8-11
8-5	Family and Subfamily Reference Months Weights, by RHTYPE (HTYPE), EFTYPE (FTYPE), and ESFTYPE (STYPE) in Wave 1 of the 1990 Panel	8-13
8-6	Calendar Month Estimation: Using a Single Core Wave File in Wave 1 of the 1991 and 1996 Panels	8-14
8-7	Calendar Month Estimation: Using Two Core Wave Files from Waves 1 and 2 of the 1991 and 1996 Panels	8-15
8-8	Calendar Year and Panel Weights, 1990–1993	8-17
8-9	Weighting Parameter Adjustment Factors for Both the Two-Panel and Three-Panel Combinations	8-21

Table **Page**

9-1 SIPP Variable Names, by File Type 9-3

9-2 Differences Among Core Wave, Topical Module, and Longitudinal Files (1990–1996 Panels) 9-11

10-1 Person-Month File Structure for the Core Wave Files 10-7

10-2 Variables Used to Uniquely Identify a Person in the Core Wave Files 10-8

10-3 How to Uniquely Identify a Person in the Core Wave Files 10-9

10-4 Variables Used to Uniquely Identify a Household or Group Quarters in the Core Wave Files 10-10

10-5 How to Uniquely Identify a Household in the Core Wave Files 10-10

10-6 Variables Used to Uniquely Identify a Family in the Core Wave Files 10-11

10-7 Uniquely Identifying Families in the Core Wave Files 10-13

10-8 Variables Describing Household and Family Composition in the Core Wave Files 10-15

10-9 The ERRP Variable in the 1996 Core Wave Files 10-17

10-10 Comparison of RRP and RRPU Variables of the Core Wave Files Prior to the 1996 Panel 10-17

10-11 Identifying Households Containing Three Generations in the Core Wave Files 10-18

10-12 Identifying Households Containing Three Generations in the Core Wave Files 10-19

10-13 How the Family-Level Variables Include the Subfamily's Information in the Core Wave Files 10-21

10-14 Identifying Movers in the Core Wave Files 10-22

10-15 Example of Household Changes and Their Effects on the ID Variables of the Core Wave Files 10-23

10-16 Variables Describing Participation in Government Transfer Programs and Health Insurance Programs in the Core Wave Files 10-27

10-17 Example of Program Units, Coverage, and Recipiency in the Core Wave Files 10-30

10-18 Topcoding Criteria for the 1996 Panel 10-32

10-19 Topcode Amounts Used for Monthly Employment Income in Wave 1 of the 1996 Panel 10-33

10-20 Example of Employment Income Topcoding in the 1996 Panel 10-35

10-21 Example of Topcoding in the Core Wave Files Prior to the 1996 Panel: Single Person Household 10-36

Table **Page**

10-22 Weight Variables in SIPP Core Wave Files for the 1996 and 1990–1993 Panels 10-38
11-1 Example of the Topical Module File Structure 11-7
11-2 Monthly Interview Status Variables in the 1984–1993 SIPP Panels 11-10
11-3 Interview Month and Reference Months for Each Rotation Group in Wave 4 of the 1993 Panel 11-10
11-4 Variables Common to the Core Wave and Topical Module Files from Wave 1 of the 1996 Panel 11-12
11-5 Examples of Same Variables with Different Names in the Core Wave and Topical Module Files Prior to the 1996 Panel 11-12
11-6 Variables Used to Uniquely Identify a Person in the Topical Module Files 11-13
11-7 How to Uniquely Identify a Person in the Topical Module Files 11-15
11-8 Variables Used to Uniquely Identify a Household or Group Quarters in the Topical Module Files 11-15
11-9 How to Uniquely Identify a Household in the Topical Module Files 11-16
11-10 Variables Used to Uniquely Identify a Family in the Topical Module Files for the 1996 Panel 11-17
11-11 Uniquely Identifying Families in the Topical Module Files in the 1996 Panel ... 11-18
11-12 Household and Family Composition Variables in the Topical Module Files 11-19
11-13 Relationship to the Household Reference Person in the Topical Module Files .. 11-20
11-14 ERRP (RRP) Coding for the Same Three-Generation Household When Two Different People Are Designated as the Reference Person in the Topical Module Files 11-21
11-15 Identifying Households Containing Three Generations in the Topical Module Files 11-22
11-16 Identifying Movers in the Core Wave Files 11-23
11-17 Example of Household Changes and Their Effects on the ID Variables in the Core Wave Files 11-25
12-1 Summary of Panels, Waves, Reference Months, and Sample Sizes 12-7
12-2 Example of the Longitudinal Research File Structure 12-8
12-3 Reference Periods for Each Rotation Group of the 1992 Panel 12-9
12-4 Monthly Data from the 1992 Panel, Realigned by Calendar Month 12-11
12-5 Variables Used to Uniquely Identify a Person in the Longitudinal Research Files 12-14

Table **Page**

12-6 How to Uniquely Identify a Person in the Longitudinal Research Files 12-15

12-7 Variables Used to Uniquely Identify a Household in the Longitudinal Research Files 12-15

12-8 How to Uniquely Identify a Household or Group Quarters in a Given Month of the Longitudinal Research Files.....12-16

12-9 Variables Used to Identify Families in the Longitudinal Research Files 12-18

12-10 How to Uniquely Identify a Family in a Given Month of the Longitudinal Research Files 12-20

12-11 Variables Used to Describe Household Composition in the Longitudinal Research Files 12-21

12-12 Relationship to the Household Reference Person in a Given Month.....12-21

12-13 Using RRP to Identify Households Containing Three Generations in the Longitudinal Research Files 12-22

12-14 Using PNSP and PNPT to Identify Households Containing Three Generations in the Longitudinal Research Files.....12-22

12-15 Family Income in the Longitudinal Research Files 12-23

12-16 How to Identify Movers in the Longitudinal Research Files.....12-24

12-17 Another Example of Household Changes and Their Effects on the ID Variables in the Longitudinal Research Files.....12-25

12-18 Household Changes and Their Effects on the Household ID (HH-ADDID$_i$) Variable in the Longitudinal Research File 12-27

12-19 Variables Describing Participation in Government Transfer Programs and Health Insurance Programs in the 1990–1993 Longitudinal Research Files.....12-29

12-20 Example of Program Units, Coverage, and Benefit Amounts in the Longitudinal Research Files 12-31

12-21 Unearned Income in the Longitudinal Research Files.....12-32

12-22 User-Created SSI and FSP Variables Using the Unearned Income Variables in the Longitudinal Research Files 12-34

12-23 Example of Topcoding in the Longitudinal Research Files.....12-37

13-1 Example of the Core Wave Person-Month File Structure 13-7

13-2 Example of the Core-Wave Wide-Record/Person File Structure (After Applying the Program in Figure 13-1 to the Data in 13-1).....13-7

13-3 Variables Identifying People in the Core Wave and Longitudinal Research Files for Panels Prior to 1996.....13-9

Table **Page**

13-4 Variables Identifying People in the Topical Module and Core Wave Files for Panels Prior to 1996 13-14
13-5 Variables Identifying People in the Topical Module and Longitudinal Research Files Prior to the 1996 Panel 13-15
13-6 Reasons for Nonmatches 13-17
B-1 Examples of Income Amounts That Need to Be Topcoded B-2
B-2 Earnings Topcodes B-4
B-3 1996 Panel Topcoding Specifications B-6
C-1 Major Groupings of Later Wave Noninterview Cells C-19
C-2 Major Groupings of Calendar Year (Panel) Noninterview Cells C-21

Figures

Figure **Page**

2-1 Following Rules 2-10
3-1 Skip Pattern Example 3-2
4-1 Sequence of Cross-Sectional Imputation and Longitudinal Editing Procedures 4-4
10-1 Excerpt from a Data Dictionary for the Core Wave Files 10-3
10-2 Corresponding SAS and FORTRAN Syntax to Read the Data from the Core Wave Files 10-5
11-1 Excerpt from the Data Dictionary for the Topical Module Files 11-3
11-2 Corresponding SAS and FORTRAN Syntax to Read Data from Topical Module Files 11-5
12-1 Excerpt from the 1993 Longitudinal Research File Data Dictionary 12-4
12-2 Corresponding SAS and FORTRAN Syntax to Read in Data from the 1993 Longitudinal Research File Data Dictionary 12-5
12-3 Algorithm for Realigning SIPP Panel Month to Calendar Months in the 1992 Panel 12-10
12-4 Constructing Family and Subfamily ID Variables in the Longitudinal Research Files 12-18
12-5 Creating Monthly Food Stamp and SSI Income Variables from the Unearned Income Variables in the Longitudinal Research Files 12-36

Figure **Page**

13-1 Sample SAS Code to Change the Core Wave Files from Person-Month Format to Person-Record Format from Wave 2 of the 1996 Panel 13-5

13-2 Sample SAS Code to Change the Longitudinal Research Files from Person-Record Format to Person-Month Format for Panels Prior to 1996 13-10

13-3 Data Dictionary Entries for Variables Identifying the Reason a Person Left the SIPP Sample 13-19

C-1 Second-Stage Cells for Hispanics C-6

C-2 Second-Stage Cells for Non-Hispanic Children C-7

C-3 Second-Stage Cells for Non-Hispanic Adults C-8

C-4 Calendar Year and Panel Weight Second-Stage Cells for Hispanics C-23

C-5 Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Children C-23

C-6 Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Adults C-24

Section I

1. Introduction

This guide is intended as a reference for analysts who need information about using the Survey of Income and Program Participation (SIPP). The main objective of SIPP is to provide accurate and comprehensive information about the income and program participation of individuals and households in the United States, and about the principal determinants of income and program participation. SIPP offers detailed information on cash and noncash income on a subannual basis. The survey also collects data on taxes, assets, liabilities, and participation in government transfer programs. SIPP data allow the government to evaluate the effectiveness of federal, state, and local programs.

This chapter and the ones that follow come under two main sections. Section I encompasses discussions of survey design and content, data editing and imputation procedures, sampling and nonsampling error, and weighting. Section II provides information about working with each of the three types of SIPP microdata files (the core wave files, topical module files, and full panel files), as well as instructions for linking SIPP files. This introduction offers a brief overview of each of those topics.

Evolution and History of SIPP

Until the advent of SIPP, the major source of data on income and program participation was the Current Population Survey (CPS) March Income Supplement. The CPS continues to be the source of all official income and poverty statistics published by the Census Bureau. The CPS, however, is designed primarily to obtain information on employment. Because income measurement was never the primary purpose of the CPS, it has certain gaps in this area. For example, CPS respondents are asked in March to recall their income during the preceding calendar year. Many respondents have difficulty in remembering sources such as property income or irregular income over the yearlong reference period. Also, the CPS does not capture the impact of changes in household composition during the year, nor does the survey explicitly measure periods of program participation. Further, the CPS does not collect data on assets and liabilities, which are needed to measure more completely a household's economic status and eligibility for program benefits. To add those items to the CPS questionnaire would dilute the main purpose of that survey and unduly increase respondent burden. Finally, the CPS is designed to be a cross-sectional survey. During the 1970s, the increasing size of government programs and their interactions with the labor market led to a need for longitudinal data.

To address those data issues, the Department of Health, Education, and Welfare (HEW) initiated the Income Survey Development Program (ISDP) in the late 1970s. In developing ISDP content and procedures, HEW focused on questionnaire length, length of reference period, and linkage of survey data to program records. The 1979 ISDP Panel was a longitudinal survey in which respondents were asked about their income, labor force participation, and other characteristics;

repondents were recontacted every 3 months to supply information on themselves and others with whom they resided; the 3-month span was the *reference period* for the interview.

The First SIPP Panels

The lessons learned from ISDP were incorporated into the initial design of SIPP, which was used for the first 10 years of the survey. The original design of SIPP called for a nationally representative sample of individuals 15 years of age and older to be selected in households in the civilian noninstitutionalized population. Those individuals, along with others who subsequently lived with them, were to be interviewed once every 4 months over a 32-month period. To ease field procedures and spread the work evenly over the 4-month reference period for the interviewers, the Census Bureau randomly divided each panel into four *rotation groups*. Each rotation group was interviewed in a separate month. Four rotation groups thus constituted one cycle, called a *wave*, of interviewing for the entire panel (Chapter 2). At each interview, respondents were asked to provide information covering the 4 months since the previous interview. The 4-month span was the reference period for the interview. The first sample, the 1984 Panel, began interviews in October 1983 with sample members in 19,878 households. The second sample, the 1985 Panel, began in February 1985. Subsequent panels began in February of each calendar year, resulting in concurrent administration of the survey in multiple panels.

The original goal was to have each panel cover eight waves. However, a number of panels were terminated early (Chapter 2) because of insufficient funding. For example, the 1988 Panel had six waves; the 1989 Panel, part of which was folded into the 1990 Panel, was halted after three waves. In addition, the intent was for each SIPP panel to have an initial sample size of 20,000 households. That target was rarely achieved; again, budget issues were usually the reason.

The 1996 redesign (discussed below) entailed a number of important changes. First, the 1996 Panel spans 4 years and encompasses 12 waves. The redesign has abandoned the overlapping panel structure of the earlier SIPP, but sample size has been substantially increased: the 1996 Panel had an initial sample size of 40,188 households (Chapter 2).

The 1996 Redesign

In 1990, the Census Bureau asked the Committee on National Statistics (CNSTAT) at the National Research Council to undertake a comprehensive review of SIPP. The resulting report, *The Future of the Survey of Income and Program Participation* (Citro and Kalton, 1993), summarizes the first 9 years of SIPP and provides recommendations for the future of the survey. Some of those recommendations were implemented with the 1996 SIPP Panel in what is known as the *1996 redesign*.

One of the goals of the 1996 redesign was to improve the quality of longitudinal estimates in order to provide better information for policy makers. Specific changes include the following:

- A larger initial sample than in previous panels, with a target of 37,000 households;
- A single 4-year panel instead of overlapping 32-month panels;
- Twelve or 13 waves instead of 8;
- The introduction of computer-assisted interviewing (CAI), which, among other improvements, permits automatic consistency checks of reported data during the interview; those checks can reduce the level of postcollection edits and imputation and thus help to maintain longitudinal consistency; and
- Oversampling of households from areas with high poverty concentrations.

The first interviews of the redesigned SIPP began in April 1996 with the 1996 Panel. Later in 1996, Congress passed the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA). That law significantly altered the nature of public transfer programs, shifting more responsibility to state governments, establishing new eligibility rules for a number of programs, and setting limits on recipiency. The existing welfare program, Aid to Families with Dependent Children (AFDC), was replaced with a new program, Temporary Assistance for Needy Families (TANF). Those changes came after interviewing for the 1996 Panel had already begun with a questionnaire designed for the array of transfer programs that existed before PRWORA was enacted. To accommodate program changes brought about by PRWORA, the Census Bureau began adapting transfer-program questions to reflect the current situation.

Uses of SIPP

SIPP produces national-level estimates for the U.S. resident population and subgroups. Although the SIPP design allows for both longitudinal and cross-sectional data analysis, SIPP is meant primarily to support longitudinal studies. SIPP's longitudinal features allow the analysis of selected dynamic characteristics of the population, such as changes in income, eligibility for and participation in transfer programs, household and family composition, labor force behavior, and other associated events.

One of the most important reasons for conducting SIPP is to gather detailed information on participation in transfer programs. Data from SIPP allow analysts to examine concurrent participation in multiple programs. SIPP data can also be used to address the following types of questions:

- How have changes in eligibility rules or benefit levels affected recipients?
- How have changes in the eligibility rules affected the program target population, that is, those eligible to receive benefits?
- How does income from other household members affect labor force participation and reasons for not working?
- How do wealth and income patterns differ for various age, gender, and racial groups?

Because SIPP is a longitudinal survey, capturing changes in household and family composition over a multiyear period, it can also be used to address the following questions:

- What factors affect change in household and family structure and living arrangements?
- What are the interactions between changes in the structure of households and families and the distribution of income?
- What effects do changes in household composition have on economic status and program eligibility?
- What are the primary determinants of turnover in programs such as Food Stamps?

The Survey

SIPP data show sample members' lives at discrete points in time, as well as a history of changes in their economic circumstances and household relationships. Understanding survey design, content, and procedures is key for analysts wishing to use SIPP data.

Design of SIPP

The adults followed in each SIPP panel come from a nationally representative sample of households in the civilian noninstitutionalized U.S. population. People selected into the SIPP sample are interviewed once every 4 months over the life of the panel. If original sample members 15 years of age or older move from their original addresses to other addresses, they are interviewed at the new addresses. The survey sample includes children residing with original sample members. If, after the first interview, other people not previously in the survey become part of a respondent's household, the new people are interviewed as long as they continue living with respondents from the first interview (Chapter 2).

SIPP Contents

Information collected in SIPP falls into two categories: core and topical. The core content includes questions asked at every interview and covers demographic characteristics; labor force participation; program participation; amounts and types of earned and unearned income received, including transfer payments; noncash benefits from various programs; asset ownership; and private health insurance. Most core data are measured on a monthly basis, although a few core items are measured only as of the interview date, once every 4 months.

Other questions produce in-depth information on specific subjects and are asked less frequently. Those topical questions are often found in *topical modules* that usually follow the core content. Topical questions probe in greater detail about particular social and economic characteristics and

personal histories. Included are such topics as assets and liabilities, school enrollment, marital history, fertility, migration, disability, and work history. Topical module questions typically collect information on events in the past or characteristics that tend to change slowly, if at all.

Data Editing and Imputation

Computer-assisted interviewing (CAI) allows some data editing to occur while the interview is in progress because the system detects inconsistencies and prompts the interviewer to ask the respondent for additional information. CAI also allows use of prior wave data for editing missing data from later waves, thus lessening the need for subsequent longitudinal editing. However, editing and imputation still occur after SIPP interviews are completed (Chapter 4). The Census Bureau edits data for consistency, imputes missing data, and creates internal data files and public use files for each wave.

After each panel is concluded, the Census Bureau creates a full panel file by stripping all edited and imputed values from the core data, linking those data, and then applying a different set of longitudinally consistent edit and imputation procedures to the resulting file. As part of that process, some data are recoded to maintain respondent confidentiality.

The Census Bureau uses several imputation procedures. Most common is some version of a sequential hot deck, in which SIPP statisticians impute missing data by searching for a "donor" respondent who is similar to the respondent with the missing data. The donor's answers are used in the assignment of missing data to the original respondent's record. Specific imputation procedures are discussed in Chapter 4. Data editing is still preferable to imputation and is used whenever a missing item can be logically inferred from other information that has been provided.

Accessing SIPP Information

Most analysts will find the published estimates from SIPP data useful. Census Bureau publications may provide required estimates, saving users the need to generate those estimates themselves. Published estimates can also provide a crosscheck for estimates prepared by analysts from the microdata files.[1]

The Census Bureau makes published estimates from SIPP data available from several sources (Chapter 5). All public use microdata files are available on magnetic media or CD-ROM, along with a full set of documentation, directly from the Census Bureau. The Inter-university Consortium for Political and Social Research (ICPSR) also provides access to SIPP microdata

[1] Prior to the 1996 Panel, the Census Bureau estimates were usually impossible to replicate exactly because they were based on internal data files that had not yet been topcoded and otherwise edited to protect the confidentiality of respondents. Although new topcoding procedures are being implemented with the 1996 and subsequent panels, to facilitate the production of comparable estimates, exact replication of some Census Bureau estimates will still be impossible.

for member institutions. In addition, the SIPP data and documentation that the Census Bureau releases are not copyrighted and thus can be shared, although users are cautioned that this provision applies only to materials written and distributed directly by federal agencies. Finally, analysts conducting exploratory work might wish to investigate the Census Bureau's on-line resources. SIPP microdata are available through two access tools—Surveys-on-Call and FERRET (Chapter 5). The home sites of both online tools can be accessed at the SIPP Web site (http://www.sipp.census.gov/sipp).

Nonsampling Errors, Sampling Errors, and Weighting

The *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a), offers an in-depth discussion of the sources and magnitude of errors in SIPP-based estimates. Although it addresses both sampling and nonsampling errors, it emphasizes the latter. This *Users' Guide* provides a summary chapter addressing nonsampling errors (Chapter 6), a chapter on sampling errors (Chapter 7), and a chapter on the use of weights (Chapter 8). In addition, Appendix C addresses weighting in detail.

Nonsampling Errors

All surveys—including SIPP—are subject to nonsampling errors from various sources. SIPP contains nonsampling errors common to most surveys, as well as errors that stem from SIPP's longitudinal design. Undercoverage in household surveys is due primarily to within-household omissions; the omission of entire households is less frequent. SIPP experiences some differential undercoverage of demographic subgroups; for example, the coverage ratio of black males over 15 years of age is much lower than that for white males in the same age group. To compensate for this differential undercoverage, the Census Bureau adjusts SIPP sample weights to population control totals. Little is known, however, about how effective those adjustments are in reducing biases.

Sample attrition is another major concern in SIPP because of the need to follow the same people over time. Attrition reduces the available sample size. To the extent that those leaving the sample are systematically different from those who remain in the sample, survey estimates could be biased.

Response errors in SIPP take on a number of forms. Recall errors are thought to be the source of the "seam phenomenon." This effect results from the respondent's tendency to project current circumstances back onto each of the 4 prior months that constitute the SIPP reference period. When that happens, any changes in respondent circumstances that occurred during that 4-month period appear to have happened in the first month of the reference period. A disproportionate

number of changes appear to occur between the fourth month of one wave and the first month of the following wave, which is the "seam" between the two waves—hence the name.

Another potential source of response error is the time-in-sample effect. This effect refers to the tendency of sample members to "learn the survey" over time. The more times a sample member is interviewed, the better he or she learns the questionnaire. The concern is that sample members will alter their responses to the survey questions in an effort to conceal sensitive information or to minimize the length of the interview.

Sampling Errors

A common mistake in the estimation of sampling errors for survey estimates is to ignore the complex survey design and treat the sample as a simple random sample (SRS) of the population. This mistake occurs because most standard software packages for data analyses assume simple random sampling for variance estimation. When applied to SIPP estimates, SRS formulas for variances typically underestimate the true variances. Chapter 7 describes how to obtain appropriate variance estimates that take into account SIPP's complex sample design.

Weighting

SIPP data analysts should understand the importance of using weights. The weight for a responding unit in a survey data set is an estimate of the number of units in the target population that the responding unit represents. In general, because population units may be sampled with different selection probabilities, and because response and coverage rates may vary across subpopulations, different responding units represent different numbers of units in the population.[2]

The combined effects of differential response, differential coverage, and differential attrition mean that unweighted analyses can produce biased results. Each SIPP file contains several alternative sets of weights that address the variety of units of analysis (such as persons, households, families, and subfamilies) and time periods for which survey estimates may be needed. It is important to understand the different weights on the files and to use those that are appropriate for a particular analysis.

The selection and use of weights in SIPP analyses are discussed in Chapter 8 and Appendix C.

[2] Most SIPP panels have not sampled different subpopulations at different rates. There are two exceptions: the 1990 and 1996 Panels. Chapter 2 discusses the oversamples included in each of those panels.

SIPP Public Use Files

There are three types of SIPP microdata files available for public use: core wave files, topical module files, and full panel files. Although content overlaps among these files, each is designed to facilitate a different kind of analysis.

Core Wave Files

SIPP core wave files contain the core labor force, income, household and family composition, and program participation data from one wave of interviews. Since the 1990 Panel, these files have been issued in a person-month format, with up to four records for each sample member. Each record contains data from one of the four reference months covered by the wave.[3]

Topical Module Files

Each topical module file contains all of the topical module subject areas that were administered during the wave in question. The files contain one record for each person who was a sample member at the time of the interview. When critical demographic and weight variables are included, the topical module files can be used independently from the core wave and full panel files. However, because topical module files contain only a small subset of the core items, users often need to merge data from either the core wave or the full panel files.

Full Panel Files

Full panel files are released after interviewing for a panel is completed. They contain one record for each original sample member, all children, and all adults who entered the sample after Wave 1. People who were not interviewed for 1 or more months over the course of the panel either have their data imputed or are identified as not in the sample, although their records remain in the file. Variables within each record correspond to the information that was collected in the core content sections of the interviews. Different variables occur with different frequency, depending upon how often certain questions were asked. For example, because a sample member's sex, date of birth, and race are unlikely to change, the variables corresponding to those attributes occur only once in each record. On the other hand, some questions from the core content, such as those about income and program participation, are asked for each month of the panel; the number of corresponding variables will reflect that fact. Similarly, SIPP-generated information can occur once (e.g., person number) or many times (e.g., monthly interview status) on each record.

[3] Prior to the 1990 Panel, core wave files were issued with a single record for each person. Each record contained data for all 4 reference months covered by the wave.

Linking Files

Before linking files, users must understand several conceptual issues: reasons for nonmatches, handling of nonmatches; data quality of matched records containing imputed data; and design of the linked file. There are five ways of linking SIPP data files: within a core wave file; core wave file to core wave file; topical module file to core wave file; topical module file to full panel file; and core wave file to full panel file. The linking process is generally the same for each type of link. However, because variable names and file structures are different, the process for each type of linkage is described in Chapter 13.

Comparison of SIPP with Other Surveys

Because there is some overlap in the content of SIPP and certain other surveys, the question arises: When should an analyst use SIPP instead of the other surveys? A brief look at selected surveys might provide some guidance (Table 1-1 compares some key points as well).

Current Population Survey

The CPS, sponsored jointly by the Census Bureau and the Bureau of Labor Statistics (BLS), is primarily a labor force survey. It is used to compute the federal government's official monthly unemployment statistics, along with other estimates of labor force characteristics. In addition to its core content, a different supplement is fielded each month. One of these, the March Annual Demographic Supplement, is currently the official source of estimates of income and poverty in the United States. Compared with SIPP, however, the CPS has gaps in the area of income measurement. A yearlong reference period means that CPS respondents are more likely than SIPP respondents to forget or misreport certain asset income or irregular income sources. The CPS does not collect data on assets and liabilities to the same extent as SIPP. The CPS is also less comprehensive in the area of program participation, sometimes missing partial-year data.

The CPS reporting unit is the person, but the sample covers housing units; whoever happens to be living at the address at the time of the interview is in the sample. When residents of a CPS housing unit move, they are not followed; instead, the new residents become sample members. Housing units spend 4 months in the sample, 8 months out, and 4 months in again. The target sample size for the CPS is 50,000 housing units each month. Like SIPP, the CPS sample covers the U.S.-resident noninstitutionalized population, although, unlike SIPP, the CPS includes people living in military barracks.

Table 1-1. Comparison of SIPP, CPS, and PSID

Feature	Survey of Income and Program Participation	CPS (March Income Supplement)	Panel Study of Income Dynamics
Sample size and design	1996 Panel: 40,188 households; new panel periodically; each original-sample adult in panel for no. of months in survey; interviews every 4 months	50,000 households; each household in sample for 8 months over 2-year period; rotation group design; monthly interviews (income supplement once per year)	9,000 families; over-represents low-income families; continuing panel with annual interviews
Sample designed to be representative within states?	No	Yes	No
Income data	Data for about 70 cash and in-kind *Source*s at each 4-month wave, with monthly reporting for most *Source*s	Data for prior calendar year for about 35 cash and in-kind *Source*s	Data for prior calendar year for about 25 cash and in-kind *Source*s with specific months received
Tax data	Information to determine federal, state, and local income taxes; payroll taxes; property taxes	None	Information to determine federal, state, and local income taxes; payroll taxes; property taxes
Asset-holdings data	Detailed inventory of real and financial assets and liabilities once each year for panels from 1996 forward and at least once per panel in prior years; more frequent measures for assets relevant for assistance programs	None, except home ownership	Regularly, information about home value and mortgage debt; occasionally, information about saving behavior and wealth
Expenditure data	Information at least once each panel before 1996 and once a year 1996 and beyond on previous month's out-of-pocket medical care costs, shelter costs (mortgage or rent and utilities), dependent care costs, and child support payments	None	Monthly rent or mortgage costs; annual utility costs; average weekly food costs; child support payments

Note: SIPP sample size and design information valid for the 1996 Panel. For information about pre-1996 SIPP panels, see Chapter 2.

Source: Citro, C.F., Michael, R.T., and Maritano, N. (eds.) (1995). *Measuring Poverty: A New Approach.* Washington, DC: National Academy Press, Appendix B.

The Panel Study of Income Dynamics

The Panel Study of Income Dynamics (PSID) was begun in 1968 as a nationally representative, longitudinal survey of the U.S. population. It initially included about 5,000 households and now has about 8,700. The University of Michigan conducts PSID on an annual basis; the focus of the

survey is economics and demographics, especially income sources and amounts, employment family composition changes, and residential location. The content is broad, however, and includes sociological and psychological measures. As of 1995, PSID had collected information from more than 50,000 individuals, spanning as much as 28 years of their lives. The sample includes individuals interviewed every year since 1968, a representative national sample of 2,000 Hispanic households added in 1990, and families formed by members of the original sample families.

Survey of Program Dynamics

The Survey of Program Dynamics (SPD) is a new longitudinal survey designed to be an annual follow-up to the 1992 and 1993 SIPP Panels. Approximately 38,000 households were in the initial sample; a second phase, initiated with the implementation of the core SPD questionnaire in 1998, was projected to include approximately 18,500 households, including all sample households with children and an overrepresentation of households in and near the poverty threshold. SPD data for 1996–2002, along with information collected from 1992 through 1995 for SIPP, will provide a combined 10 years of data measuring program eligibility, access, and participation. Analysts will be able to track welfare dependency, the beginning and end of periods of welfare, factors that may be causes of such periods, and the impacts that the changes will have on families, adults, and children over time.

Guide to This Document

The balance of this *Users' Guide* is organized as follows. Chapters 1 through 5 are introductory chapters, designed mainly for beginning SIPP users.

- Chapter 2 discusses how the SIPP survey is designed and implemented. The chapter describes the structure of the survey, sample selection, and field procedures.
- Chapter 3 examines the general nature of questions in SIPP. Discussion focuses on core and topical content, including brief descriptions of individual topical modules.
- Chapter 4 describes what happens after data collection. This chapter covers all aspects of post-data-collection processing, including consistency checks, data editing, and procedures for imputing missing data.
- Chapter 5 describes SIPP data files and supporting documentation and tells analysts where to find that information.

Chapters 6 through 8 provide more technical information on how to properly use the data and interpret the results.

- Chapter 6 discusses the types and sources of nonsampling error in SIPP, including recall error, the seam effect, time-in-sample effects, attrition bias, and sources of additional information about these topics.
- Chapter 7 defines sampling error and discusses how to calculate sampling errors for SIPP estimates.
- Chapter 8 discusses the topic of weights in SIPP, with a focus on how to choose weights.

Chapters 9 through 13 provide specific instructions for the use of the SIPP public use microdata files.

- Chapter 9 introduces this section by giving an overview of issues common to all of the SIPP data files.
- Chapter 10 describes how to use the core wave files. The chapter describes the structure of the files and how to use the accompanying technical documentation. It also discusses how the core wave files relate to the core survey instrument. Finally, the chapter provides detailed descriptions of how to use the core wave files when performing common tasks.
- Chapter 11 describes how to use the topical module files, the structure of the files, and use of the accompanying technical documentation. It also discusses how the topical module files relate to the corresponding topical module survey instruments. Finally, the chapter provides detailed descriptions of how to use the topical module files when performing common tasks.
- Chapter 12 describes how to use the full panel files, the structure of the files, and use of the accompanying technical documentation. It also discusses how the full panel files relate to the core survey instruments. Finally, the chapter provides detailed descriptions of how to use the full panel files when performing common tasks.
- Chapter 13 describes how to link core wave, topical module, and full panel files. The chapter covers both important conceptual issues and the mechanics of linking the various files.

Finally, the *Users' Guide* includes the following additional information:

- Appendixes contain in-depth discussion of weighting; tables with information about the size and number of waves, missing waves, oversampling, and additional information for selected SIPP panels; a crosswalk; and detailed information about topcoding.
- An acronym list provides a guide to the acronyms used in this manual.
- The glossary defines terms that may be unfamiliar to some users.
- The references section contains references and suggested reading for all chapters in this guide.
- An index helps users locate information quickly and easily.

Where to Go for More Information

The following sources provide expanded, specific information about various aspects of SIPP and related products.

SIPP Web Site

The SIPP homepage (located at **http://www.sipp.census.gov/sipp**) includes, among other things, this *Users' Guide* and an online tutorial that provides a hands-on introduction to SIPP. As the survey and data files evolve, the online documentation will be kept current. Also, users may subscribe at the SIPP Web site to sipp-users, a listserv for SIPP Users Group members. List members share new reports and studies, programming help, and research ideas.

SIPP Quality Profile

The *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a), summarizes what is known about the sources and magnitude of errors in estimates based on SIPP data. It presents information on errors associated with each phase of survey operations: frame design and maintenance, sample selection, data collection, data processing, estimation (weighting), and data dissemination. Some information, such as the outcome of macroevaluation studies, is addressed outside of this framework in a separate chapter. The *SIPP Quality Profile* is available at the SIPP Web site.

Bibliography

The SIPP bibliography, also available at the SIPP Web site under Publications and Analyses, is the most comprehensive, currently available online resource of published and unpublished documents related to SIPP. It includes substantive studies that use SIPP data, as well as citations to methodological research about SIPP. Documents relating to the ISDP also are included. The bibliography contains nearly 2,000 references to reports, conference papers, working papers, journal articles, dissertations, books, and book sections. Abstracts are available for selected publications.

Reports and Working Papers

The references cited in this report include several types of Census Bureau publications. The P-70 series (Current Population Reports, Household Economic Studies) presents tabulations and

analyses of SIPP data. SIPP working papers provide information about methodological aspects of the survey as well as analyses of SIPP data. The working papers are not cleared for formal publication but are readily available at the SIPP Web site. Since 1984, papers on SIPP results and methodology presented at the annual meeting of the American Statistical Association have been published in the working-paper series. Several important papers on SIPP methodology and evaluation studies have been presented and published in the proceedings of the Census Bureau's annual research conferences, which began in 1985. In addition to those sources, papers and reports with information about the quality of SIPP data have been published by numerous other agencies, organizations, and professional associations.

Technical Documentation

Technical documentation accompanies the SIPP microdata files that users acquire from the U.S. Census Bureau. The technical documentation briefly describes the contents of the particular file and includes the following items:

- A glossary of selected terms,
- Lists of codes and descriptions,
- A data dictionary and instructions on how to use it,
- A source and accuracy statement,
- A copy of the core questionnaire used for the panel in question,
- User notes, and
- File information.

2. SIPP Sample Design and Interview Procedures

This chapter provides new users of the Survey of Income and Program Participation (SIPP) with basic information about the organizing principles of SIPP, sample selection, and the data collection process. The chapter also briefly reviews interview procedures.

SIPP is a longitudinal survey that collects information on topics such as income, participation in government transfer programs, employment, and health insurance coverage. The initial survey design called for the introduction of a new sample, called a *panel*, every year; each panel was planned to cover 32 months. In practice, a number of panels have been shorter. A result of the initial design was that multiple SIPP panels were in the field simultaneously. A redesign introduced with the 1996 Panel abandoned the overlapping panel structure and extended the length of the 1996 Panel to 4 years. Subsequent panels will be 3 years in length.

Organizing Principles

SIPP is administered in *panels* and conducted in waves and *rotation groups*. Within a SIPP panel, the entire sample is interviewed at 4-month intervals. These groups of interviews are called *waves*. The first time an interviewer contacts a household, for example, is Wave 1; the second time is Wave 2, and so forth. As discussed in Chapter 3, each wave contains *core* questions that are asked each time, along with *topical* questions that vary from one wave to the next.

Sample members within each panel are divided into four subsamples of roughly equal size; each subsample is referred to as a *rotation group*. One rotation group is interviewed each month.[1] During the interview, information is collected about the previous 4 months, which are referred to as *reference months*. Thus, each sample member is interviewed every 4 months, with information about the previous 4-month period collected in each interview (see Table 2-2).

Panels

The original design of SIPP called for an initial selection of a nationally representative sample of households, with all adults in those households being interviewed once every 4 months over a 32-month period. In addition, interviews were to be conducted with any other adults living with original sample members at subsequent waves. The first sample, the 1984 Panel, began

[1] The month in which the interview takes place is called the interview month.

interviews in October 1983. The 1985 Panel began in February 1985. Subsequent panels began in February of each calendar year, resulting in concurrent administration of the survey in multiple panels. Because of budget constraints, actual panel duration has varied. The original goal was to have panels covering eight waves (32 months). In several instances, panels were terminated after seven waves (28 months). Two panels were terminated even earlier: 1988 (six waves) and 1989 (three waves).

With certain exceptions (Table 2-1), each panel overlapped part of the previous panel, with the result that there were two or three active panels at any given time. The overlap allows analysts to combine records from different panels, thus having larger samples (and lower standard errors) for cross-sectional analyses.[2] The overlapping feature of the SIPP design was dropped with the 1996 redesign. Standard errors have remained small since the redesign because the 1996 and following panels each have target sample sizes of at least 37,000 interviewed households for Wave 1, almost twice the size of two of the previous panels.

Table 2-1. Summary of the 1984–1996 SIPP Panels

Panel[a]	Date of First Interview	Date of Last Interview	Number of Wave 1 Eligible Households	Number of Wave 1 Original Sample Members	Number of Waves	Short Waves[b]
1984	Oct. 83	Jul. 86	20,897	55,400	9	2, 8
1985	Feb. 85	Aug. 87	14,306	37,800	8	2
1986	Feb. 86	Apr. 88	12,425	32,800	7	3
1987	Feb. 87	May 89	12,527	33,100	7	-
1988	Feb. 88	Jan. 90	12,725	33,500	6	
1989	Feb. 89	Jan. 90	12,867	33,800	3	
1990	Feb. 90	Sep. 92	23,627	61,900	8	
1991	Feb. 91	Sep. 93	15,626	40,800	8	
1992	Feb. 92	May 95	21,577	56,300	10	-
1993	Feb. 93	Jan. 96	21,823	56,800	9	
1996	Apr. 96	Mar. 00	40,188	95,402	13	

[a] No new panels in 1994 and 1995.

[b] Short waves contained three rotations instead of the standard four.

Source: SIPP Quality Profile, 3rd Ed. (U.S. Census Bureau, 1998a).

Although most available data predate the 1996 redesign (discussed in Chapter 1), the redesign affected the nature of some panels. In preparation for the redesign, the Census Bureau canceled the 1994 and 1995 Panels and extended the 1992 and 1993 Panels (Table 2-1). The last 1993 Panel interview took place in January 1996 to ensure that data would remain continuous. Also in 1996, the Census Bureau initiated the Survey of Program Dynamics (SPD) as an extension of SIPP. For the SPD, the Census Bureau began recontacting people in the 1992 and 1993 SIPP panels and will continue annual data collection through 2002. The plan is to yield 10 years of

[2] Combining data across panels allows for larger sample sizes and, consequently, smaller standard errors for some types of estimates. It also helps alleviate two types of bias common to longitudinal surveys: time-in-sample effects and attrition bias.

data (1992–2001) for those two panels to support analyses of changes during welfare reform and for the pre- and postreform periods (Chapter 1).

Waves and Rotation Groups

One full 4-month cycle of administering the questionnaire to the entire panel is a *wave*. The 1984 through 1993 Panels were designed to have eight waves each, although more often than not the number of waves actually administered was different (Table 2-1). The 1996 Panel has 12 waves.

Rotation groups are random subsamples of approximately equal size. Each month, the members of one rotation group are interviewed; over the course of 4 months, all rotation groups are interviewed, providing data for the full set of 4 months. For many survey items, SIPP collects data for each of the 4 calendar months preceding the interview month. Those 4 months together are called reference months, or the *reference period*. (Table 2-2 provides an illustration of the reference months for the various rotation groups in each wave of the 1996 Panel.)

The reference period length and the timing of the interviews address several concerns: respondent recall error, which increases as the recall period lengthens; respondent burden, which increases with the number of times they are interviewed; and the costs of frequent interviews. By spreading the interviews for each wave evenly over 4 months, the rotation group structure allows the Census Bureau to keep a skilled and experienced team of interviewers in the field year round. This eases management burden and allows Census Bureau interviewers to master the complexities of the SIPP questionnaire and to maintain that mastery.

Each SIPP panel prior to 1990 had fewer than eight waves or contained one wave that consisted of fewer than four rotation groups (Table 2-1). As discussed in Chapter 3, the questionnaire administered at each wave contains *core questions*, those asked at every interview, along with sections containing *topical questions* that vary from one wave to the next. Respondents in the skipped rotation groups have no gap in core data, but they do not provide core data for the full duration of the panel, and they lack topical data for the wave in which they were skipped. Analysts should be alert to the consequences of the skipped rotations: some topical information is not available for the full sample, and the length of time an analyst can follow adults from the original sample is reduced for selected rotation groups.

Reference Periods

The reference period for most core items is the 4-month period preceding the month of the interview for the given wave. Data for most core items are collected for each of the preceding 4 months. Some data on labor force characteristics are collected with weekly resolution. Subsequently, weekly labor force characteristics are recorded on a monthly basis.

Table 2-2. 1996 Panel: Rotation Groups, Waves (W), and Reference Months

Reference Month	Rotation Group 1	2	3	4
Dec. 95	W1 1			
Jan. 96	W1 2	W1 1		
Feb. 96	W1 3	W1 2	W1 1	
Mar. 96	**W1 4**	W1 3	W1 2	W1 1
April 96	W2 1	**W1 4**	W1 3	W1 2
May 96	W2 2	W2 1	**W1 4**	W1 3
June 96	W2 3	W2 2	W2 1	**W1 4**
July 96	**W2 4**	W2 3	W2 2	W2 1
Aug. 96	W3 1	**W2 4**	W2 3	W2 2
Sep. 96	W3 2	W3 1	**W2 4**	W2 3
Oct. 96	W3 3	W3 2	W3 1	**W2 4**
Nov. 96	**W3 4**	W3 3	W3 2	W3 1
Dec. 96	W4 1	**W3 4**	W3 3	W3 2
Jan. 97	W4 2	W4 1	**W3 4**	W3 3
Feb. 97	W4 3	W4 2	W4 1	**W3 4**
Mar. 97	**W4 4**	W4 3	W4 2	W4 1
April 97	W5 1	**W4 4**	W4 3	W4 2
May 97	W5 2	W5 1	**W4 4**	W4 3
June 97	W5 3	W5 2	W5 1	**W4 4**
July 97	**W5 4**	W5 3	W5 2	W5 1
Aug. 97	W6 1	**W5 4**	W5 3	W5 2
Sep. 97	W6 2	W6 1	**W5 4**	W5 3
Oct. 97	W6 3	W6 2	W6 1	**W5 4**
Nov. 97	**W6 4**	W6 3	W6 2	W6 1
Dec. 97		**W6 4**	W6 3	W6 2
Jan. 98			**W6 4**	W6 3
Feb. 98				**W6 4**

Reference Month	Rotation Group 1	2	3	4
Dec. 97	W7 1	*See Wave 6 data in bottom of first column.*		
Jan. 98	W7 2	W7 1		
Feb. 98	W7 3	W7 2	W7 1	
Mar. 98	**W7 4**	W7 3	W7 2	W7 1
April 98	W8 1	**W7 4**	W7 3	W7 2
May 98	W8 2	W8 1	**W7 4**	W7 3
June 98	W8 3	W8 2	W8 1	**W7 4**
July 98	**W8 4**	W8 3	W8 2	W8 1
Aug. 98	W9 1	**W8 4**	W8 3	W8 2
Sep. 98	W9 2	W9 1	**W8 4**	W8 3
Oct. 98	W9 3	W9 2	W9 1	**W8 4**
Nov. 98	**W9 4**	W9 3	W9 2	W9 1
Dec. 98	W10 1	**W9 4**	W9 3	W9 2
Jan. 99	W10 2	W10 1	**W9 4**	W9 3
Feb. 99	W10 3	W10 2	W10 1	**W9 4**
Mar. 99	**W10 4**	W10 3	W10 2	W10 1
April 99	W11 1	**W10 4**	W10 3	W10 2
May 99	W11 2	W11 1	**W10 4**	W10 3
June 99	W11 3	W11 2	W11 1	**W10 4**
July 99	**W11 4**	W11 3	W11 2	W11 1
Aug. 99	W12 1	**W11 4**	W11 3	W11 2
Sep. 99	W12 2	W12 1	**W11 4**	W11 3
Oct. 99	W12 3	W12 2	W12 1	**W11 4**
Nov. 99	**W12 4**	W12 3	W12 2	W12 1
Dec. 99		**W12 4**	W12 3	W12 2
Jan. 00			**W12 4**	W12 3
Feb. 00				**W12 4**

Note: The cell entry W1 1 represents Wave 1, reference month 1. The last reference month of each wave is in boldface type. For rotation group 1, the reference months for Wave 1 were Dec. 95 through Mar. 96.

After the basic demographic information, one of the first items in the SIPP interview illustrates the availability of time-specific data in SIPP. The respondent is asked if he or she had a health insurance plan at any time during the previous 4 months. If the answer is yes, SIPP asks if the

respondent had coverage in each of the individual 4 months. Thus data are collected for 4 individual months at each wave. Over the course of a 13-wave panel, data are collected for 52 consecutive months for each panel member. For the 1996 Panel, the rotation groups were interviewed in order. Specifically, for Wave 1, rotation group 1 was interviewed in April, rotation group 2 in May, rotation group 3 in June, and rotation group 4 in July. For previous panels, however, the specific months varied slightly among rotation groups. With the 1990 Panel, for instance, panel members in rotation group 2 were interviewed first; rotation group 1 was actually the fourth rotation group surveyed in that panel.[3]

Sample Design

SIPP uses a complex sample design that has important implications for the estimation of standard errors. Because the SIPP design is not a simple random sample, the standard errors reported by most off-the-shelf statistical software will underestimate the true standard errors of estimates from SIPP. (See Chapter 7 for details.) A detailed description of the SIPP sample design and standard error calculations can be found in the third edition of the *SIPP Quality Profile* (U.S. Census Bureau, 1998a).

Selection of Sampling Units

The Census Bureau employs a two-stage sample design to select the SIPP sample. The two stages are (1) selection of primary sampling units (PSUs) and (2) selection of address units within sample PSUs. Census Bureau interviewers follow an established procedure to identify sample members within the selected address units.

Primary Sampling Units

The frame for the selection of sample PSUs consists of a listing of U.S. counties and independent cities, along with population counts and other data for those units from the most recent census of population. Counties either are grouped with adjacent counties to form PSUs or constitute a PSU by themselves.

Following the formation of the PSUs, the smaller ones, called non-self-representing (NSR) PSUs, are then grouped with similar PSUs in the same region (South, Northeast, Midwest, West) to form strata; census data for a variety of demographic and socioeconomic variables are used to determine the optimum groupings. A sample of NSR PSUs is selected in each stratum to represent all PSUs in the stratum. All of the larger PSUs are included in the sample and are called self-representing (SR) PSUs.

[3] An explanation for the relabeling of rotation groups in earlier panels is provided in Chapter 2 of the 2nd edition of the *SIPP Users' Guide* (U.S. Census Bureau, 1991).

Selection of Addresses in Sample PSUs

SIPP selects addresses from five separate, non-overlapping sampling frames maintained by the Census Bureau. They are unit (formerly called the address enumeration districts [Eds] frame); area (area EDs frame); group quarters (special places frame); housing unit coverage; a coverage improvement frame, and a new-construction (or permit) frame. The first three frames are based on census counts from the most recent decennial census; unit and area frames are determined by a process called "address screening," which has been done at the block level since 1990. The unit frame lists addresses of housing units located in census blocks in areas that issue building permits and in which at least 96 percent of the addresses are complete (with street name and house number). The area frame contains addresses from the remaining census blocks that are not in permit-issuing areas, or where more than 4 percent of the addresses in the blocks are missing. Those addresses are mostly in rural areas. The group quarters frame includes boarding houses, hotel rooms, and institutions that are found in the decennial census but are not counted as housing units. Together, the three frames provide almost 90 percent of the sample addresses for each SIPP panel.

The coverage improvement frame is used to include addresses of housing units that were missed in the census count but were found in postenumeration surveys. The percentage of sample addresses from this frame is typically small (0.1 percent of the sample addresses in the 1986 Panel).

The new-construction frame is used to provide coverage of new structures for which building permits have been issued since the last decennial census in areas covered by the unit frame. This frame is updated continually, and the percentage of addresses sampled from it increases each year until data from another decennial census become available.

Within each sample PSU, the addresses in the sampling frames are grouped into clusters. The clusters are then sampled, and the selected cluster of addresses is included for interviewing.[4] In the unit frame, the 1996 Panel had clusters of one housing unit; for prior panels, clusters of two neighboring addresses were used. In the area and group quarter frames, clusters are constructed with the expectation of four housing units or housing unit equivalents. With the area frame, the sampled clusters are visited by SIPP interviewers prior to the scheduled interviewing. The interviewers list all residential addresses within the selected clusters. With the new-construction frame, the 1996 Panel has a 50-50 mixture of four- and eight-unit clusters. Previously, clusters of four housing units were formed. No clustering is used with the coverage improvement frame.

Identifying Household Members Within Sampled Addresses

At the time of the first interview, the Census Bureau interviewer visits sampled addresses, verifies the addresses, determines whether they contain occupied housing units, and identifies the housing units located at each address. A housing unit is defined as a living quarters with its own entrance and cooking facilities. The people living in a housing unit constitute a household (see below). Interviews are conducted at all households in sampled addresses. However, SIPP does

[4] In a few cases, where the clusters contain many more housing units than expected, a subsample of addresses is selected.

not treat the household as a continuous unit to be followed in the panel. SIPP is a person-based survey; as discussed below, SIPP follows original sample members regardless of household composition.

The interviewer compiles a roster for each sampled household, listing all people living or staying at the address. Next, the interviewer identifies those who are household members by determining if the address is their usual residence (Table 2-3).[5] SIPP designates all people who are considered members as original sample members. Over the course of the panel, original sample members are followed and interviewed every 4 months.[6]

Table 2-3. Household Membership

Question	YES (Is Member of Household)	NO (Not Member of Household)
Person staying at SIPP address at time of interview		
Members of family, visitors, etc.—ordinarily sleeps here	Y	
– here temporarily, no living quarters held elsewhere	Y	
– here temporarily, living quarters held elsewhere		N
In Armed Forces, stationed locally and sleeps here	Y	
In Armed Forces, stationed elsewhere and here on leave		N
Student temporarily attending school here, living quarters held elsewhere		N
– married and accompanied by own family	Y	
– student nurse attending school nearby	Y	
Absent person who usually lives at SIPP address		
Inmate in an institutional special place regardless of whether living quarters are being held here		N
Temporarily on vacation, in hospital, and living quarters held	Y	
Absent for work, living quarters held here	Y	
Absent for work, living quarters held here and elsewhere but comes here infrequently		N
Unmarried college student working away from home during break, living quarters held here	Y	
In Armed Forces, stationed elsewhere	Y	
In school elsewhere, living quarters held—not married or with own family	Y	
– married and accompanied by own family		N
– attending school overseas		N
– student nurse living at school		N
Exceptions and doubtful cases		
Person with two residences, sleeps most often in other location		N
Person with two concurrent residences, sleeps here most often	Y	
Citizen of foreign country temporarily in U.S., living on premises of an embassy, ministry, legation, chancellery, or consulate		N
Citizen of foreign country temporarily in U.S.—studying here and no other usual residence in U.S.	Y	
– living and working here and no other usual residence in U.S.	Y	
– visiting or traveling in U.S.		N

Source: SIPP Information Booklet, 1990 Panel (Waves 1–8) and 1991 Panel (Waves 1–8), Form SIPP-7004A (1-9-89).

[5] In most cases, a person is a member of a household if the sample unit is that person's usual place of residence at the time of the interview. The person may be present or temporarily absent. A person staying in the sample unit who has no usual place of residence elsewhere is a household member. A usual place of residence is the place where a person normally lives and sleeps. This must be specific living quarters held for the person to which he or she is free to return at any time.

[6] In the 1993 Panel only, SIPP followed all original sample members regardless of age. Previous panels, as well as the 1996 Panel, have followed only people 15 years of age or older who were original sample members.

Oversampling

Originally, SIPP did not oversample any groups within the population. Over the years, however, budget constraints dictated a reduction in the SIPP panel size. As a result, analysts found it difficult to conduct meaningful analyses of government programs for the low-income population because the sample sizes for the subpopulations were too small. In response to those concerns about the diminished usefulness of SIPP data, the Census Bureau pursued budget initiatives to increase the sample to its original size and to oversample the low-income population.

Oversampling occurs when certain groups or units are sampled with higher probabilities than others. Analysts then have enough cases to complete analysis of subpopulations or subgroups of the population. The share of an oversampled group in the resulting sample is greater than its share in the population from which it was drawn. Although this imbalance addresses the need for increased sample sizes for certain subpopulations, analysts looking at the entire sample will need to use weights in their analyses to redress the imbalance (Chapter 8).[7]

Oversampling in the 1990 Panel

As detailed in the *SIPP Quality Profile* and discussed in Allen et al. (1993), oversampling was used with the 1990 Panel, which included about 3,900 predominantly low-income households from the truncated 1989 Panel (see Tables 2-1 and 2-4). In the 1990 Panel, the Census Bureau included all housing units from Wave 1 of the 1989 Panel in which the head of household was black, Hispanic, or female with no spouse present living with relatives (FHNSP). Such households tend to have higher poverty rates than the general population. The 1990 Panel also included a small sample of other housing units for the 1989 Panel. Table 2-4 shows the components of the 1990 Panel.

Table 2-4. Composition of the 1990 Panel

Components	Number of Eligible Households
Households in addresses originally to be interviewed first in the 1990 Panel	19,700
Households associated with sample addresses first interviewed in February through May 1989 (in the 1989 Panel) and at the time headed by a black, Hispanic, or FHNSP[a]	2,700
Households in one-ninth of all other 1989 Panel sample addresses	1,200

[a] Female head of household with no spouse present living with relatives.

Source: Allen, Petroni, Singh, 1993.

Oversampling in the 1996 Panel

The Census Bureau also oversampled the low-income population for the 1996 Panel,[8] using 1990 decennial census information. Housing units within each PSU were split into high- and low-

[7] Weights are needed even if there is no oversampling. See Chapter 8.

[8] For a more detailed discussion of the 1996 oversample design, see Huggins and King (1997).

poverty strata. If the housing unit received the Census long form that included income questions, the unit's poverty status was determined directly; for other housing units, poverty status was assumed on the basis of responses to Census short-form items predictive of poverty rates. The Census Bureau then sampled the low-income stratum at 1.66 times the rate of the high-income stratum in each PSU. Compared with the number of cases produced without oversampling, this oversampling produced an 18 percent increase in the number of cases in and near poverty at Wave 1.[9] Even greater gains occurred in some subgroups, such as blacks and Hispanics in poverty, with a gain in the number of sample cases as high as 24 percent. However, the increases in effective sample sizes were somewhat smaller after allowance was made for the increased variance associated with differential weighting. Also, the sample sizes for the higher income and higher age groups were reduced.

Following Rules

SIPP is a true longitudinal survey that tracks people over time. With few exceptions, original sample members are interviewed every 4 months over the duration of the panel. When original sample members move to new addresses, interviewers attempt to locate them and continue to interview them every 4 months.

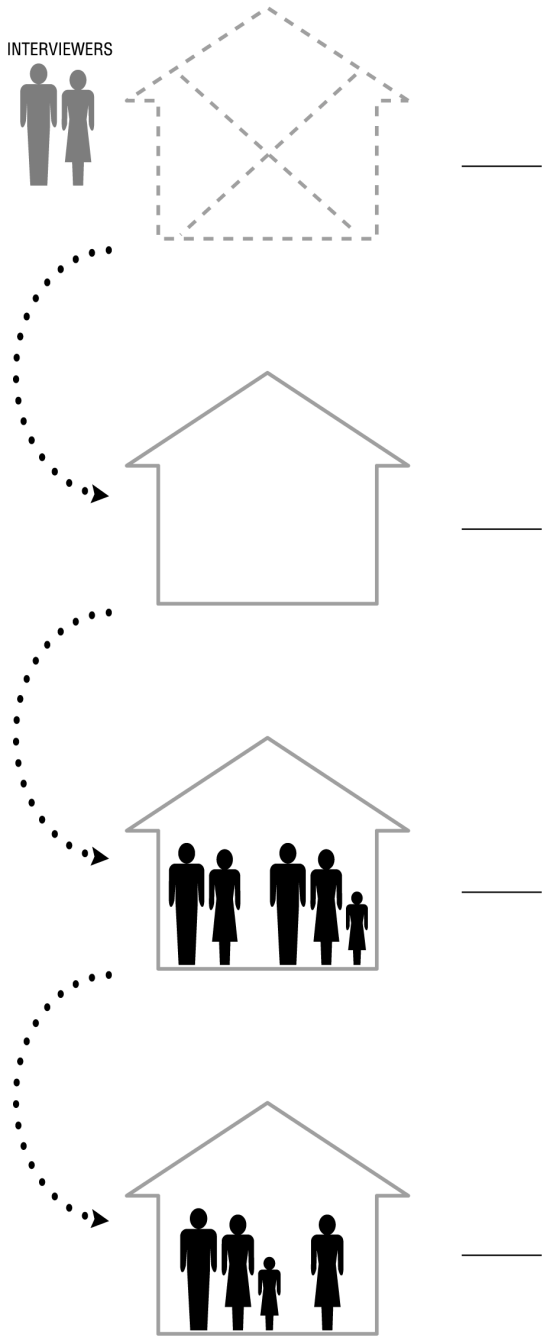
The SIPP rules call for following original sample members who move, provided they are not institutionalized, do not live in military barracks, or do not move abroad. Prior to the 1993 Panel, and resuming with the 1996 Panel, original sample members under age 15 who moved were not followed. Thus, data were collected for them in subsequent waves only if they either continued to live with an original sample member 15 years or older or were age 15 by the last day of the reference period in which they moved. With Wave 4 of the 1993 Panel, SIPP began following all children who were in original sampled households (*SIPP Quality Profile*, 1998, pp. 3–6), including babies born to sample members during the panel.

When original sample members move into households with other individuals not previously in the survey, the new individuals become part of the SIPP sample for as long as they continue to live with an original sample member. Similarly, when new individuals move in with original sample members after the first interview, they too become part of the SIPP sample for as long as they continue to live with an original sample member. If no original sample members live at an address where a previous interview was conducted, SIPP does not collect information from the new occupants of that address.

Figure 2-1 illustrates the following rules in practice.

[9] Low-income strata were sampled at a rate of 0.00062389. High-income strata were sampled at a rate of 0.00037489. The oversampling rate therefore comes to 1.6642.

Figure 2-1. Following Rules



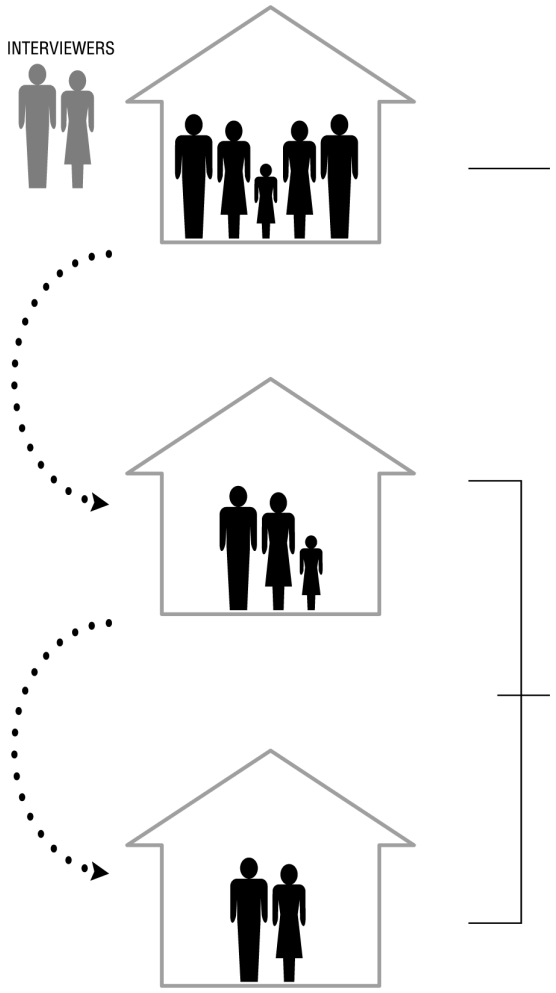
Demolished address unit – no interview.

Vacant address unit – no interview.

Five people (mom, dad, son, daughter, and cousin) reside at this address and thus constitute a household. **Wave 1** interview conducted for all five people.

Son joined Army and is living in barracks. He is not followed because military bases are outside the scope of the SIPP sample. However, a record exists in the **Wave 2** interview reflecting proxy responses by another member of the household. Interviewer takes data on the four people who remain at this address.

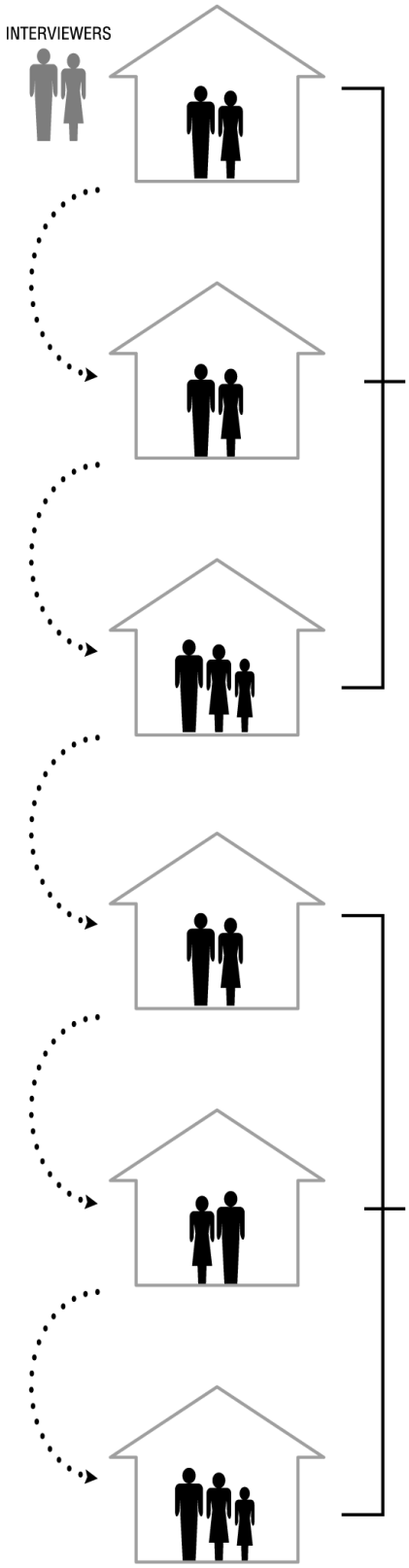
Figure 2-1. Following Rules *(continued)*



Daughter got married; she and husband live with her parents and cousin at time of **Wave 3** interview. The husband is interviewed at the same time that others in the house are interviewed. There is no further information taken on the son (who joined the Army and is living in barracks, which is outside the SIPP universe).

Daughter and her husband moved to a new address and formed their own household at the time of **Wave 4**. The interviewer takes data on mom, dad, and cousin in the first household; and daughter and daughter's husband in the second household.

Figure 2-1. Following Rules *(continued)*

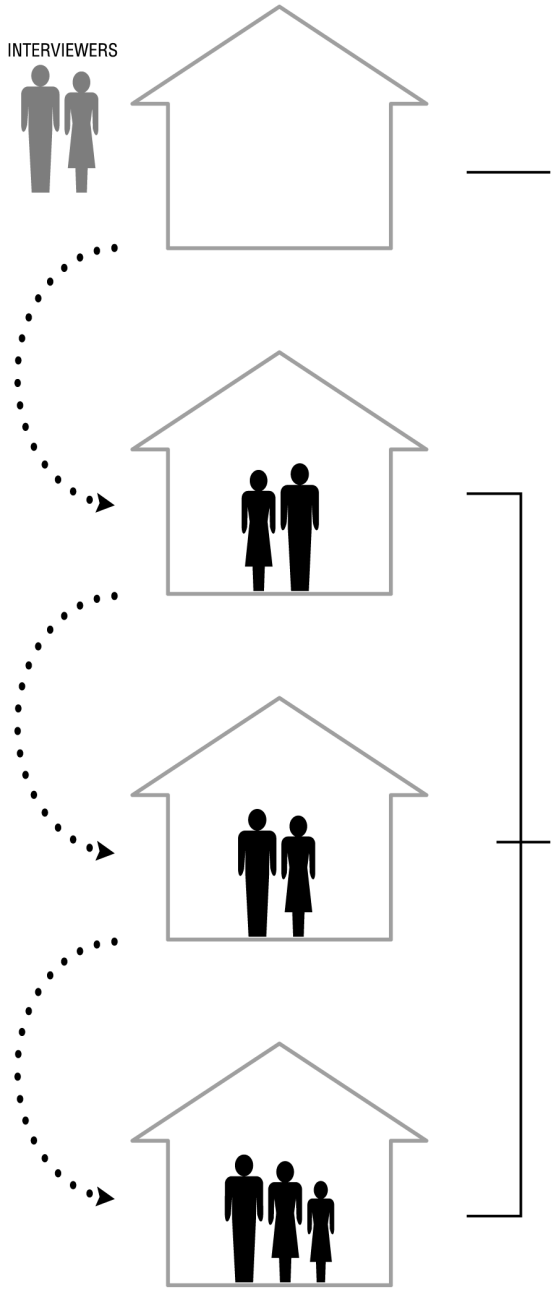


The cousin, who is over 15[a], moved and now lives with her mother and father, who were not in the sample originally. Therefore, for this **Wave 5** interview, the interviewer takes data from seven people (mom and dad in the first household, daughter and daughter's husband in the second household, and cousin, cousin's mother, and cousin's father) in the third household.

In **Wave 6**, there is no change from the previous wave.

[a] For Waves 4+ of the 1993 Panel **only**, SIPP followed original sample persons under 15 years old who moved to other households with or without another original SIPP panel member over 15. In all other panel years, SIPP did not follow original sample persons under 15 years old who moved to other households with or without another original SIPP panel member over 15. In this example, therefore, the cousin is followed because she is over 15. In the 1993 Panel, the cousin would have been followed without regard to age.

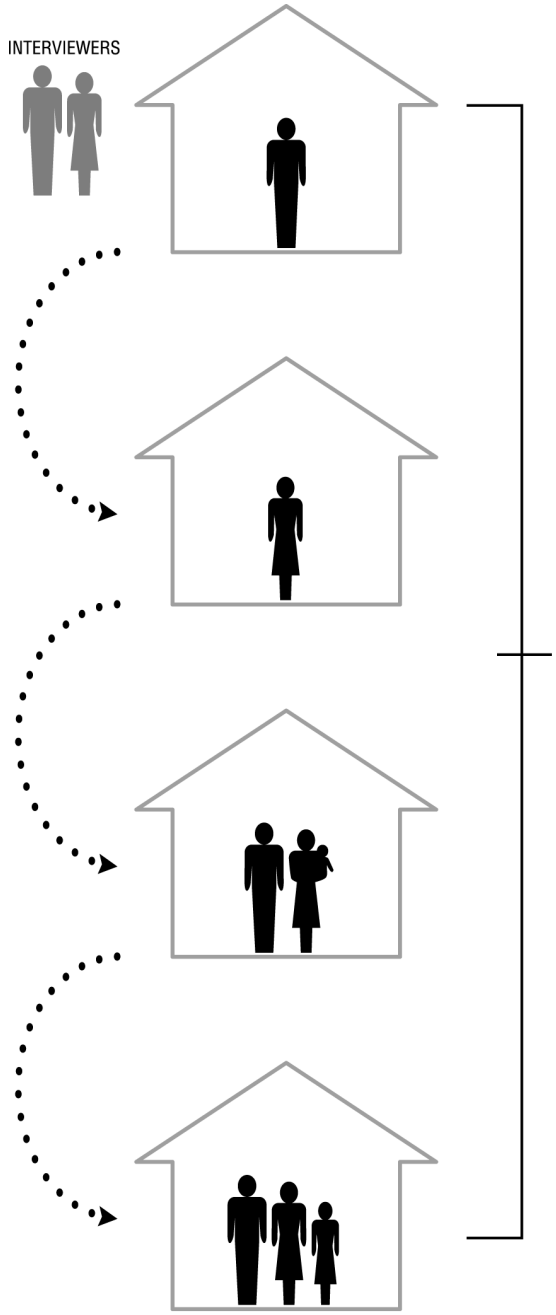
Figure 2-1. Following Rules *(continued)*



At the time of **Wave 7**, the interviewer discovers that mom and dad have moved out of their old home.

The interviewer locates mom and dad and interviews them at their new address. The daughter and her husband are interviewed at their previous address, as are the cousin and the cousin's parents. Altogether, the interviewer takes data from seven people (mom, dad, daughter, daughter's husband, cousin, cousin's mother, and cousin's father) in three households.

Figure 2-1. Following Rules *(continued)*



Mom and dad have separated at the time of **Wave 8.** Mom is in the same address as in the previous wave, but dad is in a new location; thus they form separate households. Meanwhile, the daughter and husband now have a baby and the cousin's household has remained the same. The interviewer takes data for eight people (mom, dad, daughter, daughter's husband, daughter's baby, cousin, cousin's mother, and cousin's father) in four households.

Interviewers rely on several sources of information to locate movers. At the first interview, the interviewer obtains the name, address, and telephone number of a person who could furnish the new address should the entire household move. If necessary, interviewers may contact neighbors, employers, mail carriers, real estate companies, rental agents, or postal supervisors to locate original sample members who have moved.

If an entire household moves, the interviewer tries to find the original sample members and interview them at their new address(es) if they remain in the locality. If the household relocates into or close to a different PSU, a SIPP interviewer in that area may interview them. For example, if a couple moves from Boston to Seattle, a SIPP interviewer in the Seattle area will likely interview the couple for the remaining waves of their panel. Should the entire household move more than 100 miles away from a SIPP PSU, attempts will be made to interview by telephone. If the household cannot be reached, the sample members will be dropped from the survey. Specifically, they will be treated as Type D noninterviews (Type D noninterviews are discussed later in the chapter).

If only some original sample members move, the interviewer completes interviews with all eligible household members at both the original address and the address(es) of those who have moved. If an original sample member leaves a SIPP household and the remaining original sample members cannot provide a new address, the interviewer will try to find the person through the means discussed above. Similar to what happens with a household, if an individual original sample member moves within the United States but more than 100 miles away from a SIPP PSU, a telephone interview will be attempted. When that is not possible, the person is treated as a Type D noninterview.

SIPP does not interview original sample members if they move outside the United States, become members of the military living in barracks, or become institutionalized (e.g., nursing home residents, prison inmates). The Census Bureau attempts to track such individuals, however. Should they return to the noninstitutionalized resident U.S. population, the Census Bureau will resume trying to interview them.[10]

Difference Between Movers and Those Who Are Temporarily Away

There is an important difference between a mover and a person who is temporarily away. A mover no longer lives at the sample address. On the other hand, a person is temporarily away if the household is that person's usual place of residence, according to the membership rules given in Table 2-3, and specific living quarters are held for the person to which he or she is free to return at any time. The following two examples may help to illustrate the distinction:

[10] A member of the armed forces who lives in a barracks is not eligible for an interview; a member of the armed forces who lives elsewhere is eligible.

- A college student living on campus with a room held at home is still a household member at the sample address. In this case, the interviewer would try to interview that student or obtain a proxy interview with the household reference person. If the hypothetical college student originally lived in New York and, upon graduation, moved to Los Angeles to live on his or her own, the student would be considered to have moved as of the graduation date. The student's new address in Los Angeles would become his or her new household, and, if the student was an original sample member, he or she would be treated in the same way as any other original sample member who moved to the new address.
- If a household member is in the hospital following an operation but is expected to come home, that person is still a household member at the original address. If an individual interview is not feasible, the interviewer might do a proxy interview for that person. If, however, the person moved into a nursing home, he or she would not be eligible for a SIPP interview, whether individual or proxy. At each interview, the interviewer asks the status of any primary sample member who entered an institution between Wave 1 and the current wave. If the interviewer learns that the person has returned to the noninstitutionalized population, an interview is attempted.

Interview Procedures

At Wave 1, interviews are attempted for all members of selected housing units who are 15 years of age or older.[11] The Census Bureau prefers that all SIPP sample members 15 years of age or older who are present at the time of the interview answer for themselves unless they are physically or mentally unable to do so. For those who are absent or incapable of responding, SIPP will accept a proxy interview, usually with another household respondent.

After Wave 1, the interviewer compiles (or updates) a separate household roster for each housing unit, listing all people living or staying at the unit, including anyone who may have joined the household, such as a new spouse or baby, and the dates they entered the household. The interviewer then decides whether each person is a household member by using rules that determine whether the person is a usual resident of the unit (Table 2-3).

Key to SIPP data collection is identification of a *reference person* for the household, an owner or renter of record. The interviewer lists other people in the household according to their relationship to the reference person.

Also noted are people who left the household and their dates of departure. If some—but not all—sample members have moved since the last interview, the interviewer completes interviews at the original address and also obtains the new address(es) of the individuals who moved. For those remaining at the same address, the interviewer verifies that certain previously collected information still applies, completes the questionnaire for each person 15 years of age or older,

[11] Detailed information about interview procedures is available from the Census Bureau in the SIPP interviewer's instruction manual (U.S. Census Bureau, 1993).

and collects certain information for children under age 15. Information is also collected for all new household members. Movers are interviewed at their new addresses, along with other household members they are living or staying with at the time.

Most interviews conducted through 1991 were in the form of personal visits. In 1992, SIPP switched to maximum telephone interviewing to reduce costs. Wave 1, 2, and 6 interviews were still conducted in person, but other interviews were conducted by telephone to the extent possible. SIPP telephone interviews and personal visits are carried out by the same interviewer interacting with the same respondents. Interviewers typically make phone calls from their homes. For security and confidentiality reasons, they are not allowed to use cellular or cordless telephones in the interviews. If a standard telephone is not available, the interviews must be conducted face-to-face. Repeated failure to reach a respondent by telephone may also require an in-person visit to the listed address.

When respondents are not able to furnish all requested information at the interview, interviewers arrange to get the answers by telephone if the respondents are willing. Callbacks can also help correct inconsistencies found during questionnaire editing. With the 1996 redesign, computer-assisted interviewing (CAI) was begun. Thus, automatic consistency checks for selected data occur during the interview. (For more on editing and imputation, see Chapter 4.)

The 1996 redesign included a change in the method of data collection. Prior to 1996, interviewers used a paper questionnaire. Starting in 1996, however, interviewers began conducting interviews with a laptop computer. Both the paper survey and the CAI instrument have skip patterns that help the interviewer avoid asking irrelevant questions (see Chapter 3 for more on skip patterns). In the paper survey, interviewers would encounter points at which they had to look at previously given answers before deciding whether or not to ask certain questions. With CAI, the instrument skips directly to the next applicable question.

Nonresponse

All surveys experience some degree of nonresponse. As discussed in Chapter 6, in a longitudinal survey such as SIPP, as the number of waves increases, nonresponse may result in a corresponding increase in bias. Since nonrespondents may differ from respondents in terms of the variables collected in the survey, the occurrence of nonresponse gives rise to concerns about bias in the survey results. Weighting adjustments are made in an attempt to reduce or eliminate bias (Chapter 8), but concerns about nonresponse bias remain.

The rate of sample loss[12] in SIPP generally declines from one wave to the next. The total number of sample members lost, also known as total sample attrition, always increases over time. Wave 1 nonresponse rates for SIPP have been about 7.7 percent.[13] There is usually a sizable

[12] The accumulation of cases that are no longer being interviewed because of as yet unrecovered refusals or as yet unfound movers.

[13] Nonresponse rates have not been stable, ranging from 6.70 percent for the 1984 through 1990 Panels to 8.48 percent for the 1991 through 1996 Panels.

sample loss at Wave 2, with a lower rate of additional attrition occurring at each subsequent wave. Prior to the 1992 Panel, SIPP lost roughly 20 percent of the original sample by the panel's completion. The sample loss rate for the 1996 Panel was 35.5 percent by the end of the 12th, or final, wave. Chapter 6 in this volume and the *SIPP Quality Profile* provide more detailed discussions of the implications of nonresponse for data quality. SIPP deals with the various types of nonresponse by weighting adjustments or imputation (Chapters 8 and 4). Table 2-5 shows cumulative loss rates for two types of nonresponse, discussed below.

The Census Bureau distinguishes between household and person nonresponse. Household nonresponse occurs either when the interviewer cannot locate the household or the when interviewer locates the household but cannot interview any adult household members. Person-level nonresponse occurs when at least one person in the household is interviewed and at least one other person is not—usually because that person refuses to answer the questions, or is unavailable and no proxy is taken. The Census Bureau categorizes household nonresponse as Types A and D (detailed definitions and discussion of rates follow),[14] and person-level nonresponse as Type Z.

Household Nonresponse

Type A household nonresponse occurs when the interviewer finds the household's address, but obtains no interviews. Those households contain people eligible for SIPP interviews, but every eligible member of the household is a noninterview. Examples of Type A nonresponse include the following:

- The interviewer finds no one at home despite repeated visits.
- All eligible household members are away during the entire interview period (e.g., an extended vacation).
- Household members refuse to participate in the survey.
- The interviewer cannot reach the housing unit because of impassable roads, such as from a natural disaster.
- Interviews cannot be taken because of serious illness or death in the household.

When this type of household nonresponse occurs in Wave 1, SIPP makes no attempt to interview the household members at subsequent waves. For Type A nonresponse that occurs in subsequent waves, however, interviewers try to obtain interviews on the following wave. *New* Type A noninterviews represent the first time a Type A household nonresponse occurred. *Old* Type A

[14] The Census Bureau recognizes two other types of household noninterviews. Type B occurs in Wave 1 when the address unit is vacant or in some way unfit for residence; in subsequent waves, Type B occurs when people enter institutions. Type C occurs in Wave 1 when the housing unit has been demolished or converted to some other use; in subsequent waves, Type C occurs when all sample members in a household are outside the scope of the survey, e.g., deceased, living abroad, or living in armed forces barracks.

Table 2-5. Household Noninterview and Sample Loss Rates: 1990–1996 Panels

Wave	1990 Panel			1991 Panel			1992 Panel			1993 Panel			1996 Panel		
	Type A	Type D	Loss	Type A	Type D	Loss	Type A	Type D	Loss	Type A	Type D	Loss	Type A	Type D	Loss
1	7.3	—	7.3	8.4	—	8.4	9.3	—	9.3	8.9	—	8.9	8.4	—	8.4
2	10.9	1.5	12.6	12.3	1.5	13.9	12.8	1.7	14.6	12.4	1.7	14.2	13.1	1.3	14.5
3	11.5	2.6	14.4	13.1	2.7	16.1	13.1	2.8	16.4	12.9	2.9	16.2	15.6	1.9	17.8
4	12.5	3.4	16.5	13.6	3.6	17.7	13.8	3.6	18.0	13.9	3.8	18.2	17.6	3.1	20.9
5	13.6	4.6	18.8	14.5	4.2	19.3	14.9	4.7	20.3	14.9	4.7	20.2	20.4	3.8	24.6
6	14.1	5.3	20.2	14.4	5.1	20.3	15.3	5.4	21.6	15.9	5.5	22.2	22.2	4.4	27.4
7	14.3	5.9	21.1	14.7	5.6	21.0	16.0	5.9	23.0	17.2	6.2	24.3	23.8	4.8	29.9
8	14.4	5.9	21.3	14.5	5.9	21.4	16.9	6.7	24.7	17.5	6.9	25.5	24.2	5.4	31.3
9	—	—	—	—	—	—	17.7	7.3	26.2	18.2	7.5	26.9	25.0	5.6	32.8
10	—	—	—	—	—	—	17.5	7.6	26.6	—	—	—	26.1	6.0	34.0
11	—	—	—	—	—	—	—	—	—	—	—	—	25.5	6.2	35.1
12	—	—	—	—	—	—	—	—	—	—	—	—	—	6.2	35.5

Note: The sample loss rate is the cumulative noninterview rate adjusted for unobserved growth in the Type A noninterview units (created by splits).

Source: SIPP Quality Profile, 3rd Ed. (U.S. Census Bureau, 1998a).

nonresponse represents unsuccessful attempts to convert a Type A noninterview from the previous wave. Two consecutive Type A noninterviews render the case ineligible for interviews at the following wave.[15]

Type D household nonresponse concerns original sample members who move to an unknown or uninterviewable address; it applies only to Wave 2 and beyond. Those noninterviews occur when a household or some members of a household are living at an unknown new address or at an address located more than 100 miles from a SIPP sample area and cannot be contacted by telephone.[16] For the 1996 Panel, Type D noninterviews are attempted three times before they are dropped.

Person Nonresponse

There are two forms of person-level, or Type Z, nonresponse. The first applies to those instances in which a sample person was in the household during part (or all) of the reference period and was part of the household on the date of the interview but refused to answer, or was not available for the interview and a proxy interview was not obtained. The second form of Type Z noninterview occurs when a person was part of the household during part of the 4-month reference period but then moved and was no longer a household member on the date of the interview.[17] While household nonresponse is usually handled by weighting adjustments, Type Z cases are handled by imputation (i.e., they are matched to donors, and data from the donor case are substituted for the missing interview—see discussion of imputation and weighting in Chapters 4 and 8). Nearly half of SIPP Type Z nonrespondents are not interviewed at any of the waves.

Item Nonresponse

Item nonresponse is an additional source of missing data; it occurs when a respondent does not answer one or more questions, even though most of the questionnaire is completed. Respondents might refuse to answer a particular question or set of questions. Sometimes, item nonresponse

[15] For each wave, the rate of Type A nonresponse is calculated by adding the number of Type A noninterviews for the wave to the number of Type A noninterviews dropped from the sample in prior waves and dividing that sum by the total of the number of interviewed households plus all Type A and Type D noninterviews.

[16] For each wave, the rate of Type D nonresponse is calculated by adding the number of Type D noninterviews for the wave to the number of Type D noninterviews dropped from the sample in prior waves, and dividing that sum by the total of the number of interviewed households plus all Type A and Type D noninterviews.

[17] If the person was an original sample member, information will be taken for the portion of the reference period in which he or she was still at the address, and an effort will be made to locate the person. If the person was not an original sample member, information will be taken for the portion of the reference period in which he or she was still at the address, after which the person will not be pursued.

occurs when respondents do not have the information requested.[18] Although interviewers are trained to attempt to persuade respondents to answer all applicable questions, and will call back if a respondent can provide data at a later time, those efforts are not always successful. Item nonresponse can also result from the postinterview data editing process when respondents provide inconsistent information or when an interviewer incorrectly records a response. In many cases, the Census Bureau handles item nonresponse by imputation, that is, by assigning values for the missing items (Chapter 4).

[18] The information provided may also be inconsistent with edit specifications, and the response is thus deleted during the processing stage. Or, interviewers may forget to ask for the information or record it incorrectly, resulting in an edit failure. See Chapter 4 on editing and imputation.

3. Survey Content

This chapter provides analysts using the Survey of Income and Program Participation (SIPP) with an overview of the survey content. SIPP is a longitudinal survey that collects information on topics such as poverty, income, employment, and health insurance coverage. SIPP *core content* covers demographic characteristics, work experience, earnings, program participation, transfer income, and asset income. Each interview wave contains additional *topical content*, including one or more *topical modules*, allowing the Census Bureau to address a range of subjects.[1]

The SIPP Interview

With the 1996 Panel, computer-assisted interviewing (CAI) was introduced. SIPP interviewers began using a laptop computer to collect survey data.[2] CAI presents a number of advantages over interviewing with a paper instrument, the method used in previous panels (Chapter 2). Survey elements appear seamless to both the interviewer and the respondent. In addition, the CAI instrument makes certain decisions about which questions to ask, whom to ask, and so forth, that were once left to the discretion of the interviewer. CAI also allows much of the core content from prior waves to be referenced in each interview. The CAI instrument uses responses and complicated logic from one part of the interview in subsequent parts of the interview, which permits checking for consistency and accuracy in the data while the interviewer is still in contact with the household.

This chapter will associate the word *core* with items in the survey that remain constant from one wave to the next, and the word *topical* with items that do not appear in every wave. For both the CAI instrument and the pre-1996 paper survey, data gathered every time the survey is conducted are referred to as *core content*. The *core questionnaire* collects critical labor force, income, and program participation data and is repeated at each interview. Questions asked periodically and targeted to specific topics outside the range of the core content provide *topical content* and are referred to as *topical modules*.

Cooperative, available respondents 15 years of age and older answer questions for themselves, to the extent possible. While questionnaires are not completed for household members under age 15, information is collected about them so that household members under age 15 are fully represented in the SIPP sample. When necessary, information in the CAI instrument is used to determine the next best person in the household with whom a dependent or proxy interview should be conducted; that is often, but not always, the *reference person* (Chapter 2).

[1] Analysts should consult the actual survey instrument for answers to specific questions about the ordering and wording of survey items. The technical documentation can be ordered separately (Chapter 5). The SIPP Interviewer Procedures Manual also can be ordered from the Census Bureau.

[2] Although all interviews were conducted using an automated survey instrument residing on a laptop, not all interviews were done in person. In some cases, interviews were conducted by phone from the interviewer's home.

Skip patterns within SIPP control which questions are asked of each respondent. Skip patterns tailor the questions to the circumstances of the respondent and bypass irrelevant questions. For example, if a respondent has already said that he or she did not work during the reference period, the skip pattern will prevent the interviewer from asking the person what kind of job was held during that time. The CAI instrument automatically calls up the next relevant question, making the skip patterns transparent to both interviewers and respondents. Before the introduction of CAI, interviewers followed instructions on the paper survey in order to skip inappropriate questions. Figure 3-1 illustrates the way in which skip patterns worked in the paper survey. Since CAI handles skip patterns from "behind the scenes," Figure 3-1 might also be viewed as showing what is invisible in CAI.

Figure 3-1. Skip Pattern Example

7c. **Could . . . have taken a job during those weeks if one had been offered?**	__ Yes – *Skip to 7e* __ No
7d. **What was the main reason . . . could not take a job during those weeks?** *Mark (x) only one.*	__ Already had a job __ Temporary illness __ School __ Other *(Specify)* _____

[Notes to interviewers are italicized; respondent's name is filled in; and statements read to respondents are in bold.]

Core Content

Core questions are typically asked at the start of the interview. At the beginning of each household visit, the Census Bureau interviewer completes or updates a roster listing all household members, verifies basic demographic information about each person, and checks certain facts about the household. The CAI instrument performs "behind the scenes" case management functions at the same time. Prior to the advent of CAI, that information was contained on the *control card*, which provided a mechanism for carrying information forward from one wave to the next for each sample member. Core questions covering key areas of SIPP follow the initial questions. For the most part, the 1996 Panel and prior panels cover the same content; however, the organization of the content within the 1996 CAI instrument is somewhat different.

Core Content for 1996 and Subsequent Panels

SIPP core content covers a variety of topics, including labor force status and employment, earnings, business ownership, assets, income, program participation, child support collection, health insurance, and education, among others. While CAI allows the SIPP interview to proceed seamlessly, analysts will perceive distinct sections within the core data.

Employment and Earnings

The first group of survey questions addresses employment and earnings. This section collects information about the respondent's labor force status for each week of the reference period; identifies characteristics of employers, self-employment, and businesses the respondent might own; and gathers data about earnings, whether from a job or from self-employment. Respondents are asked about their labor force status and any unemployment compensation for a time period covering the beginning of the 4-month reference period up through the date of the interview. The type of work performed and dates of employment are also noted. The interviewer asks respondents who own businesses whether they are active in its management, own it as an investment, or are involved in some combination thereof. The survey also collects data on time spent looking for work, moonlighting, and the current employment situation for up to two jobs and two businesses. Employment status is derived from information about specific jobs.

The flow of the survey is such that questions about employment and job characteristics are asked first, with amounts collected separately. Probes ensure that amounts are reasonable and that gross amounts are obtained. Respondents are asked to refer to records whenever possible.

Program, General, and Asset Income

These questions focus on income from a source other than the respondent's work situation. Many of the questions address income or benefits from programs such as Social Security or Food Stamps (and in 1996 have been adapted to capture postreform welfare benefits); the survey also collects information about retirement, disability and survivors' income, unemployment insurance and workers' compensation as well as severance pay, lump-sum payments from pension or retirement plans, child support, and alimony payments. A set of general income questions takes information collected previously and obtains more details about who is covered, how payments are received, reasons for receiving government transfer income, and other data having to do with program participation. SIPP also collects information on amounts of "roll over" retirement accounts.

To obtain information on asset income, interviewers ask respondents which assets they own, prompting the respondent from a list including U.S. savings bonds, 401(k) plans, stocks, rental property, and the like. Respondents are also asked if they have received any lump-sum or regular payments from an IRA, Keogh, 401(k), or thrift plan. Other questions address income received from assets owned, other than retirement accounts. Income for some assets is collected and

recorded within preset ranges. Most asset income is recorded in exact amounts whenever possible, however. The issue of joint ownership of assets is also addressed.

Additional Questions

SIPP core content also includes small sections that deal with health insurance ownership and coverage (Medicare coverage, Medicaid, private and employer-provided health insurance, and reasons for noncoverage), education (educational attainment, adult school enrollment, and educational assistance), and energy assistance and school lunch program participation.

Table 3-1 lists possible income and benefit sources, along with some special indicators.

Core Content for Pre-1996 Panels

Core content in the paper surveys used before the 1996 Panel was structured differently, in four very distinct sections that are described below.

Labor Force and Recipiency

The first set of survey questions addressed the respondent's labor force status, sources of any income received, participation in government transfer programs, and health insurance coverage during the 4-month reference period. Respondents were asked about any employment during each of the 4 months prior to the interview month, although detailed information about their specific jobs was not collected here. Respondents who were employed were asked about the number of hours they worked during a typical week and the number of weeks they worked. For those who did not work, SIPP interviewers asked if they were on layoff or had looked for a job. These survey questions also elicited whether any income had been received from a list of potential sources, including government programs. Respondents were asked about their ownership of assets, although this section of the interview did not include questions about amounts earned in those assets.

Earnings and Employment

This section of the SIPP core asked respondents who reported any employment during the 4-month reference period covered by the interview a more detailed series of questions about the jobs they held. Interviewers collected information for up to two different "wage and salary" jobs in each wave. For each job, data were collected on occupation, industry, and work activities and duties. Several questions aimed to determine the total pay from each job for each month of the reference period. Similar information was collected for up to two different "self-employment" jobs in each wave.

Table 3-1. Types of Income Recorded in SIPP

Wage or Salary Income	**Asset Income (General Amounts Type 2)**
Income from job 1	Regular/passbook savings accounts in a bank, savings and loan, or credit union
Income from job 2	Money market deposit accounts
Income from business 1	Certificates of Deposit or other savings certificates
Income from business 2	NOW, Super NOW, or other interest-earning checking accounts
Program and Miscellaneous Income (General Amounts Type 1)	Money market funds
Social Security	U.S. government securities
U.S. Government Railroad Retirement payments	U.S. Government Savings Bonds (E, EE)
Federal Supplemental Security Income	Municipal or corporate bonds
State Supplemental Security Income	IRA or Keogh account
State unemployment compensation	Other interest-earning assets
Supplemental Unemployment Benefits	Stocks or mutual fund shares
Other unemployment compensation	Rental property
Veterans compensation or pensions	Mortgages from which payments are received
Black Lung payments	Royalties
Worker's Compensation	Other financial investments not already mentioned
State temporary sickness or disability benefits	
Employer or union temporary sickness benefits	**Noncash Income (other than WIC and Food Stamps)**
Employer disability payments	Public housing occupancy
Severance pay	Rent subsidies
Payments from a sickness, accident, or disability insurance policy purchased on your own	Energy assistance
Aid to Families with Dependent Children/Temporary Assistance for Needy Families	Subsidized school lunches or breakfasts
General Assistance or General Relief	**Special Indicators**
Foster child care payments	Worked
Other welfare	Disabled
Women, Infants and Children nutrition programs	VA disability rating of 100%
Pass through child support payments	VA disability of less than 100%
Food Stamps	Medicare
Child support payments	Medicaid
Alimony payments	**Educational Assistance**
Pension from company or union	College work study
Federal Civil Service or other federal civilian employee pensions	Health or Nursing Grant, ROTC, NSF Grant
U.S. military retirement pay	Stafford Grant
National Guard or Reserve Forces retirement	Perkins Grant
State government pensions	SLS Grant
Local government pensions	Grant, scholarship, tuition reimbursement from school attended
Income—paid-up life insurance policies or annuities	Teaching or research assistantship from school attended
Estates and trusts	Grant or scholarship from the state, such as SSIGP, Douglas scholarships
Other payments for retirement, disability, or survivor	Grant or scholarship from some other *Source*, such as foundation, corporation, community group, National Merit scholarships
GI Bill/VEAP education benefits	PELL Grant
Other VA educational assistance	Supplemental Educational Opportunity Grants
Draw from IRA/Keogh 401(k) or thrift plan	National Direct Student Loan
Income assistance from a charitable group	Guaranteed Student Loan
Money from relatives or friends	JTPA training
Lump-sum payments	Employer assistance
Income from roomers or boarders	Fellowship/scholarship
National Guard or Reserve pay	Other financial aid
Incidental or casual earnings	
Other cash income not included elsewhere	

Amounts of Income Received

The third group of core questions addressed the amounts of income or benefits received from sources other than earnings.[3] Detailed information was also collected about participation in government transfer programs. For each nongovernment, nonasset source reported (e.g., alimony payments), respondents were asked the amount of income received during each of the prior 4 months. If benefits were received from government programs, respondents were asked the reason for program participation and who within the household was covered. Questions about asset income, from sources such as interest, dividends, rents, and royalties, sought only the total amount for the 4-month reference period. Examples of assets include money market funds, stocks, rental property, and other financial investments. An example of income earned from an asset would be the interest from a savings account.

Program Questions

The final section of the SIPP core included questions about participation in programs that provide subsidized housing, energy assistance, and school meal programs.

Topical Content

Topical questions are those that are not repeated in each wave. These questions usually appear in separate topical modules that follow the core questions. Topical modules are designed to gather specific information on a wide variety of subjects. They provide a broader picture of the types of individuals who are responding to the survey and give SIPP some flexibility in collecting data on emerging issues. Some topical modules are included in each panel but, unlike the core content, are not in each wave. The frequency and timing of these modules may vary. For example, the personal history topical modules are always administered once, in Waves 1 and 2. Other topical modules are asked multiple times within the same panel; the Assets and Liabilities module, for example, is included four times within the 1996 Panel.

In some instances, the interview flows more smoothly if topical questions are placed with core questions that relate to the same topic. For example, topical questions on asset balances are divided between items included in the core questionnaire and items included in a separate topical module. SIPP asks questions about ownership and an income amount in the core. Questions relating to asset balances appear in the asset topical module. Similarly, home-based-employment and size-of-firm data collected in the 1992 and 1993 Panels (Waves 6 and 3, respectively) are incorporated into the core questionnaire. The term *topical module*, therefore, actually refers to all topical items of the same theme, instead of those that are grouped together into a distinct module, because the frequency with which the item appears is more important than its location.

[3] As with all of SIPP, respondents include all people 15 years old and over. When children under 15 have their own income, it is recorded as having been received by an adult on their behalf.

Reference periods for items in topical modules vary widely, ranging from the respondent's status at the time of the interview to the respondent's experience over his or her entire life. When working with data from the SIPP topical modules, analysts should check question wording concepts carefully to ascertain the reference period. They should also check the universe for each question, because topical modules are not uniformly asked of all respondents. For example, only people 25 years of age or older are asked topical module questions about their retirement and pension accounts. Questions on shelter costs and energy usage are asked only of the reference person. In other modules, a screening question will determine who is and is not asked the remainder of the module—in the case of the Work Schedule module, for example, only those who worked during the previous month answer the entire set of questions.

The relationship between topical module titles and content is not perfectly consistent. Over the history of SIPP, there have been situations in which either the topical module content changed with no change in title or the topical module title changed with little change in content. In a few situations, content has "floated" from one topical module to another. And sometimes there has been significant overlap in content between two topical modules with different titles.

The actual questions are provided with the microdata technical documentation. Specific topical modules are discussed below, with the panels and waves listed in brackets (e.g., [93-3, 96-6] for a module asked in the third wave of the 1993 Panel and the sixth wave of the 1996 Panel). Chapter 5 lists topical modules and the panels and waves in which they were included in the survey. Table 3-2 groups topical modules thematically (modules may appear in more than one category).

Table 3-2. Topical Modules Grouped Thematically

Category	Topical Module
Health, Disability, & Physical Well-Being	Adult Well-Being; Children's Well-Being; Functional Limitations and Disability; Health and Disability; Health Status and Utilization of Health Care Services; Long-Term Care; Medical Expenses and Work Disability; Work Disability History
Financial	Annual Income and Retirement Accounts; Assets and Liabilities; Real Estate Property and Vehicles; Recipiency History; Retirement Expectations and Pension Plan Coverage; School Enrollment and Financing; Selected Financial Assets; Shelter Costs and Energy Usage; Support for Nonhousehold Members; Taxes
Child Care & Financial Support	Child Care; Child Support Agreements; Child Support Paid; Support for Nonhousehold Members
Education & Employment	Education and Training History; Employment History; Job Offers; School Enrollment and Financing; Work-Related Expenses; Work Schedule
Family & Household Characteristics & Living Conditions	Extended Measures of Well-Being; Family Background; Fertility History; Household Relationships; Marital History
Personal History	Education and Training History; Employment History; Fertility History; Marital History; Migration History; Recipiency History; Work Disability History
Welfare Reform	Eligibility for and Recipiency of Public Assistance; Benefits; Job Search and Training Assistance; Job Subsidies; Transportation Assistance; Health Care; Food Assistance; Electronic Transfer of Benefits; Denial of Benefits

Specific Topical Modules

Adult Well-Being. Asks the reference person about consumer durables, living conditions, crime, neighborhood conditions, community services, basic needs, and food adequacy. This topical module assesses the standard of living of SIPP respondents. It is similar to Extended Measures of Well-Being and incorporates Basic Needs information that was asked as a separate module in 93-9. [93-9, 96-8]

Annual Earnings and Benefits. Includes questions that ask people about their calendar-year wages and salaries and income from their own businesses, as well as the receipt of certain employer-provided benefits not covered elsewhere in SIPP, such as the use of a company car or truck, an expense account, or the provision of free meals and lodging. In addition, a series of questions is administered about reasons for leaving for those persons who left a job during the calendar year. Questions about calendar-year earnings, taxes, health and life insurance deductions, and retirement contributions are designed to obtain the most accurate data available, and respondents are encouraged to refer to W-2 forms and other records. This module is administered twice per panel. [84-6]

Annual Income and Retirement Accounts. Obtains respondent estimates of calendar-year business income and respondents' personal retirement plans. The module asks about businesses owned by respondents, gross income and expenses to such businesses, net income to such businesses, retirement accounts, including IRA, Keogh, and 401(k), and respondent participation in those retirement plans. [84-9, 85-5, 85-8, 86-5, 87-5, 88-5, 90-5, 90-8, 91-5, 91-8, 92-5, 92-8 93-5, 93-8, 96-4, 96-7, 96-10]

Assets, Liabilities, and Eligibility. Collects information about the value of assets and debt on assets and expands on data gathered in the core questions. The intent of this topical module is to derive a comprehensive measure of household net worth and to collect information used to determine eligibility for federal assistance programs. To that end, the topical module includes selected additional questions needed to determine program eligibility. Some of the assets included are savings accounts, stocks, mutual funds, and bonds. Data on unsecured liabilities such as loans, credit cards, and medical bills are also gathered. Assets and liabilities that are held jointly are identified to prevent double-counting. The 1996 version of this module has seven sections: value of business; interest earning accounts; stocks and mutual funds; mortgages; other assets; assets and liabilities; and real estate, shelter costs, dependent care, and vehicle ownership. (Also asked as Assets and Liabilities.) [84-4, 84-7, 85-3, 85-7, 86-4, 86-7, 87-4, 90-4, 91-7, 92-4, 93-7, 96-3, 96-6, 96-9, 96-12]

Child Care. Collects information about all child care arrangements, for all children under 15, from mothers, single fathers, or guardians, regardless of labor force status. Those with children under age 15 are asked about the type of child care arrangements, who provides the care, the number of hours of care per week, where the care is provided, and the cost of the care. The module asks whether a relative or nonrelative cared for the child, and if the child was in school. Before the 1993 Panel, the module collected information about only one to two child care arrangements from mothers, single fathers, or guardians who were either working, in school, or

looking for a job during the 4-month reference period. [84-5, 85-6, 86-3, 86-6, 87-3, 87-6, 88-3, 88-6, 89-3, 90-3, 91-3, 92-6, 92-9, 93-3, 93-6, 96-4, 96-10]

Child Support Agreements. Helps determine whether money received as child support affects participation in government programs and whether lack of support from one parent causes the other parent to need government assistance. The module collects information about characteristics of child support agreements, the annual amount and frequency of payments, and provisions for health care costs. Additional questions cover custodial arrangements, contact with public agencies for assistance in collection of child support, frequency of contact with the absent parent, current place of residence of the absent parent, and reasons for nonaward of child support. Questions about paternity establishment status are also asked about children of women with nonwritten agreements and all never married women. [85-6, 86-3, 86-6, 87-3, 87-6, 88-3, 88-6, 89-3, 90-3, 90-6, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-5, 96-11]

Child Support Paid. Serves as a counterpart to the Child Support Agreements module. It seeks information about support for children of the respondent who are under 21 years old and who live with another parent or guardian at any time during the module's reference period of 4 months. [96-3, 96-6, 96-9, 96-12]

Children's Well-Being. Asks the designated parent or guardian about the health of children in the household, care of the child by nonfamily members, activities the family does with the children (such as reading and outings), lessons and activities outside of school, rules for children's TV viewing, and the respondent's opinion about the quality of the neighborhood. The module obtains information about children in three age groups—under 6 years old, ages 6–11, and ages 12–17—for as many as seven children in each category. Certain questions target fathers or stepfathers who are not designated parents; other questions address whether the child attends a public or private school. Content of this module varies across different panels and waves; analysts should check the documentation for exact content. [92-9, 93-6, 93-9, 96-6, 96-11]

Education and Training History. Collects information about respondent's highest level of school completed or degree received, courses or programs studied, and dates of receipt of high school and postsecondary degrees or diplomas. The module determines if the respondent attended a public or a private high school. Job-related-training questions address training designed to help find or develop skills for a new job as well as to improve skills at the current or most recent job. People 15 years of age and older are asked whether they have received job training; if they have, they are asked about the duration of the training, how it was used, how it was paid for, and if it was federally sponsored.[4] (Variations are also asked as Education and Work History [84-3] and Education and Training [84-6].) [86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2]

Employer-Provided Health Benefits. Collects data on the availability of health care benefits from employers and the demographics of workers with and without employer-provided health coverage. The module asks whether the plan restricts the respondent to specified doctors,

[4] All of the "History" topical modules are designed to collect information about the respondent's experiences prior to the beginning of the SIPP panel. This information is most useful in combination with the more current longitudinal information collected during the panel.

if family members are covered, and whether any family members have pre-existing conditions not covered by the plan. The module also asks about long-term health care options. [96-5]

Employment History. Identifies patterns of employment, length of employment at certain jobs, and reasons for any periods of unemployment subsequent to the respondent's first job. Beginning with the 1996 Panel, specific questions that address type of work done, job duties, and the industry in which the respondent works were moved into the core content; previously, such questions had been part of this module. [86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-1, 93-1, 96-1]

Extended Measures of Well-Being. Assesses the standard of living of SIPP respondents. Three types of questions address the objective physical conditions in which the respondents live, respondents' ability to meet specified basic needs during the reference period, and respondents' subjective assessments of the quality of their living situations. Included under the first category are questions about the presence and condition of specified consumer durable goods in the home (e.g., clothes washers, refrigerators, air conditioners) and the physical condition of the home itself (e.g., condition of the roof and walls, state of the home's electrical wiring and plumbing). Another series of questions concerns conditions in the respondent's neighborhood, such as safety, cleanliness, and traffic. The second group of questions concerns whether members of the respondent's household had sufficient food to eat during the 4-month reference period and whether they were able to pay rent and other bills or to obtain medical care when needed. Respondents are also asked about the sources of help available when the respondent is in need (e.g., family, friends, or community). Finally, respondents rate their satisfaction with the quality of different aspects of their living conditions. Included are items such as the quality of the furnishings, convenience of the home to shopping, and the general state of repair of their home. (Some of those questions have been asked as a Basic Needs module [93-9].) [91-6, 92-3]

Family Background. Asked of people between ages 25 and 64. Obtains family characteristics at the time of the respondent's 16th birthday, including how many brothers and sisters the person had, with whom the person lived, the highest grade of school completed by the parents, and the occupations of the parents. [86-2, 87-2, 88-2]

Fertility History. Asked only of females 15 years of age and older and males 18 and older. Men are asked about the number of children they have fathered, and women are asked about their birth histories. Interviewers ask women who have had children when their first and last children were born, along with questions about their employment status during pregnancy and prior to the birth of their first child, circumstances of any absence from work before and after the first birth, and the maternity leave policies of their employers. Postbirth employment is also covered. [84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2]

Functional Limitations and Disability. Provides data that can be used to evaluate links between types of disability, the family financial situation, and program participation. This module is asked in three variations: overall, adult, and children. Adults are asked the standard Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL) battery of questions. Questions address physical and mental conditions affecting the respondent, the use of mobility aids, vision and hearing impairments, speech difficulties, lifting and aerobic difficulties, and the ability to function independently within the home. For those under age 22, the questions

are modified, referring to age-appropriate activities (e.g., questions about work activities are recast to ask about analogous school activities). Questions about children also address the use of special education services. For those under age 15, the interviewer asks the questions of the designated parent or guardian. [90-3, 90-6, 91-3, 92-6, 93-3 for overall module; 92-9, 93-6, 96-5, 96-11 for separate children and adults modules]

Health and Disability. Gathers data for all sample members about their general health, functional limitations (using the standard ADL battery of questions), work disability, and the need for personal assistance. Respondents are asked about any hospital stays during the reference period, other periods of illness, other health facilities used, and their health insurance coverage. Information on children is collected from a designated parent or guardian. (Variations are also asked as Functional Activities, Disability Status of Children, and Disability Questions.) [84-3 for Health and Disability; 88-6, 89-3 for Functional Activities; 85-6, 86-3, 87-6, 88-3, 88-6, 89-3 for Disability Status of Children; 96-4 for Disability Questions]

Health Status and Utilization of Health Care Services. Asks about hospital stays, including any in psychiatric institutions; other illnesses or injuries that left the respondent bedridden for at least most of 1 day; doctor visits and frequency of visits, dental visits and frequency of visits; where the respondent seeks health advice (doctor's office, clinic, hospital); and health insurance coverage. (Also asked as Utilization of Health Care Services.) [85-6, 86-3, 87-6, 88-3, 88-6, 89-3, 90-3, 90-6, 91-3, 92-6, 92-9, 93-3, 93-6, 96-3, 96-6, 96-9, 96-12]

Home Health Care. Asks about the type and sources of help given to respondents who needed help with their personal care, household activities, and basic errands because of a health condition. Respondents are asked if caregivers were relatives or nonrelatives, and whether or not the caregivers were household members. This module also asks about members of the household who might have given such care, on a nonprofessional level, to a person outside the household. Questions determine the relationship of the caregiver and recipient(s) and the kind of care given. [88-6, 89-3]

Household Relationships. Collects information about relationships among household members. The SIPP core questions gather extensive information about household composition for each month of the panel. This information allows for the identification of families and subfamilies and details each household member's relationship to the household reference person.[5] As extensive as this information is, it does not cover the interrelationships of all household members. For example, the SIPP core provides no information about the relationships between members of two different unrelated (to the household reference person) subfamilies residing in the same household. This topical module fills that gap by collecting complete information about how each member of the household is related to every other member of the household. Relationships are specified in detail; for example, a brother is a full brother, half

[5] The family is defined by the Census Bureau as two or more people who are living together and are related by blood, marriage, or adoption. A primary family is the family containing the household reference person; an unrelated subfamily is a family that does not contain the reference person or anyone related to the reference person. Related subfamilies are families within the primary family. A daughter and husband living with the daughter's parents would constitute a related subfamily. The reference person is the person in whose name the home is owned or rented. If the house is owned jointly by a married couple, either the husband or the wife may be listed as the reference person.

brother, stepbrother, or adoptive brother. In-law relationships are also identified. [84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2]

Housing Costs, Conditions, and Energy Usage. Collects information on mortgage payments, real estate taxes, fire insurance, principal owned, when the mortgage was obtained, and interest rates; rent; type of fuel used and heating facilities; appliances; and vehicles.[6] Questions on value of home and automobile are used in conjunction with assets and liabilities reported in the Assets and Liabilities Topical Module to calculate each individual's net worth. This topical module also helps to fulfill a need for information concerning energy usage that has resulted from increased interest in recent years over the rising costs of energy and concerns about conservation. The information can be used in analysis of the requirements of individuals and households who participate in energy assistance programs. [84-4]

Job Offers. Asks about any job offers received by respondents who were looking for work or who were on layoff during the reference period. If the respondent was offered a job and did not accept it, questions probe the reason for rejecting the job and the amount of money that was offered. [85-6, 86-3]

Long-Term Care. Focuses on health-related conditions that might cause a person to need help around the home. Specific questions address the ability of people in the household to manage their personal care, housework, meal preparation, and basic errands outside the home. The module ascertains whether or not individuals providing such assistance are household members. Additional questions ask about community services and the financial burden of acquiring assistance. The module also asks about the activities of respondents who themselves provided such assistance on a nonprofessional basis to individuals outside the household. (Also asked as Home Health Care.) [85-6, 86-3, 87-6, 88-3, 88-6, 89-3]

Marital History. Asks questions of all respondents aged 15 and older who have ever been married. The date of the present marriage is determined; for those married more than once, SIPP records the dates of their first two marriages and their last marriage, if married more than twice. If appropriate, respondents are asked when their previous marriages ended and whether they were widowed or divorced at the end of their marriages. [84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2]

Medical Expenses and Work Disability. Gathers data about out-of-pocket medical expenses, health services, doctor visits, prescription drugs, insurance reimbursement, and health and physical conditions that might affect the respondent's ability to work. The reasons for and length of any hospitalizations are determined, and respondents are asked about the types of medical professionals who delivered care. Most questions apply to both children and adults. (Also asked as Medical Expenses.) [87-7, 88-4, 89-4, 90-7, 91-4, 92-7, 93-4, 93-7, 96-3, 96-6, 96-9, 96-12]

Migration History. Asks respondents aged 15 and older where they were born, where they have lived, and how long they have lived in those places. Respondents born in a foreign country

[6] Subsequent to the 1984 Panel, questions on energy usage were combined into a separate module. Vehicles and housing values are retained together in a module entitled "Real Estate and Vehicles."

are asked about their citizenship status and when they came to the United States to stay. [84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2]

Property Income and Taxes. Collects information on rental income received during the calendar year and on interest earned and/or dividends from assets such as savings accounts, money market deposit accounts, interest-earning checking accounts, bonds, or stocks. They are also asked about federal and state income tax liabilities and certain other tax information such as type of return, use of selected schedules (for example, Schedula A, Itemized Deductions; Schedule B, Interest or Dividends; or Form 4835, Farm Rental Income), and number of exemptions. The tax questions are asked in order to develop better estimates of the distribution of after-tax income and to help build better microsimulation models of the tax and transfer system. This module is administered twice per panel. [84-6]

Real Estate Property and Vehicles. Gathers information about housing tenure and financing, other real estate ownership, and automobile ownership. Home owners are asked a series of questions that allow the estimation of net real estate equity. Questions about vehicles address ownership, type of vehicle (i.e., car, truck, motorcycle), value, and amount owed. Those questions are also used in program eligibility simulations. (A variation of this module is asked as Real Estate, Shelter Costs, Dependent Care, and Vehicles.) [84-7, 85-3, 85-7, 86-4, 86-7, 87-4, 87-7, 88-4, 90-4, 90-7, 91-4, 91-7, 92-4, 92-7, 93-4, 93-7]

Reasons for Not Working/Reservation Wage. Ascertains the reasons that persons are not in the labor force and the conditions under which persons might want to join the labor force. The reservation wage questions ask about the pay rate that a person would require in order to begin working (Ryscabage, 1987). Questions are also asked about job search and, if people have been offered but did not accept a job, the reason they refused it. This module was discontinued after the 1985 Panel. [84-5]

Recipiency History. Obtains a profile of a respondent's pattern of participation in certain government programs prior to the beginning of the SIPP panel. Specific questions address the first time a respondent participated in a particular program, the length of participation, and the number of times the respondent has been in the program. [86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-1, 93-1, 96-1]

Retirement Expectations and Pension Plan Coverage. Obtains information about the respondent's pension plan coverage for the most important current job or business, and information from persons currently receiving retirement benefits from a former job or business. Respondents are asked about their coverage and vesting in pension plans, types of plans, the reasons they are not included by or do not participate in plans, current contributions and amounts of money in their accounts if applicable, and how the money in their own plans is invested. Other questions concern loans from pension accounts and treatment of lump sums received from prior job pension plans.

Respondents currently receiving pension income are asked about the types of pension they receive, provisions for cost-of-living adjustments, and health benefits. Respondents are also asked Industry and Occupation data about the job or business from which their pensions are

received. (Also asked as Pension Plan Coverage [84-7].) [84-4, 85-7, 86-4, 86-7, 87-4, 90-4, 91-7, 92-4, 93-9, 96-7]

School Enrollment and Financing. Seeks information about basic educational attainment, enrollment in public and private schools, and whether those in government programs differ from others in terms of financing their education and their sources of educational assistance. Asked of people aged 15 and older, the module includes questions to pinpoint the grade level of people enrolled in a general, technical, or business school; their pattern of full- or part-time enrollment; amount of tuition and fees; costs of room and board; and books and supplies. Specific sources of educational assistance, such as the GI Bill or employer assistance, are also determined. (Also asked as Education Financing and Enrollment.) [84-9, 85-5, 85-8, 86-5, 87-5, 88-5, 90-5, 90-8, 91-5, 91-8, 92-5, 92-8, 93-5, 93-8, 96-5]

Selected Financial Assets. Focuses on the value of such assets as savings bonds, checking accounts, retirement accounts, life insurance, and the number of years respondents have held certain assets. [87-7, 88-4, 90-7, 91-4, 92-7, 93-4]

Shelter Costs and Energy Usage. Collects information on rent or mortgages, real estate taxes, and insurance; energy costs; and motor vehicles. The information is pertinent to the determination of eligibility for a number of federal assistance programs. (Also asked as Housing Costs, Conditions, and Energy Usage.) [84-4, 86-6, 87-3]

Support for Nonhousehold Members. Provides information about respondents' routine payments supporting people who are not current household members. Includes both child support payments for own children under 21 years of age and payments made to (or for) people who are not children of the respondents—for example, an elderly parent in a nursing home or an adult child living away from home and in an entry-level job. Questions about child support include number of children supported, type and year of agreement, annual amount and method of payment, health care provisions and custodial arrangements, and amount of contact with the absent children. Questions about support for other persons outside the household include their relationship to the respondent, living arrangement, and annual amount of support paid. [84-5, 84-8, 85-4, 85-6, 86-3, 86-6, 87-3, 87-6, 88-3, 88-6, 89-3, 90-3, 90-6, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-5]

Taxes. Includes questions about exemptions, calendar-year wages and salaries, income from businesses, itemized deductions, and earned income credits. Respondents are asked about federal and state income tax liabilities, exemptions, amounts owed for federal and property taxes, and amounts from a variety of tax schedules. To help ensure accuracy, interviewers encourage respondents to refer to income tax returns and other records. Historically, this module has been administered at least twice per panel, generally in the spring when respondents were likely to be preparing their tax returns for the prior year. (Also asked as Earnings and Benefits, and Property Income and Taxes.) [84-6, 84-9, 85-5, 85-8, 86-5, 87-5, 88-5, 90-5, 90-8, 91-5, 91-8, 92-5, 92-8, 93-5, 93-8, 96-4, 96-7, 96-10]

Time Spent Outside Work Force. Collects information about work history and reasons for not working. Asked of people 21 or older, this short module addresses up to four periods of 6 months or longer in which the respondent did not work at a paid job or business. [90-6]

Welfare History and Child Support. Collects information on how long individuals may have received aid from specific welfare programs and on child support agreements and their fulfillment. The data from the welfare history questions will be used to measure the extent to which persons and households have been dependent upon government transfer programs in their general finances and will be helpful in evaluating the effectiveness of the programs.

One series of questions in the module concerns the Food Stamp, AFDC/Temporary Assistance for Needy Families (TANF), and SSI programs. Current recipients are asked how long they have been receiving, or have been authorized to receive, these benefits. Recipients and nonrecipients are asked whether they had at any previous time applied for benefits, whether they received them, and, if so, when and for how long. This module was incorporated into a series of history modules, collectively called the Personal History Topical Module, beginning with the 1986 Panel.

The Child Support Topical Module attempts to determine whether those entitled to receive child support payments have in fact received them. The module asks whether the child support agreement was court ordered or arranged otherwise and how the payments were to be made. It also asks for the amount and regularity of payment and whether a child support enforcement office has provided any help. [84-5]

Welfare Reform. Seeks information about eligibility for and recipiency of public assistance. Specific questions address benefits, assistance that supports a respondent seeking work or acquiring training, requirements for receiving benefits (such as job hunting, drug testing, etc.), job subsidies, transportation assistance, health care, and food assistance. This module also gathers information about electronic transfer of benefits and denial of benefits to the respondent. [96-8]

Work Disability History. Asks a series of questions about chronic health conditions that may affect the amount or type of work a respondent can do. Included are any such physical, mental, or other health conditions that interfere with the respondent's ability to work for at least 3 months. Questions are asked about when the limiting condition first became an issue, whether the person was working at the time, whether the condition resulted from an accident or injury, and if so, where the accident or injury occurred. Shorter-term conditions (including pregnancy) are not included as limiting conditions. [86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2]

Work-Related Expenses. Asks about work-related expenses for each employer the respondent had during the reference period. Questions address various costs of working, such as union dues, licenses, special tools, and uniforms. Mode of transportation and mileage driven to and from work are determined, along with any parking or mass transit fees. (Also asked as Work-Related Expenses and Child Support Paid.) [84-5, 84-8, 85-4, 86-6, 87-3, 96-3, 96-6, 96-9, 96-12]

Work Schedule. Collects information about the number of hours and days worked during a typical week in the fourth reference month. Questions about whether or not the respondent worked only at home on any days are included. [87-6, 88-3, 88-6, 89-3, 90-3, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-4, 96-10]

4. Data Editing and Imputation

This chapter describes the data editing and imputation procedures applied to data from the Survey of Income and Program Participation (SIPP) after completion of the interviews. Three different approaches are used for dealing with missing data in SIPP:

- Weighting adjustments are used for some types of noninterviews;
- Data editing (also referred to as logical imputation) is used for some types of item nonresponse; and
- Statistical (or stochastic) imputation is used for some types of unit nonresponse and some types of item nonresponse.

Weighting is discussed in Chapter 8.

The chapter begins with a brief discussion of the types of missing data and the goals of imputation in SIPP. It then presents an overview of the editing and imputation procedures used to deal with missing and inconsistent data. Next, the chapter provides a detailed description of each of the major steps used by the Census Bureau when creating its internal files and the files that are released for public use. Prior to 1996 the development of cross-sectional wave files involved mainly cross-sectional editing and imputation. The longitudinal files involved longitudinal editing. Beginning with the 1996 Panel, the processing procedures for the wave files were replaced with methods that use prior wave information to inform the editing and imputation of a current wave (after wave 1). The generic imputation technique, that is, the hot-deck method, is still used in the 1996+ Panels, but the donors are now chosen on the basis of similarities in reported prior wave information when that reported information exists.

The SIPP Web site (http://www.sipp.census.gov/sipp/) supplements the information in this chapter with detailed information about all variables on the public use files.

Types of Missing Data

As in all surveys, there are two general types of missing data in SIPP: unit nonresponse and item nonresponse. Unit nonresponse occurs in SIPP when one or more of the people residing at a sample address are not interviewed and no proxy interview is obtained. This can happen for a number of reasons, described in Chapter 2. Most types of unit nonresponse are dealt with through weighting adjustments (see Chapters 2 and 8). However, the data editing and statistical imputation procedures described in this chapter are used with one type of unit nonresponse: Type Z noninterviews, which occur when an interview is obtained from at least one household member but interviews are *not* obtained from one or more other sample persons in that

household.[1] Prior to the 1996 Panel and in some instances in the 1996 Panel, the method used to adjust for person-level noninterviews in the core wave files is known as *Type Z imputation*, which is discussed below.

Item nonresponse occurs when a respondent completes most of the questionnaire but does not answer one or more individual questions. Item nonresponse data in SIPP occur under the following circumstances:

- Responding sample persons refuse or are unable to provide requested information;
- Interviewers fail to ask a question or incorrectly record a response;
- A response is inconsistent with related responses or is incompatible with response categories; and
- Interviewers make an error when recording or keying in the data.[2]

Item nonresponse data are generally imputed for core items, as well as for many topical module items.

Goals of Imputation

Missing data cause a number of problems: analyses of data sets with missing data are more problematic than analyses of complete data sets; there is a lack of consistency among analyses because analysts compensate for missing data in different ways and their analyses may be based on different subsets of data; and, in the presence of nonresponse that is unlikely to be completely random, estimates of population parameters are biased.

Because missing data are always present to some degree, analyses of survey data must be based on assumptions about patterns of missing data. When missing data are not imputed or otherwise accounted for in the model being estimated, the implicit assumption is that data are missing at random after controlling for other variables in the model. The imputation procedures used for SIPP are based on the assumption that data are missing at random within subgroups of the population (as defined by the cells of the imputation matrices described later in this chapter).

The statistical goal of imputation is to reduce the bias of survey estimates. This goal is achieved to the extent that systematic patterns of item nonresponse are correctly identified and modeled. In SIPP, the statistical goals of imputation are general, rather than specific. Instead of addressing the estimation of specific parameters, SIPP procedures are designed to provide reasonable estimates for a variety of analytical purposes.

[1] That can happen either because people refuse to be interviewed or because they are unavailable for the interview and a proxy interview is not obtained.

[2] Prior to the 1996 Panel, errors could also occur when data-entry workers were keying in results from the paper survey.

Data editing is generally preferred over statistical imputation, and it is used whenever a missing item can be logically inferred from other data that have been provided. When information exists on the same record from which missing information can logically be inferred, that information is used to replace the missing information. The advantage of data editing is that it avoids the increase in variance that occurs when missing items on one record are imputed with nonmissing responses from other records.

Assessing the Influence of Imputed Data on Analysis

Users of SIPP data interested in assessing the influence of imputed data on their analyses should consider whether SIPP imputation procedures have properties that affect their specific analytical requirements. A general discussion of the treatment of missing data in sample surveys is given in Kalton and Kaspyrzyk (1986). Sedransk (1985), Little (1986), and Jinn and Sedransk (1987) discuss properties of commonly used imputation processes. An example of the impact of imputation procedures on the distributional characteristics of a low-income population is discussed in Doyle and Dalrymple (1987).

An evaluation of the effects of imputed data should include a review of rates of unit nonresponse and an assessment of the extent of item nonresponse. Unit nonresponse tends to increase over the life of a panel, as does the likelihood that nonresponse is not a random effect. And as the percentage of eligible sample members re-interviewed decreases, the pool from which donors[3] are selected shrinks accordingly. This smaller pool of donors leads to an increased likelihood that individual donors will be used more than once, which in turn increases the variance of an estimate.

The effects of imputation will likely be small for items with low rates of missing data as long as rates of item nonresponse are not high among important subclasses. Lepkowski et al. (1987), using data from a large federal survey, provide a framework for evaluating the effect of imputed values on analyses. This framework can be readily adapted to SIPP analyses.

An Overview of the Process

There are two phases to the processing of SIPP data. At the conclusion of each wave of interviewing, the data collected during that wave are processed, creating the core wave and topical module files. That is the first phase of processing. Then, at the conclusion of the final wave of interviews, core data from all waves are linked and a new set of edit and imputation procedures is applied to the resulting full panel file. That is the second phase of processing.

[3] Cases with complete data that are the source of the imputed values placed on the records with missing data.

Figure 4-1 illustrates the steps that generate the Census Bureau's internal core wave and full panel files.

Figure 4-1. Sequence of Cross-Sectional Imputation and Longitudinal Editing Procedures

Imputation of Sample Unit Characteristics (Tenure, etc.)	**Imputation of Item Missing Data for Sample Unit Characteristics and Personal Demographic Characteristics**	**Sequence is Repeated for Each Wave in a Panel**
Imputation of Personal Demographic Characteristics (Age, Race, Marital Status)		
Type Z Imputations[a]	**Imputation of Person-Level Noninterviews**	
Imputation of Labor Force Items and Recipiency of Income and Assets	**Imputation of Item Nonresponse in Core Questions**	
Imputation for Item Nonresponse in Records for "Other" Cash Income		
Imputation for Item Nonresponse in Self-Employment Identification Sections		
Imputation for Item Nonresponse in Asset Sections (Property Income)		
Imputation for Item Nonresponse for Household Program Information		
Editing for Demographic and Household Variables, Employment Variables, General Amount Variables, and Other Variables	**Editing of Longitudinal Record**	

[a] Most Type Z records in the 1996 Panel were not handled in a separate process.

Phase 1 Summary

There are six steps in the first phase of SIPP data processing:

1. As each wave of interviewing is completed, core data collected during the wave are edited for internal consistency.
2. Following data editing, the statistical matching and hot-deck procedures described later in this chapter are used to impute missing data from the core wave file.
3. A public use version of the core wave file is then created from the resulting internal core wave file. The public use file is the same as the Census Bureau's internal file except that it has certain information suppressed or topcoded to protect the confidentiality of survey respondents (see sections on Topcoding and Suppression of Geographic Information, at the end of this chapter).
4. On a separate production track from the core data, data from the topical module file administered with the wave are edited for internal consistency. The extent of data editing varies across the topical modules, and some topical modules receive almost no editing.

5. Next, hot-deck procedures are used to impute missing data in the topical module. The extent of imputation varies across the topical modules; some topical modules have no missing data imputed.
6. A public use version of the topical module file is created from the resulting internal file. As with the public use core wave files, the public use topical module files have certain information suppressed to protect the confidentiality of survey respondents.

These steps are repeated at the conclusion of each wave of interviews. Prior to the 1996 Panel, each wave was processed independently of other waves of data. Thus, when multiple core wave files are linked, apparent changes in a respondent's status could be due to different applications of data edits and imputations to the files being combined (file linkage is the subject of Chapter 13). With the 1996 data, the hot-deck procedure was redesigned to rely on historical information reported in prior waves. In addition, other forms of longitudinal imputation, such as carryover methods, were adapted.

Phase 2 Summary

At the conclusion of the panel, the Census Bureau creates a full panel file containing core data from all waves. There are four steps to this process.

1. Core data from all waves are linked. Those data have already been subjected to the Phase 1 edit and imputation procedures.
2. A series of longitudinal edits are applied to the full panel file. Unlike the core wave edit procedures, these edits are designed to create longitudinally consistent records for each person. Both reported values and values that were imputed during the first phase of processing are subject to change. Thus, the data in a full panel file may differ from the data in the core wave files from which the full panel file was constructed.
3. A missing wave imputation procedure is then applied. Data are imputed when a sample member was absent for one or two consecutive waves but was present for the two adjacent waves. Data for the missing wave(s) are interpolated on the basis of information from the fourth month of the prior wave and the first month of the subsequent wave. The missing wave imputation procedure was introduced with the 1991 Panel. Earlier panels were not subjected to this procedure.
4. A public use version of the full panel file is created from the resulting internal file. The public use file has certain information suppressed to protect the confidentiality of survey respondents.

The balance of this chapter describes in greater detail the full sequence of data edit and imputation procedures applied to SIPP data files. Most of the material contained in this chapter is taken from Pennell (1993).

Phase 1: Data Editing and Imputation Procedures for the Core Wave Files

The data processing sequence for each wave is detailed below.

Data Entry and Initial Editing

Beginning with the 1996 Panel (Chapter 2), all of the data entry and some of the initial data editing are performed by computer-assisted interviewing while the interview is in progress. Before the 1996 Panel, the first stages of data processing involved editing the paper questionnaires for completeness, reasonableness, and consistency. Those data checks were conducted first by field representatives before they submitted their questionnaires to the regional offices and then by the regional and central offices of the Census Bureau. The next step was data entry, in which clerks keyed in the information from control cards and questionnaires. Edits were built into the data-entry program to ensure that the data were keyed in the proper sequence and that certain key identifiers, such as control number, name, and relationship to householder, were present. Following this step, the data files were transmitted electronically to Census Bureau headquarters.

Imputation for Sample Unit Characteristics and Personal Demographic Characteristics

Items in this category, including housing tenure (owned or rented), age, race, marital status, and so forth, must be present for any further data processing to take place. If these values cannot be logically derived, they are imputed. The imputation procedure is a modified version of the sequential hot-deck procedure described below.

Type Z Imputation for Core Items in the Core Wave Files

Pre-1996 Panels. Type Z imputation was the method used in the pre-1996 panels to impute core items for person-level noninterviews. There are two categories of person-level noninterviews subject to imputation for the core questions. The first category includes individuals 15 years of age and older who were members of interviewed households at the beginning of the 4-month reference period but were not original sample members or members of any SIPP-interviewed household on the date of the interview—that is, people not interviewed because they moved out of the sample household between the beginning of the reference period and the interview date. Had these people been original sample members, they would be interviewed at their new address.

Rather, these are all people who entered the SIPP sample after the first wave and were in the sample because at some point they were living with an original sample member.

The second category of imputed noninterview includes people 15 years of age or older who were members of SIPP-interviewed households on the date of the interview and during all or a portion of the 4-month reference period but who were not interviewed because they refused to cooperate or were unavailable for the interview and a proxy interview was not obtained.

The Type Z imputation procedure is based on a hierarchical sorting and merging operation that matches noninterviews with respondents on socioeconomic characteristics available for both. The variables used to match noninterviews with respondents are age, race, gender, marital status, household relationship, education, veteran status, parent/guardian status, and income and asset sources. Pennell (1993, Figure C-1) provides a table of variables used to match recipients with donors. The Type Z imputation procedure is designed to always find a match. Type Z noninterviews are imputed by assigning values from the matching donor to the noninterview record. The donor values are assigned in full, except for identification variables or other variables not relevant for the household in which the noninterview occurred. Pennell (1993) gives a complete account of Type Z imputation, including detailed descriptions of matching operations.

1996 Panel. In Waves 2–12 of the 1996 Panel, the general imputation procedure (the sequential hot-deck procedure described in the following pages) is being used to impute core items for most person-level noninterviews. That is, these types of noninterviews are no longer set aside—in the 1996 and later panels—for the specialized Type Z imputation procedure. However, the Type Z imputation procedure is still used in Wave 1 of the 1996 Panel (because there is no prior wave information to inform the imputation process) and for noninterviews for persons in Waves 2–12 for whom there is no prior wave information (because they are new to the sample).

Imputation of Item Nonresponse in Core Questions

SIPP core items are imputed in the following order:

1. Labor force participation, recipiency of income, and asset holdings;
2. Other cash income;
3. Wage, salary, and self-employment income amounts;
4. Asset income amounts; and
5. Program participation and benefits.

The Sequential Hot-Deck Imputation Procedure

The statistical imputation method used to impute missing items from the core questions and topical modules is known as a *sequential hot-deck procedure*.[4] In a general sense, the sequential hot-deck procedure, like the Type Z imputation procedure, matches a record with missing data to that of a donor with similar background characteristics and uses the donor's values. This procedure differs from data editing, which replaces missing data with inferred values based on nonmissing data from the *same* case.

The sequential hot-deck procedure used in SIPP involves five key steps:

1. Specifying cold-deck or initial donor values;
2. Sorting the sample cases;
3. Identifying records with no item nonresponse and updating hot-deck values;
4. Classifying cases into subclasses of the population, referred to as imputation classes or adjustment cells, according to values on a set of classification or auxiliary variables that are nonmissing for all cases (this step is omitted in the initial processing of the key demographic items—race, gender, etc.); and
5. Selecting replacement values from donor cases to impute item-missing data on recipient records.

Two types of sequential hot-deck imputation are used to provide values for missing items. In Wave 1 and for each sample member who is new to a subsequent wave, the hot deck is cross-sectional; only values from current wave responses are used in the definition of the hot-deck cells. Beginning with Wave 2, previous wave values are included in the definition of the hot-deck cells. In both instances, however, only current wave values from selected donors are used to replace missing items (with several exceptions, described below). Longitudinal (or "previous wave") hot-deck imputation was not performed prior to the 1996 Panel. Each wave received only the cross-sectional hot-deck imputation.

For example, the item indicating whether a person worked part-time in the reference period for the wave (a dichotomous item) uses the longitudinal hot deck for "old" sample members and the cross-sectional hot deck for new sample members. The 1996 Panel cross-sectional hot-deck imputation is based on a cell structure with 288 cells that are based on cross-classifications of sex (two categories), race (two categories), age (six categories), marital status (three categories), disability status (two categories), and presence of own children (two categories). On the basis of his or her current wave values for those categories, each new sample member in any later wave is assigned to a cell; then the donor's value in that cell is used to impute a value to the new sample member.

[4] The hot-deck procedure used in SIPP for the core questions and topical module items is sequential because the selection of replacement values is implemented one record at a time from an ordered file.

The longitudinal hot-deck imputation for the part-time work item for old sample members in Waves 2+ is based on a cell structure with 576 cells that are based on the same categories described above with one extra category: whether or not the person worked part-time in the previous wave. A donor is selected from that cell, and that value is imputed. The actual item is imputed from a donor's value of the item in the current wave; the previous wave value is used only in the assignment of the cell. That procedure guarantees that the sample member is matched to the donor who had the same value for the item in the previous wave. Therefore, sample members who worked part-time in the previous wave will be matched only to donors who also worked part-time in the previous wave. However, the actual hot-deck imputation comes from the donor's value in the current wave, which may or may not include part-time work.

Imputed values for the sample member are allowed in assigning the cell for some items. If a sample member had an imputation for part-time work in the previous wave, that imputation is used to define the cell for the longitudinal hot-deck imputation, even though it is an imputation itself. That is not done for other items, such as asset items. Only a nonimputed or logically imputed value "counts" toward the longitudinal hot deck for those items.

The part-time item is dichotomous; the previous wave imputation matrix was essentially the current wave imputation matrix with the previous wave's value of the item added to the matrix. In many cases, the differences between the two imputation matrices will be more pronounced, especially for items with several categories of answers. An example of this is the item "reasons why person worked less than 35 hours in the reference period." There are 12 categories for that item. The previous wave hot-deck imputation matrix uses the following characteristics to define cells:

Previous wave value for item (12 categories);

- Sex (two categories);
- Race (two categories);
- Age (six categories).

The current wave imputation matrix uses the following characteristics to define cells:

- Sex (two categories);
- Race (two categories);
- Age (six categories);
- Marital status (three categories);
- Disability status (two categories);
- Presence of own children (two categories).

A different type of example is the item gross pay in the first month of the reference period. For new SIPP sample members, a cross-sectional hot-deck imputation is carried out by using the following characteristics to generate cells:

- Industry and occupation category (16 categories);
- Sex (two categories);
- Hours worked (three categories);
- Education level (three categories).

For old sample members, a longitudinal hot-deck imputation is carried out by using the previous wave value for the item gross pay in the fourth month of the preceding wave's reference period.[5] This continuous value is divided into 138 categories, starting from $1 to $100, to over $50,000. Sample members are matched to donors by using the previous wave values of those categories.

For labor force items, the Census Bureau uses the following special imputation procedures when a person has no current wave information indicating whether or not he or she worked during the reference period. If the Census Bureau can infer from what it knows about the previous reference period whether the person had a job or business at the start of the current period, the Census Bureau carries out the following procedure:

1. If the person was working at the end of the prior wave, then labor force participation is imputed from a single donor for the complete current wave.
2. The Census Bureau then projects job characteristics for the person from the person's prior wave through the current wave.
3. Finally, the Census Bureau edits the job characteristics for consistency with the imputed labor force participation variables.

This procedure is known as an EPPFLAG imputation, after the name of the variable that indicates its use.

If a person was a nonworker in the prior wave or the Census Bureau cannot infer work status on the basis of prior wave data, then the person's work status is imputed. If the person is imputed as a worker in the reference period, the Census Bureau imputes the complete set of job/business characteristics variables and labor force participation variables to the person from one donor, in order to maintain consistency among the fields. That procedure is called a "little Type Z" imputation.

For some items in some cases, a direct logical or carryover imputation is made. The carryover imputation takes the previous wave's value for the item for the sample member and imputes it to the current wave. That imputation is done particularly for items that rarely (or never) change for a sample member across waves (such as sex and race) or for items that change in predictable ways (such as age).

[5] The second month of the reference period actually uses as the "previous wave value" the first month value, with the third month using the second month, and so forth, so that these imputations are really previous month rather than previous wave.

SIPP hot-deck procedures are designed to preserve the univariate distribution of each variable subjected to imputation. These procedures do not, in general, preserve the covariances among variables. Although some of those interrelationships might be preserved to a certain extent, that is not the primary intent of the hot-deck imputation procedures used by the Census Bureau. One consequence is that imputation can introduce inconsistencies into the data. For example, if a respondent has reported program participation, but his or her income is too high for that program, it is possible that the income data have been imputed. Whenever users detect inconsistencies, it is wise to check the allocation (imputation) flag to see if the inconsistent data might have been imputed. The discussion of allocation (imputation) flags later in this chapter provides more information.

Starting or Cold-Deck Values

In other surveys, cold-deck values in a sequential hot-deck procedure historically served as the initial set of replacement values for missing items in the first record processed; missing items in subsequent records typically received replacement (hot-deck) values from the current data set. In SIPP, however, cold-deck values are seldom used as replacement values for either the first or subsequent records processed. During later stages of processing, as the cold-deck values are replaced with information from the current wave, the array of cells is referred to as the hot-deck matrix. The cells in the matrix are defined by the cross-classification of auxiliary variables (Pennell, 1993, Figure 3.3). Each cell in the matrix corresponds to respondent cases with the same set of values on the classification variables. Many different matrices are defined in SIPP, and each matrix corresponds to one or more variables subject to imputation.

Sorting the Sample Cases

The records in the sample file are sorted by three geographic variables prior to imputing item-missing data. The three geographic sort variables are primary sampling unit, segment number, and serial number. The cases are sorted prior to processing and are not re-sorted at any other time during the imputation process. The sorting operation creates a file in which neighboring records represent geographically proximate households.

Preprocessing the Sample File: Initial Updating of Cold-Deck Values

Once the cases have been sorted, they are processed through a series of programs. During the first pass against the programs, the cold-deck values are updated with information from the current wave; missing data are not imputed. The initial processing is done separately for each of the five groups of related core variables listed above. During the first pass, the first record in the sorted file with consistent and nonmissing data for a particular group of variables is identified and the values from that case replace the cold-deck values for that section in the matrix. The values for each subsequent record with consistent and nonmissing information update the previous set of consistent and nonmissing values written to the matrix. The checking and updating operation continues until all records in the data file have been processed. The last values written to the matrix serve as the starting values in the subsequent sequential hot-deck

procedure. In this way, cold-deck values are rarely used as replacement values in SIPP because the initial processing usually replaces all starting values with values from the current wave of data.

Allocating Cases into Imputation Classes

In the next step of the imputation procedure, each respondent record or noninterview record in the sorted file is allocated to one of the imputation classes or adjustment cells according to its values on the set of classification, or auxiliary, variables.[6]

1. The auxiliary variables are chosen for each item or set of related items on the basis of their level of correlation with the item receiving the imputation (i.e., classification variables are chosen on the basis of their ability to explain the variability of the item or set of related items); Census Bureau researchers assign different sets of classification variables to different sets of items.
2. The auxiliary variables are either dichotomous or polychotomous categorical variables (e.g., sex, race); if they are continuous, they are categorized into a parsimonious number of levels (e.g., income, asset levels).
3. The level of the auxiliary variables then define a matrix, with the number of cells in this matrix being the product of the number of levels for each auxiliary variable. For example, an imputation defined by five variables, each with three levels, has a total of 243 cells. Any given item or set of related items may have imputation matrices with the numbers of cells ranging from under 100 to well over 1,000, depending on the matrix.

Auxiliary variables such as sex, race, and categorizations of age (with different categorizations for different items) are used frequently in the matrices, as are more specialized auxiliary variables that are relevant for particular items (such as industry and occupation category for the monthly gross pay item). Pennell (1993) gives examples of the different sets of classification variables for previous panel years.

The allocation of sample cases into imputation classes (also known as subclasses or strata) according to a set of classification variables serves several purposes. Ideally, the set of classification variables should account for a large proportion of the variance in the variable being imputed and should be associated with variations in response rates. To the extent that this is accomplished, the classification procedure creates homogeneous adjustment cells containing similar cases. In this way, donors and recipients are similar under the assumption that the nonresponse mechanism within the imputation class is not related to the item being imputed; *that is, an underlying assumption is made that item nonresponse data are distributed randomly within the subclass defined by the cross-classification of the auxiliary variables.* The selection of classification variables may also place bounds on the range of values that can be imputed and implicitly satisfy edit constraints. The implicit stratification created by the sort order of the file

[6] This step is omitted for the imputation of the primary demographic values that are imputed before the person-level noninterviews.

further improves the opportunity for better imputation to the extent that nearby cases are more similar to each other than cases that are farther apart in the file.

Imputing for Missing Data and Updating of Hot-Deck Values

The selection of replacement values for missing items is restricted to donor and recipient records within each particular cell; that is, records allocated to one cell never donate information to records in another cell with missing items. As the file is processed through the set of programs the second time, the imputations are performed and the set of hot-deck values is updated once again.

The records are processed sequentially, according to the sort order of the file. A missing item is given the value of the last corresponding item that is nonmissing from a record in that imputation class. If the value of an item in the current record is nonmissing, it replaces the previous hot-deck value for that imputation class. In this way, the hot-deck value for each imputation class is constantly being updated with the value of the last nonmissing case.

The updating is done item by item. Missing items in one record receive the current set of replacement values. Then the nonmissing values in that record are used to update the hot deck in preparation for the next record. At any point during the process, the donated values in the hot deck likely come from many different respondents, even within imputation classes. That is why this imputation procedure does not preserve covariances among the variables being imputed.

Allocation (Imputation) Flags

An allocation (imputation) flag is associated with each core item subject to imputation. When an item has been imputed, an allocation (imputation) flag for that item is set. Beginning with the 1996 Panel, allocation flags denoting either data edits or statistical imputations for all variables are included on the core wave files. For core wave files from earlier panels, imputation flags are included for most items subject to imputation.

An allocation (imputation) flag with the value 0 indicates no imputation, a value of 1 or 2 indicates a hot-deck imputation that uses only current quarter values, a value of 3 indicates a logical imputation, and a value of 4 indicates a dependent imputation. This last category includes imputations in which data have been carried over from the sample unit's previous wave data and imputations in which previous wave data are used as control variables. For detailed documentation about the coding of allocation (imputation) flags for specific variables, analysts can refer to the data dictionary for the data file with which they are working.

For items that receive Type Z imputations (in both the pre-1996 panels and the 1996 Panel) and items receiving EPPFLAG and little Type Z imputations in the 1996 Panel, the allocation (imputation) flag for a particular imputed item will not indicate by itself the imputation status of the item. For Type Z imputations, the EPPINTVW field in the 1996 Panel and the person-level INTVW field in the pre-1996 panels will indicate whether the Type Z procedure was used to impute all items for the sample person (in these cases, EPPINTVW = 3 or 4 or INTVW = 3 or

4).[7,8] The individual imputation flag for each item indicates whether or not that item was imputed during the processing of the donor's fields.

For EPPFLAG imputations, the EPPFLAG field will equal 1. When this is true, all labor force participation and job/business characteristics fields are imputed via the EPPFLAG procedure, whether or not the individual items indicate an imputation. As with the Type Z procedure, an allocation (imputation) flag with a value greater than zero for any of the labor force participation items means that the values of these items are not the original values from the donor but are processed values that are consistent with the sample person's demographics and household composition; for the job/business characteristics fields, an allocation flag with a value of "4" indicates that the sample person's values in these fields have been projected forward from the person's values for these fields in the previous wave.

To find little Type Z imputations, check the allocation (imputation) flag of the variable EPDJBTHN. If (a) EPDJBTHN = 1 (indicating that the person was a worker), (b) this item's allocation (imputation) flag is 1 or 4, and (c) EPPFLAG is not 1, then a little Type Z imputation has taken place for all of the labor force participation and job/business characteristics fields. As with the Type Z procedures, the allocation (imputation) flag for an individual item only indicates whether the item was imputed when the donor's fields were processed.

The full panel files carry only a subset of the allocation (imputation) flags carried on the core wave files. The value of an allocation (imputation) flag is set during wave processing, and, usually, it is not modified to reflect any changes in value resulting from the longitudinal editing discussed below. The Census Bureau does reset the values of some allocation flags to indicate that a longitudinal imputation has occurred.

Topical Module Imputation Procedures

When item-missing data in topical modules are imputed, the same sequential hot-deck procedure used to impute item-missing data in the SIPP core is used. Topical module data for Type Z noninterviews are also imputed item by item with the sequential hot deck. Those cases are *not* subjected to the Type Z imputation procedure that was used for core items in the pre-1996 panels.

[7] The codes for EPPINTVW and INTVW differ. In the 1996 Panel, EPPINTVW is coded as follows: 1 = Interview (self), 2 = Interview (proxy), 3 = Noninterview—Type Z, 4 = Noninterview—pseudo Type Z (left sample during the reference period), and 5 = Children under 15 during the reference period. In the pre-1996 panels, INTVW for person is coded as follows: 0 = Not applicable (children under 15), 1 = Interview (self), 2 = Interview (proxy), 3 = Noninterview—Type Z refusal, and 4 = Noninterview—Type Z other.

[8] Note that for the 1990–1993 Panels, INTVW can equal 5 on the core wave files (this value is not documented in the codebook). A value of 5 denotes persons in the sample early in the wave who were not in the sample at the time of interview. Such persons are processed as if they are a Type Z nonrespondent. Prior to the 1990 Panel, such persons are identified as those with PP-MIS5 (1 but PP-MISj ≠ 1 for j = 1, 2, 3, or 4.

Phase 2: Data Editing Procedures for the Full Panel Files

At the conclusion of each SIPP panel, core data from all waves are assembled into the full panel file. That assembly is done after all waves have been processed separately, producing the core wave files. Once all waves are linked, longitudinal edits are applied to the SIPP full panel files to ensure that the data for each respondent are consistent over time. Although the core wave files are edited for consistency, some types of inconsistencies become apparent only when looking at the data over multiple waves. Starting with the 1996 Panel, some longitudinal editing has been built into the CAI instrument. The ability to carry data across waves in the CAI environment is expected to result in better cross-wave consistency in the core wave files and in less need for subsequent longitudinal editing.[9]

Pre-1996 Full Panel Files

Because the specifications for editing the 1996 full panel files differ from those for the pre-1996 files, the following discussion refers only to pre-1996 procedures. Longitudinal edits in the pre-1996 panels were applied for selected variables. The edits were designed (1) to correct cross-wave inconsistencies, which become apparent only when multiple waves are examined together, and (2) to honor the preference to replace imputed values from one wave with reported values from another wave.

Unlike the hot-deck imputation procedures used with the core wave files, the longitudinal edits in the pre-1996 files did not replace missing data for one person with reported data from another person. When a data value was modified during longitudinal editing, the replacement value was obtained from the same record either directly (by copying a reported value from a different month) or indirectly (using some form of interpolation or extrapolation from reported values in other months). Those procedures could cause modifications both in reported and imputed values. When a data value was modified during longitudinal editing, the associated imputation flag was not changed. In addition, the core wave files were not revised to reflect changes made during longitudinal editing. Thus, the data for any given respondent may differ between the core wave files and the full panel file, and estimates based on the full panel file may differ from those based on the core wave files.

[9] Prior to CAI, a *control file* was developed at Wave 1 that contained a unique identifier for each sample person, as well as that person's age, sex, and race. In subsequent waves, the control file provided a means of detecting inconsistencies in age, sex, and race across waves. As each wave of data was received, the reported age, sex, and race of the sample person were checked against the control file and corrections were made. Also prior to CAI, income recipiency was brought forward to the subsequent wave.

The longitudinal edits in the pre-1996 files were performed independently on four groups of variables:

1. Demographic and household composition variables;
2. Earned income variables;
3. Other income variables, Food Stamp variables, WIC variables, and program coverage variables; and
4. Medical insurance variables.

In most cases, the values reported during Wave 1 were used as the standard against which inconsistencies were judged. Pennell (1993) provides detailed information about longitudinal consistency edits for specific variables.

1996 Full Panel File

The specifications for editing the 1996 full panel file are not yet complete. The basic difference between the pre-1996 and the 1996 full panel files is that the editing procedures for the 1996 panel incorporate longitudinal imputation based on prior wave information.

Missing Wave Imputation

There are many instances in which data are missing for a person in one or two consecutive waves but are present for that same person in the two adjacent waves. For example, a person may be missing in Wave 5 but have complete data for Waves 4 and 6. Beginning with the 1991 Panel, the Census Bureau began imputing those missing waves in the full panel files. Missing wave imputation is performed only when one or two consecutive missing waves are bounded on both sides by waves in which the sample member was present. If a respondent has missing data for more than two consecutive waves, the imputation is not performed.

For missing waves that are bounded on each side by interviewed waves, data are interpolated using a *random carryover* procedure. A value r is randomly assigned to each nonrespondent's household for each missing wave, where $r = 0, 1, 2, 3$, or 4. The first r reference months within the missing wave receive their imputed values from the fourth month of the preceding wave, and the remaining $4 - r$ reference months receive their imputed amounts from the first month of the subsequent wave.

Although this procedure results in data conducive to many analytic purposes, the random carryover forces stability in responses for wave nonrespondents. That stability could result in underestimation of between-wave changes. The procedure also results in imputed waves that do *not* exhibit the seam effect common to waves of reported data (Chapter 6). Williams and Bailey (1996) provide a complete account of the handling of missing wave data in SIPP.

Confidentiality Procedures for the Public Use Files

All of the editing and imputation procedures described in the preceding sections are part of the process of preparing the data for internal Census Bureau use. Before the files are released for public use, they undergo additional editing to protect the confidentiality of respondents. Two procedures are used: topcoding of selected variables (income, assets, and age) and suppression of geographic information. As a result of these procedures, estimates based on data from the public use files will differ slightly from the Census Bureau's published estimates.

Topcoding

One piece of information that might reveal a respondent's identity is a very high income. For that reason, the Census Bureau *topcodes* income before making that information publicly available, recoding any income amounts over a certain maximum value to that maximum. In other words, income on the public use data files has a ceiling value. Although income is the primary variable that is topcoded, other variables that may disclose a respondent's identity, such as age, are also topcoded. A few variables, such as starting dates for employment, may be *bottomcoded* if they pose a disclosure risk. Chapter 10 and Appendix B provide a thorough discussion of topcoding methods and procedures in SIPP.

Suppression of Geographic Information

Geographic information that can be used to directly identify survey respondents, such as an address, is removed from the public use files. In addition, states and metropolitan areas with populations less than 250,000 are not identified. Specific nonmetropolitan areas (such as counties outside of metropolitan areas) are never identified. In certain states, when the nonmetropolitan population is small enough to present a disclosure risk, a fraction of that state's metropolitan sample is recoded to nonmetropolitan status. For that reason, the SIPP data cannot be used to estimate characteristics of the population residing outside metropolitan areas. Chapter 10 provides details.

For the 1996 Panel, state-level geography is shown for 45 states and the District of Columbia. The remaining five states are combined as follows:

1. Maine, Vermont; and
2. North Dakota, South Dakota, Wyoming.

For the 1984 through 1993 Panels, state-level geography is shown for 41 individual states and the District of Columbia; the nine other states are combined into three groups:

1. Maine, Vermont;
2. Iowa, North Dakota, South Dakota; and
3. Alaska, Idaho, Montana, Wyoming.

5. Finding SIPP Information

Both the data collected in SIPP and supporting documentation are available in various forms. They include published estimates based on those data, microdata in several formats, documentation for each of the microdata files, and more general documentation about methodological issues in SIPP. The latter includes the *SIPP Quality Profile*, a series of working papers distributed by the Census Bureau, articles published in academic journals, and conference proceedings. This chapter discusses SIPP published estimates, briefly describes the data files and supporting documentation, and provides information on how to obtain them.

Published Estimates from SIPP

Published estimates from SIPP data are useful to data analysts in a number of ways. First, Census Bureau publications may already contain the estimates needed for the research project at hand, thus saving users the need to generate those estimates themselves. Second, published estimates can often provide a useful cross-check for closely related estimates prepared by analysts.

Published estimates are based on the Census Bureau's internal data files, and it is often impossible to replicate published estimates exactly. That is because the internal files have not been subjected to topcoding and other data-suppression techniques that are necessary to protect confidentiality on the public use microdata files. Chapter 4 provides information on data editing and imputation.

The Census Bureau's P-70 series of publications is the primary source for published estimates from SIPP. Table 5-1 displays the titles and publication numbers of reports in the series that are currently available from the Census Bureau. Copies of those reports can be obtained from the U.S. Government Printing Office, Washington, DC 20402. For telephone orders, users can call (202) 783-3238, or they can fax orders to (202) 783-3236. An updated list of P-70 series reports can be obtained from the SIPP Web site (http://www.bls.census.gov/sipp/); each of the reports contains a phone number the reader can call for further information or clarification. Users can reach the population division staff for demographics questions at (301) 457-2422, or they can call the SIPP information phone number: (301) 457-3242.

SIPP Public Use Microdata Files

Following data collection as described in Chapter 2 and postcollection processing as described in Chapter 4, the Census Bureau prepares data files in formats compatible with the most common methods of analysis. Those microdata are available in several file formats and can be obtained on

Table 5-1. Publications in the P-70 Series

Publication Number	Title
P-70-1	Economic Characteristics of Households in the U.S. Third Quarter 1983
P-70-2	Economic Characteristics of Households in the U.S. Fourth Quarter, 1983
P-70-3	Economic Characteristics of Households in the U.S. First Quarter,1984
P-70-4	Economic Characteristics of Households in the U.S. Second Quarter, 1984
P-70-5	Economic Characteristics of Households in the U.S. Third Quarter, 1984
P-70-6	Economic Characteristics of Households in the U.S. Fourth Quarter, 1984
P-70-7	Household Wealth and Asset Ownership, 1984
P-70-8	Disability, Functional Limitations, and Health Insurance Coverage: 1984-1985
P-70-9	Who's Minding the Kids? Child Care Arrangements: Winter 1984-1985
P-70-10	Male-Female Differences in Work Experience, Occupation, and Earnings: 1984
P-70-11	What's It Worth? Educational Background and Economic Status: Spring 1984
P-70-12	Pensions: Workers Coverage and Retirement Income, 1984
P-70-13	Who's Helping Out? Support Network Among American Families
P-70-14	Characteristics of Persons Receiving Benefits from Major Assistance Programs
P-70-15-RD-1	Transitions in Income and Poverty Status: 1984-1985
P-70-16-RD-2	Spells of Job Search and Layoff...and Their Outcomes
P-70-17	Health Insurance Coverage, 1986-1988
P-70-18	Transitions in Income and Poverty Status: 1985-1986
P-70-19	The Need for Personal Assistance with Everyday Activities: Recipients and Caregivers
P-70-20	Who's Minding the Kids? Child Care Arrangements: Winter 1986-1987
P-70-21	What's It Worth? Educational Background and Economic Status: Spring 1987
P-70-22	Household Wealth and Asset Ownership: 1988
P-70-23	Family Disruption and Economic Hardship: The Short-Run Picture for Children
P-70-24	Transitions in Income and Poverty Status: 1987-1988
P-70-25	Pensions: Worker Coverage and Retirement Benefits, 1987
P-70-26	Extended Measures of Well-Being: 1984
P-70-27	Job Creation During Late 1980's: Dynamic Aspects of Employment Growth
P-70-28	Who's Helping Out? Support Network Among American Families
P-70-29	Health Insurance Coverage: 1987 to 1990
P-70-30	Who's Minding the Kids? Child Care Arrangements: Fall 1988
P-70-31	Characteristics of Recipients and the Dynamics of Program Participation: 1987-1988
P-70-32	What's It Worth? Educational Background and Economic Status: Spring 1990
P-70-33	Americans with Disabilities: 1991-1992
P-70-34	Household Wealth and Asset Ownership: 1991
P-70-35	Monitoring the Economic Health of American Households: Average Monthly Estimates of Income, Labor Force Activity, Program Participation and Health Insurance, First Quarter 1984 to Third Quarter 1991
P-70-36	Who's Minding the Kids? Child Care Arrangements: Fall 1991
P-70-37	Dynamics of Economic Well-Being: Health Insurance, 1990-1992
P-70-38	The Diverse Living Arrangements of Children: Summer 1991
P-70-39	Dollars for Scholars: Postsecondary Costs and Financing, 1990-1991
P-70-40	Dynamics of Economic Well-Being: Labor Force and Income: 1990-1992
P-70-41	Dynamics of Economic Well-Being: Program Participation: 1990-1992

(table continues)

Table 5-1. Publications in the P-70 Series (*continued*)

Publication Number	Title
P-70-42	Dynamics of Economic Well-Being: Poverty: 1990
P-70-43	Dynamics of Economic Well-Being: Health Insurance: 1991-1993
P-70-44	The Effect of Health Insurance Coverage on Doctor and Hospital Visits: 1990-1992
P-70-45	Dynamics of Economic Well-Being: Poverty: 1991-1993
P-70-46	Dynamics of Economic Well-Being: Program Participation: 1991-1993
P-70-47	Asset Ownership of Households: 1993
P-70-48	Dynamics of Economic Well-Being: Labor Force: 1991-1993
P-70-49	Dynamics of Economic Well-Being: Income: 1991-1992
P-70-50	Beyond Poverty, Extended Measures of Well-Being: 1992
P-70-51	What's It Worth? Field of Training and Economic Status: 1993
P-70-52	What Does it Cost to Mind Our Preschoolers?
P-70-53	Who's Minding Our Preschoolers?
P-70-54	Who Loses Coverage and for How Long?
P-70-55	Dynamics of Economic Well-Being: Poverty: 1992-1993, Who Stays Poor? Who Doesn't?
P-70-56	Dynamics of Economic Well-Being: Income, 1992-1993, Moving Up and Down the Income Ladder
P-70-57	Dynamics of Economic Well-Being: Labor Force, 1992-1993—A Perspective on Low-Wage Workers
P-70-58	Dynamics of Economic Well-Being: Program Participation, 1992-1993—Who Gets Assistance?
P-70-59	My Daddy Takes Care of Me! Fathers as Care Providers
P-70-60	Financing the Future: Postsecondary Students, Costs, and Financial Aid
P-70-61	Americans with Disabilities: 1994-95
P-70-62	Who's Minding Our Preschoolers – Fall 1994 Update
P-70-63	Dynamics of Economic Well Being: Poverty, 1993-94
P-70-64	Who Loses Coverage, and For How Long?
P-70-65	Moving Up and Down the Income Ladder
P-70-66	Seasonality of Moves and Duration of Residence
P-70-67	Extended Measures of Well-Being: Meeting Basic Needs
P-70-69	Dynamics of Economic Well-Being: Program Participation, Who Gets Assistance?
P-70-70	Who's Minding the Kids? Child Care Arrangements
P-70-71	Household Net Worth and Asset Ownership, 1995
P-70-73	Americans With Disabilities: 1997

a variety of media. The following sections describe the file formats currently in use, each of which is used for somewhat different SIPP data. Information is also provided about how to obtain those data and supporting documentation.

Formats and Contents of SIPP Microdata Files

SIPP public use microdata are available in four types of files: core wave files, topical module files, and full and partial panel files. The files vary in content and structure. Analysts should be aware that their need for files depends on their particular application.

Data files are available through the Customer Services Branch, Administrative and Customer Services Division, at (301) 457-4100. Users can also extract data files by using on-line data access tools, as described later in this chapter in "Sources for Obtaining SIPP Microdata."

Core Wave Files

Core wave files contain the core labor force, income, household and family composition, and program participation data from one wave of interviews. The core wave files are currently available in *person-month* format, containing, for every person who was a member of a SIPP household for at least 1 month during the 4-month reference period for that wave, one record for each month that person was in-sample.[1] In other words, a person who was in-sample for all 4 reference months has four records—one for each reference month. A person who was in-sample for only 1 month would have just one record. The core wave files were designed to be used for cross-sectional analyses. Analysts who do not wish to wait for the release of certain files can link one or more core wave files to make their own longitudinal files. Chapter 13 discusses linking files. Table 5-2 illustrates the structure of the person-month format for core wave files.

The core wave files are the only source of monthly cross-sectional weights. When using data drawn from the full panel files for cross-sectional analyses, users must merge weights from the core wave files. Chapter 8 explains how to select and merge weights.

Topical Module Files

Each topical module file contains selected core information along with the data from the topical module administered in a given wave. As described in Chapter 2, different topical modules are administered in each wave of a SIPP panel. Table 5-3 shows which topical modules were administered for each wave of each SIPP panel. Table 5-4 lists topical areas along with the panels and waves in which they were administered. Topical module files are issued in person-record format; there is one record for each person who was a member of a SIPP household at the time of the interview for that wave. Table 5-5 illustrates the structure of a topical module file. For the topical modules, there are people for whom there is no topical information. Chapter 2 describes how the interviews are conducted and how topical module information is collected; Chapter 4 explains how missing data are handled in the files. In the 1996 Panel, the month that determines the universe for the topical module files changed to month 4.

[1] Prior to the 1990 Panel, the Census Bureau issued core wave files in a format with a single record for each person. Those files are described in earlier editions of the *SIPP Users' Guide*.

Table 5-2. Structure of the Person-Month Format Core Wave Files

SUID[a]	Person	Month	Household Vars	Family Vars	Subfamily Vars	Sample Status	Other Person Vars
1	1	1				Yes	
		2				Yes	
		3				Yes	
		4				Yes	
	2	1				Yes	
		2				Yes	
		3	Missing	Missing	Missing	No	Missing
		4	Missing	Missing	Missing	No	Missing
	3	1				Yes	
		2				Yes	
		3	Missing	Missing	Missing	No	Missing
		4				Yes	
2	1	1				Yes	
		2				Yes	
		3				Yes	
		4				Yes	
	2	1	Missing	Missing	Missing	No	Missing
		2				Yes	
		3				Yes	
		4				Yes	
3	1	1				Yes	
		2				Yes	
		3				Yes	
		4				Yes	
	2	1				Yes	
		2				Yes	
		3	Missing	Missing	Missing	No	Missing
		4	Missing	Missing	Missing	No	Missing
4	1	1				Yes	
		2				Yes	
		3				Yes	
		4				Yes	

[a] Sample unit ID number. Chapter 4 provides more information about identification numbers in SIPP.

Table 5-3. Topical Modules, by Panel and Wave

Wave	Subject Areas
1996 Panel	
1	Recipiency History, Employment History
2	Work Disability History, Education and Training History, Marital History, Migration History, Fertility History, Household Relationships
3	Assets, Liabilities, and Eligibility; Medical Expenses/Utilization of Health Care—Adults; Medical Expenses/Utilization of Health Care—Children; Work-Related Expenses; Child Support Paid
4	Annual Income and Retirement Accounts, Taxes, Work Schedule, Child Care, Disability Questions
5	School Enrollment and Financing, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability—Adults, Functional Limitations and Disability—Children, Employer-Provided Health Benefits
6	Children's Well-Being, Assets, Liabilities, and Eligibility, Medical Expenses/Utilization of Health Care—Adults, Medical Expenses/Utilization of Health Care—Children, Work-Related Expenses, Child Support Paid
7	Annual Income and Retirement Account, Taxes, Retirement and Pension Plan Coverage; Home Health Care
8	Adult Well-Being, Welfare Reform
9	Assets, Liabilities, and Eligibility; Medical Expenses/Utilization of Health Care—Adults; Medical Expenses/Utilization of Health Care—Children; Work-Related Expenses; Child Support Paid
10	Annual Income and Retirement Accounts, Taxes, Work Schedule, Child Care
11	Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability—Adults, Functional Limitations and Disability—Children
12	Assets, Liabilities, and Eligibility; Medical Expenses/Utilization of Health Care—Adults; Medical Expenses/Utilization of Health Care—Children; Work-Related Expenses; Child Support Paid; Children's Well-Being
1993 Panel	
1	Recipiency History, Employment History
2	Work Disability History, Education and Training History, Marital History, Migration History, Fertility History, Household Relationships
3	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability, Utilization of Health Care Services
4	Selected Financial Assets; Medical Expenses and Work Disability; Real Estate, Shelter Costs, Dependent Care, and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability—Adults, Utilization of Health Care Services—Adults, Functional Limitations and Disability—Children, Utilization of Health Care Services–Children, Children's Well-Being
7	Assets and Liabilities; Real Estate, Shelter Costs, Dependent Care, and Vehicles; Medical Expenses and Work Disability
8	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
9	Retirement Expectations and Pension Plan Coverage, Child Support Agreements, Child Care, Support for Nonhousehold Members, Work Schedule, Children's Well-Being, Basic Needs

(table continues)

Table 5-3. Topical Modules, by Panel and Wave (*continued*)

Wave	Subject Areas
1992 Panel	
1	Recipiency History, Employment History
2	Work Disability History, Education and Training History, Marital History, Migration History, Fertility History, Household Relationships
3	Extended Measures of Well-Being (Consumer Durables, Living Conditions, Basic Needs)
4	Assets and Liabilities, Retirement Expectations and Pension Plan Coverage, Real Estate Property and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability, Utilization of Health Care Services
7	Selected Financial Assets; Medical Expenses and Work Disability; Real Estate, Shelter Costs, Dependent Care, and Vehicles
8	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
9	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability—Adults, Utilization of Health Care Services—Adults, Functional Limitations and Disability—Children, Utilization of Health Care Services—Children, Children's Well-Being
10	No Topical Modules
1991 Panel	
1	No Topical Modules
2	Recipiency History, Employment History, Work Disability History, Education and Training History, Marital History, Migration History, Fertility History, Household Relationships
3	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability, Utilization of Health Care Services
4	Selected Financial Assets; Medical Expenses and Work Disability; Real Estate, Shelter Costs, Dependent Care, and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Extended Measures of Well-Being (Consumer Durables, Living Conditions, Basic Needs)
7	Assets and Liabilities, Retirement Expectations and Pension Plan Coverage, Real Estate Property and Vehicles
8	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
1990 Panel	
1	No Topical Modules
2	Recipiency History, Employment History, Work Disability History, Education and Training History, Marital History, Migration History, Fertility History, Household Relationships
3	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability, Utilization of Health Care Services
4	Assets and Liabilities, Retirement Expectations and Pension Plan Coverage, Real Estate Property and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Time Spent Outside Work Force, Child Support Agreements, Support for Nonhousehold Members, Functional Limitations and Disability, Utilization of Health Care Services
7	Selected Financial Assets; Medical Expenses and Work Disability; Real Estate, Shelter Costs, Dependent Care, and Vehicles
8	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing

(table continues)

Table 5-3. Topical Modules, by Panel and Wave (*continued*)

Wave	Subject Areas
1989 Panel	
1	No Topical Modules
2	Recipiency History, Employment History, Work Disability History, Education and Training History, Marital History, Migration History, Fertility History, Household Relationships
3	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Home Health Care, Disability Status and Utilization of Health Care Services, Functional Activities
4	The 1989 Panel was terminated following Wave 3.
1988 Panel	
1	No Topical Modules
2	Recipiency History, Employment History, Work Disability History, Education and Training History, Family Background, Marital History, Migration History, Fertility History, Household Relationships
3	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Long-Term Care, Disability Status of Children, Health Status and Utilization of Health Care Services
4	Selected Financial Assets; Medical Expenses and Work Disability; Real Estate, Shelter Costs, Dependent Care, and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Home Health Care, Disability Status of Children, Health Status and Utilization of Health Care Services, Functional Activities
7	No Wave 7
8	No Wave 8
1987 Panel	
1	No Topical Modules
2	Recipiency History, Employment History, Work Disability History, Education and Training History, Family Background, Marital History, Migration History, Fertility History, Household Relationships
3	Child Care Arrangements/Child Support Agreements, Support for Nonhousehold Members, Work-Related Expenses, Shelter Costs/Energy Usage
4	Assets and Liabilities, Real Estate Properties and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Work Schedule, Child Care, Child Support Agreements, Support for Nonhousehold Members, Long-Term Care, Disability Status of Children, Health Status and Utilization of Health Care Services
7	Selected Financial Assets; Medical Expenses and Work Disability; Real Estate, Shelter Costs, Dependent Care, and Vehicles
8	No Wave 8

(table continues)

Table 5-3. Topical Modules, by Panel and Wave (*continued*)

Wave	Subject Areas
1986 Panel	
1	No Topical Modules
2	Recipiency History, Employment History, Work Disability History, Education and Training History, Family Background, Marital History, Migration History, Fertility History, Household Relationships
3	Child Care Arrangements/Child Support Agreements, Support for Nonhousehold Members, Job Offers, Health Status and Utilization of Health Care Services, Long-Term Care, Disability Status of Children
4	Assets and Liabilities, Retirement Expectations and Pension Plan Coverage, Real Estate Property and Vehicles
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Child Care Arrangements/Child Support Agreements, Support for Nonhousehold Members, Work-Related Expenses, Shelter Costs/Energy Usage
7	Assets and Liabilities, Pension Plan Coverage, Real Estate Property and Vehicles
8	No Wave 8
1985 Panel	
1	No Topical Modules
2	No Topical Modules
3	Assets and Liabilities, Real Estate Property and Vehicles
4	Support for Nonhousehold Members/Work-Related Expenses, Marital History, Migration History, Fertility History, Household Relationships
5	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
6	Child Care Arrangements/Child Support Agreements, Support for Nonhousehold Members, Job Offers, Health Status and Utilization of Health Care Services, Long-Term Care, Disability Status of Children
7	Assets and Liabilities, Retirement Expectations and Pension Plan Coverage, Real Estate Property and Vehicles
8	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing
1984 Panel	
1	No Topical Modules
2	No Topical Modules
3	Education and Work History, Health and Disability
4	Assets and Liabilities; Retirement and Pension Coverage; Housing Costs, Conditions, and Energy Usage
5	Child Care, Welfare History and Child Support, Reasons for Not Working/Reservation Wage, Support for Nonhousehold Members/Work-Related Expenses
6	Earnings and Benefits, Property Income and Taxes, Education and Training
7	Assets and Liabilities, Pension Plan Coverage, Real Estate Property and Vehicles
8	Support for Nonhousehold Members/Work-Related Expenses, Marital History, Migration History, Fertility History, Household Relationships
9	Annual Income and Retirement Accounts, Taxes, School Enrollment and Financing

Table 5-4. Topical Modules, by Subject

Subject Areas	Panel and Wave[a]
Marital History	84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2
Fertility History	84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2
Household Relationships	84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2
Migration History	84-8, 85-4, 86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2
Family Background	86-2, 87-2, 88-2
Annual Income and Retirement Accounts	84-9, 85-5, 85-8, 86-5, 87-5, 88-5, 90-5, 90-8, 91-5, 91-8, 92-5, 92-8, 93-5, 93-8, 96-4, 96-7, 96-10
Taxes	84-9, 85-5, 85-8, 86-5, 87-5, 88-5, 90-5, 90-8, 91-5, 91-8, 92-5, 92-8, 93-5, 93-8, 96-4, 96-7, 96-10
Assets and Liabilities	84-4, 84-7, 85-3, 85-7, 86-4, 86-7, 87-4, 90-4, 91-7, 92-4, 93-7, 96-3, 96-6, 96-9, 96-12
Selected Financial Assets	87-7, 88-4, 90-7, 91-4, 92-7, 93-4
Retirement Expectations and Pension Plan Coverage	84-4, 85-7, 86-4, 86-7, 87-4, 90-4, 91-7, 92-4, 93-9, 96-7
Pension Plan Coverage	84-7, 86-8
Earnings and Benefits	84-6
Recipiency History	86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-1, 93-1, 96-1
Child Support Agreements	85-6, 86-3, 86-6, 87-3, 87-6, 88-3, 88-6, 89-3, 90-3, 90-6, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-5, 96-11
Child Support Paid	96-3, 96-6, 96-9, 96-12
Child Care	84-5, 85-6, 86-3, 86-6, 87-3, 87-6, 88-3, 88-6, 89-3, 90-3, 90-6, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-4, 96-10
Support for Nonhousehold Members	84-3, 84-5, 84-8, 85-4, 85-6, 86-3, 86-6, 87-3, 87-6, 88-3, 88-6, 90-3, 90-6, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-5
Welfare History and Child Support	84-5
Real Estate Property and Vehicles	84-7, 85-3, 85-7, 86-4, 86-7, 87-4, 90-4, 91-7, 92-4, 93-7
Real Estate, Shelter Costs, Dependent Care, and Vehicles	87-7, 88-4, 90-7, 91-4, 92-7, 93-4, 93-7
Shelter Costs/Energy Usage	86-6, 87-3
Property Income and Taxes	84-6
Housing Costs, Conditions, and Energy Usage	84-4
Employment History	86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-1, 93-1, 96-1
Work—Disability History	86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2
Work Schedule	87-6, 88-3, 88-6, 89-3, 90-3, 91-3, 92-6, 92-9, 93-3, 93-6, 93-9, 96-4, 96-10
Work-Related Expenses	84-5, 84-8, 85-4, 86-6, 87-3, 96-3, 96-6, 96-9, 96-12
Reasons for not Working/Reservation Wage	84-5
Time Spent Outside Work Force	90-6
Job Offers	85-6, 86-3
Home-Based Self-Employment/Size of Firm	92-6, 93-3
Education and Training History	86-2, 87-2, 88-2, 89-2, 90-2, 91-2, 92-2, 93-2, 96-2
Education and Work History	84-3
School Enrollment and Financing	84-9, 85-5, 85-8, 86-5, 87-5, 88-5, 90-5, 90-8, 91-5, 91-8, 92-5, 92-8, 93-5, 93-8, 96-5
Education and Training	84-6
Functional Limitations and Disability	90-3, 90-6, 91-3, 92-6, 93-3

(table continues)

Table 5-4. Topical Modules, by Subject (*continued*)

Subject Areas	Panel and Wave[a]
Functional Limitations and Disability—Adults	92-9, 93-6, 96-5, 96-11
Functional Limitations and Disability—Children	92-9, 93-6, 96-5, 96-11
Disability Status of Children	85-6, 86-3, 87-6, 88-3, 88-6, 89-3
Functional Activities	88-6, 89-3
Medical Expenses and Work Disability	87-7, 88-4, 90-7, 91-4, 92-7, 93-4, 93-7
Utilization of Health Care Services	90-3, 90-6, 91-3, 92-6, 93-3
Utilization of Health Care Services—Adults	92-9, 93-6, 96-5, 96-12
Utilization of Health Care Services—Children	92-9, 93-6, 96-5, 96-12
Health Status and Utilization of Health Care Services	85-6, 86-3, 87-6, 88-3, 88-6, 89-3
Long-Term Care	85-6, 86-3, 87-6, 88-3
Home Health Care	88-6, 89-3
Health and Disability	84-3
Employer-Provided Health Benefits	96-5
Disability Questions	96-4
Extended Measure of Well-Being (Consumer Durables, Living Conditions, Basic Needs)	91-6, 92-3
Adult Well-Being	96-8
Basic Needs	93-9
Welfare Reform	96-8
Children's Well-Being	92-9, 93-6, 93-9, 96-6, 96-11

[a] The number preceding the hyphen indicates the year of the panel, and the number following the hyphen indicates the wave number. Thus, 84-8 denotes that the information was collected in the 1984 Panel, during Wave 8.

Table 5-5. Structure of Topical Module Microdata File

SUID[a]	Person	Interview Status in Interview Month	Core Vars	Topical Module Vars
1	1	Yes		
	2	Yes		
	3	No	Missing	Missing
	4	Yes		
	5	No	Missing	Missing
2	1	Yes		
	2	Yes		
3	1	Yes		
4	1	Yes		
	2	No	Missing	Missing
	3	Yes		
5	1	Yes		
	2	Yes		
	3	Yes		

[a] Sample unit ID number. Chapter 4 provides more information about identification numbers in SIPP.

Full and Partial Panel Files

At the conclusion of each panel, the Census Bureau creates a single full panel file containing all data from the core wave files for every person who was a member of the SIPP sample at any time during the life of that panel.[2] To date, the full panel files have been issued in a format that contains one record for each person. That record contains either data or missing value codes for most core questionnaire items for every month of the panel.[3] Chapter 3 discusses survey content, including information about the content of the core questionnaire. At the time that this Guide was written, full panel files had been issued for all SIPP panels prior to the 1996 Panel. Because of the extended (4-year) duration of the 1996 Panel, the Census Bureau is modifying its procedures for releasing information for the full panel.

Sources for Obtaining SIPP Microdata

SIPP microdata files can be obtained from several sources. All public use microdata files can be obtained on magnetic media or CD-ROM directly from the Census Bureau. When microdata files are obtained directly from the Census Bureau, users are provided with a full set of documentation for those files, including all currently available applicable User Notes (discussed later in this chapter). Users can also be placed on a distribution list to receive information from the Census Bureau regarding any errors found in, or revisions made to, those files, by contacting the Customer Services Branch, Administrative and Customer Services Division, at (301) 457-4100.

In addition, analysts affiliated with institutions that are members of the Inter-university Consortium for Political and Social Research (ICPSR) can obtain all SIPP microdata from that source. Users should contact the ICPSR representative at their institutions for more information. Finally, SIPP data and documentation, as released by the Census Bureau, are not copyrighted. The data files and supporting documentation can therefore be freely copied and distributed to other users.[4]

There is another source of SIPP data that can be quite useful for simple exploratory work. SIPP microdata are available on-line at the Census Bureau's Web site (http://www.census.gov/) and from the SIPP Web site (http://www.sipp.census.gov/sipp/). Those Internet sites offer two data access tools—Surveys-on-Call, which is part of the Data Extraction System (DES), and FERRET, which is part of the new Census Bureau Data Access and Dissemination System (DADS).

Surveys-on-Call provides access to SIPP longitudinal files for the 1988 through 1993 Panels and for wave and topical module files for the 1990 through 1993 Panels. Surveys-on-Call allows users to define microdata extracts from the SIPP public use microdata files. Users can choose

[2] Because of the volume of data collected in the 1996 Panel, that procedure may not occur for the 1996 full panel file.

[3] In the case of items that are asked only once per interview rather than for each month of the 4-month reference period, there is a field for each interview rather than for each month.

[4] This provision pertains only to materials authored and distributed by the Census Bureau or other federal agencies. It does not imply any rights to copy and distribute material published by any other party.

data for selected years, wave files, core files, topical module files, or longitudinal files. They can also select variables of interest and use variables as selection criteria. For example, an analyst might want to extract recipiency information for females between the ages of 18 and 25 from Wave 5 of the 1993 Panel. Once defined, analysts can download those extracts to their own computers for analysis. Surveys-on-Call creates microdata extracts from the SIPP public use files only. It does not include any options for performing analyses on-line. On-line help is available at each step of the data-extraction process. Users are encouraged to explore the capabilities of this system by creating several small extracts.

SIPP data available on the Federal Electronic Research Review and Extraction Tool (FERRET) include files from the 1996 Panel and the longitudinal files from the 1992 and 1993 Panels. FERRET is the product of a joint project of the U.S. Census Bureau and the Bureau of Labor Statistics. It is a system enabling users to access and manipulate large demographic and economic data sets on-line. FERRET is designed to aid not only sophisticated researchers, but also reporters, students, government policy makers, and amateur statisticians. SIPP is one of several surveys available through FERRET.[5]

Other Sources of Information About SIPP

Other sources of information about SIPP include the *SIPP Quality Profile*, User Notes, and SIPP working papers. The SIPP Web site includes an extensive bibliography that provides references to SIPP-related research and documentation, data dictionaries, variable metadata documenting all information relevant to variables that appear on the public use microdata files, and a computer-based tutorial that introduces users to methods and concepts needed to use SIPP data.

SIPP Quality Profile

The *SIPP Quality Profile* documents data quality issues related to SIPP. It summarizes what is known about the sources and magnitude of errors in estimates based on SIPP. The *SIPP Quality Profile* covers both sampling and nonsampling error, with an emphasis on nonsampling error. There have been three editions of the *SIPP Quality Profile*. The third edition, by Kalton, Winglee, & Jabine (U.S. Census Bureau, 1998a), updates the two previous editions, by King, Petroni, & Singh (U.S. Census Bureau, 1987) and Jabine, King, & Petroni (U.S. Census Bureau, 1990). The third edition of the *SIPP Quality Profile* is available on-line at the SIPP Web site.

[5] Among the current and future topics accessible through FERRET are employment, health care, education, race and ethnicity, health insurance, housing, income and poverty, aging, marriage, and the family. FERRET allows users to quickly locate current and historical information from survey sources, get tabulations for specific information they need, make comparisons between different data sets, create simple tables, and download large amounts of data to desktop and larger computers for custom reports.

SIPP User Notes

The SIPP User Notes, issued periodically by the Census Bureau, contain updated information for specific microdata files. The User Notes include corrections to the data dictionaries, announcements of errors found in the public use data files after their release, and recommended corrections for those data errors. Analysts obtaining SIPP microdata files directly from the Census Bureau will receive all User Notes that have been issued for those files at the time of purchase. Users who obtained files from other sources should contact the Customer Services Branch, Administrative and Customer Services Division, at (301) 457-4100, to request the User Notes that have been issued for the data they plan to use. User Notes are also available at the SIPP Web site (http://www.sipp.census.gov/sipp/).

Microdata Technical Documentation

Users purchasing SIPP microdata files directly from the Census Bureau receive, along with the data files, a package of technical documentation. The technical documentation includes:

- A data dictionary, containing information about the file structure and the names, locations, and contents of all variables. The printed version of the data dictionary also includes information about the structure of the machine-readable data dictionary supplied with each file.
- A source and accuracy statement, containing detailed information about sample weights and computation of standard errors using Census Bureau generalized variance procedures. This information is specific to the panel, wave, and content of the data file. For example, the topical module file and the core wave file for Wave 7 of the 1990 Panel have different source and accuracy statements.
- A copy of the questionnaire screens and program code used to collect the information contained in the microdata file for the computer-assisted interviews for the 1996 Panel, which is available from the SIPP Web site (Chapter 2).

SIPP Working Papers

The Census Bureau publishes a series of SIPP working papers. Those papers are written by authors inside the Census Bureau and by outside analysts. The series includes research papers based on SIPP data or related to the SIPP program. SIPP working papers can be obtained from the SIPP Web site (http://www.sipp.census.gov/sipp/) or ordered from the Customer Services Branch, Administrative and Customer Services Division, at (301) 457-4100.

Bibliography

A bibliography of works related to SIPP is available on-line from the SIPP Web site. This relatively comprehensive bibliography contains references for journal articles, research papers, and working papers that use SIPP data or that discuss the SIPP survey.

Variable Metadata

Variable metadata, available in the data dictionary, provide a complete characterization of a variable's content. Variable metadata include all information relevant to variables that appear in the SIPP public use microdata files, including the variable name, a description of the variable, the concept label, data type (binary or character), suggested weight variable when applicable, descriptions of all possible values, and other data when applicable. A variable summary will be included for each public use variable. The summary identifies all edits, recodes, and imputations that affect the final edited output variable.

What's Available from the Survey of Income and Program Participation?

What's Available from the Survey of Income and Program Participation?, published by the Census Bureau, provides a complete directory of available SIPP data and publications. The directory lists materials available in both print and electronic formats. *What's Available* includes a listing of SIPP working papers, User Notes, public use microdata files, P-70 series population reports, and compilations of relevant papers published in the proceedings from the annual meetings of the American Statistical Association (ASA). *What's Available from the Survey of Income and Program Participation?* is updated periodically. Users can review the most recent edition at the Census Bureau Web site.

Table 5-6 lists telephone numbers to call for obtaining additional information about specific aspects of SIPP.

Table 5-6. Telephone Numbers for Information About Specific Aspects of SIPP

Subject Fields	Telephone Number
Adult well-being	(301) 457-2464
Aging population	(301) 457-2738
Child care	(301) 457-2416
Child well-being	(301) 457-2465
Education	(301) 457-2464
Fertility	(301) 457-2449
Health insurance	(301) 457-3215
Income	(301) 457-3212
Labor force, employment, and earnings	(301) 457-3229
Marriage and family	(301) 457-2416
Migration	(301) 457-2454
Pensions	(301) 457-3229
Poverty	(301) 457-3215
Wealth (assets)	(301) 457-3229
Women	(301) 457-2378
Methodology	**Telephone Number**
Data collection procedures	(301) 457-3819
Questionnaire design	(301) 457-3819
Estimation and weighting	(301) 457-4192
Nonsampling and sampling errors	(301) 457-4192
Survey design	(301) 457-4192

6. Nonsampling Errors

This chapter summarizes information about nonsampling errors in the Survey of Income and Program Participation (SIPP) that may affect the results of certain types of analyses. All surveys are subject to various sources of nonsampling errors, and SIPP is no exception. Nonsampling errors in SIPP include those that are found in most surveys as well as errors that arise because of SIPP's panel nature. The chapter focuses on the extent of nonsampling errors in SIPP and the impact of those errors on some survey estimates. The following topics are discussed:

- Undercoverage;
- Nonresponse;
- Measurement errors; and
- Effects of nonsampling errors on some survey estimates.

Undercoverage

One source of error in SIPP, as in other household surveys, is differential undercoverage of demographic subgroups. Black males over 15 years of age are most affected by undercoverage. The coverage ratio for this subgroup was about 0.82 in the 1990 and 1991 SIPP Panels. (Coverage ratio is computed as the survey estimate of the number in the subgroup before post-stratification, divided by a population estimate for the subgroup from population projections based on the most recent census.) For black males in their mid to late 20s, the coverage ratio was lower, about 0.65 in the same panels (*SIPP Quality Profile*, 3rd Ed. [U.S. Census Bureau, 1998a, Chapter 3]; hereinafter in this chapter, *SIPP Quality Profile*, 3rd Ed). These coverage ratios may understate the magnitude of the coverage problems because census undercounts are not reflected in the coverage ratios before 1992. Undercoverage in household surveys is attributed mainly to within-household omissions; the omission of entire households is less frequent. Shapiro et al. (1993) estimated that about 70 percent of the undercoverage for young black males consists of within-household omissions; the corresponding percentage for the white population is about 60 percent. To compensate for undercoverage, the Census Bureau uses population controls to adjust SIPP weights. Little is known about the effectiveness of the adjustments in reducing biases.

Nonresponse

Nonresponse is a major concern in SIPP because of the need to follow the same people over time. In SIPP, nonresponse can occur at several levels: household nonresponse at the first wave and thereafter; person nonresponse in interviewed households; and item nonresponse, including

complete nonresponse to topical modules. At the household level, the rate of sample loss for the 1991 Panel rose from about 8 percent at Wave 1 to more than 21 percent by Wave 8. For the same panel, 23 percent of the original sample persons who participated in Wave 1 missed one or more interviews for which they were eligible in later waves. At the item level, the nonresponse rate is typically around 10 percent or less for items on income amounts but somewhat higher for items on asset amounts. Nonresponse reduces the effective sample size (and, therefore, increases sampling error) and introduces bias in the survey estimates. The Census Bureau uses a combination of weighting and imputation methods to reduce the biasing effects of nonresponse at all three levels in SIPP. The effectiveness of those procedures remains a matter of ongoing review and research (*SIPP Quality Profile*, 3rd Ed., Chapters 4, 5, and 8).

Measurement Errors

Measurement errors are associated with the data collection phase of the survey. They may vary across SIPP panels because of changes in data collection procedures over the years. Most core survey items in SIPP are used consistently at every panel, although there have been occasional changes to improve the clarity of some items. The data collection method, which was face-to-face interviewing for the early panels, was changed to a maximum use of telephone interviewing in February 1992. Telephone interviewing was used as the primary mode of data collection between February 1992 and January 1996 for all waves except Waves 1, 2, and 6, for which face-to-face interviewing was used. The switch to telephone interviewing has had no known adverse effects on data quality.

Computer-assisted interviewing (CAI) was introduced with the 1996 SIPP Panel. The effects of CAI on survey responses have yet to be determined (*SIPP Quality Profile*, 3rd Ed., Section 11.3). For the 1996 Panel, computer-assisted personal interviewing (CAPI) was used for Waves 1 and 2. After Wave 2, the field representatives used the CAI instrument in face-to-face interviews with approximately one-third of the respondents; for the remaining interviews, the field representatives used the CAI instrument but conducted telephone interviews from their homes.

The combination of face-to-face interviews and telephone interviews used across waves is prespecified and varies for different subgroups of the sample according to the following scheme (Waite, 1996). Sample members are assigned to one of three interviewing mode subgroups. For each subgroup, a pattern of interviewing modes is designated and repeated every three waves. Thus, for Waves 3, 4, and 5, subgroup 1 is assigned the sequence face-to-face, telephone, telephone; subgroup 2, the sequence telephone, face-to-face, telephone; and subgroup 3, the sequence telephone, telephone, face-to-face. Under this scheme, which is applied with each rotation group, one-third of the sample is interviewed in person each wave and each month, and every household is interviewed in person once a year. The same sequence is repeated for Waves 6 and beyond, with a cycle of three waves (*SIPP Quality Profile*, 3rd Ed.).

Response errors in SIPP include errors of recall, errors in proxy respondents' reports, and other errors associated with the panel nature of SIPP. SIPP uses a 4-month recall period to reduce

memory error, and respondents are encouraged to use financial records and an event calendar to facilitate recall. Although the level of accuracy for self-response is generally believed to be higher than for proxy response (see Moore, 1988, for a contrary view), achieving a higher proportion of self-response would increase data collection costs and might lead to some increase in person nonresponse rates (*SIPP Quality Profile*, 3rd Ed., Section 4.5.3).

A potential source of response error that arises from the panel nature of SIPP is the time-in-sample effect (or panel conditioning). This effect occurs when the responses given at later waves are affected by the respondents' experiences of being interviewed in previous waves. The extent of this error is difficult to evaluate because it is often confounded with other sources of error, particularly attrition. Thus far, studies have found little evidence of systematic biases resulting from time-in-sample effects (Pennell and Lepkowski, 1992; McCormick et al., 1992).

Measurement errors can also occur when respondents misinterpret questions. For example, when asked about earnings, some respondents may have reported take-home pay instead of gross earnings. There is also some evidence of confusion in regard to welfare programs, such as the old Aid to Families with Dependent Children and general assistance programs.

Another response error identified through the panel nature of SIPP is the seam phenomenon. Research has consistently indicated that respondents tend to report the same status (e.g., employment or program participation) and the same amounts (e.g., Social Security income) for all 4 months within a wave, with most reported changes occurring between the last month of one wave and the first month of the subsequent wave. This phenomenon results in an overstatement of changes at the on-seam months (the boundary between interviews in successive waves of a panel) and an understatement of changes at the off-seam months. The seam phenomenon affects most variables for which monthly data are collected. As a result of the rotation group pattern, the phenomenon has relatively small effects on cross-sectional estimates based on all four rotation groups. That is because there is only one rotation group (or one-fourth of the sample) that is on seam and three rotation groups off seam for any given pair of calendar months. The effects of the seam phenomenon on longitudinal estimates are not well known (*SIPP Quality Profile*, 3rd Ed., Chapter 6).

Effects of Nonsampling Error on Survey Estimates

A considerable amount of research has been conducted to investigate the various sources of nonsampling error in SIPP. The results of the research are summarized in the *SIPP Quality Profile*, 3rd Ed.). The research includes, for example, the SIPP Record Check Studies (Marquis and Moore, 1989a,b, 1990; Marquis et al., 1990) that compared SIPP responses on program participation with administrative records. Despite the volume of this methodological research, it remains difficult to quantify the combined effects of nonsampling errors on SIPP estimates. The problem is made more complex because the effects of nonsampling error of different types on survey estimates vary, depending on the estimate under consideration. There are, however, some

findings about nonsampling error that SIPP users should bear in mind when conducting their analyses and examining their results. Those findings include the following:

- Some demographic subgroups are underrepresented in SIPP because of undercoverage and nonresponse. They include young black males, metropolitan residents, renters, people who changed addresses during a panel (movers), and people who were divorced, separated, or widowed. The Census Bureau uses weighting adjustments and imputation to correct the underrepresentation. Those procedures, however, may not fully correct for all potential biases (*SIPP Quality Profile*, 3rd Ed., Chapter 8).
- The SIPP estimates of income from Social Security, Railroad Retirement, and Supplemental Security programs represent more than 95 percent of the amounts reported by administrative sources. The SIPP estimates of unemployment income, workers' compensation income, veteran's income, and public assistance income, however, are low relative to the amounts reported by administrative sources (Coder and Scoon-Rogers, 1996).
- Evaluation studies typically find that SIPP estimates (as well as other survey estimates) of property income are generally poor. Among the different types of property income, reports of interest and dividend income are most prone to error. Respondents are often confused about those two sources of income, and both sources tend to be underreported (Coder and Scoon-Rogers, 1996).
- SIPP estimates of assets, liabilities, and wealth are low relative to estimates from the Federal Reserve Board (Eargle, 1990).
- For SIPP panels before 1996, the estimates of the percentages of people in poverty were lower than those found in the Current Population Survey (CPS) (Shea, 1995a).
- SIPP estimates of the working population differ from those produced from CPS. The differences may be explained largely by substantial conceptual and operational differences in the collection of labor force data in the two surveys (*SIPP Quality Profile*, 3rd Ed., Chapter 10).
- The SIPP estimates of people without any health insurance coverage are much lower than the CPS estimates. There are reasons to believe that the SIPP estimates are more accurate (McNeil, 1988).
- The SIPP estimates of the number of births compare favorably with the CPS estimates. Both surveys, however, provide estimates that are low relative to the records from the National Center for Health Statistics (NCHS). The SIPP estimates of the number of marriages are fairly comparable with the NCHS counts, but the SIPP estimates of the number of divorces are consistently lower than the NCHS estimates (*SIPP Quality Profile*, 3rd Ed., Chapter 10).

In spell analyses, Kalton et al. (1992) found that spell durations of multiples of 4 months (e.g., 4 months, 8 months, 12 months) were particularly common, a feature that can be explained by the seam phenomenon.

7. Sampling Error

This chapter discusses methods for obtaining the sampling error estimates derived from the Survey of Income and Program Participation (SIPP) panels. The sample selected for each SIPP panel is a stratified multistage probability sample. This complex sample design needs to be taken into account when estimating the variances of SIPP estimates. The SIPP data files contain variables, related to the sample design, that are created for the purpose of variance estimation. Several software packages are now available for computing variance estimates for a wide range of statistics based on complex sample designs. Using the variables that specify the design, these programs can calculate appropriate variances of survey estimates. The Census Bureau also provides generalized variance functions (GVFs) that can be used to obtain approximate estimates of sampling variance for SIPP estimates.

A common mistake in the estimation of sampling error for survey estimates is to ignore the complex survey design and treat the sample as a simple random sample (SRS) of the population. That mistake occurs because most standard software packages for data analyses assume simple random sampling for variance estimation. When applied to SIPP estimates, SRS formulas for variances typically underestimate the true variances. This chapter describes how appropriate variance estimates, which take into account the complex sample design, can be obtained for SIPP estimates.

The topics discussed in this chapter are:

- Direct variance estimation;
- Approximate variance estimates obtained from GVFs; and
- Variance estimation when some data are imputed.

Direct Variance Estimation

The primary sampling unit (PSU) plays a key role in variance estimation with a multistage sample design. SIPP PSUs are mostly counties, groups of counties, or independent cities (*SIPP Quality Profile*, 3rd Ed. [U.S. Census Bureau, 1998a, Chapter 3]), which are sampled with probability proportional to size within strata. The PSUs are sampled without replacement so that no PSU is selected more than once for the sample. Some PSUs are so large that they are included in the sample with certainty. Because no sampling is involved, those PSUs are, in fact, not PSUs but strata. The actual PSUs for those certainty selections are the enumeration districts and other units selected within them.

Although the SIPP PSUs are selected without replacement (as is the case with most multistage designs), for the purpose of variance estimation they are treated as if they were sampled with replacement. The with-replacement assumption greatly facilitates variance estimation since it means that variance estimates can be computed by taking into account only the PSUs and strata, without the need to consider the complexities of the subsequent stages of sample selection. This widely used simplifying assumption leads to an overestimation of variances, but the overestimation is not great.

Several software packages are available for computing variances of a wide range of survey estimates (e.g., means and proportions for the total sample and for subclasses, for differences in means and proportions between subclasses, and for regression and logistic regression coefficients) from complex sample designs. Many of these packages are listed on the Web: http://www.fas.harvard.edu/~stats/survey-soft/survey-soft.html. Lepkowski and Bowles (1996) examined eight of the packages.

These packages use a variety of methods for variance estimation. Some use an approach based on a Taylor-series approximation, or linearization, method. Others use a replication method, such as jackknife repeated replications or balanced repeated replications. Although some methods have advantages in some situations, there is generally little to recommend one method over another. The variance estimates they produce are not identical, but the differences are usually small. See Wolter (1985) and Rust (1985) for discussions of these methods.

Variance Units and Variance Strata, 1990–1993 Panels

For the 1990–1993 SIPP Panels, the sample member record contains information concerning the PSU and stratum within which the member was sampled. This information is needed as input for all of the specialized software packages. The original PSU and strata codes are not included in the SIPP public use data files, however, to avoid potential identification of small geographic areas and sampled individuals. Instead, sets of PSUs are combined across strata to produce variance units and variance strata, with two variance units in each variance stratum. Variance units and variance strata may be treated as PSUs and strata for variance estimation purposes. Their use does not give rise to any bias in the variance estimates. The variance estimates are somewhat less precise, however, than those obtained from the use of the PSUs and strata that have not been combined.

Under the complex sample design, the number of degrees of freedom for variance estimation depends on the number of variance strata. The 1984 SIPP Panel consists of 142 variance units in 71 variance strata; the panels between 1985 and 1991 have 144 variance units and 72 variance strata; and the 1992–1993 Panels have 198 variance units and 99 variance strata. As a rough approximation, the number of degrees of freedom for a variance estimate is the number of variance strata. Thus, for national estimates, the variance estimates have about 71 degrees of freedom for the 1984 Panel, 72 degrees of freedom for the 1985–1991 Panels, and 99 degrees of freedom for the 1992–1993 Panels. Regional estimates will have fewer degrees of freedom because such estimates include only some of the variance strata.

Table 7-1 displays the variable names for the variance stratum and variance unit codes in the SIPP core wave files and the SIPP full panel files. These codes can be employed as stratum and PSU codes in any of the software packages for variance estimation with complex sample designs.

Table 7-1. Variance Stratum Code and Variance Unit Code in SIPP Files, 1990–1993

Variable for Variance Estimation:	SIPP Core Wave File	SIPP Full Panel File
Variance stratum code	HSTRAT	VARSTRAT
Variance unit (or half-sample) code	HHSC	HALFSAMP

Replication Weights for the 1996 Panel

Analysts should use Fay's method for estimating variances for the 1996 SIPP Panel. Fay's method is a modified balanced repeated replication (BRR) method of variance estimation. The difference between the basic BRR method and Fay's method is that the BRR method uses replicate factors of 0 and 2, whereas Fay's method uses one factor, k, which is in the range (0, 1), with the other factor equal to $2 - k$. In Fay's method, the introduction of the perturbation factor $(1 - k)$ allows the use of both halves of the sample. Thus, Fay's method has the advantage that no subset of the sample units in a particular classification will be totally excluded. The variance formula for Fay's method is

$$\mathrm{Var}(\theta_0) = \{1/[G(1-k)^2]\} \sum_{i=1}^{G} (\theta_i - \theta_0)^2, \tag{7-1}$$

where

G = number of replicates;

$1 - k$ = perturbation factor;

i = replicate i, $i = 1$ to G;

θ_i = ith estimate of the parameter θ based on the observations included in the ith replicate;

θ_0 = survey estimate of the parameter θ based on the full sample.

The 1996 SIPP Panel uses 108 replicate weights, which are calculated on the basis of a perturbation factor of 0.5 ($k = 0.5$). Inserting those values into Equation (7-1) results in the 1996 SIPP Panel variance formula of

$$\mathrm{Var}(\theta_0) = [1/(108 * 0.5^2)] \sum_{i=1}^{108} (\theta_i - \theta_0)^2.$$

The Census Bureau used VPLX software to compute the replicate weights that are available through FERRET.

Using GVFs to Approximate Variance Estimates

The Census Bureau provides two forms for approximate variance estimation: GVFs and tables of standard errors (the square root of the variance) for different estimated numbers and percentages. The generalized estimates provide indications of the magnitude of the sampling error in the survey estimates. They serve as convenient ways to summarize the sampling errors for a broad variety of estimates.

The GVFs for SIPP were derived by modeling the standard error behavior of groups of estimates with similar standard errors. The mathematical form of the function adopted is

$$s = (ax^2 + bx)^{1/2}, \tag{7-2}$$

where s represents the standard error and x the value of an estimate. The parameters a and b are derived on the basis of a selected group of estimates. They are updated annually and are included in the source and accuracy statement that accompanies each SIPP data file for a panel. It is essential to use the parameter estimates for a specific panel and to follow the instructions to apply necessary adjustments to obtain the correct estimates for subgroups. Besides GVFs, the Census Bureau provides summary tables of general standard errors. Those estimates are also available in the source and accuracy statements. The following examples show how to use GVFs to estimate the standard errors of estimated numbers and of sample means. The use of GVFs and tables of standard errors is described in the source and accuracy statements for each panel.

Before looking at the examples, the user should note that the generalized variance estimates for estimating the standard errors of other statistics may not be accurate for small subgroups. Using the 1984 SIPP Panel, Bye and Gallicchio (1989) developed variance functions for participants of Old-Age, Survivors, and Disability Insurance (OASDI) and Supplemental Security Income (SSI) programs. They found that for estimates of less than 10 million, the generalized standard error estimates provided by the Census Bureau were 1.20 to 1.75 times larger than those obtained from the variance functions developed specifically for that subgroup.

Using GVFs for Standard Errors of Estimated Numbers

The approximate standard error, s, of an estimated number of persons (or households, and families) can be obtained by the formula

$$s = (ax^2 + bx)^{1/2}, \tag{7-3}$$

where a and b are the parameters associated with the estimate for the particular reference period, and x is the weighted estimate. This equation is appropriate for the standard errors of estimated numbers and should not be applied to estimates of dollar values.

Suppose that the number of households with monthly household income above $6,000 is estimated from Wave 1 of the 1991 Panel to be 472,000. The approximate values of a and b from Table 6 of the source and accuracy statement of the 1991 Panel are a = -0.0001005 and b = 9,286. Then, the standard error, s, of this estimated number is given by

$$s = [(-0.0001005 * 472{,}000^2) + (9{,}286 * 472{,}000)]^{1/2} = 66{,}000.$$

The approximate 90 percent confidence interval for the estimated number can be computed as $x \pm 1.64\ s$, which ranges from 364,000 to 580,000. Therefore, a conclusion that the average estimate derived from all possible samples lies within an interval computed in this way would be correct for roughly 90 percent of all samples.

Using GVFs for the Standard Error of a Mean

A mean is defined here to be the average quantity of some characteristic (other than the number of persons or households) per person or household. For example, a mean could be the average monthly household income of females 25 to 54 years of age. The formula used to estimate the standard error of a mean, $\bar{x}$, is

$$s_{\bar{x}} = \sqrt{\frac{b}{y} s^2}\ , \tag{7-4}$$

where y is the size on which the estimate is based, s^2 is the estimated population variance of the characteristic, and b is the parameter associated with the particular type of characteristic. Because of the approximations used in developing this formula, an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error.

The estimated population mean is computed with the formula

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \tag{7-5}$$

and the estimated population variance can be computed as

$$s^2 = \frac{\sum w_i (x_i - \bar{x})^2}{\sum w_i} \quad or \quad \frac{\sum w_i (x_i - \bar{x})^2}{\sum w_i - 1} \tag{7-6}$$

with the use of standard software for weighted data. Suppose that, based on Wave 1 data of the 1991 Panel, the mean monthly cash household income for females aged 25 to 54 is $2,530, the weighted number of females in this age range is y = 39,851,000, and the population variance is estimated to be s^2 = 3,159,887. When the appropriate b parameter of 7,514 from Table 6 of the

source and accuracy statement for Panel 1991 is used, the estimated standard error of this mean is

$$s_{\bar{x}} = [(7{,}514 * 3{,}159{,}887)/39{,}851{,}000]^{1/2} = \$24.$$

Thus, the 90 percent confidence interval, computed as

$$\bar{x} \pm 1.64 s_{\bar{x}},$$

ranges from $2,491 to $2,569. Therefore, a conclusion that the average estimate derived from all possible samples lies within an interval computed in this way would be correct for roughly 90 percent of all samples.

Variance Estimation with Imputed Data

Imputation methods are used to fill in several types of missing data in SIPP. They are used to complete some item nonresponse, person-level nonresponse within households (Type Z nonresponse), and some wave nonresponse (intermittent responses bounded by two responding waves). Imputation fills in gaps in the data set and makes data analyses easier. It also allows more people to be retained as panel members for longitudinal analyses. The concern, however, is that imputation fabricates data to some degree. Treating the imputed values as actual values in estimating the variance of survey estimates leads to an overstatement of the precision of the estimates (Brick and Kalton, 1996). It is important to recognize this fact when sizable proportions of values are imputed.

8. Using Sampling Weights on SIPP Files

This chapter describes the use of sampling weights in analyzing data from the Survey of Income and Program Participation (SIPP). Each SIPP file contains a number of alternative sets of weights for use in data analysis. The several different sets of weights are needed to cater to the different possible units of analysis (persons, households, families, and subfamilies) and different time periods for which survey estimates may be required.

A common mistake in the analysis of a survey like SIPP is to ignore the weights entirely, that is, to perform an unweighted analysis. This chapter explains why an unweighted analysis is likely to produce biased estimates. It is important to understand the different sets of weights on the files and to use the set that is appropriate for a particular analysis. Topics covered in this chapter include:

- What weights are and why they should be used;
- What weights are available in SIPP files;
- Which weights to use for a particular analysis;
- How weights are constructed;
- Using weights in the core wave files;
- Using weights in the topical module files;
- Using weights in the full panel files; and
- Using weights in combined panel files.

For the 1996 Panel, most variable names changed from those used in previous panels. To aid users working with files from panels prior to 1996, this chapter presents both the old and the new variable names whenever a variable is mentioned. In both the main body of the text and in tables, the old names are presented in parentheses following the new names. For example, the sample unit ID variable name, which is SSUID in the 1996 Panel, was SUID in previous panels; it is written in this chapter as SSUID (SUID).

What Weights Are and Why They Should Be Used

The weight for a responding unit in a survey data set is an estimate of the number of units in the target population that the responding unit represents. In general, since population units may be sampled with different selection probabilities and since response rates and coverage rates may

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

vary across subpopulations, different responding units represent different numbers of units in the population. The use of weights in survey analysis compensates for this differential representation, thus producing estimates that relate to the target population.

Most SIPP panels have not sampled different subpopulations at different rates (the exceptions are the 1990 and 1996 Panels). However, there are some minor variations in sampling rates in all SIPP panels and, more important, there are appreciable variations in response and coverage rates across subpopulations. As a result, there is nontrivial variation in SIPP weights (see *SIPP Quality Profile*, 3rd Ed. [U.S. Census Bureau, 1998a, Table 8.1]). For example, in Wave 1 of the 1993 Panel, the final person lower quartile weight is 4,400 and the upper quartile weight is 5,245 (the maximum weight is 28,695). A respondent with a final person weight of 4,400 represents 4,400 people in the U.S. population for the reference month, whereas a respondent with a weight of 5,245 represents 5,245 people. Because weights in SIPP vary over a sufficiently large range of values, performing unweighted analyses may produce appreciably biased estimates for the U.S. population.

Table 8-1 illustrates the effects of weighting on a selection of estimates obtained from Wave 1 of the 1990 Panel. The 1990 Panel included an oversample of households headed by blacks, Hispanics, and females with no spouse present and living with relatives. Since those groups are overrepresented in this sample, failure to use the weights would lead to overrepresentation of the groups in the population estimates based on that sample. At the household level, the unweighted percentage of households headed by females with no spouse present is 14.3 percent, whereas the weighted estimate is 11.7 percent. At the person level, the magnitude of the differences between weighted and unweighted estimates is less, but still appreciable.

Table 8-1. Weighted and Unweighted Point-in-Time Estimates of Percentages Based on Core Wave 1 of the 1990 SIPP Panel for January 1990

Characteristics	Percentage	
	Weighted[a]	Unweighted
Household-Level		
Female-headed households with no spouse present, living with relatives	11.7	14.3
Person-Level		
Female	51.3	52.2
Race/Ethnicity		
White	84.2	82.1
Black	12.4	14.4
American Indian, Eskimo, or Aleut	0.6	0.6
Asian or Pacific Islanders	2.9	2.9
Age over 65 years	10.4	10.6
Receiving Food Stamps [RCUTYP27 (FOODSTMP)]	6.7	7.7
RCUTYP20 (AFDC)	3.8	4.6

[a] Weighted by WPFINWGT (FNLWGT)—final weight for person—and WHFNWGT (HWGT)—final weight for households.

Weights Available in SIPP Files

Table 8-2 lists the weight variables in SIPP data files for the 1996 and 1990–1993 Panels. For earlier panels, the user should refer to the data dictionary for the particular file.

Table 8-2. Weight Variables in SIPP Files for the 1996 and 1990-1993 Panels

Variable Name	Description
Core Wave Files	
WPFINWGT (FNLWGT)	Reference month, final weight of person
WHFNWGT (HWGT)	Reference month, final weight of household
WFFINWGT (FWGT)	Reference month, final weight of family
WSFINWGT (SWGT)	Reference month, final weight of related subfamily
WPFINWGT (P5WGT)[a]	Interview (5th) month, final weight of person
WHFNWGT (H5WGT)[a]	Interview (5th) month, final weight of household
Topical Module Files	
WPFINWGT (FINALWGT)	Prior to 1996: interview month, final weight of person. 1996+: 4th reference month, final weight of person
Full Panel Files[b]	
WPFINWGT (FNLWGT) _x_	Calendar year *x*, final weight of people in the calendar year cohort
PNLWGT (Not kept for 1996 panel)	Final weight for people in full panel cohort

[a] Beginning with the 1996 Panel, SIPP files no longer include the interview month weights.

[b] The number of calendar year weights in the full panel file depends on the panel's duration. The 1990 full panel file contains two calendar year weights: WPFINWGT90 (FNLWGT90) and WPFINWGT91 (FNLWGT91). The 1992 full panel file has three calendar year weights: WPFINWGT92 (FNLWGT92), WPFINWGT93 (FNLWGT93), and WPFINWGT94 (FNLWGT94). The 1996 full panel file will have four calendar year weights when it is complete.

Choosing a Weight

The decision of which weight to use for a given analysis depends on the population of interest for that analysis. Useful guidance for choosing the correct set of weights is to consider to what population the results are intended to apply.

The weights in the SIPP files are constructed for sample cohorts defined by:

- Month (e.g., the reference month weights in the core wave files and interview month weights in the topical module files);
- Year (e.g., the calendar year weights in the full panel file); and
- Panel (e.g., the full panel weight in the full panel file).

Users can choose to base their analyses on:

- A cross-sectional sample at a given month;
- A longitudinal sample that provides continuous monthly data over a year;

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

- A longitudinal sample that provides monthly data over the life of a panel (about 32 months, or 48 months with the 1996 Panel); or
- A subset of the sample and/or the period in any of the above.

Monthly (cross-sectional) weights allow the use of all available data for a given month. For this type of analysis, users can choose among the following units of analysis:

- Person (e.g., WPFINWGT (FNLWGT));
- Household (e.g., WHFNWGT (HWGT));
- Family (e.g., WFFINWGT (FWGT)); and
- Related subfamily (e.g., WSFINWGT (SWGT)).

Analysts can use longitudinal samples to follow the same people over time and hence study such issues as the dynamics of program participation, lengths of poverty spells, and changes in other circumstances (e.g., household composition). The longitudinal weights allow the inclusion of all people for whom data were collected for every month of the period involved (calendar year or full panel period), including those who left the target population through death or because they moved to an ineligible address (institution, foreign living quarters, military barracks), as well as those for whom data were imputed for missing months. The Census Bureau makes nonresponse adjustments to the longitudinal weights to compensate for panel attrition and poststratification adjustments to make the weighted sample totals conform to population totals for key variables.

How Weights Are Constructed

This section describes how the weights are constructed. The basic components for all the different sets of weights are the same, namely:

- A base weight that reflects the probability of selection for a sample unit;
- An adjustment for subsampling within clusters;
- An adjustment for movers (in Waves 2 and beyond);
- A nonresponse adjustment to compensate for sample nonresponse; and
- A poststratification (second-stage calibration) adjustment to correct for departures from known population totals.

Weights

Reference month final weights are provided on the SIPP core wave files for persons, households, families, and subfamilies; interview month final weights are provided for persons and households. The special weights for persons are constructed first. The household, family, and

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

related subfamily final weights are derived from the final person weights. This section summarizes the steps involved in constructing the various sets of weights, starting with the final person weights for a reference or interview month. Appendix C provides the technical details and reasons for some of the adjustments.

The reference and interview month weights[1] for people on the core wave files are computed (i.e., are nonzero) for all responding sample members who are "in scope" (i.e., a part of the survey's universe—the resident, noninstitutional population of the United States) in the specified month.[2] A number of factors lead to fluctuations in sample size from month to month. They include births, deaths, immigration, and emigration from the population (and therefore from the sample). In addition to those population dynamics, people move into and out of the sample as a result of the changing household composition of sample members. (Chapter 2 describes the SIPP "following rules.")

In Wave 1, the weight for each sample person per month is a product of four components:

1. **Wave 1 base weight.** This weight is the inverse of the probability of a sample person's address being selected.
2. **Duplication-control factor.** This factor adjusts for the occasional subsampling of clusters. Clusters are occasionally subsampled in the field when they turn out to be much larger than expected.[3]
3. **Wave 1 nonresponse adjustment.** This adjustment compensates for different rates of household noninterview within adjustment classes. More than 500 nonresponse adjustment classes are defined based on a cross-classification of characteristics. Those characteristics include Census Region; MSA/Place Status (MSA-central city, MSA-non-central city, other place); race of reference person (black, nonblack); household tenure (owner, renter); household size (1, 2, 3, 4+ people). In addition, the within-primary-sampling-unit poverty stratum (high poverty, low poverty) was added for the 1996 Panel.
4. **Wave 1 second-stage calibration.** This adjustment brings the sample estimates into agreement with independent monthly estimates of population totals. The characteristics used for calibration include age, race, sex, Hispanic origin, family relationship, and household type. A raking procedure is used to ensure that the weights agree with all the control totals included for calibration. The adjustment is done by rotation group, with each group assigned one-fourth of the population total for the month.

In subsequent waves, each person receives an initial weight that is carried over from the preceding wave. This weight is adjusted to compensate for changes in the sample between waves resulting from movers and nonresponse, and then it is realigned to match the population totals for the reference or interview month:

[1] Interview month weights were not computed for the 1996 Panel.
[2] Persons subjected to Type Z imputation receive weights, although they are not respondents.
[3] This adjustment has been used since Wave 5 of the 1984 Panel.

- **Wave 2+ initial weight.** This is the weight from the previous wave before the second-stage calibration for each original sample person who is a reference person or is in group quarters for the current wave.
- **Wave 2+ mover's adjustment.** This adjustment is made to compensate for including people who were not in the original sample but were in the SIPP universe in Wave 1 and who moved into a sample household after Wave 1. For people in housing units that contain adult members who were not part of the original sample but were in the SIPP universe at Wave 1, the weights are decreased. For example, if a third adult moves into a household occupied by two original sample persons, all three adults would receive the initial weight of the original sample persons multiplied by a factor of two-thirds.
- **Wave 2+ nonresponse adjustment.** The nonresponse adjustment for Waves 2 and beyond is used to compensate for household nonresponse after the first interview. The nonresponse adjustment classes are defined on the basis of sample unit characteristics and personal demographic characteristics[4] from the most recent wave. The information used consists of household characteristics. Reference person characteristics are used to define some of the household characteristics. Tenure (owner/renter occupied), household type (female householder, no spouse present; 65+; other), race and Hispanic origin, and education level are defined at the household level by using reference person data. Other household characteristics include size, poverty status, type of income, type of financial assets, census division, and number of imputed items. Poverty threshold, census division, and number of imputed items are new to the 1996 Panel. Some adjustment classes are combined to ensure that the adjustment for each class does not exceed a factor of 2, and each class contains at least 30 unweighted sample households.
- **Wave 2+ second-stage calibration.** To derive this adjustment, use the same procedure as in Wave 1; that is, use the appropriate population control totals by reference month.

The reference month final weights for households, families, and subfamilies are derived from the person weights:

- The household weight is the person weight of the household reference person (renter/owner of housing unit).
- The family weight is the person weight of the family reference person.
- The subfamily weight for a related subfamily is the person weight of the related subfamily reference person (Chapter 10 explains how to identify households, families, and subfamilies).
- The interview month final household weight is the person weight of the household reference person in the interview month. (This weight does not apply to the 1996 Panel.)

[4] Known as the control card information before the 1996 Panel, when computer-assisted interviewing (CAI) began.

Final Full Panel and Calendar Year Weights

Final full panel and final calendar year weights are provided on the full panel files for eligible sample members. There is one set of final panel weights and generally more than one set of calendar year weights, one for each calendar year covered by the panel. The 1992 Panel file has three sets of calendar year weights because that panel covered 3 calendar years. The 1996 Panel file will have four sets of calendar year weights.

Final panel weights are computed only for people who are in the sample at Wave 1 of the panel and for whom data are obtained (either reported or imputed) for every month of the panel for which they were in scope for the survey. Other people in the panel file are assigned weights of zero. Most people with nonzero final panel weights have provided data for all months of the panel. However, people who missed a wave and whose missing wave data were imputed and people who provided data up to the point that they left the survey (through death or because they moved to an ineligible address) are also assigned nonzero final panel weights. (In core panels, it also includes those missing up to two consecutive waves, if the waves are bounded.)

Final calendar year weights are computed only for people who had an interview covering the control date[5] and for whom data are obtained (either reported or imputed) for every month of the calendar year for which they were in scope for the survey. Other people are assigned final calendar year weights of zero. Some people who joined the household of an original sample person after the start of the panel are assigned nonzero calendar year weights for the second calendar year, if data are obtained for that period.

The full panel weighting scheme does not assign weights to people who enter the sample universe after Wave 1. Similarly, the calendar year weighting scheme does not assign weights to people who do not have an interview covering the control date. This group consists of (a) people who enter the sample universe after the first wave of interviewing for the calendar year and (b) people who were in the sample universe in the first wave of interviewing in the calendar year but did not have an interview covering the control date. For example, newborn infants and people leaving institutions who are entering the sample universe after Wave 1 are assigned full panel and calendar year 1 weights of zero. Note that the same people will receive positive calendar year 2 (CY2) weights if they are in the sample universe in the first wave of interviewing for CY2 and have an interview covering the control date for CY2.

The final panel and calendar year weights are constructed from the following three components:

1. **Initial weight.** This weight is constructed from the components of the cross-sectional weights at the start of the panel and calendar year weighting periods before the second-stage calibration adjustment.

[5] The calendar year control dates are January 1 for the given calendar year. The exception is calendar year 1996 for the 1996 Panel. Its control date is currently March 1, 1996. This would change to January 1 should there be imputation for January and February data.

2. **Nonresponse adjustment factors.** These factors account for noninterviewed eligible sample persons not already accounted for in the noninterview adjustment component of the initial weight. The adjustment classes are similar to those used in the Wave 2+ nonresponse adjustment factors.
3. **Second-stage calibration factors.** These factors are determined by a process similar to that used for reference and interview month weighting. The control totals used for the calendar year weights are the population estimates for the control date of the relevant year. Those for the full panel weight are the population estimates for a designated date in the first wave of the panel (March 1 for most recent panels).

Using Weights in the Core Wave Files

Each core wave file contains reference month weights for persons, households, families, and subfamilies and, prior to the 1996 Panel, interview month weights for persons and households (interview month weights are not computed for families and related subfamilies).

In the 1989 and earlier panels, each person's record in a core wave file contained 18 weight variables, comprising weights for the four analysis units (persons, households, families, and subfamilies) for each of the four reference months and the person or household weights for the interview month. For the 1990 and later panels, the file structure was changed to a person-month format, as described in Chapter 10. With that format, each person-month record has only six weights, four for the four analysis units for that month and two for the two analysis units (household and family/related subfamily) for the interview month.

This section describes those weights and indicates how they should be used for different types of analysis.

Reference Month and Interview Month Weights

To understand the format of the reference month and interview month weights, analysts may find it useful to recall the SIPP survey design and the file structure for the core wave file. The full SIPP sample consists of four rotation groups; for each wave, interviewing is spread over 4 months. One rotation group is interviewed per month, with the reference months for each rotation group being the 4 months preceding the interview month. As successive rotation groups are interviewed, the 4-month reference periods advance by 1 month. Therefore, there are 4 interview months and 4 reference months per rotation group for each wave.

There are four final person reference month weights per sample person, one for each month in the reference period. Beginning with the 1990 Panel, the reference month weights are provided as one variable—that is, WPFINWGT (FNLWGT) for persons—in four separate person-month records per person. The reference month weight on each record refers to the specific month to

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

which the data relate. The core wave files for earlier panels used one record per person. On those files, the four reference month weights were shown as four separate variables.

The interview month weight for a particular rotation group represents one-quarter of the U.S. population at the month of interview. The sum of the interview month weights for the four rotation groups is an estimate of the total U.S. population across the 4 months of interviewing per wave. The interview month weight can be used to form person or household estimates that specifically refer to characteristics as of the interview month. For example, an analyst might want to estimate the number of unmarried adults living with an aged parent as of the latest observation. The interview month weight can also be used for estimating a few of the demographic characteristics, such as race and sex, and other information that appears on the file for the 4-month reference period as a whole, but not for each month.

Analysts should not use interview month weights to form estimates referring to the reference period plus the interview month. That is because characteristics at the time of the interview date are not necessarily representative of the rest of the reference period (i.e., people could move, marry, or leave the country). Beginning with the 1996 Panel, the core wave file no longer provides the interview month weight, since the focus of the data is the 4 calendar months prior to that month.

Person Reference Month and Interview Month Weights

For person-level analyses, the weights available in the core wave file are WPFINWGT (FNLWGT) (the reference month weight) and WPFINWGT (P5WGT) (the interview month weight—not applicable to the 1996 Panel). WPFINWGT (FNLWGT) is the estimated number of people in the population that the sample person represents in a specific reference month. The reference month is given by the variables RHCALMN (MONTH) and RHCALYR (YEAR), which are derived based on SROTATON (ROT) (rotation group) and SREFMON (REFMTH) (reference month). The interview month weight WPFINWGT (P5WGT) is also called the fifth-month weight. This weight shows the number of people in the population that the sample person represents at the interview month.

Table 8-3 shows the reference months and interview month weights for two hypothetical sample persons in Wave 1 of the 1991 Panel, based on the person-month format. The persons can be identified by the variables SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) (Chapter 10 describes how to identify a person). There are four records per person, one for each reference month. The first four records are for the first person, who is from rotation group 2: SROTATON = 2 (ROT = 2). Reference month 1, SREFMON = 1 (REFMTH = 1), corresponds to October 1990 (MONTH and YEAR). WPFINWGT (FNLWGT) for SREFMON (REFMTH) = 1 is 5,000, meaning that this person represents 5,000 people in the population in October 1990. The values of WPFINWGT (FNLWGT) in subsequent months are slightly different because of adjustments to the weight resulting from fluctuations in the population and in the sample. The second person is from rotation group 3. Since the month of interview for this person is different

Table 8-3. Final Person Weights for Four Reference Months and One Interview Month in Wave 1 of the 1991 Panel

SSUID (SUID)	EENTAID (ENTRY)	EPPPNUM (PNUM)	SROTATON (ROT)	SREFMON (REFMTH)	RH CALMN (MONTH)	RH CALYR (YEAR)	WPFIN WGT (FNLWGT)	WPFIN WGT (P5WGT)
123456789	11	101	2	1	10	90	5,000	5,025
123456789	11	101	2	2	11	90	5,005	5,025
123456789	11	101	2	3	12	90	5,010	5,025
123456789	11	101	2	4	01	91	5,020	5,025
321456789	11	101	3	1	11	90	6,500	6,525
321456789	11	101	3	2	12	90	6,510	6,525
321456789	11	101	3	3	01	91	6,520	6,525
321456789	11	101	3	4	02	91	6,530	6,525

from that of the first person, the reference months for this person are also different. The variables RHCALMN (MONTH) and RHCALYR (YEAR) can be used to select records with data for a particular month.

Household Reference Month and Interview Month Weights

Households in the core wave file refer to a group of people who occupy a housing unit in a specific calendar month. For each household, the household weight WHFNWGT (HWGT) is the weight of the reference person (the renter/owner of a housing unit) of the household. WHFNWGT (HWGT) shows the number of households in the population that the sample household represents in that reference month. The household interview month weight WHFNWGT (H5WGT) is the number of households in the population that the sample household represents at the month of interview (which varies within a wave over a 4-month period). Note that the household reference person can change from one month to the next, resulting in a change of WHFNWGT (HWGT). WHFNWGT (HWGT) is assigned to all household members.

Table 8-4 shows WHFNWGT (HWGT) and WHFNWGT (H5WGT) for five members of a household and their person weights. The variables SSUID (SUID) and SHHADID (ADDID) identify the household (Chapter 10 describes how to identify households). The WHFNWGTs (HWGTs) and WHFNWGTs (H5WGTs) for all members of a household are equal to the WPFINWGTs (FNLWGTs) and WPFINWGTs (P5WGTs) of the reference person in the household, respectively. In this case, the household reference person is the father. The user should note that weights for husbands and wives are equalized in the weight process. Therefore, couples (e.g., father and mother, daughter and son-in-law) have the same person weights.

Table 8-4. Household, Reference Month, and Interview Month Weights for Members of a Household for a Given Month in Wave 1 of the 1990 Panel

Household Member	SSUID (SUID)	SHHADID (ADDID)	EENTAID (ENTRY)	EPPPNUM (PNUM)	WHFN WGT (HWGT)	WHFN WGT (H5WGT)	WPFIN WGT (FNLWGT)	WPFIN WGT (P5WGT)
Father[a]	101111103	11	11	101	5,000	5,050	5,000	5,050
Mother	101111103	11	11	102	5,000	5,050	5,000	5,050
Daughter	101111103	11	11	103	5,000	5,050	4,800	4,865
Son-in-law	101111103	11	11	104	5,000	5,050	4,800	4,865
Grandchild	101111103	11	11	105	5,000	5,050	3,000	3,035

Note: Month = 01; Year = 1990.

[a] Reference person of household.

Family and Related Subfamily Reference Month Weights

All sample persons in a core wave file are assigned a family type, EFTYPE (FTYP), consisting of the following categories: primary families, unrelated subfamilies, primary individuals, and secondary individuals. A family is defined as a group of two or more persons related by birth, marriage, or adoption who reside together. A primary family is a family containing the household reference person and all of his or her relatives. An unrelated subfamily is a family in a household that is not related to the household reference person. A primary individual is a household reference person who lives alone or lives with only nonrelatives. A secondary individual is not a household reference person and is not related to any other people in the household.

Related subfamily units within primary families are identified by ESFTYPE (STYPE) (0 = not in a subfamily; 1 = in a related subfamily; 2 = in an unrelated subfamily). Related subfamilies are families that are related to, but do not include, the household reference person. For example, the daughter, son-in-law, and grandchild in Table 8-4 constitute a related subfamily within a primary family. They are members of the father and mother's primary family unit, as well as members of their own subfamily.

The SIPP core wave files provide reference month weights for families and related subfamilies. The family reference month weight WFFINWGT (FWGT) is equal to the person weight of the family reference person in that month; it is assigned to all family members. The subfamily reference month weight WSFINWGT (SWGT) is equal to the person weight of the related subfamily reference person; it is assigned to all subfamily members and is set equal to zero for people not in related subfamilies.

Primary individuals are the household reference persons and the family reference persons. For a primary individual, WFFINWGT (FWGT) = WPFINWGT (FNLWGT) = WHFNWGT (HWGT). Secondary individuals are classified as family reference persons who are not household reference persons. Therefore, for secondary individuals, WFFINWGT (FWGT) = WPFINWGT (FNLWGT) ≠ WHFNWGT (HWGT). The only exception is for people

in group quarters, RHTYPE = 6 (HTYPE = 6). The first secondary person in group quarters is labeled the household reference person; for that person, WFFINWGT (FWGT) = WPFINWGT (FNLWGT) = WHFNWGT (HWGT).

Table 8-5 shows the weights for the different analysis units by type of household, RHTYPE (HTYPE), and by type of family, EFTYPE (FTYPE). Three households are shown. The first household is a married couple family household, RHTYPE = 1 (HTYPE = 1), consisting of a primary family and a related subfamily, ESFTYPE = 1 (STYPE = 1). The WHFNWGT (HWGT) for each member of this household is equal to the person weight of the household reference person (i.e., the father in this case). Members of this household belong to one primary family. Therefore, the WFFINWGT (FWGT) for each member is equal to the person weight of the family reference person (who is also the father). Some members of this primary family belong to a related subfamily unit (i.e., daughter, son-in-law, and grandchild). The subfamily weight WSFINWGT (SWGT) for each member of the subfamily is equal to the person weight of the subfamily reference person (e.g., the daughter). WSFINWGT (SWGT) is zero for the father and mother who are not part of the subfamily.

The second household is a male-householder nonfamily household, RHTYPE = 4 (HTYPE = 4), with three unrelated individuals. The household reference person is the primary individual, EFTYPE = 34 (FTYPE = 4), and the others are secondary individuals, EFTYPE = 45 (FTYPE = 5). The WHFNWGT (HWGT) for this household is the person weight of the household reference person, and the weight is the same for all individuals. The WFFINWGT (FWGT) is different for each individual because each one is treated as his or her own family reference person.

The third household is a group-quarters household, RHTYPE = 6 (HTYPE = 6). Because there is no household reference person based on the typical definition of renter or owner, both individuals are classified as secondary individuals, EFTYPE = 45 (FTYPE = 5). The first secondary individual in a group quarters is labeled as the household reference person, and the WHFNWGT (HWGT) for each person in group quarters is the weight of that individual. The WFFINWGT (FWGT) for each individual is different because each forms an individual family.

Calendar Month Estimation: Using a Single Core Wave File

Each core wave file consists of data from 7 calendar months covered by the reference month periods for the four rotation groups. There is only 1 calendar month with complete data from all four rotation groups. As an illustration, Table 8-6 shows the calendar months within the reference periods for Wave 1 of the 1991 Panel and the number of rotation groups available per month. The table shows that data from all four rotation groups are available for January 1991 only. Data are available from three rotation groups for December 1990 and February 1991, for two rotation groups for November 1990 and March 1991, and for one rotation group for October 1990 and April 1991.

Table 8-5. Family and Subfamily Reference Months Weights, by RHTYPE (HTYPE), EFTYPE (FTYPE), and ESFTYPE (STYPE) in Wave 1 of the 1990 Panel

Household Member	SSUID (SUID)	SHH ADID (ADDID)	RFID (FID)	RFID2 (FID2)	RSID (SID)	EENT AID (ENTRY)	EPPP NUM (PNUM)	WPFIN WGT (FNLWGT)	WHFN WGT (HWGT)	WFFIN WGT (FWGT)	WSFIN WGT (SWGT)	EF TYPE (FTYPE)	ESF TYPE (STYPE)
RHTYPE = 1 (HTYPE = 1)—Married-couple family household													
Father [a,b]	101111103	11	1	1	0	11	101	5,000	5,000	5,000	0	1	0
Mother	101111103	11	1	1	0	11	102	5,000	5,000	5,000	0	1	0
Daughter[c]	101111103	11	1	0	1	11	103	4,800	5,000	5,000	4,800	1	1
Son-in-law	101111103	11	1	0	1	11	104	4,800	5,000	5,000	4,800	1	1
Grandchild	101111103	11	1	0	1	11	105	3,000	5,000	5,000	4,800	1	1
RHTYPE = 4 (HTYPE) = 4—Male-householder nonfamily													
Male 1 [a,b]	122210000	11	1	1	0	11	101	6,000	6,000	6,000	0	4	0
Person 2[b]	122210000	11	1	1	0	11	102	4,500	6,000	4,500	0	5	0
Person 3	122210000	11	1	1	0	11	103	5,500	6,000	5,500	0	5	0
RHTYPE = 6 (HTYPE = 6)—Group quarters													
Individual 1[a]	222210000	11	1	1	0	11	101	4,500	4,500	4,500	0	5	0
Individual 2	222210000	11	1	1	0	11	102	3,500	4,500	3,500	0	5	0

Notes: Month = 01; Year = 1990. RHTYPE (HTYPE)—type of household: 1 = married couple family household, 2 = male householder family household, 3 = female householder family household, 4 = male householder nonfamily household, 5 = female householder nonfamily household, 6 = group quarters; EFTYPE (FTYPE)—type of family: 1= primary family, 3 = unrelated subfamily, 4 = primary individual, 5 = secondary individual.

[a] Household reference person—see text.

[b] Family reference person.

[c] Related subfamily reference person.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-6. Calendar Month Estimation: Using a Single Core Wave File in Wave 1 of the 1991 and 1996 Panels

Rotation Group	Interview Month	Reference Months—Wave 1, 1991 Panel						
		1990 Oct.	1990 Nov.	1990 Dec.	1991 Jan.	1991 Feb.	1991 Mar.	1991 Apr.
2	Feb. 1991	1	2	3	4			
3	Mar. 1991		1	2	3	4		
4	Apr. 1991			1	2	3	4	
1	May 1991				1	2	3	4
Rotation Group Adjustment		4	2	4/3	1	4/3	2	4
Rotation Group	**Interview Month**	**Reference Months—Wave 1, 1996 Panel**						
		1995 Dec.	**1996 Jan.**	**1996 Feb.**	**1996 Mar.**	**1996 Apr.**	**1996 May**	**1996 June**
1	Apr. 1996	1	2	3	4			
2	May 1996		1	2	3	4		
3	June 1996			1	2	3	4	
4	July 1996				1	2	3	4
Rotation Group Adjustment		4	2	4/3	1	4/3	2	4

The reference month and interview month weights for each rotation group are designed to represent a quarter of the population at the month of reference or interview, respectively. The weights for each rotation group can be inflated to represent the full population. For every month, the inflation adjustment equals four divided by the number of rotation groups available. For example, the adjustment for October 1990 is 4/1 because there is only one rotation group in this month. For January 1991, the adjustment factor is 1 because all four rotation groups are available for this month.

Users are strongly encouraged to use the full sample of all four rotation groups whenever possible. The core wave files are designed to support analysis using the full sample of all four rotation groups (discussed below). While the weights can be modified to compensate for a smaller sample, estimates based on a subset of rotation groups will be less reliable than those based on the full sample.

Calendar Month and Quarterly Estimation: Using Two or More Core Wave Files

Combining data from two or more core wave files can increase the data available for making estimates for calendar months or continuations of calendar months such as quarters of the year. As an example, Table 8-7 shows the effects of cumulating calendar month data across two waves: Waves 1 and 2 of the 1991 Panel. By combining Waves 1 and 2, there are now four rotation groups for calendar month estimations from January through April 1991. To calculate

Table 8-7. Calendar Month Estimation: Using Two Core Wave Files from Waves 1 and 2 of the 1991 and 1996 Panels

Rotation Group	Interview Month	Reference Months						
		1990 Oct.	1990 Nov.	1990 Dec.	1991 Jan.	1991 Feb.	1991 Mar.	1991 Apr.
Wave 1, 1991 Panel								
2	February	1	2	3	4			
3	March		1	2	3	4		
4	April			1	2	3	4	
1	May				1	2	3	4
Wave 2, 1991 Panel[a]								
2	June					1	2	3
3	July						1	2
4	August							1
1	September							
Rotation Group Adjustment		4	2	4/3	1	1	1	1
Rotation Group	**Interview Month**	**Reference Months**						
		1995 Dec.	**1996 Jan.**	**1996 Feb.**	**1996 Mar.**	**1996 Apr.**	**1996 May**	**1996 June**
Wave 1, 1996 Panel								
1	Apr. 1996	1	2	3	4			
2	May		1	2	3	4		
3	June			1	2	3	4	
4	July				1	2	3	4
Wave 2, 1996 Panel[a]								
1	August					1	2	3
2	September						1	2
3	October							1
3	November							
Rotation Group Adjustment		4	2	4/3	1	1	1	1

[a] Not all data from Wave 2 are shown in the table.

calendar month estimates for each of those months, the user can simply select the person-month records for the month of interest from a file that pools records from Waves 1 and 2 and apply the WPFINWGT (FNLWGT) associated with each record to obtain the full sample estimate.

Quarterly estimates in the form of average month estimates also can be computed based on a combined file. For example, to calculate the percentage of people receiving food stamps in the first quarter of 1991, users can obtain the weighted number of people receiving food stamps and the weighted number of the total population in each month of the quarter. Then the percentage of people receiving food stamps is the sum across months of the weighted number of people receiving food stamps divided by the sum of the weighted number of total population in the quarter. In deriving quarterly estimates, or estimates for any time interval, from data in the core wave files, users need to include all four rotation groups in each month of the estimation.

The quarterly estimates derived by this method are cross-sectional estimates, based on the samples in each month of the quarter. When working with panels prior to 1996, users interested in extracting longitudinal characteristics (e.g., the percentage of people receiving food stamps for all 3 months, or in any of the 3 months, of the quarter) are encouraged to use the full panel file. Prior to the 1996 Panel, the editing and imputation procedures used for the core wave files could introduce artificially high rates of month-to-month transitions. With the introduction of CAI in the 1996 Panel, the use of core wave files for that kind of estimation problem is expected to be much less problematic because CAI should provide more complete and accurate data.

Using Weights in the Topical Module Files

The topical module files contain one weight variable—WPFINWGT (FINALWGT). For the 1996 Panel, this weight is the person cross-sectional weight for the fourth reference month. Prior to 1996, this weight was the person interview month weight for people who provided data for a topical module. It shows the number of people in the population represented by the sample person in the interview month.

The sample weights on the topical module files are defined in the same manner as the sample weights on the core wave files. The WPFINWGT (FINALWGT) for each rotation group is defined to represent a quarter of the population at the interview month. When all four rotation groups are used, the interview month weight for the full sample represents the population estimate averaged over the 4 months of interviewing per wave.

Using Weights in the Full Panel File

The weight variables in the full panel file are the calendar year weights, WPFINWGT (FNLWGT), and the full panel weight (PNLWGT). The number of calendar year weights on the file depends on the duration of the panel. Most panels before the 1996 Panel have two calendar year weights. The exceptions are the 1989 Panel, which has one calendar year weight—WPFINWGT89 (FNLWGT89)—and the 1992 Panel, which has three calendar year weights—WPFINWGT92 (FNLWGT92), WPFINWGT93 (FNLWGT93), and WPFINWGT94 (FNLWGT94). When the 1996 full panel file is complete, it will have four calendar year weights.

The weight variables are defined for sample persons who are in the sample for different periods of time. The calendar year weights apply to sample persons who had interviews covering the control date of the corresponding calendar year and who have complete data (either reported or imputed) for every month of the year (excluding months of ineligibility). The panel weight applies to sample persons who are in the sample in Wave 1 of the panel and who have complete data (either reported or imputed) for every month of a panel (excluding months of ineligibility). People are assigned calendar year weights equal to zero when they do not have interviews covering the control date, have missing data for one or more months of the year, or both.

Similarly, people are assigned panel weights equal to zero if they were not in sample in Wave 1, have missing data for one or more months of the panel, or both.

The population of inference for each of these weights is the population of survivors of the January (or Wave 1, depending on the weight) population. Infants born after the beginning of the panel are assigned a PNLWGT of zero. Similarly, infants born after the control date are assigned a calendar year weight of zero for that year. This weighting can have important implications for those studying young children when infants are a sizable fraction of the population. For example, the WIC program serves children under 5 years of age. Infants in their first year constitute 20 percent of that population.

The SIPP full panel file contains records for every person who was ever part of a responding SIPP household. There is one record for each such person, excluding people who may have been in the sample for only 1 month. The first number in PP-EENTAID (PP-ENTRY) and in PP-EPPPNUM (PP-NUM) indicate the wave in which the person entered the sample. Each record contains month-by-month data collected at every wave. However, records with incomplete data for a given period (year or full period of the panel) are assigned weights of zero. As discussed in Chapter 4, beginning with the 1991 Panel, a new imputation procedure was put into place to allow more people to have positive weights in the full panel files. All people with one or more missing waves, each of which was bounded on both sides by interviewed waves, have their data imputed for the bounded missing waves. With this procedure, a significant portion of the panel nonrespondent records became usable records for longitudinal analysis. Beginning with the 1996 Panel, people with two consecutive missing waves can have their data imputed for those waves if they are bounded by interviewed waves.

The variables PPID (PP-ID), PP-EENTAID (PP-ENTRY), and PP-EPPPNUM (PP-PNUM) identify people in the full panel files (Chapter 12). Table 8-8 provides examples of the weights in the 1990 full panel file. The 1990 Panel provides three weights: WPFINWGT (FNLWGT90), WPFINWGT91 (FNLWGT91), and PNLWGT. The person on the first row is a complete panel member, with all three weights greater than zero. The second person has positive calendar year weights but zero PNLWGT, which probably indicates that this person provided data for the first 2 calendar years but left before Wave 8. The third person had complete (reported or imputed) data for the first calendar year, but probably left before the end of the second calendar year. The fourth person entered the panel at Wave 4 and probably remained in sample until the end of the panel. He was eligible for only a calendar year 2 weight. The last person entered at Wave 7 and was assigned a weight of zero for all three weights on the panel file (however, this person would have had reference month and interview month weights on the Wave 7 and 8 core files).

Table 8-8. Calendar Year and Panel Weights, 1990-1993

PP-ID	PP-EENTAID (PP-ENTRY)	EPPPNUM (PP-PNUM)	WPFINWGT90 (FNLWGT90)	WPFINWGT91 (FNLWGT91)	PNLWGT
123456789	11	101	5,500	6,000	6,500
123456789	11	102	5,500	6,000	0
123456789	11	101	7,200	0	0
221456789	41	401	0	6,500	0
567891211	71	701	0	0	0

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Calendar Year Estimation: Using the Full Panel File

Although the SIPP collects most core content with monthly resolution, users may need to construct calendar year estimates of quantities such as total annual income. One way to construct such estimates is to work with the full panel files, extracting those records with positive calendar year weights. For example, to estimate average annual wages in 1991 for people over age 25 on January 1, 1991, one could identify records from the 1990 Panel with positive values on the calendar year weight FNLWGT91. The annual income amount for each sample person is the sum of the amounts received during each month of the calendar year. The aggregate income estimates for the population can be derived by multiplying each person's annual income by FNLWGT91 and summing the products across all people. An estimate of average income is this weighted total income divided by the sum of the weights (summed across the same subsample of the population).[6]

Annual estimates computed with this method are based on monthly data from the same person collected at three or four points in time (depending on the rotation group of the respondent). The shorter recall period used by SIPP is generally believed to provide estimates of annual measures with less nonsampling error than other surveys that collect annual income measured only once during a year. Chapter 6 and the *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a), provide a more detailed discussion of nonsampling error in SIPP.

Spell Estimation: Using the Full Panel File

Analysis of SIPP data that takes full advantage of the longitudinal nature of the survey can take a number of forms. In studies of the dynamics of household composition, labor force activity, and welfare recipiency, analysts have applied a set of methods that fall under the general headings of survival analysis (see Kalbfleisch and Prentice, 1980) and event-history analysis (see Tuma and Hannan, 1984). Among many other topics, researchers have studied the length of time that a woman remains single, a person remains unemployed, or a person receives food stamps before marrying, getting a job, or moving off the Food Stamp program. A spell of being single, unemployed, or receiving food stamps is a period of time during which a person's status did not change, and it is the duration of those spells that is often of interest.

In these studies, the unit of analysis is the spell. A file of spells is built from the person records in the full panel file, scanning across months to find a transition into and out of the state of interest. An example of the approach is provided by Shea (1995b). She constructed spells from the records of people with positive full panel weights (PNLWGT greater than zero), restricting her

[6] For purposes of exposition, this discussion has neglected the complication that not all persons with positive calendar year weights will have 12 months of data. For example, any person who was in the population January 1 but who spent at least 1 month during that year in an institution would have fewer than 12 months of data. If that person had complete data for the months when he or she was *not* in the institution, the person would have a positive value for FNLWGT91. This issue is particularly pertinent for studies of the elderly, since a noneligible portion of that population spend some time in a nursing home or some other type of extended care facility.

analysis to spells starting after the beginning of the panel, as is commonly done. Methods have been proposed that allow for the use of spells in progress at the start of the panel when the beginning dates of those spells are known (see Guo, 1993).

An alternative approach is to use all people in the full panel file. Spells can be constructed whenever a transition into the state of interest is observed (e.g., the birth of a child to a single woman). There are three possible outcomes that might be of interest: (1) a transition out of "single parenthood" is observed when the woman marries; (2) the spell is right-censored because the woman is lost through attrition from the sample before the end of the panel and before she marries; and (3) the spell is right-censored because the panel ends before she marries. If modeled in that way, the appropriate weight would be the woman's calendar month weight associated with the month that the spell of single parenthood began. Calendar month weights are not on the full panel file, but can be merged into that file from the appropriate core wave files.

During the course of a SIPP panel, some panel members can experience multiple spells (e.g., of participation in a given program). There are two approaches to handling this situation: (1) select only the first spells that started during the life of the panel (Ruggles and Williams, 1989), or (2) use *all* spells starting during the life of the panel (Kalton et al., 1992).

The length of spells that can be fully observed depends on the duration of a panel. SIPP panels before 1991 were designed to last 32 months. However, several panels were shorter because of budget constraints. The 1992 Panel lasted 36 months. The 1996 Panel has 48 months of data.

A note for users of spell analysis is that, in SIPP, as in other panel surveys, people tend to report a change in recipiency more often between waves than within waves (the seam effect). This suggests that it may not be possible to pinpoint changes to a specific month. More detailed discussions of the seam effect are provided in Chapter 6 and in the *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a).

Pooling Data from Two or Three Panels

Prior to the 1996 Panel, the SIPP design employed overlapping panels so that two or three panels could be in progress at a given time. Thus, users can pool data from two or three panels in order to produce larger samples, and hence more precise estimates, for a given time. Table 8-9 illustrates the wave overlap for the 1984 through 1993 Panels. One can see that Wave 6 of the 1984 Panel and Wave 2 of the 1985 Panel both cover the same period in 1985.

Note that Table 8-9 does not show Wave 1 of each panel. Users are not encouraged to pool data for Wave 1 with data from any other wave. Differences in interviewing procedures, question wording, and interviewer experience between Wave 1 and other waves call into question the comparability of Wave 1 responses relative to responses at other waves. In general, when pooling data from multiple panels, users should be sensitive to the potential impact of differences in questionnaire items, time-in-sample effects, and other nonsampling errors.

When pooling data for the sample wave from two successive panels, users need to adjust the weights; otherwise, the weights for the pooled data will run to twice the U.S. population total. One simple procedure would be to halve all the weights on the file. However, since the panels are generally of different sizes, that procedure is statistically inefficient. Another simple procedure is to reduce the weights in each sample in proportion to each sample size. Although this is not necessarily the most efficient procedure, it is the one recommended here. It should serve well for most panels.

The proposed weighting adjustment for pooling data from two successive panels $i - 1$ and i is to multiply all weights in panel $i - 1$ by the factor

$$W_{i-1} = \frac{\left(\frac{E_{i-1}}{E_{i-1} + E_i} + \frac{I_{i-1}}{I_{i-1} + I_i} \right)}{2}, \tag{8-1}$$

where E = number of eligible households, I = interviews, i = panel, and

$$W_i = 1 - W_{i-1}. \tag{8-2}$$

Use this formula for any two overlapping panels, including the scenario in which three panels overlap but the interest is in only two panels. For three overlapping panels, W_{i-2} is given below. W_{i-1} and W_i can be computed with the same formula, substituting E_{i-1}, I_{i-1} and E_i, I_i for E_{i-2}, I_{i-2}, respectively:

$$W_{i-2} = \frac{\left(\frac{E_{i-2}}{E_{i-2} + E_{i-1} + E_i} + \frac{I_{i-2}}{I_{i-2} + I_{i-1} + I_i} \right)}{2}. \tag{8-3}$$

For example, there were 17,696 eligible households in Wave 6 of the 1984 Panel and 11,105 eligible households in Wave 2 of the 1985 Panel. Thus, the weight adjustment for records in Wave 6 of the 1984 Panel is

$$w_{i-1} = 0.3856,$$

or that for Wave 2 of the 1985 Panel is

$$w_i = (1 - 0.3856).$$

Eligible households are not available on the public use files for any of the wave-panel combinations shown in Table 8-9. Table 8-9 provides the weighting factors for all combinations of two and three overlapping panels.

Table 8-9. Weighting Parameter Adjustment Factors for Both the Two-Panel and Three-Panel Combinations

Month, Year	Reference Period										W_I
	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	
Feb. 1984	2										
	3										
	4										
June 1985	6	2									0.39
	7	3									0.54
	8	4									0.50
June 1986	9	5	2								0.42/0.47[a]
		6	3								0.41
		7	4								0.48
June 1987		8	5	2							0.48/0.49[b]
			6	3							0.49
			7	4							0.50
June 1988				5	2						0.51
				6	3						0.51
				7	4						0.52
June 1989					5	2					0.46
					6	3					0.46
June 1990							2				
							3				
June 1991							5	2			0.39
							6	3			0.39
							7	4			0.39
June 1992							8	5	2		0.39 / 0.57[c]
								6	3		0.57
								7	4		0.57
June 1993								8	5	2	0.51 / 0.49[d]
									6	3	0.49
									7	4	0.50
									8	5	0.50
									9	6	0.50
									10	7	0.57
										8	
										9	

[a] The combined three-panel weighting adjustments are 0.31 for the 1985 Panel and 0.27 for the 1986 Panel.

[b] The combined three-panel weighting adjustments are 0.33 for the 1986 Panel and 0.32 for the 1987 Panel.

[c] The combined three-panel weighting adjustments are 0.26 for the 1991 Panel and 0.34 for the 1992 Panel.

[d] The combined three-panel weighting adjustments are 0.37 for the 1992 Panel and 0.36 for the 1993 Panel.

Section II

9. The SIPP Public Use Files

Section I of the *Users' Guide* is written primarily for researchers who need information to guide their use of data from the Survey of Income and Program Participation (SIPP). It describes the design and content of SIPP and the processing of SIPP data by the Census Bureau. It also discusses weighting, sampling error, and nonsampling error.

Section II addresses the mechanics of using the SIPP public use files. The chapters in this section are written for the analyst needing guidance on how to accomplish a variety of common tasks. This section contains minimal discussion of underlying concepts (such as the relationship between waves, rotation groups, and reference months), which are examined in Section I.

There are five chapters in Section II: this chapter provides a general introduction to the public use files; one chapter is devoted to each of the three types of SIPP data files, and a final chapter discusses merging multiple SIPP data files. After reading the current chapter, the user working with just one type of SIPP data file may wish to turn to the chapter on that type of file. For the 1996 Panel, most variable names changed from those of previous panels. To aid users working with files from panels prior to 1996, the chapters in Section II present both the pre- and post-1996 Panel variable names when the text applies to both 1996 and pre-1996 panel files (when the 1996 Panel names are available). In the main body of the text, the pre-1996 Panel names are presented in parentheses following those from the 1996 Panel. For example, the sample unit ID variable name in the core wave files, which is "SSUID" in the 1996 Panel, was SUID in previous panels. The variable name is written in this chapter as SSUID (SUID). In tables, a variety of methods are used to present both sets of names.

The balance of this chapter provides an overview of the chapters that follow. Those chapters offer more detailed discussions, complete with specific examples and samples of programming code. This introduction highlights points that are common to all SIPP data files. It also highlights important differences.

Types of SIPP Data Files

There are three types of public use files containing SIPP data: core wave files, topical module files, and full panel longitudinal research files (referred to as either longitudinal files or full panel files):

- Core wave files are currently issued in person-month format. These files contain up to four records for each primary sample member and each person who lived with a primary sample

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

member at any time during the 4-month reference period covered by the wave. Each of the records contains data from one of the four reference months covered by the wave.[1]

- Topical module files for the 1996 Panel contain one record for each person who was a sample responding (or Type Z nonresponding) member of a SIPP household during the fourth month of the reference period for the wave. Topical module files from earlier panels contain one record for each primary sample member and each person who lived with a primary sample member at the time of the interview for the wave in which the topical module was administered.
- Full panel longitudinal research files contain one record for each primary sample member and for each person who ever lived with a primary sample member at any time during the SIPP panel—a period of up to 4 years.

Understanding the ID Variables in SIPP

Because different files contain different information, the capacity to identify people across those files is important. SIPP is a longitudinal survey designed to allow researchers to track people over time; other critical functions include identifying individuals over time and identifying when a person is present in the sample. Finally, because the relationships among people change over time, identification of those relationships at any specific time is important. The key to these tasks lies in understanding how SIPP ID variables are used to identify persons, families, and households.[2]

The most basic ID variables in SIPP have different variable names in the different types of public use files issued by the Census Bureau. Table 9-1 displays those variables and shows the names they are given in the different files.

Sample Unit IDs

When initial Wave 1 interviews are conducted, each physical dwelling unit is assigned a unique (random) sample unit ID.[3] The sample unit ID assigned to a person never changes: in all

[1] Prior to the 1990 Panel, core wave files were issued with a single record for each person. Each record contained data for all four of the reference months covered by the wave. The structure of the file was similar to the longitudinal files issued by the Census Bureau. Earlier editions of this *Users' Guide* provide details.

[2] Other variables are used to identify people who are members of related subfamilies, unrelated subfamilies (also known as secondary families), and transfer program units such as food stamp units.

[3] The sample unit ID is a random recode of three other variables in the Census Bureau internal files: the respondent's sampling area, the cluster of housing units within that area (called a segment), and a sequentially assigned serial number. Because the variables in the Census Bureau's internal files contain detailed information about the location of the dwelling unit, those variables are suppressed in the public use files to protect the confidentiality of survey respondents.

Table 9-1. SIPP Variable Names, by File Type

File Type	Sample Unit ID	Current Address ID	Entry Address ID	Person Number
Panels Prior to the 1996 Panel				
Core Wave Person-Month Files	SUID	ADDID	ENTRY	PNUM
Topical Module Files	ID	ADDID	ENTRY	PNUM
Full Panel (and Partial-Panel) Longitudinal Research Files	PP-ID	HH-ADDID	PP-ENTRY	PP-PNUM
1996 Panel				
Core Wave Person-Month Files	SSUID	SHHADID	EENTAID (No longer needed to identify persons)	EPPPNUM
Topical Module Files	SSUID	SHHADID	EENTAID (No longer needed to identify persons)	EPPPNUM
Full Panel (and Partial-Panel) Longitudinal Research Files	File not yet available. Current plans call for using the same ID variable names in all files from the 1996 Panel.			

subsequent interviews, the Wave 1 primary sample persons carry their sample unit IDs with them. This means that if they move to different addresses, they keep the same sample unit IDs. If new people join those original sample members at their original addresses, they become secondary sample members by virtue of their association with the primary sample person with whom they are living. Secondary sample persons are all assigned the sample unit ID of the primary sample member with whom they are living. At the conclusion of the panel, all people who have ever lived with a member of a given original sample unit share the same sample unit ID. That sample unit ID is their common link to the original sample unit.

Current Address IDs

The current address ID identifies each housing unit occupied by one or more original sample members in any given month.[4] Current address IDs are assigned within sample units (they are unique only when combined with the sample unit ID variable), and they have two parts. The first part (one digit for all but the 1992 and 1996 Panels, two digits for the 1992 and 1996 Panels) identifies the wave in which one or more original sample members were first scheduled to be interviewed at the address. The second part of the ID is one digit, and it is used to sequentially number addresses for households that split into two or more households as a result of a move to a different location by original sample persons. All Wave 1 households have a current address ID of 11. Any new addresses that are occupied in Wave 2 are numbered 21, 22, and so on; new addresses occupied during the Wave 3 reference period are numbered 31, 32, 33, and so on. The

[4] A house, an apartment or other group of rooms, or a single room is regarded as a housing unit if it is occupied or intended for occupancy as separate living quarters.

current address ID is a monthly variable, the value of which changes in the month in which an individual moves to a new address.

Entry Address IDs

The entry address ID is the current address ID that a sample member occupied when he or she first entered the SIPP sample. It is used in conjunction with the person number to uniquely identify persons within the sample unit and does not change even if the person moves.

Person Numbers

All primary and secondary sample members are assigned a person number when they first enter the SIPP panel. Those numbers are assigned sequentially, within each wave and within each household (current address). The first part of the person number (two digits for the 1992 and 1996 Panels, one digit for all others) indicates the wave in which the person originally entered the sample. Thus, primary sample persons have person numbers in the 100 series, beginning with 101; secondary sample members have person numbers beginning with 201 if they enter the sample in Wave 2, 301 if they enter the sample in Wave 3, 401 if they enter the sample in Wave 4, and so on.

Identifying Persons and Their Relationships

Each person in SIPP can be uniquely identified by the combination of a sample unit ID, an entry address ID,[5] and a person number. These ID variables are useful when linking the records for a single person across multiple SIPP data files. They also contain substantive information that may be useful in some situations.

Using the Monthly Interview Status Variable

The monthly interview status variable helps determine whether the data for a person in a given month should be used. This variable is labeled PP-MIS in the pre-1996 longitudinal files, in the (older) person-record-format core wave files, and in older topical module files. It is labeled

[5] For the 1996 Panel, the entry address is not necessary to uniquely identify individuals in SIPP. Its continued use will not create any problems; it just provides additional information.

PPMIS in newer pre-1996 topical module files.[6] This variable has three possible values: 0, 1, and 2. ***When using the older person-record-format core wave files, the topical module files for panels prior to 1996, and the longitudinal files, analysts need to understand that the monthly interview status is the only reliable guide as to whether the data for a given person should be used in a given month. Analysts should use data for only those months in which a person's interview status is equal to 1.*** Any data present for months when a person's interview status is coded either 0 or 2 should be ignored. A code of 0 indicates that the person was not in the sample for that month, and a code of 2 indicates a noninterview for that month.[7]

When working with other data sources, analysts often identify which cases will be used in an analysis by examining either the weight variable or the variables used in the analysis itself. In the first case, the rule is generally to use all cases with positive weights and ignore the rest. In the second case, the rule is generally to use all cases with nonmissing data. Each of those rules can lead the SIPP user astray, as illustrated below.

The presence of a zero weight is not a reliable guide to whether a person should be excluded from the planned analysis. Although those people will not enter into any weighted tabulations, they may provide important contextual information about people who do enter into those (weighted) tabulations. For example, a person with a calendar year weight of zero who is a member of the same household as a positive-weight person for only 3 months provides information about the positive-weighted person's household (including, for example, household size, composition, income, and program participation) for the 3-month period that he or she was a household member. It is for this reason that records for zero-weighted persons are retained in the SIPP data files.[8]

The presence of data in analysis fields for any given month is also not a reliable guide to whether the person should be included in the planned analyses. Data are collected for all months of the reference period for a given wave, even if the interviewed person was in the sample for only part of the reference period. For example, on the topical module and longitudinal files for panels prior to 1996, 4 months' worth of data will generally be present for a person who was a member of a SIPP household for only the last 2 months of the wave. However, only those last 2 months of data should be used.[9]

---

[6] The person-month-format core wave files contain records only for those months that a person has an interview status code of 1. The monthly interview status variable is not included in those files because it is not needed. The topical module files for the 1996 Panel contain records only for those with an interview status code of 1 in the fourth month of the wave's core reference period. Although the interview status variable is included on the topical module files from the 1996 Panel, it need not be used with them.

[7] For those months when a noninterviewed person was both in scope for the survey and had data imputed (this includes the Type Z imputations and the missing wave imputations), the variable is set to 1. In those cases, the data can be used in the same manner as any of the other imputed data in the SIPP public use files.

[8] Other important situations also arise. For example, infants are assigned a calendar year weight of zero for the year of their birth even though they have an interview status of 1 from their birth month forward. Also, a person who dies during the year will have a positive calendar year weight even though, past the month of death, he or she will have an interview status of 0 or 2. In neither case does the weight variable reflect the presence or absence of the person, or data associated with the person.

[9] The person-month-format core wave files will have only two records for that person. The topical module files for the 1996 Panel will have information only about month 4 of the wave's core reference period.

Determining Monthly Household Composition

A *household*, as the term is used in Census Bureau publications, consists of all people who occupy a housing unit, regardless of their relationships to each other.[10] For many purposes, a household can be thought of as people living at a common address. A person's current address ID in any given month, together with his or her sample unit ID, identifies the household in which that person is a member for that month. Members of the same household in a given month always have an interview status of 1 and share the same sample unit ID and current address ID. Figure 2-1 (pp. 2-10–2-14) provides an illustration of changes in houshold composition.

Determining Monthly Family Composition

The term *family*, as used in Census Bureau publications, refers to a group of two or more people related by birth, marriage, or adoption who reside together; all such people are considered members of one family. For example, if the son of the person who maintains the household and the son's wife are members of the household, they are treated as members of the parent's family. Every family must include a reference person. Two or more people living in the same household who are related to each other but not to the *household* reference person form an *unrelated subfamily* (also referred to as *secondary families*).

The labels *primary individual* and *secondary individual* as used by the Census Bureau refer to people in households who are not related to any other household members. For many purposes, they can be thought of as one-person families, and the Census Bureau sometimes refers to them as pseudo-families.

Methods for identifying the interrelationships among the household members that define these groups vary, depending on the data file being used. The topical module files do not contain any of the information needed to directly identify the different types of families.[11] When it is necessary to identify family membership in an analysis that uses information from a topical module, it is also necessary to merge data from the topical module file with either a core wave file or a longitudinal file. Procedures for merging files are discussed in Chapter 13.

Identifying family membership is easiest when working with the person-month-format core wave files. The Census Bureau has two principal methods for distinguishing families.

- The first method defines a family as all persons who are related and living together. The family ID variable RFID is used with this definition. RFID groups the household reference person with all related household members by assigning them the same ID number. This family group corresponds to the Census Bureau's definition of a primary family. RFID

[10] The one exception to this definition is people living in group quarters.

[11] The one exception is the Wave 2 topical module, which collects detailed information about all of the relationships among all of the people who are household members at the time of the Wave 2 interview.

groups members of each unrelated subfamily (and primary and secondary individuals) separately.

- The second method is similar to the first in defining a family, but the family excludes members of related subfamilies. The family ID variable RFID2 is used with this definition. RFID2 equals zero for members of related subfamilies. RFID2 groups members of each unrelated subfamily (and primary and secondary individuals) in the same way as RFID—each group has a unique number.

Analysts who want to analyze multigenerational families would use RFID2 (FID2) and the variable RSID (SID). RSID (SID) treats related subfamilies as distinct family units by assigning members of related subfamilies nonzero values. Analysts can easily distinguish unrelated subfamilies from other family units when they use these variables and numbering schemes.

Chapter 10 discusses the use of these variables in greater detail. More work is involved when using the longitudinal files or the (older) person-record-format core wave files. When working with those files, analysts must create a unique family ID from several components. A number of different strategies can be used, one of which is described in Chapter 12. Other approaches are described in earlier editions of this Guide.

Determining Monthly Transfer Program Unit Composition

Some analyses involve summarizing data for units other than households or families. The SIPP core data contain sufficient information to identify program units for participants in a range of transfer programs, including Medicare; Medicaid; Aid to Families with Dependent Children (AFDC); Temporary Assistance for Needy Families (TANF);[12] General Assistance (GA); Railroad Retirement; Social Security; Veterans Compensation and Pensions; Food Stamps; and the Women, Infants, and Children nutrition program (WIC).

The SIPP data contain fields for each adult and child, indicating whether the individual received benefits (either directly or by virtue of his or her relationship to another person designated as the principal recipient) from each of these programs in each month. The SIPP data also contain information that permits identification of program units within households. One person in each program unit is identified as a principal recipient, and variables identifying that principal recipient are included on the records of the people who are part of the program unit. People who are members of a common program unit in a given month can then be identified as those who are

[12] In August 1996, the Personal Responsibility and Work Opportunity Reconciliation Act was signed into law. This legislation replaced the old welfare system, Aid to Families with Dependent Children (AFDC), with a new program, Temporary Assistance for Needy Families (TANF). In the 1996 Panel, the questions for income type 20 referred to the AFDC program prior to Wave 4 and to the TANF program beginning in Wave 4. In Wave 9, the questions were expanded somewhat to capture the larger array of program types that could exist under TANF.

in the sample in that month (interview status = 1) with common values of:

- The sample unit ID,
- The current address ID, and
- The primary recipient ID.

Constructing Household, Family, and Program Unit Level Variables

The public use files contain selected characteristics of monthly households and families that can be used directly in planned analyses. Data needs may require analysts to construct characteristics of households, families, or program units that do not already exist on the public use files created by the Census Bureau. Analysts can use the monthly ID variables described in the preceding section to construct monthly characteristics from the public use files.

Choosing Appropriate Weight(s)

Because SIPP uses a sample design in which different households (and people) are sampled at different rates, weights generally must be used when the user desires (approximately) unbiased estimates of population characteristics. In general, the appropriate weight to use for an analysis can be identified by answering two questions:

1. Which (sub)sample of SIPP is the estimate based on?
2. What population does the sample represent?

Weights for each of the calendar months covered by a panel can be found on the core wave files. A single weight appears on the topical module files. Before 1996, the interview month was a frequent reference period for topical module questions, and the weight on the pre-1996 topical module files is the person interview month weight for people who provided data for a topical module. But, as noted earlier, starting with the 1996 Panel the interview month is no longer used as a reference month; the weight on the topical module file for the 1996 Panel is the person cross-sectional weight for the fourth reference month. Weights for estimates that refer to a calendar year—or, more accurately, the January population as it appears through the balance of the calendar year—are on the longitudinal files.[13]

Chapter 8 provides detailed information about SIPP weights and how to use them.

[13] The calendar year weights are based on all sample members who are present in January and interviewed (or imputed) for every month of the year that they were "in scope" for the survey. In other words, the weights include people who died during the year if they were interviewed until they died, but they do not include people who left the sample during the year. Because they are not members of the population on January 1, infants receive a calendar weight of zero for the year in which they are born.

Working with Multiple Files

There are a number of reasons that SIPP users commonly use data from more than one file:

1. The overlapping-wave/rotation-group structure of the survey creates many situations in which data for a single calendar reference month are contained on two different core wave files.
2. The overlapping-panel structure of the pre-1996 SIPP created many situations in which data covering a single calendar year could be found on data files from two or sometimes three different panels.[14]
3. There are many research problems in which reference to a specific calendar date is not crucial and a desire for increased sample size can lead to the use of data from multiple panels (or waves) that do not overlap.
4. Many analyses of data collected in the SIPP topical modules entail merging topical module data with files containing core data (the core wave files or the longitudinal research files).
5. Since the release of a longitudinal file cannot occur until after the final interview of the final wave of a panel, researchers requiring longitudinal data from more than one wave prior to the release of the longitudinal file must create their own linked data files from the available core wave files. As of this writing, longitudinal files are available for all but the 1996 SIPP Panel, so this procedure pertains primarily to users of data from the 1996 Panel.

Chapter 13 discusses each of these situations and describes procedures for using data from multiple files to construct estimates.

The Balance of Section II

The balance of Section II is organized as follows:

- Chapter 10 describes how to use the core wave files.
- Chapter 11 describes how to use the topical module files.
- Chapter 12 describes how to use the full panel longitudinal research files.
- Chapter 13 describes how to link the different file types.

Because many users work with only a single type of file, Chapters 10, 11, and 12 are written so that they stand alone: each chapter can be used independently, without reference to the other two chapters. Differences across the three file types in their structure and in names for common

[14] Chapter 2 discusses the overlapping wave and panel structure of SIPP.

variables make this a natural way to organize the material presented here. The advantage of this organization is that an analyst working with only a single type of file will find a complete discussion of that file type in a single chapter.

However, there is substantial overlap in the types of things that analysts will be called upon to do with each of the file types. Thus, many ideas are repeated across the three chapters. Crucial differences do exist among the chapters, however. Those differences are found in the variable names used to accomplish certain common tasks and in the ways of working with data files built around different organizational principles. While the text of a chapter may seem familiar, there are often important differences in the details.

Table 9-2 summarizes some of the more important differences among the three file types. Table 9-2 is intended primarily for users who have already worked with at least one type of SIPP data file. Analysts new to SIPP should skip the table and proceed to the chapter that discusses the type of data file with which they are working. When working with a different type of SIPP file, experienced analysts can use Table 9-2 in conjunction with the chapter that discusses that new file type; the table will help to highlight differences that might otherwise be overlooked in the general discussion.

Table 9-2. Differences Among Core Wave, Topical Module, and Longitudinal Files (1990–1996 Panels)

Topic	1996 Panel Core Wave Files	Pre-1996 Core Wave Files	1996 Panel Topical Module Files	Pre-1996 Topical Module Files	Pre-1996 Longitudinal Files
File Structure	Person-month records Table 10-1	Person-month records Table 10-1	Person records Table 11-1	Person records Table 11-1	Person records Table 12-2
Data Dictionary	Size and begin position Figure 10-1	Size and begin position Figure 10-1	Size and begin position Figure 11-1	Size and begin position Figure 11-1	1992–1993 Panels Size, begin, field length, and number of fields 1990–1991 Panels Size, begin, index, and length Figure 12-1
Importance of Monthly Interview Status Variables	Not needed on the person-month files—they contain records only for months in which the respondent is present and in scope.	On the person-month files: not needed. Person-month files contain records only for months in which the respondent's interview status equals 1. On the older person-record format files: very important. See earlier editions of this *Users' Guide* for details.	Not needed. Topical module files contain records only for people for whom EPPMIS4 = 1.	PP-MIS Very important Table 11-2	PP-MIS Very important Table 12-2
How to Identify a Person	SSUID, EPPPNUM	SUID, ENTRY, PNUM Table 10-3	SSUID, EPPPNUM Table 11-6	ID, ENTRY, PNUM Table 11-7	PP-ID, PP-ENTRY, PP-PNUM Table 12-6
How to Identify a Household	SSUID, SHHADID	SUID, ADDID Table 10-5	SSUID, SHHADID Table 11-8	ID, ADDID Table 11-9	PP-ID, HH-ADDID Table 12-8
Identification of "Merged Households"	Merged households cannot be identified in files from the 1996 Panel.	PWSUID, PWENTRY, or PWPNUM > 0	Merged households cannot be identified in files from the 1996 Panel	PNUM is between ×80 and ×99, inclusively, and *x* varies from 1 to 10. Can identify the person only after the move; need to go to the core wave file to identify the person before the move.	PP-PNUM is between ×80 and ×99, inclusively, and *x* varies from 1 to 10. Can identify the person only after the move; need to go to the core wave file to identify the person before the move.

(table continues)

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 9-2. Differences Among Core Wave, Topical Module, and Longitudinal Files (1990–1996 Panels) (*continued*)

Topic	1996 Panel Core Wave Files	Pre-1996 Core Wave Files	1996 Panel Topical Module Files	Pre-1996 Topical Module Files	Pre-1996 Longitudinal Files
Handling of "Merged Households"	Not Applicable	If the move took place after the first reference month, there will be two records for each person whose ID information changed. One record reflects what happened before the move and contains the original ID information. The other record reflects what happened after the move and contains the new ID information. If the move took place in the first reference month, there will be only one record for each person whose ID information changed. That record reflects what happened after the move and contains the new ID information.	Not applicable	No matter when the move takes place, there will be one record for each person whose ID information changed. That record reflects what happened after the move and contains the new ID information.	No matter when the move takes place, there will be two records for each person whose ID information changed. One record reflects what happened before the move and contains the original ID information. The other record reflects what happened after the move and contains the new ID information.
How to Identify a Family	SSUID, SHHADID and RFID or RFID2 or RSID or [RFID2 and RSID)]	(SUID and ADDID) and [FID or FID2 or SID or (FID2 and SID)] Table 10-7	Not in the file	Not in the file	Create the family ID variables using PP-ID, HH-ADDID, and FAMTYP Table 12-10
Working with Family-Level Income Variables	Variables for the primary family include the related subfamily in them. Separate variables for the related subfamily. Table 10-9	Variables for the primary family include the related subfamily in them. Separate variables for the related subfamily. Table 10-10	Not applicable	Not applicable	Variables for the primary family include the related subfamily in them. No separate variables for the related subfamily. Table 12-12

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 9-2. Differences Among Core Wave, Topical Module, and Longitudinal Files (1990–1996 Panels) (*continued*)

Topic	1996 Panel Core Wave Files	Pre-1996 Core Wave Files	1996 Panel Topical Module Files	Pre-1996 Topical Module Files	Pre-1996 Longitudinal Files
Variables Describing Household and Family Composition	RHNF RHNFAM RHNSF EHREFPER EHHNUMPP RHTYPE EFREFPER EFTYPE EFKIND ESFT ESFRFPER	HNF HNFAM HNSF HREFPER HNP HTYPE FREFPER FTYPE FKIND			
		FAMTYP FAMREL			FAMTYP FAMREL
	ERRP	RRP RRPU	ERRP	RRP	RRP
	EPNSPOUS	PNSP	EPNSPOUS	PNSP	ENTID-PNSP PNSP
		PNPT		PNPT	ENTID-PNPT PNPT
	EPNMOM EPNDAD EPNGUARD Table 10-8	PNGDU Table 10-8	EPNMOM EPNDAD EPNGUARD Table 11-12	Table 11-12	Table 12-11

(table continues)

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 9-2. Differences Among Core Wave, Topical Module, and Longitudinal Files (1990–1996 Panels) (*continued*)

Topic	1996 Panel Core Wave Files			Pre-1996 Core Wave Files			1996 Panel Topical Module Files	Pre-1996 Topical Module Files	Pre-1996 Full Panel Files		
	Coverage	Authorized Recipient	Person-Level Amount	Coverage	Authorized Recipient	Person-Level Amount			Coverage	Authorized Recipient	Person-Level Amount
Identifying Program Units											
Social Security	RCUTYP01	RCUOWN01	T01AMTA T01AMTK	SOCSEC	SSPNUM	S01AMTA S01AMTK			SOC-SEC	SS-PIDX	*Source*s are identified in G1SRC1 – G1SRC10.
Railroad	NA		T02AMT	RAILRD	RRPNUM	S02AMTA S02AMTK S03AMT	Not in topical module files	Not in topical module files	RAILROAD	RR-PIDX	
Fed SSI	RCUTYP03	RCUOWN03	T03AMTA T03AMTK	SSICOVRG							Amounts are located in the monthly arrays G1AMT1 – G1AMT10
Veteran's Admin.	RCUTYP08	RCUOWN08	T08AMT	VETS	VETNUM	S08AMT			VETS	VA-PIDX	
AFDC/TANF	RCUTYP20	RCUOWN20	T20AMT	AFDC	AFDCPNUM	S20AMT			AFDC	AFDCPIDX	
General Assistance	RCUTYP21	RCUOWN21	T21AMT	GENASST	GAPNUM	S21AMT			GEN-ASST	GA-PIDX	
Foster Child Care	RCUTYP23	RCUOWN23	T23AMT	FOSTKID	FKPNUM	S23AMT			FOST-KID	FOSTPIDX	
Other Welfare	RCUTYP24	RCUOWN24	T24AMT	OTHWELF	OWPNUM	S24AMT			OTH-WELF	OTH-PIDX	
WIC	RCUTYP25	RCUOWN25	T25AMT	WICCOV	WICPNUM	WICVAL			WICCOV	WIC-PIDX	
Food Stamps	RCUTYP27	RCUOWN27	T27AMT	FOODSTMP	FSPNUM	S27AMT			FOODSTMP	FS-PIDX	
Medicare		ECRMTH		CARECOV	MCDPNUM				CARECOV		
Medicaid	RCUTYP57	RCUOWN57		CAIDCOV					CAIDCOV		
CHAMPUS or CHAMPVA		RCHAPPM		CHAMP	CHPNUM				CHAMP		
Health Insurance	RCUTYP58	RCUOWN58		HIIND	HIPNUM						
	Table 10-16			Tables 10-17 and 10-18						Tables 12-19 and 12-20	

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names

Table 9-2. Differences Among Core Wave, Topical Module, and Longitudinal Files (1990–1996 Panels) (*continued*)

Topic	1996 Panel Core Wave Files	Pre-1996 Core Wave Files	1996 Panel Topical Module Files	Pre-1996 Topical Module Files	Pre-1996 Longitudinal Files
Imputed Data:					
The whole record is imputed	If no prior wave data and EPPINTVW = 3, 4	If MIS5 = 2 and MISj = 1 for j = 1, 2, 3, 4 or INTVW = 3, 4	If EPPMISA = 2 or EPPINTVW = 3, 4	If PP-MIS5 = 2 and PP-MISj = 1 for j = 1, 2, 3, 4 or INTVW = 3, 4	If WAVFLG > 0 or INTVW = 3, 4
The corresponding wave of information is imputed	If the corresponding imputation flag indicates imputation.	If the corresponding imputation flag indicates imputation.	If the corresponding imputation flag and calculation flags indicate imputation.	If the corresponding imputation flag and calculation flags indicate imputation.	If the corresponding imputation flag indicates imputation.
The variable's value is imputed	Almost all person-level variables have imputation flags. There are no imputation flags on household and family aggregates. Use the person-level imputation flags of household and family members to identify aggregate amounts that include imputed values.	Almost all person-level variables have imputation flags. There are no imputation flags on household and family aggregates. Use the person-level imputation flags of household and family members to identify aggregate amounts that include imputed values.	Most person-level variables have imputation flags. There are no imputation flags on household and family aggregates. Use the person-level imputation flags of household and family members to identify aggregate amounts that include imputed values.	Most person-level variables have imputation flags. There are no imputation flags on household and family aggregates. Use the person-level imputation flags of household and family members to identify aggregate amounts that include imputed values.	Limited set of imputation flags. There are no imputation flags on household and family aggregates. Use the person-level imputation flags of household and family members to identify aggregate amounts that include imputed values.
Topcoding	Yes	Yes	Yes	Yes	Yes
How to Identify States	TFIPSST	HSTATE	TFIPSST	STATE	GEO-STE
Weight Variables					
Household	WHFNWGT	HWGT H5WGT			
Family Subfamily	WFFINWGT WSFINWGT	FWGT SWGT			
Person	WPFINWGT	FNLWGT P5WGT	WPFINWGT	FINALWGT	FNLWGTyy, where yy is the calendar year PNLWGT
Metropolitan Areas	TMETRO TMSA	HMETRO	Not on the file	Not on the file	Not on the file

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

10. Using the Core Wave Files

This chapter discusses procedures for working with data from the core wave public use data files of the Survey of Income and Program Participation (SIPP). It describes the documentation that accompanies the core wave public use files obtained from the Census Bureau. Discussion then turns to the data files themselves. The data file structure is described, and detailed explanations are provided about how to use the core wave files when performing common tasks, including (among others):

- Identifying persons, households, families, and program units;
- Understanding the effects of topcoding;
- Using imputation flags; and
- Identifying states and metropolitan areas.

Before reading this chapter, users should read Chapter 9 for an introduction to Section II. Analysts using only one core wave file should also read about the use of sample weights (Chapter 8) and the computation of standard errors (Chapter 7). Those planning on merging data from multiple core wave files, from full panel files, or from topical module files should read Chapter 11 for information about the topical module files, Chapter 12 for information about the full panel files, and Chapter 13 for information about linking SIPP public use files.

This chapter focuses on the core wave files. It is written so that it can be used independently from the chapters describing the topical module files and the full panel files. Although there are many similarities across the three types of files, important differences do exist. Because those differences are sometimes subtle, users familiar with the topical module and full panel files should read this chapter carefully, paying close attention to information about variable names and file structures. Table 9-2 summarizes the differences among the core wave, topical module, and full panel longitudinal research files.

For the 1996 Panel, most variable names changed from those used in previous panels. To aid users working with files from panels prior to 1996, this chapter presents both the old and the new variable names when the text applies to both 1996 and pre-1996 panel files. In the main body of the text, the old names are presented in parentheses following the new names. For example, the sample unit ID variable name, which is SSUID in the 1996 Panel, was SUID in previous panels; it is written in this chapter as SSUID (SUID). In tables, a variety of methods are used to present both the old and the new names.

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Using the Technical Documentation of the Core Wave Files

Each data file received from the Census Bureau has an accompanying set of technical documentation and a data dictionary. The technical documentation includes:

- The item booklet (for the 1996 Panel);
- The paper survey instrument (for panels prior to the 1996 Panel);
- A glossary of selected terms;
- A cross-walk, mapping reference months into calendar months for each rotation group;
- A source and accuracy statement describing the sample weights and the computation of standard errors; and
- User Notes.

The survey instrument is vital to understanding what questions were asked, how they were asked, the order in which they were asked, to whom they were asked, and the way in which the answers were recorded. Some questions employ skip patterns (Chapter 3), so users should pay particular attention to which questions were skipped for which respondents. The skip patterns are best understood by consulting the survey instruments. With the introduction of computer-assisted interviewing (CAI) in the 1996 Panel, documentation of instrument screens and program code is now available from the SIPP Web site (http://www.sipp.census.gov/sipp/).

The source and accuracy statements provide information about the weights on the files, when and how to make adjustments to the weights, and one approach to computing standard errors for some common types of estimates. More extensive discussions of those topics are provided in Chapters 7 and 8 of this Guide.

The data dictionary provides a detailed description of each variable on the file. It describes four aspects of each variable:

1. The definition;
2. The sample universe of the corresponding survey question;
3. The ranges for all legal values; and
4. The location (and size) in the file.

A machine-readable version of the data dictionary accompanies each data file. It can also be downloaded from the Internet (http://www.sipp.census.gov/sipp/).

The data dictionary is formatted to facilitate processing by user-written computer programs. As shown in Figure 10-1, a "D" in the first column signifies that the next few lines define the variable: (1) the variable name; (2) the size (i.e., how many digits it contains); and (3) the

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

starting position. A "U" in the first column signifies that the next words describe the universe.[1] A "V" in the first column indicates that the next number and phrase describe one of the values of the variable. An asterisk in the first column denotes a comment. A period (.) before a word denotes the start of the value label. In the dictionaries for files from the 1996 Panel, lines beginning with a "T" contain short variable descriptions that can be used by many software packages as variable labels.

Figure 10-1. Excerpt from a Data Dictionary for the Core Wave Files

Wave 1 of the 1996 Panel

```
D EENTAID     3    506
T PE: Address ID of hhld where person entered
  Sample
     Address ID of the household that this
     person belonged to at the time this
     person first became part of the sample
U All persons
V     11:129 .Entry address ID


D EPPPNUM     4    509
T PE: Person number
     Person number. This field differentiates
     persons within the sample unit. Person
     number is unique within the sample.
U All persons
V   101:1299 .Person number


D EPPINTVW    2    513
T PE: Person's interview status
U All persons
V          1 .Interview (self)
V          2 .Interview (proxy)
V          3 .Noninterview – Type Z
V          4 .Nonintrvw = pseudo Type Z.
V            .Left sample during the
V            .reference period
V          5 .Children under 15 during
V            .reference period
```

(figure continues)

[1] The universe definitions included in the data dictionaries prior to the 1996 Panel were not always accurate. Users of pre-1996 SIPP Panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

Figure 10-1. Excerpt from a Data Dictionary for the Core Wave Files (*continued*)

Wave 9 of the 1992 Panel

```
D ENTRY       2    457
     Edited entry address ID
     Address ID of the household that this
     person belonged to at the time this
     person first became part of the sample
     Range=(11:99)
U All persons, including children

D PNUM        3    459
     Edited person number
     Range=(101:998)
U All persons, including children

D INTVW       1    462
     Person's interview status
     Range=(0:5)
U All persons, including children
V          0 .Not applicable (children
V            .under 15)
V          1 .Interview (self)
V          2 .Interview (proxy)
V          3 .Noninterview – Type Z refusal
V          4 .Noninterview – Type Z other
V          5 .Noninterview – left before
V            .interview month
```

Figure 10-2 shows sample SAS and FORTRAN syntax for reading the data described by the codebook fragment in Figure 10-1. Additional SAS program code could be used to associate value labels (SAS "formats") with the variables.

Relationship of the Core Wave Data Files to the SIPP Survey Instrument

Because the core wave data dictionary does not replicate the survey instrument, analysts should keep a few things in mind when using the data:

- The variables on the data files do not correspond one-to-one with the questionnaire items—the variables are listed in a different order, some variables are not included in the core wave files at all, and some variables are created from a combination of other variables;

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Figure 10-2. Corresponding SAS and FORTRAN Syntax to Read the Data from the Core Wave Files (See Figure 10-1 for Data Dictionary)

Wave 1 of the 1996 Panel

SAS

```
INPUT
   @506  EENTAID   3.
         EPPPNUM   4.
         EPPINTVW  2.
         ;

LABEL EENTAID = “Adrs ID where person entered sample”
      EPPPNUM = “Person number”
      EPPINTVW = “Person’s interview status”
      ;
```

FORTRAN

```
       READ(infile,1000) EENTAID, EPPPNUM, EPPINTVW
1000   FORMAT(T506,I3,I4,I2))
```

Wave 9 of the 1992 Panel

SAS

```
INPUT
   @457  ENTRY   2.
         PNUM    3.
         INTVW   1.
         ;

   LABEL ENTRY = “Edited Entry Address ID”
         PNUM = “Edited Person Number”
         INTVW = “Person’s Interview Status”
         ;
```

FORTRAN

```
       READ(infile,1000) ENTRY, PNUM, INTVW
1000   FORMAT(T457,I2,I3,I1)
```

- The range of possible values of the variables on the data files does not always correspond one-to-one with the response categories shown on the survey instrument or in the data dictionary;[2]

[2] For example, in the 1996 Panel the response categories on the instrument for CLWRK are (1) a government organization, (2) a private, for-profit company, (3) a nonprofit organization ..., (4) a family business or farm. The response categories for the corresponding edited variable ECLWRK in the data dictionary are 1 = private for-profit employee, 2 = private not-for-profit employee, 3 = local government worker, 4 = state government worker, 5 = federal government worker, 6 = family worker without pay.

- The variable name in the data dictionary may not readily indicate the variable's content;[3] and
- The complexity of the skip patterns will not be apparent by simply looking at the data dictionary.[4]

To avoid potential problems and confusion, analysts should become familiar with the survey instrument before using the data. When working with the data, analysts should refer to both the survey instrument and the data dictionary.

Structure of the Core Wave Files

Beginning with the 1990 Panel, the core wave files have been issued in person-month format, with one record per person for each month of the 4-month reference period the person is in the sample.[5] A person who was in the sample for all 4 months of the wave has four records. A person who was in the sample for 1 month has only one record. Records for persons interviewed by proxy are included in the files, as are records for persons for whom the data are imputed. The files also contain records for all children residing with original panel members.

As Table 10-1 illustrates, person number 0101 (101) was in the sample all 4 months, person number 0102 (102) was also in the sample all 4 months, person number 0201 (201) was in the sample for 2 months, and person number 0202 (202) was in the sample for 1 month. Users may find it helpful to review Figure 2-1 (pp. 2-10-2-14), which illustrates movement into and out of the sample.

Identifying Persons

There are many occasions when a user may need to identify which records belong to which individual in the SIPP data files. This need arises, for example, when:

- Merging data from topical module or full panel files to core wave files;
- Combining data from two or more core wave files;

[3] Although an attempt was made in the 1996 Panel to give all variables meaningful names, the eight-character limitation imposed by many software packages places severe constraints on the degree to which this can be done. Prior to the 1996 Panel, the situation was more pronounced since numeric sequencing was used to name variables (e.g., in the paper survey, SE22318 is the variable that indicates the total number of employees working for the second business; in CAI, that variable is TEMPB2). In the 1996 Panel, variable names beginning with a "T" have been topcoded to protect respondent confidentiality.

[4] The universe definitions included in the data dictionaries prior to the 1996 Panel were not always accurate. Users of pre-1996 SIPP Panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

[5] Prior to the 1990 Panel, core wave files had one record per person. Each record contained four occurrences of each monthly variable. For more information, see earlier editions of the *SIPP Users' Guide*.

Table 10-1. Person-Month File Structure for the Core Wave Files

1996 Panel					
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Person Number (EPPPNUM)**	**Rotation Group (SROTATION)**	**Reference Month (SREFMON)**	**Calendar Month (RHCALMN)**
123451000123	011	0101	2	1	2
123451000123	011	0101	2	2	3
123451000123	011	0101	2	3	4
123451000123	011	0101	2	4	5
123451000123	011	0102	2	1	2
123451000123	011	0102	2	2	3
123451000123	011	0102	2	3	4
123451000123	011	0102	2	4	5
123451000123	021	0201	2	1	2
123451000123	021	0201	2	2	3
123451000123	022	0202	2	4	5
Prior to the 1996 Panel					
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Person Number (PNUM)**	**Rotation Group (ROT)**	**Reference Month (REFMTH)**	**Calendar Month (MONTH)**
123451000	11	101	2	1	2
123451000	11	101	2	2	3
123451000	11	101	2	3	4
123451000	11	101	2	4	5
123451000	11	102	2	1	2
123451000	11	102	2	2	3
123451000	11	102	2	3	4
123451000	11	102	2	4	5
123451000	21	201	2	1	2
123451000	21	201	2	2	3
123451000	22	202	2	4	5

- Linking husbands and wives;
- Linking parents and children; and
- Identifying which person received government transfer income on behalf of the family.

To uniquely identify a person in the core wave files, analysts should employ the three variables shown in Table 10-2. Users should note that in the 1996 Panel, the entry address ID is no longer needed for unique identification. Its continued use will not create any problems; it is simply redundant information. That is a change from earlier panels in which the entry address ID was key to uniquely identifying persons.

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 10-2. Variables Used to Uniquely Identify a Person in the Core Wave Files

Variable Name	Description
SSUID (SUID)	Sample unit ID
EENTAID (ENTRY)	Entry address ID (Not required for identification in the 1996 Panel)
EPPPNUM (PNUM)	Person number

The variables in Table 10-2 have the following characteristics:

- SSUID (SUID) uniquely identifies each initially sampled dwelling unit.[6] Every person in a core wave file was either a member of one of those units (an original sample member) or lives with someone who was a member of an initially sampled dwelling unit. A person's connection to that unit is an attribute of that person and does not change over time.[7] This means that as people move from address to address, their SSUID (SUID) stays the same. As new people join the homes of original sample members, they receive the SSUID (SUID) of the original sample members.
- EENTAID (ENTRY) identifies the address where the person lived at the time she or he was first interviewed. It does not change even if the person moves.[8] Prior to the 1996 Panel, it was used in conjunction with the person number and sample unit ID to uniquely identify persons within the sampling unit. It is not needed to uniquely identify persons in the 1996 panel. Values for this variable are unique only within sample units. The entry address ID has two components. The first part of the ID number (two digits in the 1992 and 1996 Panels, and one digit in all others) identifies the wave in which SIPP interviews were first conducted at the address. The second part of the number (one digit in all panels) sequentially numbers addresses within a sample unit [SSUID (SUID)] that enter the sample in the same wave. See Chapter 9 for a more complete discussion.
- Prior to the 1996 Panel, PNUM uniquely identified a person within the sample unit and entry address ID. In the 1996 Panel, EPPPNUM uniquely identifies a person within the sample unit. EPPPNUM (PNUM) does not change even if the person moves.[9] The first part of EPPPNUM (PNUM) (two digits in the 1992 and 1996 Panels, one digit in all others) indicates the wave in which the person was first interviewed.[10] The remaining two digits are sequentially assigned within the household. Thus, original sample members are assigned person numbers ranging from 100 to 199. Individuals who enter the SIPP sample in Wave 2

[6] The SSUID (SUID) is a random recode of three other variables in the Census Bureau's internal (not public use) files: the respondent's sampling area (PSU), the cluster of housing units within that area (called the "segment"), and a sequentially assigned serial number. Those variables are omitted from the public use files to protect the confidentiality of the respondents.
[7] There is one rare exception to this rule for Panels prior to 1996, which is described in the section entitled "Identifying Movers" later in this chapter.
[8] See footnote 6.
[9] See footnote 6.
[10] For Wave 10 of the 1992 Panel and for the 1996 Panel, the first two digits of PNUM instead of the first digit identify the wave in which the person entered sample.

are assigned a person number ranging from 200 to 299. Those who enter in Wave 10 are assigned person numbers ranging from 1001 to 1099.

Table 10-3 illustrates how the combination of SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) uniquely identifies people and provides information about when they first entered the SIPP sample. In this example, there are eight individuals: five are original sample members, one person joined the SIPP sample in Wave 3, one joined in Wave 4, and another joined in Wave 7. Note that the person who joined the sample in Wave 3 (pre-1996 Panel) was assigned a person number of 301, but an entry address ID of 21 (not 31). That is because the first part of the entry address ID indicates the wave in which that address was first occupied by any SIPP sample member, which is not necessarily the wave in which a given member entered the sample.

Table 10-3. How to Uniquely Identify a Person in the Core Wave Files

1996 Panel			
Sample Unit ID (SSUID)	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	011	0101	Original sample member
123456789123	011	0102	Original sample member
123456789123	022	0301	Enters SIPP sample in Wave 3
123456789123	011	0401	Enters SIPP sample in Wave 4
123456789123	071	0701	Enters SIPP sample in Wave 7
321456789123	011	0101	Original sample member
321456789123	011	0102	Original sample member
321456789123	011	0103	Original sample member
Prior to the 1996 Panel			
Sample Unit ID (SUID)	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**	**Notes**
123456789	11	101	Original sample member
123456789	11	102	Original sample member
123456789	21	301	Enters SIPP sample in Wave 3
123456789	11	401	Enters SIPP sample in Wave 4
123456789	71	701	Enters SIPP sample in Wave 7
321456789	11	101	Original sample member
321456789	11	102	Original sample member
321456789	11	103	Original sample member

Identifying Households

The term *household*, as used in Census Bureau publications, refers to a group of persons who occupy a housing unit. A house, an apartment or other group of rooms, or a single room is regarded as a housing unit if it is occupied or intended for occupancy as separate living quarters. That is, the occupants do not live and eat with any other persons in the structure and there is

direct access from the outside or through a common hall. A group of friends sharing an apartment constitutes a household. Noninstitutional group quarters, such as rooming and boarding houses, college dormitories, convents, and monasteries, are classified as *group quarters* rather than households.

To uniquely identify a household or group quarters in the core wave files, analysts should use the two variables shown in Table 10-4.

Table 10-4. Variables Used to Uniquely Identify a Household or Group Quarters in the Core Wave Files

Variable Name	Description
SSUID (SUID)	Sample unit ID
SHHADID (ADDID)	Current address ID

People with the same SSUID (SUID) and SHHADID (ADDID) values live in the same household (or group quarters). The six individuals in Table 10-5 make up three households. The first household contains the first four individuals. The second household contains one person. The third household contains one person.

Table 10-5. How to Uniquely Identify a Household in the Core Wave Files

1996 Panel			
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	071	0101	Four persons in this household
123456789123	071	0102	
123456789123	071	0401	
123456789123	071	0701	
321456789123	031	0101	One person in this household
321456789123	032	0102	One person in this household
Prior to the 1996 Panel			
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Person Number (PNUM)**	**Notes**
123456789	71	101	Four persons in this household
123456789	71	102	
123456789	71	401	
123456789	71	701	
321456789	31	101	One person in this household
321456789	32	102	One person in this household

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Each household contains one reference person. The household reference person is the person in whose name the home is owned or rented. If the house is owned or rented jointly by more than one person (such as a married couple or some roommate situations), any of those people may be listed as the "reference person." Users may find it helpful to refer to Figure 2-1 (pp. 2-10-2-14), which illustrates the concepts of household and changes in household composition.

Identifying Families

The term *family*, as used in Census Bureau publications, refers to a group of two or more people related by birth, marriage, or adoption who reside together; all such individuals are considered members of one family.

There are several types of families that the Census Bureau distinguishes:

- A *primary family* is a family containing the household reference person and all of his or her relatives. This means that a household composed of a husband and wife, their son, and their son's wife (i.e., the daughter-in-law) is classified as a primary family containing four people.
- A *related subfamily* is a nuclear family that is related to but does not include the household reference person. For example, the son and his wife (i.e., the daughter-in-law) in the preceding example are a related subfamily.
- An *unrelated subfamily* (sometimes called a secondary family) is a nuclear family that is not related to the household reference person. Thus, a husband and wife who live in a friend's house are classified as an unrelated subfamily. A mother and daughter who live in the mother's boyfriend's apartment are classified as an unrelated subfamily.
- A *primary individual* is a household reference person who lives alone or lives with only nonrelatives. Primary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.
- A *secondary individual* is not a household reference person and is not related to any other people in the household. Secondary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.

To uniquely identify a family, analysts should use the variables shown in Table 10-6.

Table 10-6. Variables Used to Uniquely Identify a Family in the Core Wave Files

Variable Name	Description
SSUID (SUID)	Sample unit ID
SHHADID (ADDID)	Current Address ID
and one of the following:	
RFID (FID)	Family ID
RFID2 (FID2)	Family ID, excluding related subfamily members
RSID (SID)	Family ID, for both related and unrelated subfamilies

The Census Bureau has two principal methods for distinguishing families.

- The first method defines a family as all persons who are related and living together. The family ID variable RFID is used with this definition. RFID groups the household reference person with all related household members by assigning them the same ID number. This family group corresponds to the Census Bureau's definition of a primary family. RFID groups members of each unrelated subfamily (and primary and secondary individuals) separately.
- The second method is similar to the first in defining a family, but the family excludes members of related subfamilies. The family ID variable RFID2 is used with this definition. RFID2 equals zero for members of related subfamilies. RFID2 groups members of each unrelated subfamily (and primary and secondary individuals) in the same way as RFID—each group has a unique number.

Analysts who want to analyze multigenerational families would use RFID2 (FID2) and the variable RSID (SID). RSID (SID) treats related subfamilies as distinct family units by assigning members of related subfamilies nonzero values. Analysts can easily distinguish unrelated subfamilies from other family units when they use these variables and numbering schemes.

Table 10-7 illustrates the difference between the RFID (FID), RFID2 (FID2), and RSID (SID) variables. Those variables are set to new numbers in each month. For example, a mother, a father, and a child would be family 1 with RFID (FID) = 1 in month 1, RFID (FID) = 2 in month 2, RFID (FID) = 3 in month 3, and RFID (FID) = 4 in month 4, even though family composition remains the same. The first household in the table contains a primary family of five people. The primary family contains *two* related subfamilies. RFID (FID) and RFID2 (FID2) mask the fact that there are two related subfamilies; only RSID (SID) provides that information: RSID (SID) has nonzero values for those related subfamilies.

The second "household" is actually a group of three households, each containing a primary family, that originally formed one household. The third household contains a primary family and two unrelated subfamilies. The fourth household contains a primary individual and an unrelated subfamily. The fifth household contains only a primary individual. The sixth household is a group quarters containing two people.

The needs of the analysis will help to determine which family classification to use. The following guide may prove helpful:

- To group people into families in the same way that the Census Bureau does, use SSUID (SUID), SHHADID (ADDID), and RFID (FID).
- To analyze people in related subfamilies, include only those records with RSID (SID) greater than zero and ESFTYPE (FTYPE) equal to 2.
- To analyze all families and to keep subfamilies separate from primary families, use SSUID (SUID), SHHADID (ADDID), RFID2 (FID2), and RSID (SID) to uniquely identify each family.

Table 10-7. Uniquely Identifying Families in the Core Wave Files

1996 Panel								
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Person Number (EPPPNUM)**	**Family ID, Including Related Subfamily (RFID)**	**Family ID, Excluding Related Subfamily (RFID2)**	**Related Subfamily ID (RSID)**	**Family Type (EFTYPE)[a]**	**Related Subfamily Type (ESFTYPE)**	**Notes**
110011111123	011	0101	1	1	0	1	0	This household contains a primary family of five people. The primary family contains two subfamilies.
110011111123	011	0102	1	0	2	1	2	
110011111123	011	0103	1	0	2	1	2	
110011111123	011	0104	1	0	3	1	2	
110011111123	011	0105	1	0	3	1	2	
110077777723	011	0101	1	1	0	1	0	Three households formed by people who were originally members of the same originally sampled household (SSUID of 110077777723). Two subfamilies split off from the original household to become two new primary families at addresses 21 and 22.
110077777723	021	0102	1	1	0	1	0	
110077777723	021	0103	1	1	0	1	0	
110077777723	022	0104	1	1	0	1	0	
110077777723	022	0105	1	1	0	1	0	
122210000123	011	0101	1	1	0	1	0	This household contains a primary family and two unrelated subfamilies.
122210000123	011	0104	1	1	0	1	0	
122210000123	011	0305	2	2	0	3	0	
122210000123	011	0306	2	2	0	3	0	
122210000123	011	0307	3	3	0	3	0	
122210000123	011	0308	3	3	0	3	0	
555555555123	021	0101	1	1	0	4	0	This household contains a primary individual and an unrelated subfamily.
555555555123	021	0201	2	2	0	3	0	
555555555123	021	0202	2	2	0	3	0	
555555555123	021	0203	2	2	0	3	0	
610000000123	032	0101	1	1	0	4	0	Primary individual.
897454644123	011	0101	1	1	0	5	0	Group quarters with two secondary individuals.
897454644123	011	0102	2	2	0	5	0	

[a] EFTYPE = 1 means the person belongs to a primary family (including related subfamily members). EFTYPE = 3 means the person belongs to an unrelated subfamily. EFTYPE = 4 means the person is a primary individual. EFTYPE = 5 means the person is a secondary individual.

(table continues)

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 10-7. Uniquely Identifying Families in the Core Wave Files (*continued*)

Pre-1996 Panel								
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Person Number (PNUM)**	**Family ID, Including Related Subfamily (FID)**	**Family ID, Excluding Related Subfamily (FID2)**	**Related Subfamily ID (SID)**	**Family Type (FAMTYP)[b]**	**Related Subfamily Type (ESFTYPE)**	**Notes**
110011111	11	101	1	1	0	1		This household contains a primary family of five people. The primary family contains two subfamilies.
110011111	11	102	1	0	2	1		
110011111	11	103	1	0	2	1		
110011111	11	104	1	0	3	1		
110011111	11	105	1	0	3	1		
110077777	011	101	1	1	0	1	0	Three households formed by people who were originally members of the same originally sampled household (SUID of 110077777). Two subfamilies split off from the original household to become two new primary families at addresses 21 and 22.
110077777	021	102	1	1	0	1	0	
110077777	021	103	1	1	0	1	0	
110077777	022	104	1	1	0	1	0	
110077777	022	105	1	1	0	1	0	
122210000	33	101	1	1	0	1		This household contains a primary family and two unrelated subfamilies.
122210000	33	104	1	1	0	1		
122210000	33	305	2	2	0	3		
122210000	33	306	2	2	0	3		
122210000	33	307	3	3	0	3		
122210000	33	308	3	3	0	3		
555555555	21	101	1	1	0	4		This household contains a primary individual and an unrelated subfamily.
555555555	21	201	2	2	0	3		
555555555	21	202	2	2	0	3		
555555555	21	203	2	2	0	3		
610000000	11	101	1	1	0	4		Primary individual.
897454644	11	101	1	1	0	5		Group quarters with two secondary individuals.
897454644	11	102	2	2	0	5		

[b] FAMTYP = 1 means the person belongs to a primary family (including related subfamily members). FAMTYP = 3 means the person belongs to an unrelated subfamily. FAMTYP = 4 means the person is a primary individual. FAMTYP = 5 means the person is a secondary individual.

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Other Variables Describing Household and Family Composition

Table 10-8 shows the primary core wave variables summarizing household and family composition.[11]

Table 10-8. Variables Describing Household and Family Composition in the Core Wave Files

Variable Name		
1996 Panel	**Prior to the 1996 Panel**	**Description**
RHNF	HNF	Number of families, subfamilies, and pseudo-families in household
RHNFAM	HNFAM	Number of families and pseudo-families but excluding related subfamilies in household
RHNSF	HNSF	Number of related subfamilies in household
EHREFPER	HREFPER	Household reference person (ENTRY concatenated with PNUM)
EHHNUMPP	HNP	Number of persons in household
RHTYPE	HTYPE	Type of household (e.g., married-couple family, male householder family, etc.)
EFREFPER	FREFPER	Family reference person (ENTRY concatenated with PNUM)
EFTYPE	FTYPE	Type of family (e.g., primary family, unrelated subfamily, etc.)
EFKIND	FKIND	Head of family (e.g., husband and wife, male reference person, etc.)
ESFT	FAMTYP	Type of family to which this person belongs (e.g., primary family, related subfamily, etc.)
ESFR[a]	FAMREL	Family relationship (e.g., reference person, spouse of family reference person, child of family reference person, etc.)
ERRP	RRP	Recoded relationship to the household reference person (e.g., household reference person living with relatives, child of household reference person, etc.)
Not a variable for the 1996 Panel	RRPU	Unedited relationship to the household reference person (e.g., stepchild of household reference person, grandchild of household reference person, etc.)
EPNSPOUS	PNSP	Person number of spouse
EPNGUARD	PNGDU	Person number of guardian
EPNMOM		Person number of mother
EPNDAD		Person number of father
	PNPT	Person number of parent

[a] ESFR (edited subfamily relationship) is defined the same as FAMREL, but it applies only to subfamilies (both related and unrelated).

[11] Detailed information about the relationships between members is collected in the Household Relationships topical module (see Chapter 3 for a discussion of topical module content). See those data for extensive information about household composition.

Identifying Household and Family Reference Persons

The EHREFPER (HREFPER) variable's value identifies the household reference person. As explained in Chapter 2, the *household* reference person is the owner or renter of record. Prior to the 1996 Panel, the variable identified the household reference person by concatenating ENTRY with PNUM. For the 1996 Panel, the variable simply contains the person number of the household reference person (EHREFPER = EPPPNUM). Prior to the 1996 Panel, the household reference person was the one for whom:

- HREFPER = ENTRY * 1000 + PNUM (for Waves 1-9) or
- HREFPER = ENTRY * 10000 + PNUM (for Wave 10 of the 1992 Panel).

The EFREFPER (FREFPER) variable identifies the *family* reference person. For the 1996 Panel, the variable simply contains the person number of the family reference person (EFREFPER = EPPPNUM). Prior to the 1996 Panel, the family reference person was the one for whom:

- FREFPER = ENTRY * 1000 + PNUM (for Waves 1-9) or
- REFPER = ENTRY * 10000 + PNUM (for Wave 10 of the 1992 Panel)

Using the Relationship to Reference Person [ERRP (RRP)] Variable

For the 1996 Panel, ERRP describes how each person is related to the *household* reference person. As seen in Table 10-9, the new variable provides information about several household relationship categories that were not available from earlier panels. However, as in earlier panels, this variable summarizes the relationship to the *household* reference person, *not* to the *family* reference person.

Prior to the 1996 Panel, both edited and unedited versions of the RRP variable were included on the core wave files. As shown in Table 10-10, RRP (the edited version of the variable) summarized the values of RRPU (the unedited variable). The RRPU variable can distinguish whether someone is a grandchild, stepchild, foster child, or natural/adopted child of the *household* reference person. What it cannot do, however, is distinguish the type of child within each family: RRPU is the relationship to the *household* reference person, not the relationship to the *family* reference person. For example, using records with RRPU = 6 will *not* identify all foster children, because some could be in an unrelated subfamily. The variable FAMREL summarizes the relationship of the person to the family reference person (as reference person of family, spouse, or child).

Table 10-9. The ERRP Variable in the 1996 Core Wave Files
Edited Relationship to the Household Reference Person (ERRP)

Edited Relationship to the Household Reference Person (ERRP)	Description
1	Household reference person, living with relatives
2	Household reference person, living alone or with nonrelatives
3	Spouse of household reference person
4	Child of household reference person
5	Grandchild of household reference person
6	Parent of household reference person
7	Brother or sister of household reference person
8	Other relative of household reference person
9	Foster child of household reference person
10	Unmarried partner of household reference person
11	Housemate or roommate
12	Roomer or boarder
13	Other nonrelative of household reference person

Table 10-10. Comparison of RRP and RRPU Variables of the Core Wave Files
Prior to the 1996 Panel

Edited Relationship to the Household Reference Person (RRP)	Description	Relationship to the Household Reference Person (RRPU)	Notes
1	Household reference person, living with relatives	1	Same as code 1 under RRP
2	Household reference person, living alone or with nonrelatives	2	Same as code 2 under RRP
3	Spouse of household reference person	3	Same as code 3 under RRP
4	Child of household reference person	4	Natural/adopted child of household reference person
		5	Stepchild of household reference person
5	Other relative of household reference person	7	Grandchild of household reference person
		8	Parent of household reference person
		9	Brother/sister of household reference person
		10	Other relative of household reference person
6	Nonrelative of household reference person, but related to other members of the household	11	Same as code 6 under RRP
7	Nonrelative of all members of the household	6	Foster child of household reference person
		12	Partner/roommate of household reference person
		13	Other type of nonrelative of household reference person

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

The ERRP (RRP) variable contains summary information about each person's relationship to the household reference person. Analysts should bear in mind that the household description depends upon the identity of the household reference person. For example, the household in Table 10-11 contains a mother, her daughter, and her daughter's son. If the mother is the household reference person [ERRP = 1 (RRP = 1)], her daughter is listed as a child of the household reference person [ERRP = 4 (RRP = 4)], and the daughter's son is listed as a grandchild of the reference person in the 1996 Panel (ERRP = 5), but as another relative of the household reference person in earlier panels (RRP = 5, but the same value has a different meaning from that of the 1996 Panel variable). If the daughter is the reference person, her son is listed as a child of the household reference person (RRP = 4), and her mother is listed as the parent of the reference person in the 1996 Panel (ERRP = 6), but as another relative of the household reference person in earlier panels (RRP = 5).[12] Users should note that the identity of the household reference person can change from one month to the next; thus, the household description could also change.

Table 10-11. Identifying Households Containing Three Generations in the Core Wave Files

1996 Panel		
Household Member	**Relationship to Household Reference Person (ERRP)**	**Notes**
Mother as Household Reference Person		
Mother	1	Reference person
Daughter	4	Child of reference person
Daughter's son	5	Grandchild of reference person
Daughter as Household Reference Person		
Daughter	1	Reference person
Daughter's son	4	Child of reference person
Mother	6	Parent of reference person
Panels Prior to 1996		
Household Member	**Relationship to the Household Reference Person (RRP)**	**Notes**
Mother as Household Reference Person		
Mother	1	Reference person
Daughter	4	Child of reference person
Daughter's son	5	Other relative of reference person
Daughter as Household Reference Person		
Daughter	1	Reference person
Daughter's son	4	Child of reference person
Mother	5	Other relative of reference person

[12] Because it is impossible to anticipate all of the different living arrangements found in SIPP sample households, and in some cases more than one rule for identifying a reference person may apply, some interviewer discretion in identifying the reference person is inevitable. For that reason, the resulting choices can sometimes appear to the data analyst to be somewhat arbitrary.

Identifying a Person's Spouse, Parent, or Guardian

Four other variables on the core wave files (three prior to the 1996 Panel) can also be used to describe household and family composition. They are EPNSPOUS (PNSP), EPNDAD or EPNMOM (PNPT), and EPNGUARD (PNGDU). These variables identify the person number of the spouse, the father or mother (just one parent is identified in files from panels prior to 1996), and guardian of the person, respectively. In each case, the relative is identified only if she or he is living at the same address as the person. By building from these variables, analysts can identify a variety of family configurations. For example, these variables can be used to identify households containing three generations. Table 10-12 displays one household containing a mother and her two children. One child, EPPPNUM = 0102 (PNUM = 0102), has a son, and the other child, EPPPNUM = 0104 (PNUM = 0104), has a spouse.

Table 10-12. Identifying Households Containing Three Generations in the Core Wave Files

1996 Panel					
Household Member	**Person Number (EPPPNUM)**	**Recoded Relationship to Household Reference Person (ERRP)**	**Spouse (EPNSPOUS)**	**Parent (EPNMOM)**	**Notes**
Mother	0101	1	9999	9999	Mother
Daughter #1	0102	4	9999	0101	Child
Daughter #1's Son	0103	5	9999	0102	Grandchild
Daughter #2	0104	4	0105	0101	Child
Spouse of Daughter #2	0105	8	0104	9999	Spouse of child
Panels Prior to 1996					
Household Member	**Person Number (PNUM)**	**Recoded Relationship to Household Reference Person (RRP)**	**Spouse (PNSP)**	**Parent (PNPT)**	**Notes**
Mother	101	1	999	999	Mother
Daughter #1	102	4	999	101	Child
Daughter #1's Son	103	5	999	102	Grandchild
Daughter #2	104	4	105	101	Child
Spouse of Daughter #2	105	5	104	999	Spouse of child

Note: Value of 999 or 9999 means not applicable.

Using Family-Level Income Variables

The core wave files contain a number of family-level income variables. The family income variables on these files include the income of all related subfamily members. *In other words, primary family members, including related subfamily members, are treated as one family by the Census Bureau when calculating family-level income amounts.* The core wave files also contain

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

related subfamily income variables. These variables pool the income of all persons who are members of the same related subfamily.

Table 10-13 illustrates how the family income variables on the core wave files include the income of related subfamily members. From the previous example of a primary family of five people, the primary family contains two related subfamilies. Total family income, TFTOTINC (FTOTINC), is $4,200. The first related subfamily has a total income, TSTOTINC (STOTINC), of $1,000. The second related subfamily has $2,000 in total income.

More About Using the SIPP ID Variables: Identifying Movers

When a person moves, the current address field, SHHADID (ADDID), changes. The SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) values remain the same. The first part (two digits in the 1992 Panel and the 1996 Panel, one digit in all others) of SHHADID (ADDID) indicate(s) the wave in which a household is first interviewed at that new address. The remaining digits sequentially number the households that split into two or more households, as a result of a move to a different location by original sample members. Thus, new addresses in Wave 2 are numbered 021 (21), 022 (22), and so on. New addresses in Wave 3 are numbered 031 (31), 032 (32), and so on.

Table 10-14 shows that persons 0101 (101) and 0102 (102) in the first household are original sample members. Person 0401 (401) moved into the home of persons 0101 (101) and 0102 (102) in Wave 4. In Wave 7, all three of them moved to a new location and were joined by person 0701 (701). In the second household, person 101 is an original sample member who moved to a new location in Wave 3. In the third household, person 0102 (102) is an original sample member who used to live with persons 0101 (101) and 0103 (103) of the same sample unit ID, but moved to a new location in Wave 3 [to a different location from person 0101 (101)]. In the fourth household, person number 0103 (103) is an original sample member who used to live with persons 0101 (101) and 0102 (102) of the same sample unit ID number. All but two people moved from their original location [i.e., only two people have SHHADID (ADDID) equal to EENTAID (ENTRY)].

The next example (Table 10-15) further illustrates how the ID system works as people move to new addresses, additional people move in with them, and households split. A review of Figure 2-1 may help in understanding the various household changes.

- In Wave 1, there is a five-person household consisting of a husband, wife, daughter, son, and cousin. Since this is the first wave, the current address number is 011 (11), indicating address 1 of Wave 1, and the entry address number for each member of the household is the same as the current address number. Since they are assigned in Wave 1, the person numbers are in the 0100 (100) series and are numbered sequentially, beginning with 0101 (101).

Table 10-13. How the Family-Level Variables Include the Subfamily's Information in the Core Wave Files

1996 Panel									
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Person Number (EPPPNUM)**	**Family ID, Including Subfamily (RFID)**	**Subfamily ID (RSID)**	**Number of Persons in Family (EFNP)**	**Total Family Income (TFTOTINC)**	**Number of Persons in Related Subfamily (EFNP)**	**Total Related Subfamily Income (TSTOTINC)**	**Total Primary Family Income Net of Related Subfamily**
110011111123	11	0101	2	0	5	$4,200	0	$0	$1,200
110011111123	11	0102	2	2	5	$4,200	2	$1,000	NA
110011111123	11	0103	2	2	5	$4,200	2	$1,000	NA
110011111123	11	0104	2	3	5	$4,200	2	$2,000	NA
110011111123	11	0105	2	3	5	$4,200	2	$2,000	NA
Prior to the 1996 Panel									
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Person Number (PNUM)**	**Family ID, Including Subfamily (FID)**	**Subfamily ID (SID)**	**Number of Persons in Family (FNP)**	**Total Family Income (FTOTINC)**	**Number of Persons in Related Subfamily (SNP)**	**Total Related Subfamily Income (STOTINC)**	**Total Primary Family Income Net of Related Subfamily**
110011111	11	101	2	0	5	$4,200	0	$0	$1,200
110011111	11	102	2	2	5	$4,200	2	$1,000	NA
110011111	11	103	2	2	5	$4,200	2	$1,000	NA
110011111	11	104	2	3	5	$4,200	2	$2,000	NA
110011111	11	105	2	3	5	$4,200	2	$2,000	NA

Note: NA equals not applicable.

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 10-14. Identifying Movers in the Core Wave Files

1996 Panel				
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	071	011	0101	Persons 0101 and 0102 are the original sample members. Person 0401 begins to live with them in Wave 4. All three people move in Wave 7 and person 0701 joins them.
123456789123	071	011	0102	
123456789123	071	011	0401	
123456789123	071	071	0701	
321456789123	031	011	0101	Person 0101 is an original sample member who moved in Wave 3.
321456789123	032	011	0102	Person 0102 is an original sample member who moved in Wave 3 to a different location from person 0101.
Prior to the 1996 Panel				
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**	**Notes**
123456789	71	11	101	Persons 101 and 102 are the original sample members. Person 401 begins to live with them in Wave 4. All three people move in Wave 7 and person 701 joins them.
123456789	71	11	102	
123456789	71	11	401	
123456789	71	71	701	
321456789	31	11	101	Person 101 is an original sample member who moved in Wave 3.
321456789	32	11	102	Person 102 is an original sample member who moved in Wave 3 to a different location from person 101.

Table 10-15. Example of Household Changes and Their Effects on the ID Variables of the Core Wave Files

1996 Panel				
Household Members	**Sample Unit ID (SSUID)**	**Current Address ID (SHHADID)**	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**
Wave 1				
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
Daughter	101111103123	011	011	0103
Son	101111103123	011	011	0104
Cousin	101111103123	011	011	0105
Wave 2				
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
Daughter	101111103123	011	011	0103
Son	101111103123	011	011	0104
Cousin	101111103123	011	011	0105
Wave 3				
Father	101111103123	011	011	0101
Mother	101111101233	011	011	0102
Daughter	101111103123	011	011	0103
Son-in-Law	101111103123	011	011	0301
Cousin	101111103123	011	011	0105
Wave 4	**Parent's Household**			
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
	Daughter's Household			
Daughter	101111103123	041	011	0103
Son-in-Law	101111103123	041	011	0301
	Cousin's Household			
Cousin	101111103123	042	011	0105
Uncle	101111103123	042	042	0401
Wave 10	**Parent's Household**			
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
	Daughter's Household			
Daughter	101111103123	101	011	0103
Son-in-Law	101111103123	101	011	0301
Newborn	101111103123	101	041	1001

(table continues)

Table 10-15. Example of Household Changes and Their Effects on the ID Variables of the Core Wave Files (*continued*)

Panels Prior to 1996				
Household Member	**Sample Unit ID (SUID)**	**Current Address ID (ADDID)**	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**
Wave 1				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 2				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 3				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son-in-Law	101111103	11	11	301
Cousin	101111103	11	11	105
Wave 4	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
	Cousin's Household			
Cousin	101111103	42	11	105
Uncle	101111103	42	42	401
Wave 10[a]	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Newborn	101111103	41	41	1001

[a] Prior to the 1996 Panel, only the 1992 Panel had 10 or more waves. The Wave 2 core wave file of the 1992 Panel has expanded address ID and person ID fields (3 and 4 digits, respectively) to accommodate Wave 10 of the 1992 Panel.

- During Wave 2, the son joins the Army, moves into the military barracks, and therefore leaves the SIPP sample. For the son's record, person number 0104 (104), the person-month file, will contain a Wave 1 record for him and a Wave 2 record containing information (either imputed or provided by proxy) on his characteristics in the months of Wave 2 that he was still in the sample. If he does not return to the sample during the remainder of the panel, there will be no records for him beyond Wave 2.
- During Wave 3, the daughter marries and her husband moves into the household. The current address number where the mother, father, cousin, daughter, and son-in-law live remains the same since it is the same address. The son-in-law's entry address number is 011 (11), since he first enters the SIPP sample at an address coded 011 (11). The person number for the son-in-law is in the 0300 (300) series [0301 (301)] since he joins the SIPP sample in Wave 3.
- During Wave 4, the daughter and son-in-law move into a new house. Their current address number changes to 041 (41) to indicate that a new address has been established in Wave 4. Meanwhile, the cousin, who is over age 15, moves in with an uncle.[13] The cousin's current address number changes to 042 (42) (i.e., the second new household formed in the fourth wave from this sample unit). The assignment of address number 041 (41) to the daughter and 2 (42) to the cousin is arbitrary—it could be the other way around. The uncle enters the SIPP sample and receives an address number of 042 (42) and an entry address number of 042 (42). The uncle's person number is in the 0400 (400) series [0401 (401)], since he joins the survey in Wave 4.
- No changes in household composition are observed during Waves 5–9.
- During Wave 10,[14] the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. The newborn's entry address is 041 (41) because that is the current address ID of the daughter and son-in-law at the time of birth. The newborn's person number is 1001, reflecting the fact that the newborn came into the SIPP sample in Wave 10. Meanwhile, the cousin moves to Europe and therefore leaves the SIPP sample. The uncle, even though he did not move to Europe with the cousin, also leaves the SIPP sample because he no longer resides with an original SIPP sample member. Their records are no longer listed.

Prior to the 1996 Panel, there were two extremely rare occasions when the original SUID, ENTRY, and PNUM values were modified by the Census Bureau:

1. The first occasion was when two separate sampling units, each containing original sample members, were merged, perhaps because of a marriage. In this situation, one of the original sets of SUID and ENTRY values was retained and the other set was changed to agree with that retained set. The person-number values (PNUM) of the changed set were modified further to be between 180 and 199, inclusive.

[13] In the 1993 Panel, all original sample members were followed, no matter what their age. In all other panels (including the 1996 Panel), only those age 15 or older were followed when they moved to new addresses.
[14] Prior to the 1996 Panel, only the 1992 Panel had 10 or more waves.

2. The second occasion was when a household split into two new households (in which each new household gained a new sample person) and later the households recombined. For example, suppose that a married couple separated in Wave 3, each moving in with a sibling. Both siblings were assigned a person number of 301 because they entered the sample in Wave 3 at different addresses (thus, ADDID = 31 and 32). If the husband and wife reunited in Wave 6, bringing the siblings with them, one sibling's person number would have been changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699, inclusive).

Those two occasions were the only times when SUID, ENTRY, and PNUM changed. When it did occur, the old ID variables were stored in the previous wave variables (PWSUID, PWENTRY, and PWPNUM).[15]

When the merge occurred after the first month of a reference period, the members of the merged household (whose ID variables were modified) were assigned two sets of monthly records in the core wave file. The first set of records contained the original ID information and identified the person as having exited the sample at the time of the merge. The second set contained the new ID information and identified the person as having entered the sample at the time of the merge. When the merge occurred at the start of the reference period, only the second set of records was retained in the core wave files.

Because merged households were very rare prior to the 1996 Panel, information about them will no longer be carried on the core wave files from the 1996 Panel. When either of those two kinds of events occur in the 1996 Panel, one or more original sample members will appear to leave the sample when the merge takes place, and new people will appear to enter the sample when the merged household forms. There is no indication in the data files that the "new" sample members were previously members of the SIPP sample with different ID values.

Identifying Program Units

Besides household and family composition, the core wave files contain detailed information about participation in health insurance and various government transfer programs. For most programs, three characteristics are recorded (Table 10-16):

1. Whether the person is covered;
2. Who received the income or benefit; and
3. The amount of the income or benefit.

[15] In the 1993 Panel, merged households are identified with the variables PWSUID, PWENTRY, and PWPNUM. Before the 1993 Panel, they were identified with the variables PREV-ID, SC0064, and SC0066.

Table 10-16. Variables Describing Participation in Government Transfer Programs and Health Insurance Programs in the Core Wave Files

1996 Panel				
Program	**Coverage**	**Authorized Recipient**	**Recipiency**	**Amount**
Social Security—Adults	RCUTYP01	RCUOWN01	ER01A	T01AMTA
Social Security—Children			ER01K	T01AMTK
Railroad Retirement—Adults			ER02	T02AMT
Federal Supplemental Security Income	RCUTYP03	RCUOWN03	ER03	T03AMT
Veteran's Benefits	RCUTYP08	RCUOWN08	ER08	T08AMT
Aid to Families with Dependent Children/ Temporary Assistance for Needy Families[a]	RCUTYP20	RCUOWN20	ER20	T20AMT
General Assistance	RCUTYP21	RCUOWN21	ER21	T21AMT
Foster Child Care	RCUTYP23	RCUOWN23	ER23	T23AMT
Other Welfare	RCUTYP24	RCUOWN24	ER24	T24AMT
Women, Infants and Children (WIC)	RCUTYP25	RCUOWN25	ER25	T25AMT
Food Stamps	RCUTYP27	RCUOWN27	ER27	T27AMT
Medicare		ECRMTH		
Medicaid	RCUTYP57	RCUOWN57	ER57	
CHAMPUS		RCHAMPM		
Other Health Insurance	RCUTYP58	RCUOWN58	ER58	
Panels Prior to 1996				
Program	**Coverage**	**Authorized Recipient**	**Recipiency**	**Amount**
Social Security—Adults	SOCSEC	SSPNUM	R01A	S01AMTA
Social Security—Children			R01K	S01AMTK
Railroad Retirement—Adults	RAILRD	RRPNUM	R02A	S02AMTA
Railroad Retirement—Children			R02K	S02AMTK
Federal Supplemental Security Income	SSICOVRG[b]		R03	S03AMT
Veteran's Benefits	VETS	VETNUM	R08	S08AMT
Aid to Families with Dependent Children	AFDC	AFDCPNUM	R20	S20AMT
General Assistance	GENASST	GAPNUM	R21	S21AMT
Foster Child Care	FOSTKID	FKPNUM	R23	S23AMT
Other Welfare	OTHWELF	OWPNUM	R24	S24AMT
Women, Infants and Children (WIC)	WICCOV	WICPNUM	R25	WICVAL
Food Stamps	FOODSTMP	FSPNUM	R27	S27AMT
Medicare	CARECOV			
Medicaid	CAIDCOV	MCDPNUM		
CHAMPUS	CHAMP	CHPNUM		
Other Health Insurance	HIIND	HIPNUM		

[a] In August 1996, the Personal Responsibility and Work Opportunity Reconciliation Act was signed into law. This legislation replaced the old welfare system, Aid to Families with Dependent Children (AFDC), with a new program, Temporary Assistance for Needy Families (TANF). In the 1996 Panel, the questions for income type 20 referred to the AFDC program prior to Wave 4 and to the TANF program beginning in Wave 4. In Wave 9, the questions were expanded somewhat to capture the larger array of program types that could exist under TANF.

[b] During the 1990s, SSI was extended to children with disabilities. Consequently, beginning with the 1992 Panel, SSICOVRG was added to the core wave data files.

The coverage variables identify whether the income or benefit covers that person. In other words, when a person is flagged as covered by food stamps, RCUTYP27 (FOODSTMP) = 1, the person received the benefits either directly (because he or she was the authorized food stamp recipient) or indirectly (because he or she was in the same food stamp unit as the authorized recipient). The coverage variables also allow users to determine situations in which the program unit is a subset of the family or household.[16]

The authorized recipient variables identify the people who actually received the income or benefit for the people in their program units. In the 1996 Panel, the variables identifying the authorized recipient use only the person number, EPPPNUM. Prior to the 1996 Panel, the variables identifying the authorized recipient were constructed by concatenating the entry address, ENTRY, with the person number, PNUM.

Individuals who are members of a common program unit can be identified by using the sample unit ID, SSUID (SUID), and the authorized recipient variable. For example, members of a common food stamp unit are those with common values of SSUID (SUID) and RCUOWN27 (FSPNUM). Identifying members of common units is often necessary because most programs allow more than one program unit in a household. Medicare, however, is a person-based program in which each participant is an authorized recipient, so no additional authorized recipient variable for that program is included on the files. Prior to the 1996 Panel, there was also no authorized recipient variable for SSI on the core wave files.

There are some exceptions to these rules:

- Social Security, Railroad Retirement (prior to 1996), WIC, AFDC, and Medicaid can offer benefits solely to children. When that happens, an adult receives the income on behalf of the children. The adult, therefore, is flagged as the authorized recipient but is not flagged as covered by the program. The children are flagged as covered and have nonzero benefits.
- Most SSI recipients are elderly and disabled adults, but they can also be disabled children. In the 1990s, the definition of qualifying disabling conditions was expanded. That change in definition resulted in a rapid expansion of the child SSI caseload. Consequently, the SSICOVRG variable was included (beginning with the 1992 Panel). This variable indicates on the recipient's (the adult's) record whether the children, the adults, or both, within a family are covered by the income. Prior to the 1996 Panel, however, SSICOVRG did not flag each person individually, like the other coverage variables. Only the recipient will have had a nonzero SSI income. Beginning with the 1996 Panel, two new variables identify each individual covered by federally administered SSI (RCUTYP03) or state-administered SSI (RCUTYP04).

[16] In the 1984 and 1985 Panels, WIC coverage was imputed to children under 6 years old if a mother reported participation in the WIC program. Beginning with the 1986 Panel, WIC coverage is assessed directly for all sample members.

- The medical insurance variables simply reflect who is enrolled in which type of program. There are no associated amount variables.

These rules and exceptions are illustrated in Table 10-17. The household contains one AFDC unit and two food stamp units. The mother is covered by Social Security and SSI. The mother of the disabled child receives WIC benefits and SSI on behalf of her child, but she did not receive WIC or SSI for herself. Everyone in the household is enrolled in Medicaid. The coverage variables are set to 2 whenever the person is not covered by the particular program; the one exception (for panels prior to 1996) is SSI coverage—a value of 2 means that only the children are covered.

Users should note that, except for WIC, no amounts of income or benefit from government transfer and health insurance programs are listed in the records of children under age 15. Thus, in the case of WIC, users need to sum the amounts over all persons, including children, to get the proper WIC unit total. For all other programs, users will find the unit total benefit in the recipient's record.

Income Topcoding in the 1996 Panel

To protect the confidentiality of SIPP respondents, the Census Bureau topcodes very high incomes on the SIPP public use data files. New income topcoding procedures were instituted with the 1996 Panel. As in the past, summary income variables for persons, families, and households are the sums of the component variables *after* they have been topcoded. The summary variables are not independently topcoded. Thus, a person, family, or household with high income from several sources (multiple jobs, businesses, property) could have aggregate monthly income well over the topcode threshold for each source.

Topcoding Unearned Income in the 1996 Panel

When the total amount of asset income or of certain types of general income for a wave exceeds the established ceiling, the monthly amounts in excess of the monthly threshold are replaced by monthly topcode values. For example:

- When the amount of interest on joint municipal/corporate bonds exceeds $10,000 for the wave, each monthly amount in excess of $2,500 is recoded to $2,500.
- When the amount of interest on self-owned municipal/corporate bonds exceeds $12,800 for the wave, each monthly amount in excess of $3,200 is recoded to $3,200.

Table 10-17. Example of Program Units, Coverage, and Recipiency in the Core Wave Files

1996 Panel						
	Mother	**Daughter #1**	**Daughter #1's Son**	**Daughter #2**	**Spouse of Daughter #2**	**Daughter #2's Pregnant Daughter**
EPPPNUM	0101	0102	0103	0104	0105	0106
TAGE	70	21	4	35	36	16
AFDC/TANF						
RCUTYP20	2	1	1	2	2	2
RCUOWN20	0	0102	0102	0	0	0
ER20	0	1	0	0	0	0
T20AMT	0	123	0	0	0	0
Food Stamps						
RCUTYP27	2	1	1	1	1	1
RCUOWN27	0	0102	0102	0104	0104	0104
ER27	0	1	0	1	0	0
T27AMT	0	160	0	130	0	0
SSI						
RCUTYP03	1	2	1	0	0	0
ER03	1	1	0	0	0	0
T03AMT	188	122	0	0	0	0
WIC						
RCUTYP25	2	2	1	2	2	1
RCUOWN25	0	0	0102	0	0	0106
ER25	0	1	0	0	0	1
WICVAL	0	30.12	0	0	0	27.50
Medicaid						
RCUTYP57	1	1	1	1	1	1
RCUOWN57	0101	0102	0102	0104	0104	0106
Social Security						
RCUTYP01A	1	2	2	2	2	2
RCUOWN01A	0101	0	0	0	0	0
R01A	1	0	0	0	0	0
T01AMTA	470	0	0	0	0	0

(table continues)

Table 10-17. Example of Program Units, Coverage, and Recipiency in the Core Wave Files (*continued*)

Panels Prior to 1996						
	Mother	**Daughter #1**	**Daughter #1's Son**	**Daughter #2**	**Spouse of Daughter #2**	**Daughter #2's Pregnant Daughter**
PNUM	101	102	103	104	105	106
AGE	70	21	4	35	36	16
AFDC						
AFDCCOV	2	1	1	2	2	2
AFDCPNUM	0	11102	11102	0	0	0
R20	0	1	0	0	0	0
S20AMT	0	123	0	0	0	0
Food Stamps						
FOODSTMP	2	1	1	1	1	1
FSPNUM	0	11102	11102	11104	11104	11104
R27	0	1	0	1	0	0
S27AMT	0	160	0	130	0	0
SSI						
SSICOVRG	1	2	1	0	0	0
R03	1	1	0	0	0	0
S03AMT	188	122	0	0	0	0
WIC						
WICCOV	2	2	1	2	2	1
WICPNUM	0	0	11102	0	0	11106
R25	0	1	0	0	0	1
WICVAL	0	30.12	0	0	0	27.50
Medicaid						
CAIDCOV	1	1	1	1	1	1
MCDPNUM	11101	11102	11102	11104	11104	11106
Social Security						
SOCSEC	1	2	2	2	2	2
SSPNUM	11101	0	0	0	0	0
R01A	1	0	0	0	0	0
R01K	0	0	0	0	0	0
S01AMTA	470	0	0	0	0	0
S01AMTK	0	0	0	0	0	0

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Not all income sources are topcoded. For example, the amount of food stamp income is not topcoded. For a complete list of topcoded income variables with the topcode amounts for the 1996 Panel, users should refer to Appendix B (Topcoding).

Topcoding Employment Income in the 1996 Panel

Three different sources of monthly employment income are identified in the SIPP public use files: (1) wage and salary income, (2) self-employed earnings, and (3) other worker arrangements. Each of these three sources is topcoded separately. For each source, monthly amounts over $12,500 (one-twelfth of the $150,000 annual benchmark) are topcoded if the total income from those sources from all 4 months in the wave is greater than $50,000 (one-third of $150,000). Table 10-18 provides examples of employment income amounts that require topcoding.

Table 10-18. Topcoding Criteria for the 1996 Panel

Example	Reported Monthly Earned Income Amounts				Sum for the Wave	Is the Sum Greater than $50,000?	Topcoding Procedure
	Month 1	Month 2	Month 3	Month 4			
1	$ 3,000	$ 4,000	$ 5,000	$ 5,000	$17,000	No	None
2	$0	$0	$0	$55,000	$55,000	Yes	Topcode month 4
3	$15,000	$10,000	$10,000	$12,000	$52,000	Yes	Topcode month 1
4	$12,000	$15,000	$15,000	$15,000	$60,000	Yes	Topcode months 2, 3, and 4
5	$0	$0	$0	$49,000	$49,000	No	None
6	$15,000	$15,000	$15,000	$15,000	$60,000	Yes	Topcode all 4

When topcoding is required because the reported value exceeds the acceptable threshold, the value assigned to the variable can be determined in one of two ways: it can be set equal to the threshold, or it can be set equal to the mean of the reported amounts above the threshold. In the second case, the topcode value that is assigned is based on the respondent's gender, race/ethnic origin, and employment status (full or part year, full or part time). Table 10-19 illustrates the procedure. It shows the topcodes used in Wave 1 of the 1996 Panel for employment income. Those Wave-1-based topcodes are adjusted for inflation and real growth in earned income (see Box 10-1) and then used for all later waves of the panel.

Because of the way in which the topcode values were computed (explained in the next paragraph), the values listed for each cell are greater than the monthly value that is tested ($12,500). This method of computation may result in instances in which use of the topcode values results in total amounts for the wave (summed across all 4 months) that are greater than $50,000.

Table 10-19. Topcode Amounts Used for Monthly Employment Income in Wave 1 of the 1996 Panel

Example	Sex	Race	Worker Status	Earned Income Topcode
1	Male	Nonblack, non-Hispanic	Full year; full time	$29,660
2	Male	Nonblack, non-Hispanic	Not full year; full time	$38,270
3	Male	Black, non-Hispanic	Full year; full time	$17,530
4	Male	Black, non-Hispanic	Not full year; full time	$24,015
5	Male	Hispanic, any race	Full year; full time	$26,250
6	Male	Hispanic, any race	Not full year; full time	$24,015
7	Female	Nonblack, non-Hispanic	Full year; full time	$21,990
8	Female	Nonblack, non-Hispanic	Not full year; full time	$49,450
9	Female	Black, non-Hispanic	Full year, full time	$24,015
10	Female	Black, non-Hispanic	Not full year; full time	$24,015
11	Female	Hispanic, any race	Full year; full time	$24,015
12	Female	Hispanic, any race	Not full year; full time	$24,015

Box 10-1. Computing Earned Income Topcode Amounts for Waves 2–12 in the 1996 Panel

The topcode amount for wave *k* is computed as:

$$Topcode_{Wave\ k} = Topcode_{Wave\ 1} * 1.019^{k-1}$$

Example: Nonblack, non-Hispanic male employed full year, full time.
Wave 1 Topcode (from Table 10-19) = $29,660
Wave 7 Topcode = $29,660 * $1.019^{(7-1)}$ = $29,660 * 1.120 = $32,206

The topcode values were computed from data collected in Wave 1 of the 1996 Panel. The topcode values are the unweighted mean amounts from records identified for topcoding in Wave 1 of the 1996 Panel. A separate topcode value was computed for each of the 12 cells of Table 10-19. Each topcode value is based on amounts from all three employment income sources, and the same topcode is used for all three employment income sources. The algorithm used to calculate the assigned topcode amount is as follows:

1. Add the four monthly amounts of wage and salary income. If the sum is greater than $50,000, store the monthly amounts greater than $12,500 in the 12-cell matrix.
2. Add the four monthly amounts of self-employed earnings. If the sum is greater than $50,000, store the monthly amounts greater than $12,500 in the 12-cell matrix.
3. Add the four monthly amounts of contingent worker earnings. If the sum is greater than $50,000, store the monthly amounts greater than $12,500 in the 12-cell matrix.

On the basis of the amounts accumulated, compute a mean amount within each of the 12 cells of the matrix. That mean amount is the topcode value shown in Table 10-19.

The amounts shown in Table 10-19 were computed with data from Wave 1. Current plans call for using these amounts, adjusted for inflation and real growth in earned income by 1.019 percent *per wave* for all remaining waves of the 1996 Panel. This is equivalent to an annual increase of 5.8 percent. The mean amounts will *not* be recomputed from microdata for later waves. The formula to compute the topcode amounts for earned income in later waves is shown in Box 10-1.

The following three examples and Table 10-20 illustrate employment income topcoding:

- A black male software consultant works full time for the entire year and reports an annual salary of $196,600. His salary income varies from month to month, however, sometimes dramatically. For this wave, it is $57,100, above the first test of $50,000. The earned income topcode value for black males who work full time, full year is $17,530 (see Table 10-19: example 3, last column). That value will be used instead of the consultant's reported monthly earned income for the 1 month in which his earned income exceeded $12,500.
- A Hispanic female attorney normally works full time, the full year, with an annual income of about $300,000. In the middle of this wave, she has returned from a 6-month maternity leave; for the first 2 months of the wave, she has no earned income. Her income for the wave in question is $51,000, just over the threshold value of $50,000. The earned income topcode value for Hispanic women who work full time, full year is $24,015 (see Table 10-19: example 11, last column). That is the value that will be used as the attorney's monthly earned income for the months in which her income exceeds $12,500.
- A white male psychiatrist spends the month of August at his beach house. While on vacation, he has no earned income. When he returns to the city in September his income returns to its usual level of $20,000 for the next 3 months. His income for the wave is $60,000, exceeding the $50,000 threshold. The earned income topcode for nonblack, non-Hispanic males is $38,270 (see Table 10-19: example 2, last column). That value is used for the 3 months the psychiatrist reported income over $12,500, resulting in a total earned income for the wave of $114,810. That total, after topcoding, is substantially higher than $50,000.
- A white television actress does not work during her series' hiatus. When the series is in production, she works full time. Her annual earned income is $880,000; her income for the wave in question is $160,000. She has earned nothing in the first 3 months of the wave, and $160,000 for the fourth month. The SIPP matrix topcode for nonblack, non-Hispanic women who work full time but less than full year is $49,450 for each month (see Table 10-19: example 8, last column). That value will be assigned for the 1 month of the wave in which the actress reported earned income.

Table 10-20 Example of Employment Income Topcoding in the 1996 Panel

Worker Characteristics	Income	Reported Monthly Income Amounts				Sum for the Wave
		Month 1	Month 2	Month 3	Month 4	
Black, non-Hispanic male, working full time, full year	Reported	$10,000	$10,000	$12,300	$ 24,800	$ 57,100
	Topcoded	$10,000	$10,000	$12,300	**$ 17,530**	$ 49,830
Hispanic female, working full time, full year	Reported	$0	$0	$25,000	$ 26,000	$ 51,000
	Topcoded	$0	$0	$24,015	**$ 24,015**	$ 48,030
Nonblack, non-Hispanic male working full time, part year	Reported	$0	$20,000	$20,000	$ 20,000	$ 60,000
	Topcoded	$0	**$38,270**	**$38,270**	**$ 38,270**	$114,810
Nonblack, female, not full year	Reported	$0	$0	$0	$160,000	$160,000
	Topcoded	$0	$0	$0	**$ 49,450**	$ 49,450

Topcoding Prior to the 1996 Panel

Prior to the 1996 Panel, the data dictionary indicates a topcode of $33,332 for monthly income; that is also the income topcode for the wave. That topcode is, therefore, rarely used for a single month. In most cases, the monthly income is topcoded at $8,333 (one-fourth of $33,332), which actually represents $8,333 or more. Individual amounts above $8,333 may occasionally be shown if the respondent's income varied considerably from month to month. For example, if a respondent's income from a single job was concentrated in only 1 of the 4 reference months, SIPP could show a figure as high as $33,332.

Summary income variables on the person, family, and household records are simply the sums of the component variables after they have been topcoded. The summary variables are not independently topcoded. Thus, a person with high income from several sources (multiple jobs, businesses, property) could have aggregate monthly income well over the topcode for each source and yet SIPP could still be greatly understating the person's true income.

As shown in Table 10-21, person 101 has wages topcoded. The person received considerably more money in December than in the other months. In addition, total family income and total household income are the sum of the income amounts (in this case, WS1AMT+S01AMT) after they have been topcoded.

Table 10-21. Example of Topcoding in the Core Wave Files Prior to the 1996 Panel: Single Person Household

Person Number (PNUM)	Calendar Month (MONTH)	Household Total Income (HTOTINC)	Family Total Income (FTOTINC)	Topcoded Wages (WS1AMT)	Social Security (S01AMT)	Actual Wages
101	10	$9,333	$9,333	$8,333	$1,000	$ 8,333
101	11	$9,333	$9,333	$8,333	$1,000	$ 8,333
101	12	$9,333	$9,333	$8,333	$1,000	$12,123
101	01	$9,583	$9,583	$8,333	$1,250	$ 9,456

Using Allocation (Imputation) Flags

As described in Chapter 4, the Census Bureau often imputes information when a person does not respond to the survey or to a particular question.

1. Prior to the 1996 Panel, the whole record may have been imputed because the person refused to be interviewed (and no proxy interview was obtained) or because the person left the sample in the middle of the wave and no interview was conducted. If that happened, INTVW will be 3 or 4.[17]
2. A variable of interest may be imputed. In the core wave files prior to the 1996 Panel, there is an allocation (imputation) flag for almost all of the person-level variables. Beginning with the 1996 Panel, there is an allocation (imputation) flag associated with every variable subject to imputation. For example, AEDUCATE is the allocation (imputation) variable that identifies whether EEDUCATE is imputed.

For labor force items, the Census Bureau uses the following special imputation procedures when a person has no current wave information indicating whether or not he or she worked during the reference period.[18] If the Census Bureau can infer from what it knows about the previous reference period whether the person had a job or business at the start of the current period, the Census Bureau carries out the following procedure:

1. If the person was working at the end of the prior wave, then labor force participation is imputed from a single donor for the complete current wave.
2. The Census Bureau then projects job characteristics for the person from the person's prior wave through the current wave.

[17] For cases in the 1996 Panel for whom prior wave information did not exist for a person-level noninterview (such as in Wave 1 or in Waves 2–12 when the person was new to the sample), the whole record may have been imputed. To identify such cases, users need to check both person number (to distinguish wave of entry into the sample) and EPPINTVW, which will be 3 or 4 for these cases.

[18] Chapter 4 contains a discussion of how analysts can determine whether these special imputation procedures were used.

3. Finally, the Census Bureau edits the job characteristics for consistency with the imputed labor force participation variables.

This procedure is known as an EPPFLAG imputation, after the name of the variable that indicates its use.

If a person was a nonworker in the prior wave or the Census Bureau cannot infer work status on the basis of prior wave data, then the person's work status is imputed. If the person is imputed as a worker in the reference period, the Census Bureau imputes the complete set of job/business characteristics variables and labor force participation variables to the person from one donor, in order to maintain consistency among the fields. That procedure is called a "little Type Z" imputation.

For some items in some cases, a direct logical or carryover imputation is made. The carryover imputation takes the previous wave's value for the item for the sample member and imputes it to the current wave. That imputation is done particularly for items that rarely (or never) change for a sample member across waves (such as sex and race) or for items that change in predictable ways (such as age).

Variables are imputed and the allocation (imputation) flags are set *before* composite variables are created. For example, if income is imputed for one member of a household, that person's allocation (imputation) flag is set. However, total household income is computed *after* that imputation; if any household member had any income imputed, then total household income is based, in part, on imputed information. There is no direct indication on the records of other household members that any information has been imputed.

Because the edit and imputation procedures used in the core wave files and in the full panel longitudinal research files are different, data from the two sources will not always agree. See Chapter 4 for a more detailed discussion of the SIPP edit and imputation procedures.

Using Weights

The core wave files include a number of alternative reference month weights for use in data analysis. Table 10-22 includes examples of the weights for the 1996 and the 1990–1993 Panel core wave files. The choice of the appropriate weight for a given analysis depends on the population of interest for that analysis—person, household, family, or related subfamily. Suggestions for which weights to use and how to use them are included in the source and accuracy statements that accompany files ordered from the Census Bureau. Also, Chapter 8 of the Guide contains a full discussion of how to use weights in the core wave files.

Table 10-22. Weight Variables in SIPP Core Wave Files for the 1996 and 1990–1993 Panels

Variable Name	Description
WPFINWGT (FNLWGT)	Reference month, final weight of person
WHFNWGT (HWGT0)	Reference month, final weight of household
WFFINWGT (FWGT)	Reference month, final weight of family
WSFINWGT (SWGT)	Reference month, final weight of related subfamily
WPFINWGT (P5WGT)[a]	Interview (5th) month, final weight of person
WHFNWGT (H5WGT)[a]	Interview (5th) month, final weight of household

[a] Beginning with the 1996 Panel, SIPP files no longer include the interview month weights.

Identifying States

For the 1996 Panel, the variable TFIPSST identifies 45 states and the District of Columbia. To help protect the confidentiality of respondents, the Census Bureau combined the remaining five states as follows:

1. Maine, Vermont; and
2. North Dakota, South Dakota, Wyoming.

The core wave files from panels prior to the 1996 Panel contain the variable HSTATE, which identifies 41 individual states and the District of Columbia; the nine other states are combined into three groups:

1. Maine, Vermont;
2. Iowa, North Dakota, South Dakota; and
3. Alaska, Idaho, Montana, Wyoming.

Even though it is possible to identify most states, the SIPP sample was not designed to be representative at the state level and should not be used to produce direct state-level estimates. The state variable is included on the public use files to allow examination of how state-level characteristics affect national estimates. For example, a user could apply the state-specific eligibility criteria for a means-tested program in order to arrive at a national estimate of the number of people eligible for the program. Because some states are not uniquely identified, some method of allocating the state-specific eligibility rules to sample persons in those states would need to be devised.

Identifying Metropolitan Areas

The core wave files include two variables useful in identifying metropolitan areas. The first variable, TMETRO (HMETRO), identifies residences located in metropolitan areas. It can be used to produce national estimates of the metropolitan population. However, it *cannot* be used to produce estimates of the nonmetropolitan population. To protect respondent confidentiality, the Census Bureau recoded and identified a small random sample of metropolitan households in the public use files as nonmetropolitan. The remaining metropolitan sample should still produce (approximately) unbiased estimates of the metropolitan population. However, the procedure "contaminates" the nonmetropolitan sample, and estimates of nonmetropolitan characteristics based on that sample will be biased (the magnitude of the bias depends on the specific analysis being performed).

A second variable, TMSA (HMSA), identifies 93 MSAs (Metropolitan Statistical Areas) and CMSAs (Consolidated Metropolitan Statistical Areas), as defined by the Office of Management and Budget.

11. Using Topical Module Files

This chapter discusses procedures for working with data from the topical module public use files from the Survey of Income and Program Participation (SIPP). The chapter begins by describing the documentation that accompanies the topical module public use files obtained from the Census Bureau. The discussion then turns to the data files themselves. The data file structure is described, and detailed explanations are provided about how to use the topical module files when performing common tasks. Those tasks include:

- Using the monthly interview status variables;
- Identifying people, households, and families;
- Using imputation flags; and
- Identifying states and metropolitan areas.

Before reading this chapter, users should read Chapter 9, "The SIPP Public Use Files," for an introduction to Section II. Analysts using only one topical module file also should read about the use of sample weights (Chapter 8) and the computation of standard errors (Chapter 7). Those planning on merging data from a topical module to data from the core wave or full panel files should also read Chapter 10 for information about the core wave files, Chapter 12 for information about the full panel files, and Chapter 13 for information about linking SIPP public use files.

This chapter focuses on the topical module files. It is written so that it can be used independently of the chapters describing the core wave and full panel files. Although there are many similarities across the three types of SIPP public use data files, important differences do exist. Because those differences are sometimes subtle, users familiar with the core wave and full panel files should read this chapter carefully, paying close attention to information about variable names and file structures. Tables 9-2 and 9-3 summarize the differences between the core wave, topical module, and full panel longitudinal research files.

For the 1996 Panel, most variable names changed from those used in previous panels. To aid users working with files from panels prior to 1996, this chapter presents both the old and the new variable names when the text applies to both 1996 and pre-1996 panel files. In the main body of the text, the old names are presented in parentheses following the new names. For example, the sample unit ID variable name, which is SSUID in the 1996 Panel, was SUID in previous panels; it is written in this chapter as SSUID (SUID). In tables, a variety of methods are used to present both the old and the new names.

Using the Technical Documentation of the Topical Module Files

Each data file received from the Census Bureau comes with a set of technical documentation and a data dictionary. The technical documentation includes:

- The item booklets (for the 1996 Panel);
- The paper survey instrument (for panels prior to 1996);
- A glossary of selected terms;
- A cross-walk, mapping reference months into calendar months for each rotation group;
- A source and accuracy statement describing the sample weights and the computation of standard errors; and
- User Notes.

The survey instrument is vital to understanding what questions were asked, how they were asked, the order in which they were asked, to whom they were asked, and the way in which the answers were recorded. Some questions employ skip patterns (Chapter 3), so users should pay particular attention to which questions were skipped for which respondents. The skip patterns are best understood by consulting the survey instruments. With the introduction of computer-assisted interviewing (CAI) in the 1996 Panel, questionnaire documentation is now available from the SIPP Web site (http://www.sipp.census.gov/sipp/).

The source and accuracy statements provide information about the weights on the files, when and how to make adjustments to the weights, and one approach to computing standard errors for some common types of estimates. More detailed discussions of those topics are provided in Chapters 7 and 8 of this Guide.

The data dictionary provides a detailed description of each variable on the file. It describes four aspects of each variable:

1. The definition,
2. The sample universe of the corresponding survey question,
3. The ranges for all legal values, and
4. The location (and size) in the file.

A machine-readable version of the data dictionary accompanies each data file. It can also be downloaded from the Internet (http://www.sipp.census.gov/sipp/).

The data dictionary is formatted to facilitate processing by user-written computer programs. The upper panel of Figure 11-1 shows an excerpt from the data dictionary for the topical module from Wave 1 of the 1996 Panel. A "D" in the first column signifies that the next few lines define the variable: (1) the variable name; (2) the size (i.e., how many digits it contains); (3) the starting position; and (4) the definition. Lines beginning with a "T", added with the 1996 Panel, contain short variable descriptions that can be used by many software packages as variable labels.

Figure 11-1. Excerpt from the Data Dictionary for the Topical Module Files Wave 1 of the 1996 SIPP Panel

Wave 1 of the 1996 SIPP Panel

```
D EENTAID   3   45
T PE: Address ID of hhld where person entered Sample
    Address ID of the household that this person belonged to at the time this
    person first became part of the sample. Address ID in a specific wave should
    never be greater than (WAVE * 10 + 9).
U All persons
V   11:129 .Entry address ID

D EPPPNUM   4   48
T PE: Person number
    Person number. This field differentiates persons within the sample unit.
    Person number is unique within the sample unit across all waves of a panel.
    Person number for a specific wave should never be greater than
    (WAVE * 100 + 99).
    U All persons
V  101:1299 .Person number

D EPOPSTAT  1   52
T PE: Population status based on age in fourth ref. Month
    Population status. This field identifies whether or not a person was
    eligible to be asked a full set of questions, based on his/her age in
    the fourth month of the reference period.
U All persons
V     1 .Adult (15 years of age or older)
V     2 .Child (Under 15 years of age)

D EPPINTVW  2   53
T PE: Person's interview status at time of interview
U All persons
V     1 .Interview (self)
V     2 .Interview (proxy)
V     3 .Noninterview - Type Z
V     4 .Nonintrvw - pseudo Type Z. Left sample during the reference
V     5 .Children under 15 during reference period
```

(figure continues)

Figure 11-1. Excerpt from the Data Dictionary for the Topical Module Files (*continued*)

Wave 3 of the 1993 SIPP Panel

```
D ENTRY    2   30
 Entry address ID
    Address of the household that person belonged to at the time person
    first became part of the sample
U All persons, including children

D PNUM    3   32
 Person number
U All persons, including children

D FILLER   3   35
 Filler

D FINALWGT  9   38
 Person weight (interview month)
    There are four implied decimal places.
U All persons, including children
```

A "U" in the first column signifies that the next words describe the sample universe.[1] A "V" in the first column indicates that the next number and phrase describe one of the values of the variable. A blank in the first column denotes either a variable description or other comment. A period (.) before a word denotes the start of the value label.

Prior to the 1996 Panel, the dictionaries had a different format, shown in the second panel of Figure 11-1. A "D" in the first column signifies that the next few lines define the variable: (1) the variable name; (2) the size (i.e., how many digits it contains); (3) the starting position; and (4) the definition. A "U" in the first column signifies that the next words describe the sample universe.[2] A "V" in the first column indicates that the next number and phrase describe one of the values of the variable. An asterisk in the first column denotes a comment. A period (.) before a word denotes the start of the value label.

Figure 11-2 shows sample SAS and FORTRAN syntax for reading the data described by the codebook fragments in Figure 11-1. Additional SAS program code could be used to associate value labels (a SAS "format") with the INTVW variable.

[1] The universe definitions included in the data dictionaries prior to the 1996 Panel were not always accurate. Users of pre-1996 SIPP Panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

[2] See footnote 1.

Figure 11-2. Corresponding SAS and FORTRAN Syntax to Read Data from Topical Module Files

Wave 1 of the 1996 Panel

SAS

```
Input
  @45  EENTAID  3.
     EPPPNUM  4.
     EPOPSTAT 1.
     EPPINTVW 2.
     ;
LABEL EENTAID = "Adrs ID where person entered sample"
   EPPPNUM = "Person number"
   EPOPSTAT = "Population status based on age in fourth"
   EPPINTVW = "Person's interview status"
   ;
```

FORTRAN

```
   READ(INFILE,1000) EENTAID EPPPNUM EPOPSTAT EPPINTVW
1000 FORMAT(T45,I3,I4,I1,I2)
```

Wave 3 of the 1993 SIPP Panel

SAS

```
Input
  @30  ENTRY   2.
     PNUM    3.
  @38  FINALWGT  9.4
  ;
LABEL ENTRY  = "Entry address ID'
   PNUM   = "Person number"
   FINALWGT = "Person weight (interview month)"
   ;
```

FORTRAN

```
    READ(infile,1000) ENTRY, PNUM, INTVW
1000  FORMAT(T457,I2,I3,I1)
```

Relationship of the Topical Module Data Files to the Survey Instrument

Each wave's survey instrument includes one or more topical modules,[3] as described in Chapter 3. The questions in those modules are often asked after the core survey questions and can be found toward the end of the survey instrument. The data from the topical modules are usually combined into one topical module data file for each SIPP wave.

The topical module data dictionary does not replicate the survey instrument. Thus, analysts should keep a few things in mind when using the data:

- The variables on the data files do not correspond one-to-one with the questionnaire items—the variables are listed in a different order, some are not included in the public use files, and some are created from a combination of other variables;
- The range of possible values of the variables on the data files does not always correspond one-to-one with the response categories shown on the survey instrument or in the data dictionary;
- The variable name in the data dictionary may not readily indicate the variable's content;
- Prior to the 1996 Panel, some variable names were used in different topical module files for different variables. For example, in the 1990 Panel, TM8400 was used in the Wave 2 topical module for a variable that indicates whether the respondent completed 12th grade. The same variable name was used in the Wave 6 topical module to indicate whether the respondent was a parent of children under 21 years of age living in the respondent's household.
- The complexity of the skip patterns may not be apparent just by looking at the data dictionary. Many questions were administered only to the household reference person, or to adults (age 15 years or older), or to people 25 years or older, or to some other subset of survey respondents.[4]

To avoid potential problems and confusion, analysts should become familiar with the survey instrument before using the data. When working with the data, refer to both the survey instrument and the data dictionary.

[3] Prior to the 1992 Panel, there were no topical modules administered with the Wave 1 interview, although some topical content was included in the Wave 1 core questionnaire for the purpose of obtaining historical information. As of the 1992 Panel, Wave 1 has had topical modules.

[4] The universe definitions included in the data dictionaries prior to the 1996 Panel were not always accurate. Users of pre-1996 SIPP panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

Structure of the Topical Module Files

The topical module files for the 1996 Panel contain one record for each person who was in the sample with a completed (or imputed) interview in the fourth month of the wave's reference period (the month immediately prior to the interview). This arrangement is similar to the person-month format of the core wave files, but only records for month four are included in the topical module files. Prior to the 1996 Panel, the topical module files contained one record for each person who was interviewed or for whom an interview was attempted in that wave (Table 11-1 shows one record for each such person; compare with Table 10-1, which shows up to four records per sample person in the core wave files).[5]

In general, each topical module file contains data for all of the topical module subject areas administered during a particular wave.[6] Each topical module file also contains selected information from the SIPP core; thus, for some analyses, those files can be used independently from the core wave and full panel data files. When more detailed information from the SIPP core is needed, data from the topical modules must be merged with data from the core wave or full panel files. Chapter 13 provides a detailed discussion of merging SIPP files.

Table 11-1. Example of the Topical Module File Structure

1996 Panel			
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**
123456789123	021	011	0101
123456789123	021	011	0102
123456789123	021	021	0201
123456789123	021	021	0202
Panels Prior to 1996			
Sample Unit ID (ID)	**Current Address ID (ADDID)**	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**
123451000	21	11	101
1234551000	21	11	102
123451000	21	21	201
123451000	21	21	202

[5] The variables shown—sample unit ID, current address ID, entry address ID, and person number—are discussed in detail later in this chapter.

[6] Chapter 3 offers a detailed listing of the topical modules administered with each wave of each SIPP panel.

The topical module file structure differs from that of the core wave files in the following ways:

- For the 1996 Panel, the topical module files contain one record for each person who was a SIPP sample member during month four of the wave; the core wave files contain one record per person *for each month* the person is in the sample.
- Prior to the 1996 Panel, the topical module files contain one record per person for each person present in a SIPP household *at the time of the interview*; the core wave files contain one record per person *for each month* the person was in the sample during the *previous 4 months*.
- Prior to the 1996 Panel, the topical module files include records for people whose entire household refused to be interviewed or left the sample;[7] those people are excluded from the core wave files.
- Prior to the 1996 Panel, the structure of the topical module files was roughly similar to that of the full panel files, containing one record per person.

Reference Periods and Samples

Sample definitions and reference periods in the topical modules vary across panels, across topical modules within panels, and even within topical modules. Users should pay careful attention to those details in the topical module files they are using.

In the 1996 Panel, most topical module questions were asked only of people who were in the SIPP sample during the fourth month of the wave's reference period. People who were members of SIPP households at the time of the interview (month five) but who were not members of SIPP households during the previous month were not asked the topical module questions in the 1996 Panel. In the 1996 Panel, many of the questions refer to just that month (month four). However, some topical module questions, and in some cases entire topical modules, refer to longer periods of time, such as the previous 4 months, the previous year, or, in the various history topical modules administered with Wave 1, the person's life before SIPP.

Prior to the 1996 Panel, most topical module questions were asked of people who were in the SIPP sample at the time of the interview (month five). This included people who were household members at the time of the interview but who were not members of SIPP households at any time during the previous 4 months, the reference period for SIPP core questions in that wave.[8] Many questions asked about "current" (month five) conditions, although some asked about longer periods in the past.

[7] 7 Panels that included topical modules in Wave 1, such as the 1993 and 1996 Panels, exclude those people from the Wave 1 topical module files.

[8] This has important implications for procedures used to merge the topical modules to data from the core. Core data that correspond to the same reference month as a topical module must often be merged from the *subsequent* wave rather than from the same wave as the topical module, as discussed in Chapter 13.

Using a Person's Monthly Interview Status Variables

A person's monthly interview status variable is used to determine whether the data for that person in a given month should be used. Some analysts refer to it as the *in sample* variable to distinguish it from the household interview status variable, EOUTCOME (ITEM36B), and another variable that indicates the type of interview or noninterview for the person, EPPINTVW (INTVW). The interview status variable has three possible values: 0, 1, and 2. A value of 1 indicates that the person was both in-scope for the survey (a member of the population that the SIPP sample is intended to represent) and, aside from some item nonresponse, provided complete answers to the SIPP core questions for the reference month in question.[9]

Monthly Interview Status in the Topical Module Files from the 1996 Panel

There is only one interview status variable in the topical module files from the 1996 Panel. That variable, EPPMIS4, identifies a person's status in the fourth reference month of the wave. Because the topical module files from the 1996 Panel contain records only for people for whom this variable is equal to 1 (and so equals 1 on all records in the file), EPPMIS4 can be safely ignored when working with topical module files from the 1996 Panel.

Monthly Interview Status in the Topical Module Files from Panels Prior to 1996

The topical module files for panels prior to 1996 are different. On those files, a person's interview status variable is labeled PP-MIS1, PP-MIS2, PP-MIS3, PP-MIS4, and PP-MIS5. These variables refer to the four reference months of the wave (PP-MIS1 to PP-MIS4) and the interview month itself (PP-MIS5).

The monthly interview status is the only reliable guide to whether the data for a given person should be used in a given month. Analysts should use data for only those months in which a person's interview status (PP-MIS) is equal to 1.[10]

[9] The only exception is for Type Z noninterviews. For Type Z noninterviews prior to the 1996 Panel, complete records for the SIPP core were imputed and the monthly interview status variable was set to 1, indicating that, for most analytic purposes, the responses should be treated as though they were provided by the respondent. This exception is handled similarly in the 1996 Panel when there is no prior wave information. When prior wave information exists, items are imputed using the same hot-deck methods applied to instances of item nonresponse.

[10] As a safeguard against inadvertently using data for months when PP-MIS is not equal to 1, all monthly variables in the user's data extract should be set to a missing value for months when PP-MIS is not equal to 1. Most statistical packages allow certain values to be flagged as missing. Once flagged, those values are excluded from computations.

Any data present for months when a person's interview status is coded either 0 or 2 should be ignored. A code of 0 indicates that the person was not in the sample that month, and a code of 2 indicates a noninterview for that month.

On the topical module files for panels prior to 1996, the topical module questions were asked only of sample members with PP-MIS5 equal to 1:[11] that is, the topical module questions were asked only of those who were in the SIPP sample at the time of the interview. Because the reference periods of the topical module questions vary, some topical module questions contain information about people who had been secondary sample members during previous months, even though they were no longer part of the SIPP sample at the time of the interview. The variables PP-MIS1 to PP-MIS4 are useful when working with topical module questions that refer to previous months. The four variables are also useful when merging topical module data with data from the core, a topic discussed in Chapter 13.

Four sample members are shown in Table 11-2. Two were present in the interview month (PP-MIS5 = 1), and two were not present (PP-MIS5 = 2). Analysts interested in just the interview month should use data only for people with PP-MIS5 = 1. In this example, only persons 101 and 201 would be included.

Table 11-2. Monthly Interview Status Variables in the 1984-1993 SIPP Panels

Sample Unit ID (ID)	Current Address ID (ADDID)	Entry Address ID (ENTRY)	Person Number (PNUM)	Rotation Group (ROTATION)	PP-MIS				
					1	2	3	4	5
123451000	11	11	101	1	1	1	1	1	1
123451000	11	11	102	1	1	1	2	2	2
123451000	11	11	201	1	2	2	2	2	1
123451000	11	11	202	1	0	0	2	2	2

If the research focuses on January, analysts should use data only for people with PP-MIS*x* = 1, where *x* corresponds to the reference month that contains information about January (which varies by wave and rotation group). Assuming an analyst is interested in January 1994, the example represents Wave 4 and rotation group 1 of the 1993 Panel (see Table 11-3 for the reference months); the analyst would use only the people with PP-MIS1 = 1. Thus, only persons 101 and 102 would be included.

Table 11-3. Interview Month and Reference Months for Each Rotation Group in Wave 4 of the 1993 Panel

Rotation Group	Reference Months for Core Questions	Interview Month
2	Oct., Nov., Dec. 1993; **Jan. 1994**	Feb. 1994
3	Nov., Dec. 1993; **Jan.**, Feb. 1994	Mar. 1994
4	Dec. 1993; **Jan.**, Feb., Mar. 1994	Apr. 1994
1	**Jan.**, Feb., Mar., Apr. 1994	May 1994

[11] In some cases, questions are asked of all household members over 14 years old. In other cases, they may be asked only of the household reference person. There are also topical modules in which other subsets of household members are interviewed.

As demonstrated by this example, the topical module files for panels conducted before 1996 contain a record for each person for whom no interview data were collected, either because the person refused to be interviewed (and no proxy interview was obtained) or because the person left the survey sample (e.g., died or entered the Armed Forces or an institution). Those individuals have PP-MIS5 = 2 and PP-MISj = 1 for j = 1, 2, 3, or 4 or INTVW = 3 or 4. Their demographic information was gathered from the previous time that they were successfully interviewed; if they have topical module information, it was completely imputed by the Census Bureau.

Comparison of Variables in the Topical Module and Core Wave Files

The topical module files contain a number of variables that are also present in the core wave files. These include variables needed to identify the household and the person. Also included are selected background (demographic) characteristics. In the 1996 Panel, the values for the background characteristics correspond to the month-four values in the core wave file for the same wave for the 1996 Panel. Variables common to the core wave and topical module files are generally given the same names in both files. For example, SSUID is used for the sample unit identifier, SHHADID is the current address ID, and EPPPNUM is the person number on both files.[12] Among the background variables, TAGE is used on both files for the respondent's age, and EMS is used for the respondent's marital status. Table 11-4 shows the 27 variables that are common to the core wave file and topical module file from Wave 1 of the 1996 Panel.

Prior to the 1996 Panel, the demographic data on the topical module files corresponded to the interview month (month five), not to any of the 4 reference months for the core interview. For that reason, the information in variables such as AGE, RRP, and MS (the respondent's age, relationship to the household reference person, and marital status) could differ from the core wave file variables of the same names for the wave in which the topical module was administered. This would indicate that a change occurred between the last month of the reference period (month four) and the interview month (month five). Some variables included on both the core wave and topical module files have different names. As shown in Table 11-5, sample unit ID, rotation group, state, interview status in month five, and the person-level weight are contained in both files but have different variable names.

[12] Use of common names facilitates merging of the core wave and topical module files from the 1996 Panel. Merging files is discussed extensively in Chapter 13.

Table 11-4. Variables Common to the Core Wave and Topical Module Files from Wave 1 of the 1996 Panel

Variable Name	Description
EEDUCATE	Highest degree received or grade
EENTAID	Address ID of household where person entered
EMS	Marital status
EORIGIN	Origin of this person
EOUTCOME	Interview status code for this household
EPNDAD	Person number of father
EPNGUARD	Person number of guardian
EPNMOM	Person number of mother
EPNSPOUS	Person number of spouse
EPOPSTAT	Population status based on age
EPPINTVW	Person's interview status
EPPPNUM	Person number
ERACE	Race of this person
ERRP	Household relationship
ESEX	Gender of this person
RDESGPNT	Designated parent or guardian flag
RFID	Family ID number for this month
RFID2	Family ID excluding related subfamily
SHHADID	Household address ID—differentiates households
SPANEL	Sample code—indicates panel year
SROTATON	Rotation of data collection
SSUID	Sample unit identifier
SSUSEQ	Sequence number of sample unit — primary
SWAVE	Wave of data collection
TAGE	Age as of last birthday
TFIPSST	FIPS state code
WPFINWGT	Person weight

Table 11-5. Examples of Same Variables with Different Names in the Core Wave and Topical Module Files Prior to the 1996 Panel

Description	Variable Name in the Core Wave File	Variable Name in the Topical Module File
Sample unit ID	SUID	ID
Rotation group	ROT	ROTATION
State of residence	HSTATE	STATE
Monthly interview status in the interview month	MIS5	PP-MIS5
Person-level weight in the interview month	P5WGT	FINALWGT

Identifying People

There are many occasions when it is necessary to identify which records belong to each individual in the SIPP data files. This need arises, for example, when

- Merging data from topical module files to data from the core wave or full panel files,
- Merging data from two or more topical module data files,
- Linking husbands and wives, and
- Linking parents and children.

In the 1996 Panel, two variables are needed to uniquely identify a person: the sample unit ID and the person number.[13] For files from panels prior to 1996, three variables are needed to uniquely identify a person: the sample unit ID, entry address ID, and person number. Table 11-6 shows the variable names used in the topical module files for the 1996 Panel and for the pre-1996 Panels.

Table 11-6. Variables Used to Uniquely Identify a Person in the Topical Module Files

Variable Name	Description
SSUID (ID)	Sample unit ID
EENTAID (ENTRY)	Entry address ID (not needed in the 1996 panel)
EPPPNUM (PNUM)	Person number

The variables can be described as follows:

- SSUID (ID) uniquely identifies each initially sampled dwelling unit.[14] Every person in a core wave file was either a member of one of those units (an original sample member) or lives with someone who was a member of an initially sampled dwelling unit. A person's connection to that unit is an attribute of that person and does not change over time.[15] This means that as people move from address to address, their SSUID (ID) stays the same. As new people join the homes of original sample members, they receive the SSUID (ID) of the original sample members.

[13] Users should note that in the 1996 Panel, the entry address ID is no longer needed for unique identification. Its continued use will not create any problems; it is simply redundant information. That is a change from earlier panels, in which the entry address ID was key to uniquely identifying a person.

[14] The SSUID (ID) is a random recode of three other variables in the Census Bureau's internal (not public use) files: the respondent's sampling area (primary sampling unit), the cluster of housing units within that area (called the "segment"), and a sequentially assigned serial number. Those three variables are omitted from the public use files to protect the confidentiality of the respondents.

[15] There is one rare exception to this rule for panels prior to 1996, which is described in the section entitled "Identifying Movers" later in this chapter.

- EENTAID (ENTRY) identifies the address where the person lived at the time he or she was first interviewed. It does not change even if the person moves.[16] Prior to the 1996 Panel, it was used in conjunction with the person number and the sample unit ID to uniquely identify people within the sampling unit. It is not needed to uniquely identify people in the 1996 Panel. Values for this variable are unique only within sample units. The entry address ID has two components. The first part of the ID number (two digits in the 1992 and 1996 Panels, and one digit in all others) identifies the wave in which SIPP interviews were first conducted at the address. The second part of the number (one digit in all panels) sequentially numbers addresses within a sample unit [SSUID (ID)] that enter the sample in the same wave. See Chapter 10 for a more complete discussion.
- Prior to the 1996 Panel, PNUM uniquely identified a person within the sample unit and entry address ID. In the 1996 Panel, EPPPNUM uniquely identifies a person within the sample unit. EPPPNUM (PNUM) does not change even if the person moves.[17] The first part of EPPPNUM (PNUM) (two digits in the 1992 and 1996 Panels, and one digit in all others) indicates the wave in which the person was first interviewed.[18] The remaining two digits are sequentially assigned within the household. Thus, original sample members are assigned person numbers ranging from 100 to 199. Individuals who enter the SIPP sample in Wave 2 are assigned a person number ranging from 200 to 299. Those who enter in Wave 10 are assigned person numbers ranging from 1001 to 1099.

Table 11-7 illustrates how the combination of SSUID (ID), EENTAID (ENTRY), and EPPPNUM (PNUM) uniquely identifies people and provides information about when they first entered the SIPP sample. In this example, there are eight individuals: five are original sample members, one person joined the SIPP sample in Wave 4, one person joined in Wave 7, and one person joined in Wave 10.

To uniquely identify a household or group quarters in the topical module files, analysts should use the two variables shown in Table 11-8.

People with the same SSUID (ID) (sample unit ID) and SHHADID (ADDID) (current address ID) values live in the same household (or group quarters location) in the relevant month. For the 1996 Panel, household membership refers to month four of the wave's reference period. For panels prior to 1996, household membership refers to the interview month. The eight individuals shown in Table 11-9 make up four households. The first household contains the first four individuals. The second household contains one person. The third household contains one person. The fourth household contains two people. (Users may find it helpful to refer to Figure 2-1 [pp. 2-10-2-14], which illustrates the concepts of household and changes in household.)

[16] 16 See footnote 7.

[17] For cases in the 1996 Panel for whom prior wave information did not exist for a person-level noninterview (such as in Wave 1 or in Waves 2–12 when the person was new to the sample), the whole record may have been imputed. To identify such cases, users need to check both person number (to distinguish wave of entry into the sample) and EPPINTVW, which will be 3 or 4 for these cases.

[18] Chapter 4 contains a discussion of how analysts can determine whether these special imputation procedures were used.

Table 11-7. How to Uniquely Identify a Person in the Topical Module Files

1996 Panel				
Sample Unit ID (SSUID)	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**	**Current Address ID (SHHADID)**	**Notes**
123456789123	011	0101	071	Original sample member
123456789123	011	0102	071	Original sample member
123456789123	011	0401	071	Enters SIPP sample in Wave 4
123456789123	071	0701	071	Enters SIPP sample in Wave 7
321456789123	011	0101	031	Original sample member
321456789123	011	0102	032	Original sample member
321456789123	011	0103	101	Original sample member
321456789123	101	1001	101	Enters SIPP sample in Wave 10
Prior to the 1996 Panel				
Sample Unit ID (ID)	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**	**Current Address ID (ADDID)**	**Notes**
123456789	11	101	71	Original sample member
123456789	11	102	71	Original sample member
123456789	11	401	71	Enters SIPP sample in Wave 4
123456789	71	701	71	Enters SIPP sample in Wave 7
321456789	11	101	31	Original sample member
321456789	11	102	32	Original sample member
321456789	11	103	101	Original sample member
321456789	101	1001	101	Enters SIPP sample in Wave 10 (1992 Panel)

[a] Not needed to uniquely identify a person in the 1996 Panel.

Table 11-8. Variables Used to Uniquely Identify a Household or Group Quarters in the Topical Module Files

Variable Name	Description
SSUID (ID)	Sample unit ID
SHHADID (ADDID)	Current address ID in month 4 (in month 5)

Table 11-9. How to Uniquely Identify a Household in the Topical Module Files

1996 Panel			
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	071	0101	Four people in this household
123456789123	071	0102	
123456789123	071	0401	
123456789123	071	0701	
321456789123	031	0101	One person in this household
321456789123	032	0102	One person in this household
321456789123	101	0103	Two people in this household
321456789123	101	1001	
Panels Prior to 1996			
Sample Unit ID (ID)	**Current Address ID (ADDID)**	**Person Number (PNUM)**	**Notes**
123456789	71	101	Four people in this household
123456789	71	102	
123456789	71	401	
123456789	71	701	
321456789	31	101	One person in this household
321456789	32	102	One person in this household
321456789	101	103	Two people in this household
321456789	101	1001	

Identifying Families

The term *family*, as used in Census Bureau publications, refers to a group of two or more people related by birth, marriage, or adoption who reside together; all such individuals are considered members of one family.

The Census Bureau distinguishes among several types of families:

- A *primary family* is a family containing the household reference person and all of his or her relatives. This means that a household composed of a husband and wife, their son, and their son's wife (i.e., the daughter-in-law) is classified as a primary family containing four people.
- A *related subfamily* is a nuclear family that is related to but does not include the household reference person. For example, the son and his wife (i.e., the daughter-in-law) in the preceding example are a related subfamily.
- An *unrelated subfamily* (sometimes called a secondary family) is a nuclear family that is not related to the household reference person. Thus, a husband and wife who live in a friend's house are classified as an unrelated subfamily. A mother and daughter who live in the mother's boyfriend's apartment are classified as an unrelated subfamily.

- A *primary individual* is a household reference person who lives alone or lives with only nonrelatives. Primary individuals are sometimes treated by the Census Bureau as families of only one person and are referred to as pseudo-families.
- A *secondary individual* is not a household reference person and is not related to any other people in the household. Secondary individuals are sometimes treated by the Census Bureau as families of only one person and are referred to as pseudo-families.

In the topical module files for the 1996 Panel, the variables shown in Table 11-10 can be used to uniquely identify families.

Table 11-10. Variables Used to Uniquely Identify a Family in the Topical Module Files for the 1996 Panel

Variable Name	Description
SSUID	Sample unit ID
SHHADID	Current address ID
and one of the following:	
RFID	Family ID in month four of the wave
RFID2	Family ID in month four (excluding related subfamily members; RFID2=0 for related subfamily members)

The Census Bureau has two principal methods for distinguishing families that are based on the variables and numbering schemes shown in Table 11-10. Analysts must remember to choose which type of family classification they want and then use the appropriate method.

- The first method defines a family as all persons who are related and living together. The family ID variable RFID is used with this definition. RFID groups the household reference person with all related household members by assigning them the same ID number. This family group corresponds to the Census Bureau's definition of primary family. RFID groups members of each unrelated subfamily (and primary and secondary individuals) separately.
- The second method is similar to the first in defining a family, but the family excludes related subfamilies. The family ID variable RFID2 is used with this definition. RFID2 equals zero for related subfamilies. RFID2 groups members of each unrelated subfamily (and primary and secondary individuals) in the same way as RFID—each group has a unique number.[19]

Table 11-11 illustrates the difference between the RFID and RFID2 variables. Those variables refer to month four of the wave's reference period. For example, a mother, a father, and a child would be family 1 (RFID = 1). The first household in the table contains a primary family of five people. The primary family contains members of related subfamilies. However, the topical

[19] The variables included on the topical module files do not allow analysts to distinguish among different related subfamilies living in the same household. If needed, the RSID variable (which groups each related and unrelated subfamily separately) can be merged from the core wave files. Chapter 10 discusses the core wave files, and Chapter 13 discusses the merging of multiple SIPP files.

Table 11-11. Uniquely Identifying Families in the Topical Module Files in the 1996 Panel

Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Family ID, Including Related Subfamily (RFID)	Family ID, Excluding Related Subfamily (RFID2)	Person Number (EPPPNUM)	Notes
110011111123	11	1	1	0101	This household contains a primary family of five people. The primary family contains one or more related subfamilies.
110011111123	11	1	0	0102	
110011111123	11	1	0	0103	
110011111123	11	1	0	0104	
110011111123	11	1	0	0105	
110077777723	11	1	1	0101	Three households formed by people who were originally members of the same originally sampled household (SSUID of 110077777723). Two subfamilies split off from the original household to become two new primary families at addresses 21 and 22.
110077777723	21	1	1	0102	
110077777723	21	1	1	0103	
110077777723	22	1	1	0104	
110077777723	22	1	1	0105	
122210000123	11	1	1	0101	This household contains a primary family and two unrelated subfamilies.
122210000123	11	1	1	0104	
122210000123	11	2	2	0305	
122210000123	11	2	2	0306	
122210000123	11	3	3	0307	
122210000123	11	3	3	0308	
555555555123	21	1	1	0101	This household contains a primary individual and an unrelated subfamily.
555555555123	21	2	2	0201	
555555555123	21	2	2	0202	
555555555123	21	2	2	0203	
610000000123	32	1	1	0101	Primary individual.
897454644123	11	1	1	0101	Group quarters with two secondary individuals.
897454644123	11	2	2	0102	

module files for the 1996 Panel do not contain the variables needed to determine whether all subfamily members are members of the same subfamily. To determine that, an analyst would need to merge the RSID variable from the month four records in the core wave file.

The second "household" is actually three households, each containing a primary family, that originally formed one household. The third household contains a primary family and two unrelated subfamilies. The fourth household contains a primary family and two unrelated subfamilies. The fifth household contains a primary individual and an unrelated subfamily. The fifth household contains only a primary individual. The sixth household is a group quarters containing two people.

Other Variables Describing Household and Family Composition

The topical module files contain several additional variables from the SIPP core that describe household and family composition.[20] The household composition variables included in the topical module files from the 1996 Panel and from panels prior to 1996 are shown in Table 11-12. Additional variables from the core wave files and the full panel files can be merged with data from the topical module files when added detail is needed (Chapters 10, 12, and 13).

Table 11-12. Household and Family Composition Variables in the Topical Module Files

1996 Panel	
Variable Name	**Description**
ERRP	Relationship to household reference person in month four
EMS	Marital status in month four
EPNMOM	Person number of mother in month four
EPNDAD	Person number of father in month four
EPNGUARD	Person number of guardian in month four
EPNSPOUS	Person number of spouse in month four
RDESGPNT	Designated parent or guardian in month four
Panels Prior to 1996	
RRP	Revised relationship to the household reference person (living with relatives, child of household reference person, etc.)
PNSP	Person number of spouse
PNPT	Person number of parent

Using the Relationship to Reference Person [ERRP (RRP)] Variable

As Table 11-13 shows, ERRP (RRP) provides a summary description of how each individual is related to the *household reference person*.[21]

[20] Detailed information about the relationships between members is collected in the Household Relationships topical module. For the 1996 Panel, those data provide extensive information about household composition during month four of the wave's reference period. For earlier panels, the topical module provides information about household composition at the time of the interview.

[21] Prior to the 1996 Panel, the RRPU variable, available in the core wave files, provides additional detail not contained in the RRP variable. When needed, RRPU can be merged to data from the topical module files (Chapters 10 and 13).

Table 11-13. Relationship to the Household Reference Person in the Topical Module Files

1996 Panel	
ERRP	**Description**
1	Reference person w/related people in household
2	Reference person w/out related people in household
3	Spouse of reference person
4	Child of reference person
5	Grandchild of reference person
6	Parent of reference person
7	Brother or sister of reference person
8	Other relative of reference person
9	Foster child of reference person
10	Unmarried partner of reference person
11	Housemate or roommate
12	Roomer or boarder
13	Other nonrelative of reference person
Panels Prior to 1996	
Revised Relationship to the Household Reference Person (RRP)	**Description**
1	Household reference person, living with relatives
2	Household reference person, living alone or with nonrelatives
3	Spouse of household reference person
4	Child of household reference person
5	Other relative of household reference person
6	Nonrelative of household reference person, but related to other members of the household
7	Nonrelative of all members of the household

The ERRP (RRP) variable contains summary information about each person's relationship to the household reference person. Analysts should bear in mind that the household description depends upon the identity of the household reference person. For example, the household in Table 11-14 contains a mother, her daughter, and her daughter's son. If the mother is the household reference person [ERRP = 1 (RRP = 1)], her daughter is listed as a child of the household reference person [ERRP = 4 (RRP = 4)] and the daughter's son is listed as a grandchild of the reference person in the 1996 Panel (ERRP = 5), but as another relative of the household reference person in earlier panels (RRP = 5, but the same value has a different meaning from that of the 1996 Panel variable). If the daughter is the reference person, her son is listed as a child of the household reference person (RRP = 4) and her mother is listed as the parent of the reference person in the 1996 Panel (ERRP = 6), but as another relative of the household reference person in earlier panels (RRP = 5).[22] Users should note that the identity of the household reference person can change from one month to the next; thus, the household description could also change.

[22] Because it is impossible to anticipate all of the different living arrangements found in SIPP sample households, and in some cases more than one rule for identifying a reference person may apply, some interviewer discretion in identifying the reference person is inevitable. For that reason, the resulting choices can sometimes appear somewhat arbitrary to the analyst.

Table 11-14. ERRP (RRP) Coding for the Same Three-Generation Household When Two Different People Are Designated as the Reference Person in the Topical Module Files

Designated Reference Person	Relationship to the Household Reference Person [ERRP (RRP)]	Meaning of ERRP (RRP) Value
Mother as Household Reference Person		
Mother	1 (1)	Reference person (Reference person)
Daughter	4 (4)	Child of reference person (Child of reference person)
Daughter's son	5 (5)	Grandchild of reference person (Other relative of reference person)
Daughter as Household Reference Person		
Mother	6 (5)	Parent of reference person (Other relative of reference person)
Daughter	1 (1)	Reference person (Reference person)
Daughter's son	4 (4)	Child of reference person (Child of reference person)

Identifying a Person's Spouse, Parent, or Guardian

Four other variables on the topical module files from the 1996 Panel can be used to describe household and family composition. They are EPNSPOUS, EPNDAD or EPNMOM, and EPNGUARD. These variables identify the person number of the spouse, the father or mother (just one parent is identified in files from panels prior to 1996), and guardian of the person, respectively. On the topical module files from panels prior to 1996, only two variables are found: PNPT and PNSP, the person numbers of the person's parent and spouse, respectively. In each case, the relative is identified only if she or he is living at the same address as the person.

By building from these variables, the analyst can identify a variety of family configurations. For example, these variables can be used to identify households containing three generations. Table 11-15 displays one household containing a mother and her two children. One child, EPPPNUM = 0102 (PNUM = 102), has a son; the other child, EPPPNUM = 0104 (PNUM = 104), has a spouse.

More About Using the SIPP ID Variables: Identifying Movers

Most of the SIPP topical modules collect information that pertains to a single month—generally month four of the wave's core reference period in the 1996 Panel, and month five (the interview month) for prior panels. However, some topical modules collect information about longer reference periods, most commonly either the previous 4 months (the same period as the core questions but often not with monthly resolution), the year prior to the interview (e.g., some items in the child and adult well-being topical modules), or the prior calendar year (e.g., the annual income and retirement accounts topical module of the 1996 Panel). In instances such as these, it

Table 11-15. Identifying Households Containing Three Generations in the Topical Module Files

1996 Panel					
Household Member	**Person Number (EPPPNUM)**	**Recoded Relationship to Household Reference Person (ERRP)**	**Spouse (EPNSPOUS)**	**Parent (EPNMOM)**	**Notes**
Mother	0101	1	9999	9999	Mother
Daughter #1	0102	4	9999	0101	Child
Daughter #1's Son	0103	5	9999	0102	Grandchild
Daughter #2	0104	4	0105	0101	Child
Spouse of Daughter #2	0105	8	0104	9999	Spouse of child
Panels Prior to 1996					
Household Member	**Person Number (PNUM)**	**Recoded Relationship to Household Reference Person (RRP)**	**Spouse (PNSP)**	**Parent (PNPT)**	**Notes**
Mother	101	1	999	999	Mother
Daughter #1	102	4	999	101	Child
Daughter #1's Son	103	5	999	102	Grandchild
Daughter #2	104	4	105	101	Child
Spouse of Daughter #2	105	5	104	999	Spouse of child

Note: Value of 999 or 9999 means not applicable.

is sometimes useful to know something about household composition during the reference period of the topical module.[23] This section of the Users' Guide is primarily for users who need to know how to access that kind of information. This section may also be helpful to those who wish to gain a better understanding of the SIPP ID variables for other reasons.

When a person moves, the current address field, SHHADID (ADDID), changes. The SSUID (ID), EENTAID (ENTRY), and EPPPNUM (PNUM) values remain the same. The first part (two digits in the 1992 Panel and the 1996 Panel, one digit in all others) of SHHADID (ADDID) indicates the wave in which a household is first interviewed at that new address. The remaining digit sequentially numbers the households that split into two or more households, as a result of a move to a different location by original sample members. Thus, new addresses in Wave 2 are numbered 021 (21), 022 (22), and so on. New addresses in Wave 3 are numbered 031 (31), 032 (32), and so on.

[23] For example, a person who joined the SIPP sample in Wave 4 of the 1996 Panel could not have contributed to the household income (at least not as a household member) of the prior calendar year.

Table 11-16 shows that persons 0101 (101) and 0102 (102) in the first household are original sample members. Person 0401 (401) moved into the home of persons 0101 (101) and 0102 (102) in Wave 4. In Wave 7, all three of them moved to a new location and were joined by person 0701 (701). In the second household, person 101 is an original sample member who moved to a new location in Wave 3. In the third household, person 0102 (102) is also an original sample member who used to live with persons 0101 (101) and 0103 (103) of the same sample unit ID, but moved to a new location in Wave 3 [to a different location from person 0101 (101)]. In the fourth household, person number 0103 (103) is an original sample member who used to live with persons 0101 (101) and 0102 (102) of the same sample unit ID number. All but two people moved from their original location [i.e., only two people have SHHADID (ADDID) equal to EENTAID (ENTRY)].

Table 11-16. Identifying Movers in the Core Wave Files

1996 Panel				
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	071	011	0101	Persons 0101 and 0102 are the original sample members. Person 0401 begins to live with them in Wave 4. All three people move in Wave 7 and person 0701 joins them.
123456789123	071	011	0102	
123456789123	071	011	0401	
123456789123	071	071	0701	
321456789123	031	011	0101	Person 0101 is an original sample member who moved in Wave 3.
321456789123	032	011	0102	Person 0102 is an original sample member who moved in Wave 3 to a different location from person 0101.
Panels Prior to 1996				
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**	**Notes**
123456789	71	11	101	Persons 101 and 102 are the original sample members. Person 401 begins to live with them in Wave 4. All three people move in Wave 7 and person 701 joins them.
123456789	71	11	102	
123456789	71	11	401	
123456789	71	71	701	
321456789	31	11	101	Person 101 is an original sample member who moved in Wave 3.
321456789	32	11	102	Person 102 is an original sample member who moved in Wave 3 to a different location from person 101.

The next example (Table 11-17) further illustrates how the ID system works as people move to new addresses, additional people move in with them, and households split. (Users may also find it helpful to review Figure 2-1 [pp. 2-10–2-14], which illustrates changes in household composition.)

- In Wave 1, there is a five-person household consisting of a husband, a wife, a daughter, a son, and a cousin. Since this is the first wave, the current address number is 011 (11), indicating address 1 of Wave 1, and the entry address number for each member of the household is the same as the current address number. Since they are assigned in Wave 1, the person numbers are in the 0100 (100) series and numbered sequentially, beginning with 0101 (101).
- During Wave 2, the son joins the Army, moves into the military barracks, and therefore leaves the SIPP sample. For the son's record, person number 0104 (104), the person-month file will contain a Wave 1 record for him and a Wave 2 record containing information (either imputed or provided by proxy) on his characteristics in the months of Wave 2 that he was still in the sample. If he does not return to the sample during the remainder of the panel, there will be no records for him beyond Wave 2.
- During Wave 3, the daughter marries and her husband moves into the household. The current address number where the mother, father, cousin, daughter, and son-in-law live remains the same since it is the same address. The son-in-law's entry address number is 011 (11), since he first enters the SIPP sample at an address coded 011 (11). The person number for the son-in-law is in the 0300 (300) series [0301 (301)] since he joins the SIPP sample in Wave 3.
- During Wave 4, the daughter and son-in-law move into a new house. Their current address number changes to 041 (41) to indicate that a new address has been established in Wave 4. Meanwhile, the cousin, who is over age 15, moves in with an uncle.[24] The cousin's current address number changes to 042 (42) (i.e., the second new household formed in the fourth wave from this sample unit). The assignment of address number 041 (41) to the daughter and 042 (42) to the cousin is arbitrary—it could be the other way around. The uncle enters the SIPP sample and receives an address number of 042 (42) and an entry address number of 042 (42). The uncle's person number is in the 0400 (400) series [0401 (401)] because he joins the survey in Wave 4.
- No changes in household composition are observed during Waves 5 through 9.
- During Wave 10,[25] the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. The newborn's entry address is 041 (41), since that is the current address ID of the daughter and son-in-law at the time of birth. The newborn's person number is 1001, reflecting the fact that the newborn came into the SIPP sample in Wave 10. Meanwhile, the cousin moves to Europe and therefore leaves the SIPP sample. The uncle, even though he did not move to Europe with the cousin, also leaves the SIPP sample because he no longer resides with an original SIPP sample member. Their records are no longer listed.

[24] In the 1993 Panel, all original sample members were followed, regardless of age. In all other panels (including the 1996 Panel), only those aged 15 or older were followed when they moved to new addresses.

[25] Prior to the 1996 Panel, only the 1992 Panel had more than nine waves.

Table 11-17. Example of Household Changes and Their Effects on the ID Variables in the Core Wave Files

1996 Panel				
Household Member	**Sample Unit ID (SSUID)**	**Current Address ID (SHHADID)**	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**
Wave 1				
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
Daughter	101111103123	011	011	0103
Son	101111103123	011	011	0104
Cousin	101111103123	011	011	0105
Wave 2				
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
Daughter	101111103123	011	011	0103
Son	101111103123	011	011	0104
Cousin	101111103123	011	011	0105
Wave 3				
Father	101111103123	011	011	0101
Mother	101111101233	011	011	0102
Daughter	101111103123	011	011	0103
Son-in-Law	101111103123	011	011	0301
Cousin	101111103123	011	011	0105
Wave 4	**Parent's Household**			
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
	Daughter's Household			
Daughter	101111103123	041	011	0103
Son-in-Law	101111103123	041	011	0301
	Cousin's Household			
Cousin	101111103123	042	011	0105
Uncle	101111103123	042	042	0401
Wave 10	**Parent's Household**			
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
	Daughter's Household			
Daughter	101111103123	101	011	0103
Son-in-Law	101111103123	101	011	0301
Newborn	101111103123	101	041	1001

(table continues)

Table 11-17. Example of Household Changes and Their Effects on the ID Variables in the Core Wave Files (*continued*)

Prior to 1996 Panel				
Household Member	**Sample Unit ID (ID)**	**Current Address ID (ADDID)**	**Entry Address ID (ENTRY)**	**Person Number (PNUM)**
Wave 1				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 2				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 3				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son-in-Law	101111103	11	11	301
Cousin	101111103	11	11	105
Wave 4	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
	Cousin's			
Cousin	101111103	42	11	105
Uncle	101111103	42	42	401
Wave 10[a]	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Newborn	101111103	41	41	1001

[a] Prior to the 1996 Panel, only the 1992 Panel had 10 or more waves. Wave 2 of the 1992 Panel of the core wave files has expanded address and person ID fields (3 and 4 digits, respectively) to accommodate Wave 10 of the 1992 panel.

Prior to the 1996 Panel, there were two extremely rare occasions when the original ID, ENTRY, and PNUM values were modified by the Census Bureau:

1. The first occasion was when two separate sampling units, each containing original sample members, were merged, perhaps because of a marriage. In this situation, one of the original sets of ID and ENTRY values was retained and the other set was changed to agree with that retained set. The person-number values (PNUM) of the changed set were modified further to be between 180 and 199, inclusive.
2. The second occasion was when a household split into two new households (in which each new household gained a new sample person) and later the households recombined. For example, suppose that a married couple separated in Wave 3, each moving in with a sibling. Both siblings were assigned a person number of 301 because they entered the sample in Wave 3 at different addresses (thus, ADDID = 31 and 32). If the husband and wife reunited in Wave 6, and brought the siblings with them, one of the sibling's person numbers would have been changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699, inclusive).

Those two occasions were the only times when ID, ENTRY, and PNUM changed. When it did occur, the old ID variables were stored in the previous wave variables (PWSUID, PWENTRY, and PWPNUM), found only on the core wave files.[26]

When the merge occurred after the first month of a reference period, the members of the merged household (whose ID variables were modified) were assigned two sets of monthly records in the core wave file. The first set of records contained the original ID information and identified the person as having exited the sample at the time of the merge. The second set contained the new ID information and identified the person as having entered the sample at the time of the merge. When the merge occurred at the start of the reference period, only the second set of records was retained in the core wave files.

Because merged households were very rare prior to the 1996 Panel, information about them will no longer be carried on the topical module files from the 1996 Panel. When either of those two kinds of events occur in the 1996 Panel, one or more original sample members will appear to leave the sample when the merge takes place, and new people will appear to enter the sample when the merged household forms. There is no indication in the data files that the "new" sample members were previously members of the SIPP sample with different ID values.

Topcoding

To protect the confidentiality of SIPP respondents, the Census Bureau topcodes characteristics available on the topical module files that might allow a user to recognize the identity of a SIPP

[26] In the 1993 Panel, merged households are identified with the variables PWSUID, PWENTRY, and PWPNUM. Before the 1993 Panel, they were identified with the variables PREV-ID, SC0064, and SC0066.

respondent. The topcoding procedures used in the topical module files are similar to those used in the core wave files.[27] Generally, topcodes for continuous variables that apply to the total universe include at least ½ of 1 percent of all cases. For income variables that apply to subpopulations, topcodes include either 3 percent of the appropriate cases or ½ of 1 percent of all cases, whichever is the higher topcode. Any discrete information that is topcoded in the core wave files is topcoded in a consistent manner in the topical module files.

Characteristics that are frequently topcoded in SIPP topical module files include income and expense values, including those for a broad range of assets and liabilities. For example, the following groups of topical module variables appear in Wave 3 of the 1996 Panel: assets and liabilities, interest earnings, medical expenses, mortgage amounts, other financial assets, real estate, rental properties, stocks and mutual funds, value of business, and work-related expenses and child support paid. The documentation for the variables included in these groups indicates whether the values are topcoded and the value ranges for the variables.

Using Allocation (Imputation) Flags

As described in Chapter 4, the Census Bureau often imputes information when a person does not respond to the survey or to a particular question. A variable of interest may be imputed. In the topical module files prior to the 1996 Panel, there is an allocation (imputation) flag for almost all of the person-level variables. Beginning with the 1996 Panel, there is an allocation (imputation) flag associated with every variable subject to imputation. For example, AEDUCATE is the allocation (imputation) variable that identifies whether EEDUCATE is imputed.

Variables are imputed and the allocation (imputation) flags are set before composite variables are created. For example, if income is imputed for one member of a household, that person's allocation (imputation) flag is set. However, total household income is computed *after* that imputation; if any household member had any income imputed, total household income is based, in part, on imputed information. There is no direct indication on the records of other household members that any information has been imputed.

Using Weights

The topical module files contain one weight variable—WPFINWGT (FINALWGT). For the 1996 Panel, this weight is the person cross-sectional weight for the fourth reference month. Prior to 1996, this weight was the person interview month weight for people who provided data for a topical module. It shows the number of people in the population represented by the sample person in the interview month.

[27] Chapter 10 contains a discussion of both the new income topcoding procedures used in the 1996 Panel core wave files and the income topcoding procedures used in the pre-1996 core wave files. See also Appendix B: SIPP Topcoding Specifications.

The source and accuracy statements that accompany all SIPP topical module files ordered from the Census Bureau provide suggestions on how to use the topical module weight variable. Also, Chapter 8 of this Guide contains a full discussion of how to use weights in SIPP data files.

Identifying States

For the 1996 Panel, the variable TFIPSST identifies 45 states and the District of Columbia. The remaining five states are combined as follows:

1. Maine, Vermont; and
2. North Dakota, South Dakota, Wyoming.

The topical module files from panels prior to the 1996 Panel contain a variable STATE that identifies the state in which the household resides. The variable identifies 41 individual states and the District of Columbia; the nine other states are combined into three groups:

1. Maine, Vermont;
2. Iowa, North Dakota, South Dakota; and
3. Alaska, Idaho, Montana, Wyoming.

Even though it is possible to identify most states, SIPP was not designed to be representative at the state level and should not be used to produce state-level estimates. The state variable is included on the public use files to allow examination of how state-level characteristics affect national estimates. For example, a user could apply the state-specific eligibility criteria for a means-tested program in order to arrive at a national estimate of the number of eligible participants. Because some states are not uniquely identified, some method of allocating the state-specific eligibility rules to sample people in those states would need to be devised.

Identifying Metropolitan Areas

The topical module files do not contain any variables identifying metropolitan areas. Those needing that information should merge it from the core wave files or the full panel files. Analysts should see Chapters 10 and 12 for discussions of the core wave files and the full panel files, respectively. Chapter 13 discusses how to merge multiple SIPP public use files.

12. Using the 1990–1993 Full Panel Longitudinal Research Files

This chapter discusses procedures for working with data from the full panel longitudinal research files for the 1990 through 1993 Panels of the Survey of Income and Program Participation (SIPP). Because the full panel longitudinal research file for the 1996 Panel was still under development at the time this chapter was written, it is not yet possible to describe procedures for using that file. A revised version of this chapter will be available once the longitudinal research file for the 1996 Panel is released to the public.

The chapter begins by describing the documentation that accompanies the full panel public use files obtained from the Census Bureau. The discussion then turns to the data files themselves. The data file structure is described, and detailed explanations are provided about how to use the longitudinal research files when performing common tasks, including:

- Realigning the data by calendar month;
- Using the monthly interview status variables;
- Identifying persons, households, families, and program units;
- Working with the unearned income data;
- Understanding the effects of topcoding;
- Using imputation flags; and
- Identifying states and metropolitan areas.

Before reading this chapter, users should read Chapter 9 for an introduction to Section II. Analysts using only one longitudinal research file should also read about the use of sample weights (Chapter 8) and the computation of standard errors (Chapter 7). Those planning on merging data from a longitudinal research file to data from the core wave or topical module files should read Chapter 10 for information about the core wave files, Chapter 11 for information about the topical module files, and Chapter 13 for information about linking SIPP public use files.

This chapter focuses on the longitudinal research files. It is written so that it can be used independently of the chapters describing the core wave files and topical module files. Although there are many similarities across the three types of files, important differences do exist. Because those differences are sometimes subtle, users familiar with the core wave and topical module files should read this chapter carefully, paying close attention to information about variable

names and file structures. Table 9-2 summarizes the differences between the core wave, topical module, and longitudinal research files.[1]

Using the Technical Documentation of the 1990–1993 Longitudinal Research Files

Each data file received from the Census Bureau comes with a set of technical documentation and a data dictionary. The technical documentation includes:

- The paper survey instrument;
- A glossary of selected terms;
- A cross-walk, mapping reference months into calendar months for each rotation group;
- A source and accuracy statement describing the sample weights and the computation of standard errors; and
- User Notes.

The survey instrument is vital to understanding what questions were asked, how they were asked, the order in which they were asked, to whom they were asked, and the way in which the answers were recorded. Some questions employ skip patterns (Chapter 3), so users should pay particular attention to which questions were skipped for which respondents. These skip patterns are best understood by consulting the survey instruments.[2]

The source and accuracy statements provide information about the weights on the files, when and how to make adjustments to the weights, and one approach to computing standard errors for some common types of estimates. More detailed discussions of those topics are provided in Chapters 7 and 8 of this Guide.

The data dictionary provides a detailed description of each variable on the file. It describes four aspects of each variable:

1. The definition;
2. The sample universe of the corresponding survey question;

[1] Some of this information will change once the 1996 longitudinal research file becomes available. At that time, this guide will be updated to reflect the differences.

[2] With the introduction of CAI (computer-assisted interviewing) in the 1996 Panel, questionnaire documentation is now available at the SIPP Web site at http://www.sipp.census.gov/sipp/.

3. The ranges for all legal values; and
4. The location (and size) in the file.

A machine-readable version of the data dictionary accompanies each data file. It can also be downloaded from the Internet (http://www.sipp.census.gov/sipp/).

The data dictionary is formatted to facilitate processing by user-written computer programs.[3] As shown in Figure 12-1, a "D" in the first column signifies that the next few lines define the variable: (1) the variable name, (2) the total number of columns occupied by the variable, (3) the starting position, (4) the number of occurrences of that variable, and (5) the size of each occurrence of the variable.[4] A "U" in the first column indicates that the next words describe the universe.[5] A "V" in the first column indicates that the next number and phrase describe one of the values of the variable. An asterisk in the first column denotes a comment. A period (.) before a word denotes the start of the value label.[6]

The format of the data dictionary for the longitudinal research files is different from that used for the core wave and topical module files. The full panel data dictionary includes two extra fields on the line with a "D" in the first column. The first extra field contains the number of occurrences of the variable, and the second extra field contains the number of digits for each occurrence of the variable. These fields are needed because some variables in the longitudinal research file occur *x* times, depending on the number of waves, or *y* times, depending on the number of months in the panel.

HH-ADDID in Figure 12-1 is a monthly variable containing two digits (monthly because it occurs 36 times). PP-MIS is also a monthly variable, but its length is one digit. PP-INTVW appears once per wave (because it occurs nine times), and PP-ENTRY, PP-PNUM, SU-TOTPP, and PP-RCSEQ occur once for the entire panel.

Figure 12-2 shows sample SAS and FORTRAN syntax for reading the data described by the codebook fragment in Figure 12-1. Additional SAS program code could be used to associate variable labels and value labels (SAS "formats") with the PP-MIS and PP-INTVW variables.

---

[3] The data dictionaries for the longitudinal research files use a different format from that used for the core wave and topical module files. Users who have worked with the core wave and topical module files should take care to note those differences. In addition, the formats of the data dictionaries for the 1996 Panel core wave and topical module files, as well as the variable names used in those files, have changed in the 1996 Panel. This chapter uses variable names from the 1990–1993 SIPP Panels. When longitudinal research files are released from the 1996 Panel, a revised version of this chapter will be available with updated information. Users will be able to download that version from the SIPP Web site at http://www.sipp.census.gov/sipp/.

[4] The data dictionary for the 1992 longitudinal research file used a different format from that used in the other pre-1996 longitudinal research files. In the 1992 data dictionary, the first line for each new variable, labeled with a "D" in column 1, has the following fields: variable name, total size (number of characters), start location, the length of a single occurrence of the variable, the number of occurrences of the variable, and the number of implied decimals.

[5] The universe definitions included in the data dictionaries prior to the 1996 Panel were often inaccurate. Users of pre-1996 SIPP Panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

[6] The data dictionary for the 1992 longitudinal research file also has a line labeled with an "R" in column 1. This line provides the range of values for the variable.

Figure 12-1. Excerpt from the 1993 Longitudinal Research File Data Dictionary

```
D PP-ENTRY     2     17     1     2
    Range = (11:99)
    Edited entry address ID
    Address ID of the household that this person belonged to at the time this
      person first became part of the sample

D PP-PNUM     3     19     1     3
    Range = (101:999)
    Edited person number

D SU-TOTPP     2     22     1     2
    Range = (1:60)
    Total number of person records for this sample unit

D PP-RCSEQ     2     24     1     2
    Range = (1:60)
    Sequence number of person record within sample unit

D HH-ADDID     72     26     36     2
    Range = (0:99)
    Address ID. —— This field identifies the household this person lived in
      this month

D PP-INTVW     9     98     9    1
    Range = (0:4)
    Person's interview status for the relevant interview
V          0 .Not applicable (children under .15), not in sample, nonmatch
V          1 .Interview (self)
V          2 .Interview (proxy)
V          3 .Noninterview - Type Z refusal
V          4 .Noninterview - Type Z other

D PP-MIS     36     107     36     1
    Range = (0:2)
    Person's interview status for this month
V          0 .Not matched or not in sample
V          1 .Interview
V          2 .Non-interview
```

Figure 12-2. Corresponding SAS and FORTRAN Syntax to Read in Data from the 1993 Longitudinal Research File Data Dictionary

SAS

```
Input
   @17   PP_ENTRY   2.
         PP_PNUM    3.
         SU_TOTPP   2.
         PP_RCSEQ   2.
         (ADDID1-ADDID36)   (2.)
         (INTVW1-INTVW9)    (1.)
         (PP_MIS1-PP_MIS36) (1.)
         ;
```

FORTRAN

```
       INTEGER*2 PP_ENTRY
       INTEGER*2 PP_PNUM
       INTEGER*1 SU_TOTPP
       INTEGER*1 PP_RCSEQ
       INTEGER*1 HH_ADDID(36)
       INTEGER*1 PP_INTVW(9)
       INTEGER*1 PP_MIS(36)

       READ(infile,1000) PP_ENTRY, PP_NUM, SU_TOTPP,
      $         PP_RCSEQ, HH_ADDID, PP_INTVW, PP_MIS
1000   FORMAT(T17, I2, I3, I2, I2, 36I2, 9I1, 36I1)
```

Relationship of the Longitudinal Research Data Files to the SIPP Survey Instrument

The data dictionaries for the longitudinal research files do not replicate the survey instruments. Analysts should keep a few things in mind when using the data:

- The variables on the longitudinal research files do not correspond one-to-one with the questionnaire items. The variables are listed in a different order, some are not included in the longitudinal research file at all, and some are created from a combination of other variables.
- The range of possible values of the variables does not always correspond one-to-one with the response categories shown on the survey instrument or in the data dictionary;
- The variable name may not readily indicate its meaning; and

- The complexity of the skip patterns may not be apparent just by looking at the data dictionary.[7]

To avoid potential problems and confusion, users should become familiar with the survey instrument before using the data. When working with the data, analysts should refer to both the survey instrument and the data dictionary.

Structure of the Longitudinal Research Files

The longitudinal research files contain one record for each person who was ever in the SIPP sample for that panel. Even if the person was in the sample for just 1 month, there will be a record for that person. There are records for children as well as for adults, and there are records for people who entered the sample after the first wave.

Within each record, the variables correspond to the information that was collected in the core interviews. While most of the core items are included in the longitudinal research files, some items are not, and not all of the constructed variables found on the core wave files are included on the longitudinal research files. In addition, no items from any of the topical modules are included on the longitudinal research files. When items from the core wave or topical module files are needed, those variables must be merged with data from the longitudinal research files. Chapter 13 provides a detailed discussion of merging SIPP files.

The longitudinal research file structure differs from that of the core wave files. The longitudinal research files contain just one record per person, while the core wave files contain one record per person per month. Because some attributes do not change over the course of the panel, those variables appear once on each record (e.g., rotation group, sample unit ID, person number, sex, race, and ethnic origin). Some questions were asked once during each wave, so they appear x times on each record, where x equals the number of waves for that panel (e.g., highest grade attended, and participation in school breakfast and lunch programs). Most of the core questions were asked for each month of the panel. They appear y times on each record, where y equals the number of months for that panel (e.g., current address ID, monthly interview status, relationship to the reference person, income, and program participation).

Table 12-1 shows that the 1992 Panel has 10 waves (or 40 months) of data. The 1993 Panel has nine waves (or 36 months) of data. Thus, the interview status variable (PP-MIS) appears 40 times in the 1992 longitudinal research file, and it appears 36 times in the 1993 longitudinal research file.

[7] See footnote 5.

Table 12-1. Summary of Panels, Waves, Reference Months, and Sample Sizes

Panel Year	Reference Months	Number of Waves	Number of Months	Wave 1 Eligible Households
1984	Jun. 83 – Jun. 86	9	36	20,897
1985	Oct. 84 – Jul. 87	8	32	14,306
1986	Oct. 85 – Mar. 88	7	28	12,425
1987	Oct. 86 – Apr. 89	7	28	12,527
1988	Oct. 87 – Dec. 89	6	24	12,725
1989	Oct. 88 – Dec. 89	3	There is no longitudinal research file for the 1989 SIPP.	
1990	Oct. 89 – Aug. 92	8	32	23,627
1991	Oct. 90 – Aug. 93	8	32	15,626
1992	Oct. 91 – Mar. 95	10	40	21,577
1993	Oct. 92 – Dec. 95	9	36	21,823

Source: SIPP Quality Profile, 3rd Ed. (U.S. Census Bureau, 1998a).

Table 12-2 illustrates the longitudinal research file structure. In this example, there are five people. Sample unit ID (PP-ID), person number (PP-PNUM), and entry address ID (PP-ENTRY) appear once on each record because they are permanent characteristics of those people. Monthly interview status (PP-MIS), a monthly variable, appears 40 times because the 1992 Panel had 10 waves and each wave collected information about the 4 months prior to the interview month.

People who were not interviewed (in person or by proxy) for 1 or more months over the course of the panel either have their data imputed[8] or are identified as not in the sample (PP-MIS equal to either 0 or 2) for the months when they were not in the sample. The discussion of the PP-MIS variable later in this chapter provides additional information.

How to Align Data by Calendar Month

It is frequently useful to realign the SIPP data by calendar month instead of reference month. For example, researchers often want to analyze data for a specific calendar year (January through December) or federal fiscal year (October through September).[9] To do this, the analyst must

[8] Imputation would be by Type Z and missing-wave imputations. Chapter 4 discusses imputation methods.

[9] The longitudinal research files do not contain calendar month weights. Those weights would be needed for some types of longitudinal analyses, such as analyses of the dynamics of program participation, where the unit of analysis is a spell of program participation (Chapter 8 provides a discussion of this example). Data from the longitudinal research files can also be used for cross-sectional estimation, and they are often preferable to the data from the core wave files because the edit and imputation procedures used for the longitudinal research files are believed to result in less imputation error than the procedures used for the core wave files. The format of the file is sometimes easier to work with, even for cross-sectional applications. In those instances, the calendar month weights must be merged from the core wave files. Chapter 8 provides a detailed discussion of weighting procedures in the SIPP. Chapter 13 provides a detailed discussion of linking SIPP files.

Table 12-2. Example of the Longitudinal Research File Structure

PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	PP-MIS																			
				Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
				Month				Month				Month				Month				Month			
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
123912879	11	201	3	0	0	0	0	0	1	1	1	1	1	1	1	2	2	1	1	1	1	1	0
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	1	1	1	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
788723892	11	102	4	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	0	0	0	0
788723892	11	301	4	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
890987123	11	101	1	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	2

PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	PP-MIS																			
				Wave 6				Wave 7				Wave 8				Wave 9				Wave 10			
				Month				Month				Month				Month				Month			
				21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	2	1	1	1	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0
123912879	11	201	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	1	1	1	1	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	102	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	301	4	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
890987123	11	101	1	2	2	1	1	1	1	1	1	1	1	2	2	2	1	1	1	0	0	0	0

know the reference period for each rotation group of the panel. That information is included with the technical documentation that accompanies the longitudinal research files.

Table 12-3 shows the reference period for each rotation group of the 1992 Panel. It shows that the reference period for rotation group 2 is October 1991–January 1995. The reference period for rotation group 3 is November 1991–February 1995. The reference period for rotation group 4 is December 1991–March 1995. The reference period for rotation group 1 is January 1992–December 1994 (interviews were not conducted in Wave 10 for this rotation group).

Table 12-3. Reference Periods for Each Rotation Group of the 1992 Panel

Rotation Group (ROT)	Reference Period
2	October 1991–January 1995
3	November 1991–February 1995
4	December 1991–March 1995
1	January 1992–December 1994

The following algorithm (Figure 12-3), written for the 1992 Panel, illustrates one approach to realigning the SIPP reference months to common calendar months. The mapping depends on the panel and rotation group and must be applied to each person. The first step establishes the displacement or realignment of the months. The second step initializes each monthly variable to –9 to distinguish the calendar months in which the variable is not relevant.[10] The loop goes from 1 to 42 because in the 1992 Panel the first reference month was October 1991 and the last reference month was March 1995, which means that there were 42 calendar months covered by the panel. The third part of the algorithm realigns the input data to be based on the calendar month. Table 12-4 displays the data after the realignment.

Using the Monthly Interview Status (PP-MIS) Variables

The monthly interview status variable helps to determine whether the data for a person in a given month should be used. In the longitudinal research files, this variable is labeled PP-MIS, and it has one occurrence for each reference month of the SIPP panel. Some people refer to it as the *in-sample* variable to distinguish it from the interview status variable (PP-INTVW). The PP-MIS variables have three possible values: 0, 1, and 2.

[10] If –9 is a possible value for the variables being realigned (e.g., self-employed income can be negative), a different starting value must be used.

Figure 12-3. Algorithm for Realigning SIPP Panel Month to Calendar Months in the 1992 Panel

```
/*
Create a variable that identifies the number of months each
rotation group differs from the baseline
*/
If ROT = 2
   DISPLACEMENT = 0
Else if ROT = 3
   DISPLACEMENT = 1
Else if ROT = 4
   DISPLACEMENT = 2
Else if ROT = 1
    DISPLACEMENT = 3
End if

/*
Initialize the new, re-aligned variable. This is not needed in SAS.
When this step is used, an initial value should be chosen that
is not a legal value for the variable in the actual data.
*/
For each calendar month (for CALMM = 1 to 42):
   NEW-PP-MIS(CALMM) = -9
End loop

/*
Create the newly re-aligned variable
*/
For each reference month (for MONTH = 1 to 40):
   CALMM = MONTH + DISPLACEMENT
   NEW-PP-MIS(CALMM) = PP-MIS(MONTH)
End loop
```

The monthly interview status is the only reliable guide to whether the data for a given person should be used in a given month. Analysts should use only data for those months in which a person's interview status (PP-MIS) is equal to 1.[11]

Any data present for months in which a person's interview status is coded either 0 or 2 should be ignored. A code of 0 indicates that the person was not in the sample that month, and a code of 2 indicates a noninterview for that month.[12]

[11] As a safeguard against inadvertently using data for months when PP-MIS is not equal to 1, all monthly variables in the user's data extract should be set to a missing value for months when PP-MIS is not equal to 1. Most statistical packages allow certain values to be flagged as "missing." Once flagged, those values are excluded from computations.

[12] Beginning with the 1991 Panel, new "missing wave" imputation procedures were instituted for the longitudinal research files. Whenever data for a wave are imputed (the WAVFLG variable), PP-MIS is recoded to 1 on the longitudinal research files, indicating that the data for those months should be used. In some cases, these people will have records in the core wave files that were created during the Type Z imputation processing (see Chapter 4 for details). In some of these instances, however, the longitudinal research file will have data for people who are not present on the associated core wave data files.

Table 12-4. Monthly Data from the 1992 Panel, Realigned by Calendar Month

				NEW-PP-MIS														
				1991			1992											
PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
987913389	11	101	3	-9	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	-9	1	1	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	201	3	-9	0	0	0	0	0	1	1	1	1	1	1	1	2	2
874943283	11	101	4	-9	-9	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	-9	-9	1	1	1	0	0	1	1	1	1	1	1	1	0
788723892	11	102	4	-9	-9	1	1	1	1	1	1	1	1	1	1	1	1	2
788723892	11	301	4	-9	-9	0	0	0	0	1	1	1	1	1	1	1	1	1
788723892	11	1001	4	-9	-9	0	0	0	0	0	0	0	0	0	0	0	0	0
763483873	11	101	1	-9	-9	-9	1	1	1	1	1	1	1	1	1	1	1	1
890987123	11	101	1	-9	-9	-9	1	1	1	1	1	1	1	1	1	2	2	2

				NEW-PP-MIS											
				1993											
PP-ID	PP-ENTRY	PP-PNUM	PP-ROT	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1
112987122	11	101	2	0	0	0	0	0	0	0	0	0	0	0	0
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1
123912879	11	101	3	1	1	1	1	1	2	2	1	1	1	0	0
123912879	11	201	3	1	1	1	1	1	0	0	0	0	0	0	0
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	0	1	1	1	1	1	1	1	1	1	1	2
788723892	11	102	4	2	2	2	0	0	0	0	0	0	0	0	0
788723892	11	301	4	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1
890987123	11	101	1	1	1	1	1	1	1	1	2	2	2	1	1

(table continues)

Table 12-4. Monthly Data from the 1992 Panel, Realigned by Calendar Month (*continued*)

				NEW-PP-MIS														
				1994												1995		
PP-ID	**PP-ENTRY**	*PP-PNUM*	**PP-ROT**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**
112612345	11	101	2	1	1	1	1	1	1	1	1	1	1	1	1	1	–9	–9
112987122	11	101	2	0	0	0	0	0	0	0	0	0	0	0	0	0	–9	–9
987913389	11	101	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	–9
123912879	11	101	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	–9
123912879	11	201	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	–9
874943283	11	101	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
788723892	11	101	4	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	102	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	301	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
788723892	11	1001	4	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
763483873	11	101	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
890987123	11	101	1	1	1	1	1	1	1	2	2	2	1	1	1	0	0	0

The presence of data in analysis fields for any given month is not a reliable guide to whether the person should be included in the planned analyses. Data are collected for all months of the reference period for a given wave, even if the interviewed person was in the sample for only part of the reference period. Data are also present even if the person was not interviewed. Information from the questionnaire is imputed when the person was in sample for at least 1 month of the reference period but not actually interviewed. That includes people who moved out of scope (as defined in Chapter 2), people who died, and people who refused to be interviewed. The entire questionnaire was imputed for Type Z noninterviews (people who refused to be interviewed, living in households where other members were successfully interviewed). Chapter 4 examines imputation procedures; Chapter 8 provides information on weighting. Data are collected for all months of the reference period even if the interviewed person was in the sample for only part of the reference period.

The presence of a positive weight is also not a reliable guide to whether a person should be included in the planned analysis. Although people with zero weights will not enter into any weighted tabulations, they may provide important contextual information about people who do enter into those (weighted) tabulations. For example, a zero-weight person who is a member of the same household as a positive-weight person for only 3 months provides information about the positive-weighted person's household (including, for example, household size, composition, income, and program participation) for that 3-month period. That is why records for these zero-weighted people are retained in the SIPP full panel data files.[13]

Identifying Persons

There are many occasions when a user may need to identify which records belong to each individual in the SIPP data files. That need arises, for example, during the following procedures:

- Merging data from topical module or full panel files to core wave files;
- Combining data from two or more core wave files;
- Linking husbands and wives;
- Linking parents and children; and
- Identifying which person received government transfer income on behalf of the family.

To uniquely identify a person in the longitudinal research files, analysts should use the three variables shown in Table 12-5.[14]

[13] Using the PP-MIS variable shown in Table 12-2, one can see that the first person within each rotation group was in sample every month of the panel. The second person shown in the table left the sample before the third interview (information was probably collected by proxy interview for that wave) and did not return to the sample. The eighth person left the sample in month 13. The tenth person entered the sample in month 38 (the last wave).

[14] Beginning with the 1996 Panel, the entry address ID will no longer be needed: person numbers will be unique within sample units. Continued use of the entry address ID will not create any problems. It is simply redundant information.

Table 12-5. Variables Used to Uniquely Identify a Person in the Longitudinal Research Files

Variable Name	Description
PP-ID	Sample unit ID
PP-ENTRY	Entry address ID
PP-PNUM	Person number

- PP-ID uniquely identifies each initially sampled dwelling unit.[15] Every person in the longitudinal research file was either a member of one of those units (an original sample member) or lived with someone during the life of the panel who was a member of an initially sampled dwelling unit. A person's connection to that unit is an attribute of that person and does not change over time.[16] This means that as people move from address to address, their PP-ID stays the same. As new people join the homes of original sample members, they receive the PP-ID of the original sample members.
- PP-ENTRY identifies the address where the person lived at the time he or she was first interviewed. It does not change even if the person moves.[17] It is used in conjunction with the person number and the sample unit ID to uniquely identify persons within the sampling unit. Values for this variable are unique only within sample units. The entry address ID has two components. The first part of the ID number (two digits in the 1992 Panel, and one digit in all others) identifies the wave in which SIPP interviews were first conducted at the address. The second part of the number (one digit in all panels) sequentially numbers addresses within a sample unit (PP-ID) that enter the sample in the same wave.
- PP-PNUM uniquely identifies a person within the sample unit ID and entry address ID. PP-PNUM does not change even if the person moves.[18] The first part of PP-PNUM (two digits in the 1992 Panel, and one digit in all others) indicates the wave in which the person was first interviewed.[19] The remaining two digits are sequentially assigned within the household. Thus, original sample members are assigned person numbers ranging from 100 to 199. Individuals who enter the SIPP sample in Wave 2 are assigned a person number ranging from 200 to 299. Those who enter in Wave 10 are assigned person numbers ranging from 1001 to 1099.

Table 12-6 illustrates how the combination of PP-ID, PP-ENTRY, and PP-PNUM uniquely identifies people and provides information about when they first entered the SIPP sample. In this example, there are eight individuals: five are original sample members; one person joined the

[15] The PP-ID is a random recode of three other variables in the Census Bureau's internal (not public use) files: the respondent's sampling area (PSU), the cluster of housing units within that area (called the "segment"), and a sequentially assigned serial number. Those three variables are omitted from the public use files to protect the confidentiality of the respondents.

[16] There is one rare exception to this rule, which is described in the section entitled "Identifying Movers" later in this chapter.

[17] See footnote 16.

[18] See footnote 16.

[19] For Wave 10 of the 1992 Panel and for the 1996 Panel, the first two digits of PNUM instead of the first digit identify the wave in which the person entered the sample.

SIPP sample in Wave 4, one person joined in Wave 7, and one person joined in Wave 10 (of the 1992 Panel).

Table 12-6. How to Uniquely Identify a Person in the Longitudinal Research Files

Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Notes
123456789	11	101	Original sample member
123456789	11	102	Original sample member
123456789	11	401	Enters SIPP sample in Wave 4
123456789	71	701	Enters SIPP sample in Wave 7
321456789	11	101	Original sample member
321456789	11	102	Original sample member
321456789	11	103	Original sample member
456789123	101	1001	Enters SIPP sample in Wave 10 of the 1992 Panel

Identifying Households

The term *household*, as used in Census Bureau publications, refers to a group of people who occupy a housing unit. A house, an apartment or other group of rooms, or a single room is regarded as a housing unit if it is occupied or intended for occupancy as separate living quarters. That is, the occupants do not live and eat with any other people in the structure and there is direct access from the outside or through a common hall. A group of friends sharing an apartment constitutes a household. Rooming and boarding houses, college dormitories, convents, and monasteries are classified as *group quarters* rather than households.

To uniquely identify a household or group quarters in the longitudinal research files in a given month, analysts should use the variables shown in Table 12-7.[20]

Table 12-7. Variables Used to Uniquely Identify a Household in the Longitudinal Research Files

Variable Name	Description
PP-ID	Sample unit ID
HH-ADDID$_i$	Current address ID in the *i*th month
PP-MIS$_i$	Person's interview status in the *i*th month

[20] Since household composition changes from one month to the next, it is generally not possible to construct "longitudinal households." Users should not infer commonality across months based solely on place of residence in one month. The characteristics of the household to which a given person belongs (such as household size and household income) should be evaluated separately for each month, based on just those people who reside together in each specific month. Similar caution should be exercised when dealing with the characteristics of the family and, when applicable, the subfamily to which a person belongs.

People with the same PP-ID and HH-ADDID$_i$ values and with a PP-MIS value of 1 live in the same household (or group quarters) in the *i*th month of the reference period. The eight individuals shown in Table 12-8 make up four households. The first household contains the first four individuals. The second household contains one person. The third household contains one person. The fourth household contains two people.

This example depicts the households in the *i*th month. These people could belong to different households in other months. (Users may find it helpful when reading the following pages to refer to Figure 2-1 [pp. 2-10–2-14], which illustrates changes in household composition.)

Table 12-8. How to Uniquely Identify a Household or Group Quarters in a Given Month of the Longitudinal Research Files

Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PNUM)	Person's Interview Status (PP-MIS)	Current Address ID (HH-ADDID$_i$)	Notes
123456789	11	101	1	71	Four people in this household
123456789	11	102	1	71	
123456789	11	401	1	71	
123456789	71	701	1	71	
321456789	11	101	1	31	One person in this household
321456789	11	102	1	32	One person in this household
321456789	11	103	1	101	Two people in this household [a]
321456789	101	1001	1	101	

[a] Because this example includes a person with an entry address of 101, we know that the example refers to a month from Wave 10 of the 1992 Panel (the only panel prior to 1996 with 10 or more waves).

Identifying Families

The term *family*, as used in Census Bureau publications, refers to a group of two or more people related by birth, marriage, or adoption who reside together; all such individuals are considered members of one family.[21]

- A *primary family* is a family containing the household reference person and all of his or her relatives. This means that a household composed of a husband and wife, their son, and their son's wife (i.e., the daughter-in-law) is classified as a primary family containing four people.

[21] As with households (see footnote 20), because family composition changes from one month to the next, it generally is not possible to construct longitudinal families. Users should not infer commonality across months based solely on family membership in one month. The characteristics of the family to which a person belongs (such as family size and family income) should be evaluated separately for each month, and should be based on just those people who reside together and are members of the same family in each specific month. Similar caution should be exercised when dealing with the characteristics of the household and, when applicable, the subfamily (related or unrelated) to which a person belongs.

- A *related subfamily* is a nuclear family that is related to but does not include the household reference person. For example, the son and his wife (i.e., the daughter-in-law) in the preceding example are a related subfamily.
- An *unrelated subfamily* (sometimes called a secondary family) is a nuclear family that is not related to the *household* reference person. Thus, a husband and wife who live in a friend's house are classified as an unrelated subfamily. A mother and daughter who live in the mother's boyfriend's apartment are classified as an unrelated subfamily.
- A *primary individual* is a household reference person who lives alone or lives with only nonrelatives. Primary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.
- A *secondary individual* is not a household reference person and is not related to any other people in the household. Secondary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.

Unlike the core wave files, the longitudinal research files do not contain family identification variables (e.g., FID, FID2, and SID). Analysts needing family identification variables must either merge them from the core wave files (Chapters 10 and 13) or create them.[22] Because family composition can change over time, these are monthly variables. The algorithm in Figure 12-4 shows one approach to creating functional equivalents of the variables contained on the core wave files.[23]

The variables created by this algorithm are functionally equivalent to the variables with the same names on the core wave files: they will group people into the same family and subfamily groups. However, the actual values assigned by this algorithm to these variables generally will not equal the values found in the variables from the core wave files.

With these monthly variables (FID_i, $FID2_i$, and SID_i), users can identify common family membership in each month.[24] The Census Bureau has two principal methods for distinguishing families that are based on the variables and numbering schemes shown in Table 12-9. Analysts must remember to choose which type of family classification they want and then use the appropriate method.

- The first method defines a family as all persons who are related and living together. The family ID variable FID_i is used with this definition. FID_i groups the household reference person with all related household members by assigning them the same ID number.

[22] In most cases, it is also possible to merge these variables from the core wave files. However, beginning with the 1991 Panel, a missing wave imputation procedure was applied to the longitudinal research files: data were imputed for people with missing data for a wave but with valid data for the two adjacent waves. Although these people have data in the longitudinal research file for imputed waves, some have no data in the core wave files (some of these people are subject to Type Z imputation procedures that create records in the core wave files). For these people, merging the family ID variables from the core wave files is not an option.

[23] This algorithm uses the following (monthly) variables found on the longitudinal research files: FAMTYP and FAMNUM. These variables are discussed in greater detail in the next section.

[24] See footnotes 20 and 21.

Figure 12-4. Constructing Family and Subfamily ID Variables in the Longitudinal Research Files

```
For each person (index = ip):
      For each month (index = mo):
            If PP-MIS(mo, ip)= 1 then do:       <i.e., interview status>
                  If FAMTYP(mo, ip) = 0         <i.e., primary family>
                     then FID(mo, ip)  = 1
                        FID2(mo, ip) = 1
                        SID(mo, ip) = 0
                  Else if FAMTYP(mo, ip) = 1    <i.e., secondary individual>
                     then FID(mo, ip) = 10000 + ip
                        FID2(mo, ip) = 10000 + ip
                        SID(mo, ip) = 0
                  Else if FAMTYP(mo, ip) = 2    <i.e., unrelated subfamily>
                     then FID(mo, ip) = 100 + FAMNUM(mo, ip)
                        FID2(mo, ip) = 100 + FAMNUM(mo, ip)
                        SID(mo, ip) = 0
                  Else if FAMTYP(mo, ip) = 3    <i.e., related subfamily>
                     then FID(mo, ip) = 1
                        FID2(mo, ip) = 0
                        SID(mo, ip) = FAMNUM(mo, ip)
                  Else if FAMTYP(mo, ip) = 4     <i.e., primary individual>
                     then FID(mo, ip) = 10000 + ip
                        FID2(mo, ip) = 10000 + ip
                        SID(mo, ip) = 0
                  End if
            End "PP-MIS = 1" Block
      End month loop
End person loop
```

Table 12-9. Variables Used to Identify Families in the Longitudinal Research Files

Variable Name	Description
PP-ID	Sample unit ID
HH-ADDID$_i$	Address ID in the *i*th month
PP-MIS$_i$	Person's interview status in the *i*th month
And one of the following created variables:	
FID$_i$	Family ID in the *i*th month
FID2$_i$	Family ID in the *i*th month, excluding related subfamily members (FID2$_i$ equals zero for related subfamily members)
SID$_i$	Family ID in the *i*th month for related subfamily members (SID$_i$ assigns nonzero values only to members of related subfamilies)
FID2$_i$ and SID$_i$	Family ID in the *i*th month, separating related subfamilies from the primary family

Note: Variables FID$_i$, FID2$_i$, and SID$_i$ are not included on the longitudinal research files. They can be created by using the algorithm shown in Figure 12-4 or merged from the core wave files.

This family group corresponds to the Census Bureau's definition of a primary family. FID_i groups members of each unrelated subfamily (and primary and secondary individuals) separately.

- The second method is similar to the first in defining a family, but the family excludes related subfamilies. The family ID variable $FID2_i$ is used with this definition. $FID2_i$ equals zero for related subfamilies.

Analysts who want to analyze multigenerational families would use $FID2_i$ and the variable SID_i. SID_i treats related subfamilies as distinct family units by assigning them nonzero values. Analysts can easily distinguish unrelated subfamilies from other family units when they use these variables and numbering schemes.

Table 12-10 illustrates the difference between FID_i, $FID2_i$, and SID_i for a single month. In the month shown, the first household contains a primary family of five people. The primary family contains two related subfamilies. FID_i and $FID2_i$ mask the fact that there are two related subfamilies; only SID_i provides that information. SID_i has nonzero values only for members of related subfamilies. The second household contains a primary family and two unrelated subfamilies. The third household contains a primary individual and an unrelated subfamily. The fourth household contains only a primary individual. The fifth household is group quarters containing two people. This example depicts those families in the *i*th month. These people could belong to different families in other months.[25]

The specific analysis being planned will inform the choice of which family classification to use. To group people into families in the same way that the Census Bureau does, analysts should use PP-ID, $PP\text{-}MIS_i$, $HH\text{-}ADDID_i$, and FID_i. To analyze primary families excluding related subfamily members, analysts should include only those records with $FID2_i$ greater than zero. To analyze related subfamilies as distinct family units, analysts should use only those records with SID_i greater than zero. To uniquely identify (1) primary families excluding related subfamilies and (2) related subfamilies treated as distinct family groups, analysts should use PP-ID, $PP\text{-}MIS_i$, $HH\text{-}ADDID_i$, $FID2_i$, and SID_i. In those analyses, it is easy to distinguish unrelated families from other families.

Variables Describing Household and Family Composition

Table 12-11 shows the variables contained on the longitudinal research files summarizing household and family composition.[26]

[25] See footnote 18.

[26] More detailed information about the relationships between members is collected in the Household Relationships topical module. Those data provide extensive information about household composition at the time of the topical module interview.

Table 12-10. How to Uniquely Identify a Family in a Given Month of the Longitudinal Research Files

Sample Unit ID (PP-ID)	Current Address ID (HH-ADDID$_i$)	Person's Interview Status (PP-MIS$_i$)	Family ID, Including Subfamily (FID$_i$)	Family ID, Excluding Subfamily (FID2$_i$)	Subfamily ID (SID$_i$)	Family Type (FAMTYP$_i$)	Person Number (PP-PNUM)	Notes
110011111	11	1	1	1	0	0	101	This household contains a primary family of five people. The primary family contains two related subfamilies.
110011111	11	1	1	0	2	3	102	
110011111	11	1	1	0	2	3	103	
110011111	11	1	1	0	3	3	104	
110011111	11	1	1	0	3	3	105	
122210000	33	1	1	1	0	0	101	This household contains a primary family and two unrelated subfamilies.
122210000	33	1	1	1	0	0	104	
122210000	33	1	101	101	0	2	305	
122210000	33	1	101	101	0	2	306	
122210000	33	1	102	102	0	2	307	
122210000	33	1	102	102	0	2	308	
555555555	21	1	1001	1001	0	4	101	This household contains a primary individual and an unrelated subfamily.
555555555	21	1	101	101	0	2	201	
555555555	21	1	101	101	0	2	202	
555555555	21	1	101	101	0	2	203	
610000000	11	1	1001	1001	0	4	101	Primary individual.
897454644	11	1	1001	1001	0	1	101	Group quarters with two secondary individuals.
897454644	11	1	1002	1002	0	1	102	

Notes: Variables FID$_i$, FID2$_i$, and SID$_i$ are not part of the longitudinal research files. They can be merged from the core wave files or created using the algorithm shown in Figure 12-4. FAMTYP = 0 means the person belongs to a primary family. FAMTYP = 1 means the person is a secondary individual. FAMTYP = 2 means the person belongs to an unrelated subfamily. FAMTYP = 3 means the person belongs to a related subfamily. FAMTYP = 4 means the person is a primary individual.

Table 12-11. Variables Used to Describe Household Composition in the Longitudinal Research Files

Variable Name	Description
$FAMTYP_i$	Type of family in the *i*th month (e.g., primary family, related subfamily)
$FAMREL_i$	Family relationship in the *i*th month (e.g., reference person, spouse of family reference person, child of family reference person)
RRP_i	Recoded relationship to the household reference person in the *i*th month (e.g., household reference person living with relatives, child of household reference person)
$ENTID\text{-}SP_i$	Entry address ID of spouse in the *i*th month
$PNSP_i$	Person number of spouse in the *i*th month
$ENTID\text{-}PT_i$	Entry address ID of parent in the *i*th month
$PNPT_i$	Person number of parent in the *i*th month
$U\text{-}PNG_j$	Person number of guardian in the *j*th wave
$ENTID\text{-}GD_j$	Entry address ID of guardian in the *j*th wave

As Table 12-12 shows, RRP_i summarizes the relationship of each person to the *household reference person* in month *i*.

Table 12-12. Relationship to the Household Reference Person in a Given Month

Edited Relationship to the Household Reference Person (RRP_i)	Description
1	Household reference person, living with relatives
2	Household reference person, living alone or with nonrelatives
3	Spouse of household reference person
4	Child of household reference person
5	Other relative of household reference person
6	Nonrelative of household reference person, but related to other members of the household
7	Nonrelative of all members of the household

The household description depends on the identity of the reference person. For example, in Table 12-13, the household contains a mother, her daughter, and her daughter's son. If the mother is the household reference person (RRP_i = 1), her daughter is listed as a child of the household reference person (RRP_i = 4) and the daughter's son is listed as other relative of the household reference person (RRP_i = 5). If the daughter is the reference person, her son is listed as a child of the household reference person (RRP_i = 4) and her mother is listed as other relative of the household reference person (RRP_i = 5). Users should note that the household reference person can change from one month to the next; thus, the household description could also change.

Table 12-13. Using RRP to Identify Households Containing Three Generations in the Longitudinal Research Files

Household Reference Person	Relationship to the Household Reference Person (RRP_i)	Notes
Mother as Household Reference Person		
Mother	1	Reference person
Daughter	4	Child of reference person
Daughter's son	5	Other relative of reference person
Daughter as Household Reference Person		
Daughter	1	Reference person
Daughter's son	4	Child of reference person
Mother	5	Other relative of reference person

Six other variables in the longitudinal research file can be used to describe household and family composition: $PNSP_i$, $ENTID\text{-}SP_i$, $PNPT_i$, $ENTID\text{-}PT_i$, $U\text{-}PNG_j$, and $ENTID\text{-}GD_j$. These six variables identify the person number and entry address ID of the spouse, parent, or guardian living at the same address as the person in the *i*th month or *j*th wave (in the last two cases).[27] By building from these variables, the analyst can identify a variety of family configurations. For example, these variables can be used to identify households containing three generations. Table 12-14 displays one household containing a mother and her two children. One child (PP-PNUM = 102) has a son, and the other child (PP-PNUM = 104) has a spouse.

Table 12-14. Using PNSP and PNPT to Identify Households Containing Three Generations in the Longitudinal Research Files

Household Member	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Relationship to Household Reference Person (RRP_i)	Entry Address ID of Spouse ($ENTID\text{-}SP_i$)	Spouse ($PNSP_i$)	Entry Address ID of Parent ($ENTID\text{-}PT_i$)	Parent ($PNPT_i$)	Notes
Mother	11	101	1	11	999	11	999	Mother
Daughter #1	11	102	4	11	999	11	101	Child
Daughter #1's son	11	103	5	11	999	11	102	Grandchild
Daughter #2	11	104	4	11	105	11	101	Child
Spouse of Daughter #2	11	105	5	11	104	11	999	Spouse of child

Note: Value of 999 means not applicable.

[27] Parents and spouses always share the same sample unit ID (PP-ID) as the respondent. The variables are assigned values only in the months that people are living together. For example, a couple living together in Wave 1 would have values in the PNSP and ENTID-SP variables that pointed to each other. However, if they separate (and remain married) in Wave 2, the PNSP and ENTID-SP variables will be assigned values of 999 (indicating that the variables are not applicable).

Using Family-Level Income Variables

The longitudinal research files contain a number of family-level income variables. The family income variables on the longitudinal research files include the income of all related subfamily members. *In other words, primary family members and related subfamily members are treated as one family by the Census Bureau when calculating family-level income amounts*. The longitudinal research files do not contain any subfamily income variables. If family income variables are needed that do *not* pool related subfamilies with primary families, those income variables must be created. That is done by looping over persons with PP-MIS$_i$ of 1 and with common PP-ID, HH-ADDID$_i$, FID2$_i$, and SID$_i$ for each month.[28]

Table 12-15 illustrates how the family income variables on the longitudinal research files include the income of related subfamily members. From the previous example of a primary family of five people, the primary family contains two related subfamilies. Total family income (FF-INC$_i$) is $3,100. The incomes of all subfamily members are included in that amount.

Table 12-15. Family Income in the Longitudinal Research Files

Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Person Interview Status (PP-MIS$_i$)	Current Address ID (HH-ADDID$_i$)	Family ID, Including Subfamily (FID$_i$)	Sub-family ID (SID$_i$)	Total Family Income (FF-INC$_i$)	Person-Level Income (PP-INC$_i$)
110011111	11	101	1	11	1	0	$3,100	$ 100
110011111	11	102	1	11	1	2	$3,100	$ 500
110011111	11	103	1	11	1	2	$3,100	$ 500
110011111	11	104	1	11	1	3	$3,100	$1,000
110011111	11	105	1	11	1	3	$3,100	$1,000

More About Using the SIPP ID Variables: Identifying Movers

When a person moves, the current address field (HH-ADDID$_i$) changes. The PP-ID, PP-ENTRY, and PP-PNUM values remain the same. The first digit (or first two digits in the 1992 Panel) of HH-ADDID$_i$ indicate(s) the wave in which a household is first interviewed at that new address. The remaining digits sequentially number the households that split into two or more households, as a result of a move to a different location by original sample members. Thus, new addresses in Wave 2 are numbered 21, 22, and so on. New addresses in Wave 3 are numbered 31, 32, and so on. New addresses in Wave 10 are numbered 101, 102, and so on. (Readers may wish to refer to Figure 2-1 [pp. 2-10–2-14], which illustrates movement into and out of households.)

[28] FID$_i$ and SID$_i$ are not included on the longitudinal research files. They can be merged from the core wave files or created by using the algorithm shown in Figure 12-4.

Table 12-16 shows that persons 101 and 102 in the first household are original sample members. Person 401 moved into the home of persons 101 and 102 in Wave 4. In Wave 7, all three moved to a new location and were joined by person 701. In the second household, person 101 is an original sample member who moved to a new location in Wave 3. In the third household, person 102 is an original sample member who used to live with persons 101 and 103 of the same sample unit ID (PP-ID), but moved to a new location in Wave 3 (to a different location from person 101). In the fourth household, person number 103 is an original sample member who used to live with persons 101 and 102 of the same sample unit ID number. Person 103 moved to a new location in Wave 10 and was joined by person 1001, who just entered the SIPP sample. All but two people moved from their original location (i.e., only two people have HH-ADDID$_i$ equal to PP-ENTRY).

Table 12-16. How to Identify Movers in the Longitudinal Research Files

Wave	Sample Unit ID (PP-ID)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)	Person Interview Status (PP-MIS$_i$)	Current Address ID (HH-ADDID$_i$)	Notes
1	123456789	11	101	1	11	Persons 101 and 102 are the original sample members
	123456789	11	102	1	11	
4	123456789	11	101	1	11	Person 401 begins to live with them in Wave 4.
	123456789	11	102	1	11	
	123456789	11	401		11	
7	123456789	11	101	1	71	All three people move in Wave 7 and person 701 joins them
	123456789	11	102	1	71	
	123456789	11	401	1	71	
	123456789	71	701		71	
1	321456789	11	101	1	11	Person 101, person 102, and person 103 are original sample members.
	321456789	11	102	1	11	
	321456789	11	103	1	11	
3	321456789	11	101	1	31	Person 101 moved in Wave 3. Person 102 moved in Wave 3 to a different location from person 101. Person 103 remained with person 101.
	321456789	11	102	1	32	
	321456789	11	103	1	31	
10	321456789	11	101	1	31	Person 103 is an original sample member who used to live with persons 101 and 102 of the same ID. In Wave 10, person 103 lives in a new location with person 1001, who just entered the SIPP sample.
	321456789	11	102	1	32	
	321456789	11	103	1	101	
	321456789	101	1001	1	101	

The next example (Table 12-17) further illustrates how the ID system works as people move to new addresses, additional people move in with them, and households split. A review of Figure 2-1 (pp. 2-10–2-14) may help in understanding the various household changes.

- In Wave 1, there is a five-person household consisting of a husband, a wife, a daughter, a son, and a cousin. Because this is the first wave, the current address number is 11, indicating

Table 12-17. Another Example of Household Changes and Their Effects on the ID Variables in the Longitudinal Research Files

Household Member	Sample Unit ID (PP-ID)	Current Address ID (HH-ADDID$_i$)	Entry Address ID (PP-ENTRY)	Person Number (PP-PNUM)
Wave 1				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 2				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son	101111103	11	11	104
Cousin	101111103	11	11	105
Wave 3				
Father	101111103	11	11	101
Mother	101111103	11	11	102
Daughter	101111103	11	11	103
Son-in-Law	101111103	11	11	301
Cousin	101111103	11	11	105
Wave 4	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
	Cousin's Household			
Cousin	101111103	42	11	105
Uncle	101111103	42	42	401
Wave 10	**Parent's Household**			
Father	101111103	11	11	101
Mother	101111103	11	11	102
	Daughter's Household			
Daughter	101111103	41	11	103
Son-in-Law	101111103	41	11	301
Newborn	101111103	41	41	1001

address 1 of Wave 1, and the entry address number for each member of the household is the same as the current address number. Because they are assigned in Wave 1, the person numbers are in the 100 series and are numbered sequentially, beginning with 101.

- During Wave 2, the son joins the Army, moves into military barracks, and therefore leaves the SIPP sample.[29] The son's record, person number 104, will contain information (either

[29] Members of the armed forces are included in the SIPP sample only if they are living state-side in private housing. Those living overseas or in military barracks are not included in the SIPP sample universe.

imputed or provided by proxy) on his characteristics for the time in Wave 2 that he was still in the sample. If he does not return to the sample during the remainder of the panel, there will be no records for him beyond Wave 2.

- During Wave 3, the daughter marries and her husband moves into the household. The current address number where the mother, father, cousin, daughter, and son-in-law live remains the same because it is the same address. The son-in-law's entry address number is 11 because he first enters the SIPP sample at an address coded 11. The person number for the son-in-law is in the 300 series (301) because he joins the SIPP sample in Wave 3.
- During Wave 4, the daughter and son-in-law move into a new house. Their current address number changes to 41 to indicate that a new address has been established in Wave 4. Meanwhile, the cousin, who is over age 15, moves in with an uncle.[30] The cousin's current address number changes to 42 (i.e., the second household added into the SIPP sample in the fourth wave). The assignment of address number 41 to the daughter and 42 to the cousin is random. It could be the other way around. The uncle enters the SIPP sample and receives an address number of 42 and an entry address number of 42. The uncle's person number is in the 400 series (401) since he joins the survey in Wave 4.
- No changes in household composition are observed during Waves 5–9.
- During Wave 10, the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. The newborn's entry address is 41, since that is the current address ID of the daughter and son-in-law at the time of birth. The newborn's person number is 1001, reflecting the fact that the newborn came into the SIPP sample in Wave 10. Meanwhile, the cousin moves to Europe and therefore leaves the SIPP sample. The uncle, even though he did not move to Europe with the cousin, also leaves the SIPP sample because he no longer resides with an original SIPP sample member. Their records are no longer listed.

Table 12-18 displays this example again, but this table depicts how the HH-ADDID$_i$ variable changes over time to reflect the household composition changes. The table also illustrates the structure of the full panel data files.

There are two extremely rare occasions in which the original PP-ID, PP-ENTRY, and PP-PNUM values are modified:

1. The first occasion is when two separate sampling units, each containing original sample members, are merged, perhaps because of a marriage. In this situation, one of the original set of PP-ID and PP-ENTRY values is retained and the other set is changed to agree with the retained set. The person number values (PP-PNUM) of the changed set are modified further to be between 180 and 199, inclusive.

[30] In the 1993 Panel, all original sample members were followed, no matter what their ages. In all other panels, only people 15 years of age or older were followed when they moved to new addresses.

Table 12-18. Household Changes and Their Effects on the Household ID (HH-ADDIDi) Variable in the Longitudinal Research File

				HH-ADDID$_i$																			
				Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
	PP-	PP-		Month				Month				Month				Month				Month			
PP-ID	ENTRY	PNUM	Notes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
101111103	11	101	Father	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	102	Mother	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	103	Daughter	11	11	11	11	11	11	11	11	11	11	11	11	41	41	41	41	41	41	41	41
101111103	11	104	Son	11	11	11	11	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101111103	11	105	Cousin	11	11	11	11	11	11	11	11	11	11	11	11	11	42	42	42	42	42	42	42
101111103	11	301	Son/law	0	0	0	0	0	0	0	0	0	11	11	11	41	41	41	41	41	41	41	41
101111103	42	401	Uncle	0	0	0	0	0	0	0	0	0	0	0	0	42	42	42	42	42	42	42	42
101111103	41	1001	Newborn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

				HH-ADDID$_i$																			
				Wave 6				Wave 7				Wave 8				Wave 9				Wave 10			
	PP-	PP-		Month				Month				Month				Month				Month			
PP-ID	ENTRY	PNUM	Notes	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
101111103	11	101	Father	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	102	Mother	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
101111103	11	103	Daughter	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41
101111103	11	104	Son	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101111103	11	105	Cousin	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	0	0	0	0
101111103	11	301	Son/law	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41
101111103	42	401	Uncle	42	42	42	42	42	42	42	42	42	42	42	42	42	42	42	0	0	0	0	0
101111103	41	1001	Newborn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	41	41	41

2. The second occasion is when a household splits into two new households (in which each new household gains a new sample person) and later the households recombine. For example, assume that a married couple separate in Wave 3, each moving in with a sibling. Both siblings are assigned a person number of 301, because they entered the sample in Wave 3 at different addresses (thus, HH-ADDID$_i$ = 31 and 32). If the husband and wife reunite in Wave 6, and bring the siblings with them, one sibling's person number would be changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699, inclusive).

Because a record in the longitudinal research file describes the person throughout the entire panel and because the sample unit ID (PP-ID) cannot change on this record, each person in a merged household whose ID values were changed is assigned two full panel records. The first record contains the original ID information of the person before the merge and identifies the person as having exited the sample at the time of the merge. The second record contains the new ID information and identifies the person as having entered the sample at the time of the merge. There is no way to link the two records in the longitudinal research files.[31]

Identifying Program Units

Besides household and family composition data, the longitudinal research files contain detailed information about participation in health insurance and various government transfer programs. For most programs, three characteristics are recorded (Table 12-19):

1. Whether the person is covered;
2. Who received the income or benefit; and
3. The amount of the income or benefit.

The coverage variables identify whether the income or benefit covers that person in month i. In other words, when a person is flagged as covered by food stamps (FOODSTMP$_i$ = 1), the person either received the benefits directly (because he or she was the authorized food stamp recipient) or indirectly (because he or she was in the same program unit as the authorized recipient). The coverage variables also allow users to determine each person's membership in each program unit. That is useful because program units often exclude some members of the family or household.[32] Also, as with households and families, membership in program units can change from one month to the next. For that reason, program unit membership and characteristics of the unit should be evaluated for each month.

[31] If needed, this information can be merged from the core wave files. Chapters 10 and 13 provide details.

[32] In the 1984 and 1985 Panels, coverage for the Women, Infants, and Children (WIC) nutrition program was imputed to children under 6 years old if their mother reported participation in the WIC program. Beginning with the 1986 Panel, WIC coverage has been assessed directly for all sample members.

Table 12-19. Variables Describing Participation in Government Transfer Programs and Health Insurance Programs in the 1990–1993 Longitudinal Research Files

Program	Coverage	Authorized Recipient	G1 Source Code	Amount
Social Security	SOC-SEC	SS-PIDX	1	Locate one of the amount variables: G1AMT1–G1AMT10, using the corresponding source variables: G1SRC1–G1SRC10
Railroad Retirement	RAILROAD	RR-PIDX	2	
Federal Supplemental Security Income	—	—	3	
Veteran's Benefits	VETS	VA-PIDX	8	
Aid to Families with Dependent Children	AFDC	AFDCPIDX	20	
General Assistance	GEN-ASST	GA-PIDX	21	
Foster Child Care	FOST-KID	FOSTPIDX	23	
Other Welfare	OTH-WELF	OTH-PIDX	24	
WIC Benefits	WICCOV	WIC-PIDX	25	
Food Stamps	FOODSTMP	FS-PIDX	27	
Medicare	CARECOV	—	—	
Medicaid	CAIDCOV	—	—	
CHAMPUS	CHAMP	—	—	

The authorized recipient variables identify the people who actually received the income or benefit for the people in their program units. In the longitudinal research files, those variables do not use the entry address and person number values. Instead, they use the sequence number of the person within the sample unit (PP-RCSEQ) to identify authorized recipients. In other words, the authorized food stamp recipient is the person for whom FS-PIDX$_i$ in month i equals PP-RCSEQ.

Individuals who are members of a common program unit in a given month (i) can be identified by using the sample unit ID (PP-ID), the person's interview status in month i (PP-MIS$_i$), and the authorized recipient variable in month i. For example, members of a common food stamp unit in month i are those with PP-MIS$_i$ of 1 and common values of PP-ID (a value that does *not* change from month to month) and FS-PIDX$_i$ (a value that *does* change from one month to the next). The SIPP longitudinal research files do not include authorized recipient variables for Medicare and SSI programs.[33]

There are some exceptions to the rules:

- Social Security, Railroad Retirement, WIC, and AFDC can offer benefits solely to children. When that happens, an adult will receive the income on behalf of the children. The adult, therefore, is flagged as the authorized recipient and the income amounts appear on the record of the adult. The adult authorized recipient, however, is not flagged as being covered by the program. The children are flagged as covered.

[33] In effect, each person covered by these two programs is an authorized recipient, and the program units are the people themselves.

- Most SSI recipients are elderly and disabled adults, but they can also be children with disabilities.[34] Even so, the SSI amount is recorded on an adult's record, not on the child's record. Unlike the core wave files, the longitudinal research files have no coverage variable indicating whether or not the child, adult, or both, were covered. If needed, this information can be merged from the core wave files. Chapter 13 provides a detailed discussion of merging SIPP files.
- The medical insurance variables simply reflect who is enrolled in which type of program. There are no associated amount variables.

These rules and exceptions are illustrated in Table 12-20. The household contains one AFDC unit and two food stamp units. The mother is covered by Social Security and SSI. The mother of the (disabled) child receives SSI on behalf of her child. The grandchild receives WIC. Everyone in the household is enrolled in Medicaid. The coverage variables are set to 2 whenever the person is *not* covered by the particular program. The indicators for the authorized recipients do *not* use the PP-ENTRY and PP-PNUM values. Instead, they are based on the "line number" of the authorized recipient on the household roster. That is very different from the indicators used on the core wave files.

Using the Unearned Income Variables

To save space, the Census Bureau organizes the unearned income variables differently in the longitudinal research files than in the core wave files. As shown in Table 12-21, 10 variables on each person's record identify up to 10 different sources of unearned income (G1SRC1–G1SRC10). For each source identified, there is a corresponding amount variable ($G1AMT1_i$–$G1AMT10_i$). Income amounts are recorded with monthly resolution. The person in Table 12-21 periodically receives $500 in federal SSI and $125 in food stamps. The person does not receive any other source of unearned income.

When using these fields, analysts often find it helpful to realign the unearned income into new income-specific variables.[35]

[34] In the 1990s, the definition of qualifying disabling conditions was expanded. That change in definition resulted in a rapid expansion of the child SSI caseload.

[35] For example, Table 12-22 includes monthly variables for SSI and food stamps that were created by using the algorithm in Figure 12-5.

Table 12-20. Example of Program Units, Coverage, and Benefit Amounts in the Longitudinal Research Files

Variable	Mother	Daughter #1	Daughter #1's Son	Daughter #2	Spouse of Daughter #2
PP-PNUM	101	102	103	104	105
PP-RCSEQ	1	2	3	4	5
AGE_i	70	21	4	25	26
AFDC					
$AFDC_i$	2	1	1	2	2
$AFDCPIDX_i$	0	2	2	0	0
Food Stamps					
$FOODSTMP_i$	2	1	1	1	1
$FS\text{-}PIDX_i$	0	2	2	4	4
SSI					
This only appears in the General Amounts (G1) section.					
WIC					
$WICCOV_i$	2	2	1	2	2
$WIC\text{-}PIDX_i$	0	2	2	0	0
Medicaid					
$CAIDCOV_i$	1	1	1	1	1
Social Security					
$SOC\text{-}SEC_i$	1	2	2	2	2
General (G1) Sources and Amounts					
G1SRC1	3	20	0	27	0
$G1AMT1_i$ ($)	188	123	0	130	0
G1SRC2	1	27	0	0	0
$G1AMT2_i$ ($)	470	160	0	0	0
G1SRC3	0	3	0	0	0
$G1AMT3_i$ ($)	0	122	0	0	0
G1SRC4	0	25	0	0	0
$G1AMT4_i$ ($)	0	30.12	0	0	0

[a] These codes are explained in the next section of text.

Income Topcoding

The Census Bureau topcodes each income variable to protect against the possibility that a user might identify a SIPP respondent with very high income.[36] While the data dictionary indicates a topcode of $33,332 for monthly income, that is also the income topcode for the wave. That topcode is, therefore, rarely used for a month. In most cases, the monthly income is topcoded at $8,333, which actually represents $8,333 or more. Individual amounts above $8,333 may occasionally be shown if the respondent's income varied considerably from month to month

[36] New topcoding procedures are being implemented with the 1996 Panel. When a longitudinal research file for the 1996 Panel is available, this discussion will be revised to describe those new procedures. At present, users should note that this description does *not* pertain to the core wave files from the 1996 Panel.

Table 12-21. Unearned Income in the Longitudinal Research Files

		PP-MIS																			
		Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
		Month				Month				Month				Month				Month			
Variable		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	0	0	0	0
G1SRC1	3																				
G1AMT1($)		500	500	500	500	0	0	0	500	500	500	500	500	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2($)		0	0	0	0	0	0	0	125	125	125	125	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0																				
G1AMT10 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 12-21. Unearned Income in the Longitudinal Research Files (*continued*)

		PP-MIS																			
		Wave 6			Wave 7			Wave 8				Wave 9			Wave 10						
		Month			Month			Month				Month			Month						
Variable		**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**29**	**40**
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC1	3																				
G1AMT1($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0																				
G1AMT10 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 12-22. User-Created SSI and FSP Variables Using the Unearned Income Variables in the Longitudinal Research Files

		PP-MIS																			
		Wave 1				Wave 2				Wave 3				Wave 4				Wave 5			
		Month				Month				Month				Month				Month			
Variable		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	0	0	0	0
G1SRC1	3																				
G1AMT1 ($)		500	500	500	500	0	0	0	500	500	500	500	500	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2 ($)		0	0	0	0	0	0	0	125	125	125	125	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0																				
G1AMT10 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SSI ($)		500	500	500	500	0 [a]	0	0	500	500	500	500	500	–99	–99	–99	–99	–99	–99	–99	–99
FSP ($)		0	0	0	0	0	0	0	125	125	125	125	0	–99	–99	–99	–99	–99	–99	–99	–99

[a] In SAS, the unassigned values would have a "system missing" value displayed as a ".".

Table 12-22. User-Created SSI and FSP Variables Using the Unearned Income Variables in the Longitudinal Research File (*continued*)

		PP-MIS																			
		Wave 6				Wave 7				Wave 8				Wave 9				Wave 10			
		Month				Month				Month				Month				Month			
Variable		**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**
PP-ID	7887																				
PP-PNUM	102																				
PP-MIS		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC1	3																				
G1AMT1 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC2	27																				
G1AMT2 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC3	0																				
G1AMT3 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC4	0																				
G1AMT4 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC5	0																				
G1AMT5 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC6	0																				
G1AMT6 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC7	0																				
G1AMT7 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC8	0																				
G1AMT8 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC9	0																				
G1AMT9 ($)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1SRC10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G1AMT10 ($)																					
SSI ($)		–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99
FSP ($)		–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99	–99

Figure 12-5. Creating Monthly Food Stamp and SSI Income Variables from the Unearned Income Variables in the Longitudinal Research Files

```
For each person:
      /*
            This step is not needed in SAS
      */
      For each month (index = mo):
            If PP-MIS (mo) = 1 Then do
                  SSI(mo) = 0
                  FSP(mo) = 0
            End If PP-MIS (mo) = 1
            Else do
                  SSI(mo) = -99
                  FSP(mo) = -99
            End Else
      End month loop
      /*
            Begin here for SAS
      */
      For each G1SRC (index=i):
            If G1SRC(i)=3 Then do
                  For each month (index=mo)
                        If PP-MIS (mo) = 1 Then do SSI(mo)=G1AMT(i,mo)
                        End If PP-MIS (mo) = 1
                  End month loop
            End If G1SRC(i)=3
            Else if G1SRC(i)=27 Then do
                  For each month (index=mo)
                        If PP-MIS (mo) = 1 Then do FSP(mo)=G1AMT(i,mo)
                        End If PP-MIS (mo) = 1
                  End month loop
            End if G1SRC(i)=27
      End G1SRC loop
```

within a wave. For example, if a respondent's income from a single job was concentrated in only one of the four reference months, a figure as high as $33,332 could be shown.

Summary income variables on the person, family, and household records are simply the sums of the component variables after they have been topcoded. The summary variables are not independently topcoded. Thus, a person with high income from several sources (multiple jobs, businesses, property) could have aggregate monthly income well over the topcode for each source, and yet the data could still be greatly understating the person's true income.

As shown in Table 12-23, person 101 has wages topcoded. The person received considerably more money in December than in the other months. Also, total family income and total household income are the sum of the income amounts (in this case, WS-ERN-AMT1$_i$ + G1AMT1$_i$) after they have been topcoded.

Table 12-23. Example of Topcoding in the Longitudinal Research Files

Person Number (PP-PNUM)	Calendar Month	Household Total Income (HH-INC$_i$)	Family Total Income (FF-INC$_i$)	Wages (WS-ERN-AMT1$_i$)	Child Support Payments (G1AMT1$_i$)
101	10	$ 9,333	$ 9,333	$ 8,333	$1,000
101	11	$ 9,333	$ 9,333	$ 8,333	$1,000
101	12	$13,123	$13,123	$12,123[a]	$1,000
101	01	$ 5,793	$ 5,793	$ 4,543	$1,250

[a] This figure can exceed the nominal *monthly* topcode of $8,333 because the person's total earnings for the *wave* were below $33,332.

Using Allocation (Imputation) Flags

As described in Chapter 4, the Census Bureau often imputes information when a person does not respond to the survey or to a particular question. Two sources identify whether information has been imputed:

1. Beginning with the 1991 Panel, all data for a wave are imputed if a person was not successfully interviewed in one wave but had complete information (from either a successful interview or a proxy interview) in the two adjacent waves. In those cases, the value of WAVFLG will be greater than zero and INTVW will be 3 or 4.
2. A variable of interest may be imputed. In the longitudinal research files, allocation (imputation) flags are included for the earned income, asset income, and unearned (transfer) income variables.

Other variables are also subject to editing and imputation. The edit and imputation procedures used for the longitudinal research files differ from those used for the core wave files. The procedures used for the longitudinal research files make use of the full set of longitudinal data for a person. Because the core wave files are processed individually, the edit and imputation procedures applied to those files have, at most, 4 months of observations for a person. The procedures applied to the core wave files make greater use of cross-observation imputation methods than do those applied to the longitudinal research files.[37]

Using Weights

The full panel longitudinal research files include the calendar year weights (FNLWGTs) and the full panel weight (PNLWGT). The number of calendar year weights depends on the duration of

[37] The edit and imputation procedures applied to the core wave files from the 1996 Panel make greater use of retrospective information than procedures used in earlier panels. See Chapters 4 and 10 for details.

the panel; the number varies from one calendar year weight for the 1989 Panel to three calendar year weights for the 1993 Panel. When the 1996 full panel file is available, it will have four calendar year weights.

The source and accuracy statements that accompany all SIPP full panel files ordered from the Census Bureau provide suggestions on how to use the weight variables in those files. Also, Chapter 8 of this Guide contains a full discussion of how to use weights in full panel files.

Identifying States

The longitudinal research file contains a variable (GEO-STE) that identifies 41 individual states and the District of Columbia; the nine other states are suppressed into three groups:

1 Maine, Vermont;

2. Iowa, North Dakota, South Dakota; and

3. Alaska, Idaho, Montana, Wyoming.

Even though it is possible to identify most states, the SIPP sample was not designed to be representative at the state level and should not be used to produce direct state-level estimates. The state variable is included on the public use files to allow examination of how state-level characteristics affect national estimates. For example, a user could apply the state-specific eligibility criteria for a means-tested program in order to arrive at a national estimate of the number of people eligible for the program. Because some states are not uniquely identified, some method of allocating the state-specific eligibility rules to sample persons in those states would need to be devised.

Identifying Metropolitan Areas

The longitudinal research files do not contain any variables identifying metropolitan areas. Analysts who need this information should merge it from the core wave files. Chapter 11 provides details about how to use the variables identifying metropolitan areas. Chapter 13 provides instructions for merging data from multiple SIPP public use files.

13. Linking Core Wave, Topical Module, and Longitudinal Research Files

In many situations, a single Survey of Income and Program Participation (SIPP) data file will not contain the information needed for a project. Because only limited core information is included on the topical module files, analysts often need to merge data from the core wave or longitudinal research files with topical module information. Also, they may need to link two or more topical module files, each containing data on a different topic and collected in different waves. And there are situations in which it is necessary to merge data from the core wave files with data from the longitudinal research files. Those situations arise because not all of the core wave content is included on the longitudinal research files (e.g., calendar month weights are only on the core wave files).[1] This chapter describes procedures for linking core wave, topical module, and full panel data files.

This chapter assumes a working knowledge of the files that will be linked.[2] Analysts who are not familiar with those files should read the following before proceeding with this chapter:

- Chapter 9 for an overview of the SIPP data files;
- Chapter 10 for a discussion of the core wave files;
- Chapter 11 for a discussion of the topical module files; and
- Chapter 12 for a discussion of the longitudinal research files.

In all cases, this chapter describes procedures for linking person records across files. It does not discuss procedures for linking households or families because those procedures become problematic when working with longitudinal data.[3]

[1] Even when the same variables are on both the core wave and longitudinal research files, the data may not be the same. Different edit and imputation procedures are used for these two types of files. Prior to the 1996 Panel, all edit and imputation procedures applied to the core wave files worked entirely *within* the given file. Information from previous waves or later waves was not used. Beginning with the 1996 Panel, edit and imputation procedures applied to the core wave files make greater use of information from previous waves. However, because the core wave files are processed as the data become available, it is not possible to make use of information from future waves. The edit and imputation procedures applied to the longitudinal research files, however, make use of each person's full longitudinal record. There are many times when the preferred data for a study will be on the longitudinal research files but the weights will be on the core wave files.

[2] This chapter does not discuss the longitudinal research file from the 1996 Panel because, as of this writing, it is not available. That information will be added to an updated version of this chapter once the file becomes available. In the interim, the only information included in this chapter on the 1996 longitudinal research file is the new variable names being used in the 1996 Panel data files.

[3] Difficulties arise when unit composition changes over time. In those situations, there is no unambiguous way to define longitudinal households and families, and many *ad hoc* procedures run the risk of introducing biases into

This chapter begins with a discussion of the mechanics involved in linking SIPP data files. The procedures are straightforward and easily implemented. In each case there are three basic steps:

1. Create data extracts from each of the files to be linked;
2. Sort the files in common order by using the variables identified as match keys; and
3. Merge the files.

There are two general formats that the final files can take. This chapter refers to these as person-month format (the format of the current core wave files) and person-record format (the format of the longitudinal research files).[4] The choice of format will be a function of the planned analysis and the software that will be used for that analysis. Where appropriate, procedures for generating each type of data file are described.

After discussing the mechanics of linking SIPP files, this chapter discusses why nonmatches occur and suggests ways to deal with them.

For the 1996 Panel, most variable names changed from those of previous panels. To aid users working with pre-1996 panel files, this chapter presents both the old and the new variable names when the text applies to both. In the main body of the text, the old names are presented in parentheses following the new names. For example, the sample unit ID variable name, which is SSUID in the 1996 Panel, was SUID in previous panels; it is written in this chapter as SSUID (SUID). In tables, a variety of methods are used to present both the old and the new names.

Procedures for Linking Files

There are six types of merges that SIPP users commonly need to perform:

1. Person-month records within a core wave file can be linked, creating a single wide record for each person rather than a record for each person for each month;[5]
2. Two or more core wave files can be linked together;
3. Core wave files can be linked to longitudional research files;

analyses of those units. The alternative approach that has gained acceptance in the research community involves assigning to people the characteristics of the households or families to which they belong at each point in time. Subjects can then be followed over time, as can the characteristics of the households or families to which they belong. One exception to the longitudinal household problem is with program units (e.g., food stamp units), where program rules can be used to define when changing composition constitutes the formation of a new unit (as opposed to changed composition of an existing unit). For discussions of the issues involved in studying longitudinal households and families, see McMillen and Herriot (1985), Duncan and Hill (1985), Citro et al. (1986), and Kalton et al. (1987).

[4] Some software (e.g., Stata) refers to this as "wide" format, while the person-month format is referred to as "long."

[5] This procedure transforms the current format of the core wave files into a format similar to that used prior to the 1990 Panel, a format analogous to that used for the longitudinal research files.

4. Two or more topical module files can be linked to each other;
5. Topical module files can be linked to core wave files; and
6. Topical module files can be linked to longitudinal research files.

This chapter addresses each of these merges in turn.

Linking Within a Core Wave File—Transforming the Person-Month Format into the Person-Record Format

This procedure transforms the person-month-format core wave files (with one record per person per month) into a single wide record per person (the format used for the core wave files before the 1990 Panel). As well as being useful in its own right, reformatting is often a necessary first step when merging core wave files with data from either the topical module files or from the longitudinal research files.

Two approaches for this link are described. Programmers using third-generation languages, such as FORTRAN and PL/1, typically use the first approach. Programmers using fourth-generation languages, such as SAS and SPSS, typically use the second approach.

The first approach (using FORTRAN) contains four steps:

1. Sort the file by person and reference month, using the following variables: sample unit ID [SSUID (SUID)], entry address ID [EENTAID (ENTRY)], person number [EPPPNUM (PNUM)], and reference month [SREFMON (REFMTH)].[6] This is the sort order the Census Bureau uses for the core wave files. If the file being used is in its original sort order, this step can be skipped.
2. Define and initialize monthly variable arrays to some "missing data" code. Users should be careful to choose initial values outside the range of legal values for the variables of interest. For example, the variable TAGE (AGE) would be defined as an array of four elements, and each element could be initialized to –9 (an age that no one can have); the variable TPTOTINC (TOTINC) would be defined as an array of four elements and each element could be initialized to –999999 (a negative value outside the range of the variable), and so on.
3. Read each person's corresponding person-month record and put the information into the appropriate element of the array.
4. Write the person-based record from the information stored in the arrays.

The second approach (using SAS) also contains four steps:[7]

[6] In the 1996 Panel, the entry address is no longer needed to uniquely identify people. Its continued use will not create any problems; it is simply redundant information for purposes of identifying SIPP sample members.

[7] An alternative procedure that may be useful in many cases uses SAS Proc Transpose. Stata also has a procedure—reshape—that can accomplish this task.

1. Sort the file by person and reference month, using the following variables: sample unit ID [SSUID (SUID)], entry address ID [EENTAID (ENTRY)], person number [EPPPNUM (PNUM)], and reference month [SREFMON (REFMTH)]. This is the sort order used by the Census Bureau for the core wave files. If the file being used is in its original sort order, this step can be skipped.
2. Write out four files, each one containing the person ID variables and the variables for 1 of the 4 months. For example, file1 would have the person ID variables [SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM)] and the variables for month one, file2 would have the person ID variables and the variables for month two, and so on.
3. Rename the (monthly) variables in each of the four files to unique names. For example, the variable names in file1 might be TAGE1 (AGE1) and PTOTINC1[8] (TOTINC1); in file2 the variable names might be TAGE2 (AGE2) and PTOTINC2 (TOTINC2).
4. Merge the four files together, using SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) as the match keys.

The SAS code in Figure 13-1 performs the above steps.

The person-month format of the core wave files (before reformatting) is illustrated in Table 13-1. Person number 101 is in the sample all 4 months, person number 102 is in the sample all 4 months, person number 201 is in the sample for 2 months, and person number 202 is in the sample for 1 month. The person-record format (after reformatting) is illustrated in Table 13-2. Missing data are indicated by a single period, the default missing data code in SAS. For the FORTRAN example, the missing data would have codes of –9 and –999999.

Linking Two or More Core Wave Files

There are three reasons to link two or more core wave files:

1. To create an analysis file for one or more calendar months containing data from all four rotation groups. For example, data for March 1994 are contained in the Wave 7 file (of the 1992 Panel) for rotation groups 4 and 1, and in the Wave 8 file for rotation groups 2 and 3. (Data for the same calendar month are also in Waves 4 and 5 of the 1993 Panel.)
2. To create an analysis file containing more than 4 months of information for each person. This linkage is of primary interest to users of the 1996 Panel, beause longitudinal research files for all other panels are available from the Census Bureau.
3. As preparation for merging core wave data with data from either the topical module files or the longitudinal research files.

[8] Because variable names in SAS are limited to eight characters, the monthly variable name is shortened from TPTOTINC1 (nine characters) to PTOTINC1 (eight characters).

Figure 13-1. Sample SAS Code to Change the Core Wave Files from Person-Month Format to Person-Record Format from Wave 2 of the 1996 Panel

```
/*
   this creates the initial extract from the full core wave file
*/
data allmnths;
   set corewv962
      (keep =
         ssuid
         eentaid
         epppnum
         srefmth
         tage
         tptotinc
      );
run;

/*
   sort the data - if the master file was in its original order, this
   step is not needed
*/
proc sort;
   by ssuid eentaid epppnum srefmth;
run;

/*
   write out 1 file for each of the four months, renaming variables in
   the process
*/
data
   file1
      (rename =
         (tage = tage1
          tptotinc = ptotinc1
          srefmth = srefmth1
         )
      )
   file2
      (rename =
         (tage = tage2
          tptotinc = ptotinc2
          srefmth = srefmth2
         )
      )
   file3
      (rename =
         (tage = tage3
          tptotinc = ptotinc3
          srefmth = srefmth3
         )
      )
```

(figure continues)

Figure 13-1. Sample SAS Code to Change the Core Wave Files from Person-Month Format to Person-Record Format from Wave 2 of the 1996 Panel (*continued*)

```
   file4
      (rename =
         (tage = tage4
          tptotinc = ptotinc4
          srefmth = srefmth4
         )
      )
   ;

   set allmnths;

   select (srefmth);
      when (1) output file1;
      when (2) output file2;
      when (3) output file3;
      when (4) output file4;
   end;
run;

/*
   merge the 4 "monthly" files together, forming the final file
*/
data newfile;
   merge
      file1
      file2
      file3
      file4
      ;
   by ssuid eentaid epppnum;
run;
```

Creating files in the person-month format is straightforward. In this instance, the files from each of the contributing core wave files simply need to be sorted and interleaved to create the final analysis file. The final sort order would likely be based on SSUID (SUID), EENTAID (ENTRY), EPPPNUM (PNUM), SWAVE (WAVE), and SREFMON (REFMTH).

If a person-record format (with just one record per person) is desired, the first step is interleaving the files to create the person-month-format file. Then, using that as the input file, analysts can apply the procedures described in the preceding section to generate a file with a single wide record for each person. There will be up to 4 months of data for each wave used. In the example from Tables 13-1 and 13-2, if three waves of data are being combined, the final file will have 12 values for SREFMON (REFMTH), TAGE (AGE), and TPTOTINC (TOTINC). In the SAS program code, the names would likely be REFMTH1–REFMTH12, TAGE1–TAGE12, and TOTINC1–TOTINC12.

Users attempting to create their own longitudinal databases from the core wave files should proceed cautiously. The edit and imputation procedures applied to the core wave files for the

Table 13-1. Example of the Core Wave Person-Month File Structure

Sample Unit ID [SSUID (SUID)]	Entry Address ID [(EENTAID (ENTRY)]	Person Number [EPPPNUM (PNUM)]	Reference Month [(SREFMON (REFMTH)]	Age [TAGE (AGE)]	Total Income [(TPTOTINC (TOTINC)]
123456781000	011 (11)	0101 (101)	1	42	$2000
123456781000	011 (11)	0101 (101)	2	42	$2100
123456781000	011 (11)	0101 (101)	3	42	$2000
123456781000	011 (11)	0101 (101)	4	43	$2000
123456781000	011 (11)	0102 (102)	1	41	$ 500
123456781000	011 (11)	0102 (102)	2	41	$ 500
123456781000	011 (11)	0102 (102)	3	41	$ 0
123456781000	011 (11)	0102 (102)	4	41	$ 0
123456781000	011 (11)	0201 (201)	2	18	$ 200
123456781000	011 (11)	0201 (201)	3	18	$ 200
123456781000	011 (11)	0201 (201)	4	18	$ 200
123456781000	011 (11)	0202 (202)	2	2	$ 0
123456781000	011 (11)	0202 (202)	3	2	$ 0
123456781000	011 (11)	0202 (202)	4	2	$ 0

Table 13-2. Example of the Core-Wave Wide-Record/Person File Structure (After Applying the Program in Figure 13-1 to the Data in Table 13-1)

Sample Unit ID [SSUID (SUID)]	Entry Address ID [EENTAID (ENTRY)]	Person Number [EPPPNUM (PNUM)]	Reference Month (SREFMTH)[a]				Age (TAGE)[b]				Total Income (PTOTINC)[c]			
			1	2	3	4	1	2	3	4	1	2	3	4
123456781000	011 (11)	0101 (101)	1	2	3	4	42	42	42	43	$ 2000	$ 2100	$ 2000	$ 2000
123456781000	011 (11)	0102 (102)	1	2	3	4	41	41	41	41	$ 500	$ 500	$ 0	$ 0
123456781000	011 (11)	0201 (201)	.	2	3	4	.	18	18	18	.	$ 200	$ 200	$ 200
123456781000	011 (11)	0202 (202)	.	2	3	4	.	2	2	2	.	$ 0	$ 0	$ 0

Note: . = missing.

[a] 1 = SREFMTH1, 2 = SREFMTH2, 3 = SREFMTH3, 4 = SREFMTH4.

[b] 1 = TAGE1, 2 = TAGE2, 3 = TAGE3, 4 = TAGE4.

[c] 1 = PTOTINC1, 2 = PTOTINC2, 3 = PTOTINC3, 4 = PTOTINC4.

SIPP panels prior to the 1996 Panel were all "within wave" procedures. This means that the edits and imputations applied to a person's records in one wave were independent of those in other waves. Imputation procedures for most of the core wave files from the 1996 Panel are different. The new procedures do make use of information from the preceding wave. When linking data across waves, apparent changes in income, program participation, labor force behavior, or most other outcomes could be due to real changes reported by the respondent, or they could be an artifact of the data editing and imputation performed by the Census Bureau. Although this problem arises primarily with the core wave files from panels prior to 1996, it is also true of the 1996 Panel.[9]

[9] The new imputation procedures for the 1996 Panel are expected to introduce less error than procedures used for earlier panels. Thus, the number and magnitude of spurious changes (as well as falsely imputed stability) should be reduced. Even so, imputation errors will occur, and caution is advised when using the core wave files for longitudinal research.

There are two ways to identify cases with edited or imputed data. In panels prior to 1996, the entire record was imputed if (1) MIS5 = 2 and MISj = 1 for j = 1, 2, 3, or 4 or (2) INTVW = 3 or 4. The record was imputed in the 1996 Panel if EPPINTVW = 3 or 4. In the 1996 Panel, persons with Type Z noninterviews with prior wave information have their items imputed with procedures that use their prior wave responses. The relatively few cases with no prior wave information (those in Wave 1 and those in Waves 2–12 who are new to the sample) have their records imputed with the Type Z procedure used in the pre-1996 files. For all panels, if the record was not imputed, it is necessary to check the allocation (imputation) flags associated with the variables of interest. Once identified, users might need to implement some form of longitudinal editing and imputation or distinguish in their analyses between "real" changes and those that may result from the core wave data processing procedures.

Basic demographic information, such as age, race, and sex, can also appear to change from one wave to the next. In these instances, changes reflect corrections made in later interviews to information collected in earlier interviews; it is generally safe to assume the most recent data are correct.

When using the core wave files for longitudinal research, analysts should also note that the sample weights included on the core wave files are calendar month specific. These weights may not be appropriate for the planned longitudinal analyses. Chapter 8 has a detailed discussion of how to use the sample weights provided with the SIPP files.

Linking Core Wave Files to Longitudinal Research Files

There are relatively few circumstances in which the core wave and full panels files need to be linked because, for the most part, they contain the same information.[10] In general, if the same information is available from both the core wave and longitudinal research files, the information from the longitudinal research files is preferable because the edit and imputation procedures used for the longitudinal research files are believed to introduce less error than the procedures used for the core wave files.[11] However, some core information is contained only on the core wave files, and, therefore, at times it will be necessary to merge the core wave and longitudinal research files.

The following steps are necessary to link data from the core wave files with data from the full panel files:

1. Create data extracts from the core wave and longitudinal research files;
2. Put the two extracts into the same format (either person-month format or person-record format);

[10] Because the 1996 longitudinal research file is not complete yet, the discussion in this section pertains only to files for earlier panels. A revised version of this chapter will be available on the Census Bureau SIPP Web site (http://www.sipp.census.gov/sipp/) when the 1996 longitudinal research file is completed.

[11] See footnote 1.

3. Sort the extracts into the same order; and
4. Merge the extracts, creating the final file.

The variables that uniquely identify people in the core wave and longitudinal research files have different names. Table 13-3 shows the names for the three variables needed to match people across those files for panels prior to 1996.[12]

Table 13-3. Variables Identifying People in the Core Wave and Longitudinal Research Files for Panels Prior to 1996

Variable	Core Wave Files		Longitudinal Research Files
Sample Unit ID	SUID	is matched to	PP-ID
Entry Address ID	ENTRY	is matched to	PP-ENTRY
Person Number	PNUM	is matched to	PP-PNUM

If the final file will be in person-record format, these are the only variables needed for the sort and merge operations (steps 3 and 4, above). If the final file will be in person-month format, then WAVE and REFMTH are also needed.

Figure 13-2 shows the SAS code to transform data from the longitudinal research files in wide-record format into the person-month format used in the core wave files. The program creates a person-month format file from the 1993 longitudinal research file.

Because SAS does not allow variable names with embedded dashes, the "-" characters in the variable names have been replaced with underscore ("_") characters. The 1993 Panel had 10 waves, so the output file will have up to 40 monthly records for each person: no records are written for any months when pp_mis is not equal to 1. The program creates a data set with seven variables: SUID (renamed from PP_ID), ENTRY (renamed from PP_ENTRY), PNUM (renamed from PP_PNUM), REFMTH (which ranges from 1 to 4), WAVE (which ranges from 1 to 10), AGE, and TOTINC.

The REFMTH variable is computed as modulus (i/4) if it is not equal to 0, or 4 if is equal to 0. The modulus is the remainder from the division, so in month six of the panel the quantity is modulus (6/4) = 2, in month seven it is modulus (7/4) = 3, and in month eight it is 4 (since the remainder from the division of 8 by 4 is 0).

The wave is computed as the first integer greater than or equal to i/4. For month one, i/4 = 0.25, so wave = 1. For month four, i/4 = 1, so wave = 1. For month 17, 17/4 = 4.25, so wave = 5.

The file created by the program in Figure 13-2 could be merged with an extract from the core wave files from the 1993 Panel, using SUID, ENTRY, PNUM, WAVE, and REFMTH as the match keys. If the longitudinal research file was in its original sort order, the file created by the program in Figure 13-2 will already be sorted by this set of match keys.

[12] Current plans call for using consistent variable names across all files from the 1996 Panel.

Figure 13-2. Sample SAS Code to Change the Longitudinal Research Files from Person-Record Format to Person-Month Format for Panels Prior to 1996

```
Data pmonth
   (keep =
      pp_id
      pp_entry
      pp_pnum
      refmth
      wave
      age
      totinc
    rename =
      (pp_id = suid
       pp_entry = entry
       pp_pnum = pnum
      )
   );

   /*
      this example works with the 1993 SIPP panel - 10 waves
   */
   set sipp93fp
      (keep =
         pp_id
         pp_entry
         pp_pnum
         pp_mis1 - pp_mis40
         age1 - age40
         totinc1 - totinc40
      );

   /*
      define arrays to ease the programming burden
   */
   array ages {40} age1 - age40;
   array totincs {40} totinc1 - totinc40;
   array pp_mis {40} pp_mis1 - pp_mis40;

   do i = 1 to 40;                  /* for each month */
      if (pp_mis{i} eq 1) then do;  /* if pp_mis is 1, use the data */
         age = ages{i};             /* the age in this month */
         totinc = totincs{i};       /* total income this month */

         j = mod(i,4);
         if (j eq 0) then refmth = 4;/* the reference month */
         else refmth = j;

         wave = ceil(i/4);          /* the wave */
         output;                    /* write out the record */
      end;
   end;
run;
```

When text copy applies to both 1996 and pre-1996 panel files, pre-1996 variable names appear in parentheses following 1996 variable names.

Values for AGE and TOTINC from the core wave and longitudinal research files will not match for all people in all months because the core wave files and the longitudinal research files are subjected to different edit and imputation procedures.

In addition, beginning with the 1991 Panel, a missing wave imputation procedure has been applied to the longitudinal research files: people who had missing data from one wave but complete data from the two adjacent waves had data imputed for the missing wave in the longitudinal research files.[13] This means that some people will have data in the longitudinal research files for months in which they have no records in the associated core wave files (those who were not Type Z nonrespondents).

Linking Two or More Topical Module Files

At times it will be necessary to merge data from two or more topical module files. Any project that studies the relationship between subject areas covered by different topical modules will require such a merge. One example might be a study of the relationship between the use of health care services (collected in Wave 3 of the 1993 Panel) and medical expenses (collected in Wave 4 of the 1993 Panel).

The mechanical process of linking topical module files is relatively straightforward. The topical module files all have the same format (one record per person) and variable names, for the ID variables are consistent across the topical module files: individuals are uniquely identified by the combination of SSUID (ID), EENTAID (ENTRY), and EPPPNUM (PNUM).

However, a number of cautions should be noted:

1. Prior to the 1996 Panel, there were instances in which the same variable name was used in different topical module files for different variables. For example, in the 1990 Panel, TM8400 was used in the Wave 2 topical module for a variable that indicates whether the respondent completed 12th grade. The same variable name was used in the Wave 6 topical module to indicate whether the respondent was a parent of children under 21 years of age living in his or her household.
2. Not all people with records in one topical module file will have records in another topical module file. In the topical module files from the 1996 Panel, there will generally be a record for each person who was a responding SIPP household member in the fourth month of the wave's core reference period. Prior to the 1996 Panel, all household members in the interview month have topical module records for a given wave. However, household composition changes from one wave to the next: some people leave SIPP households and others join SIPP

[13] Many of these situations arise with Type Z nonrespondents: nonresponding people who live in households with other responding sample members. Type Z nonrespondents in the pre-1996 core wave files and those in the 1996 Panel files with no prior wave information were subjected to a whole-record imputation procedure, described in Chapter 10. These people would have records in the core wave files, but different information—because it was imputed using different procedures—in the longitudinal research files.

households, and this changing composition is reflected in the topical module files. Also, in the 1996 Panel, some people who were nonrespondents in month four of one wave may have been respondents in month four of another wave. Thus, when topical module files are merged, there will be a nontrivial number of nonmatches: people with data from only one of the topical modules. Nonmatches are addressed in greater detail later in this chapter.

3. Choosing appropriate weights is complicated by the fact that there are a substantial number of nonmatches across topical modules. One solution is to use one of the weights from the longitudinal research files. Chapter 8 gives a detailed discussion of the SIPP weights.

Often it will be necessary to merge additional information (such as sample weights) from the core wave or longitudinal research files when working with multiple topical modules.

Users interested in measuring change with data from the topical module files (such as changes in asset holdings, or changes in health or disability status) should proceed with caution. First, in some instances measurement error is large relative to the actual changes that have taken place. One example is found in the topical modules that measure levels of household assets and liabilities.[14] Although the topical modules can provide estimates of aggregate-level changes in those instances, users should not attempt to measure those changes at the individual level. Also, the edit and imputation procedures applied to the topical module files are all "within wave" procedures. This means that the edits and imputations applied to a person's records in one wave are independent of those in other waves. When data are linked across waves, apparent changes could be due to real changes reported by the respondent or they could be artifacts of the data editing and imputation performed by the Census Bureau.

There are two ways to identify cases with edited or imputed data. In panels prior to 1996, the entire record was imputed if (1) PP-MIS5 = 2 and PP-MISj = 1 for j = 1, 2, 3, or 4 or (2) INTVW = 3 or 4. In the 1996 Panel, the record was imputed if (1) EPPMIS4 = 2 or (2) EPPINTVW = 3 or 4. In the 1996 Panel, persons with Type Z noninterviews who have prior wave information have their records imputed with procedures that use their prior wave responses. For persons with no prior wave information (those in Wave 1 and those in Waves 2–12 who are new to the sample), the Type Z imputation procedure is used. On all panels, users should check the imputation flags associated with the variables of interest.

Linking Topical Module Files to Core Wave Files

Because the topical module files contain only limited information from the SIPP core, there will be many times when it is necessary to merge data from the topical module files with data from the SIPP core. One source of these data is the core wave files.[15]

[14] See the *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a) and SIPP Working Paper series for discussions of this issue as it relates to this and other SIPP topical modules.

[15] The next section describes procedures for merging topical module files with data from the longitudinal research files.

The first decision that must be made is which core wave file to use. Special attention should be paid to the reference periods for the topical module items of interest. In the 1996 Panel, topical module questions refer to either month four of the wave's core reference period, or to a longer period in the past (such as the preceding 12 months or the prior calendar year). In those instances, information would come from the month-four records of the core wave files from the same wave (and possibly from earlier months and waves). Prior to the 1996 Panel, many topical module items referred to conditions in the interview month. The interview month, however, is not included as a separate record in the core wave file for the same wave as the topical module.[16] Rather, core information for the interview month of one wave is found in the month-one information from the following wave. For example, the interview month for Wave 3 is month 13 in the SIPP panel, and core data for month 13 are collected as the first reference month of Wave 4.[17] Commonly used reference periods for topical module items are the current (interview) month (month one of the *next* wave), the previous month (month four of the current wave), the previous 4 months (the full reference period for the current wave), and the previous year.

The topical module files have one record per person, while the core wave files have up to four records for each person (one record per person for each month the person was a SIPP sample member). There are at least three options available when merging topical modules with data from the SIPP core wave files:[18]

1. Pick a single month from the core wave files. For example, if the topical module items use the interview month as their reference period, it may make sense to use records for month one from the core wave files from the *next* wave.
2. Spread the topical module data across all records from the core wave file. That results in a final file in person-month format.
3. Create a single record for each person from the appropriate core wave file and merge the topical module data to that record. This results in a final file in the person-record format with the same monthly detail as in the second option described above.

The steps involved are as follows:

1. Create an extract from the core wave file(s) of interest.
2. If a single record for each person is desired, apply the algorithm in Figure 13-1, which is described in the section entitled Linking Within a Core Wave File—Transforming the Person-Month Format into the Person-Record Format.

[16] Some of the interview month information is contained on the records for the four reference months of the wave. But in the person-month-format file there is no separate record for the interview month itself.

[17] Information collected during the interview month of one wave may not match the information collected about the same calendar month in the subsequent wave. In the 1996 Panel, dependent interviewing techniques and other checks made possible with CAI are used to help resolve those inconsistencies.

[18] Yet another option is to create a single record from the core wave files containing aggregate measures for the reference period of interest. For example, it might make sense to create a single record from the "current" core wave file with total income received during all 4 months of the wave's reference period. Or the average number of hours worked per week during the previous 4 months might be appropriate. Once the aggregate record is created, the merge step is similar to the others described in this section.

3. Sort the core wave extract using SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) as the sort keys. These three variables uniquely identify people in the core wave files. If the core wave extract is in the person-month format, include SREFMON (REFMTH) as the final sort key.
4. Create an extract from the topical module file of interest. Sort the topical module extract using SSUID (ID), EENTAID (ENTRY), and EPPPNUM (PNUM) as the sort keys.
5. For the 1996 Panel, merge the core wave extract with the topical module extract; use SSUID, ENTAID, and EPPPNUM as the sort keys. For panels prior to 1996, merge the core wave extract with the topical module extract; use the sort keys shown in Table 13-4.

Table 13-4. Variables Identifying People in the Topical Module and Core Wave Files for Panels Prior to 1996

Variable	Topical Module Files		Core Wave Files
Sample Unit ID	ID	is matched to	SUID
Entry Address ID	ENTRY	is matched to	ENTRY
Person Number	PNUM	is matched to	PNUM

When data from panels prior to 1996 are used, there will likely be a nontrivial number of nonmatches between the core wave files and the topical module files. That will be true even when a topical module is merged with core data from the same wave, because people who were members of a SIPP household in the interview month but not during the previous 4 months will have records in the topical module files but not in the core wave files.

Linking Topical Module Files to Longitudinal Research Files from Pre-1996 Panels

While topical module files can be linked with data from the core wave files, there are many times when it will be necessary or desirable to use the longitudinal research files instead.[19] For example, if the full panel weights[20] are needed for the planned analysis, they must come from the longitudinal research files. When the same core items are available from the core wave and the longitudinal research files, analysts may prefer to use the longitudinal research files because the edit and imputation procedures used for them are believed to introduce less error than the procedures used for the core wave files.

[19] Because the full panel longitudinal research file for the 1996 SIPP was still under development at the time this chapter was written, it is not yet possible to describe procedures for using that file. A revised version of this chapter will be available once the longitudinal research file for the 1996 Panel is released to the public.

[20] Chapter 8 discusses the SIPP weights, their derivation, and use.

The steps involved are as follows:

1. Create an extract from the longitudinal research file.
2. If a file in the person-month format is desired, apply the algorithm described in the section above, Linking Core Wave Files to Longitudinal Research Files. The example in Figure 13-2 can be adapted to that purpose, but the ID variables would need to be renamed to match those used in the topical module files rather than in the core wave files (Table 13-5).
3. Sort the full panel extract; use PP-ID, PP-ENTRY, and PP-PNUM as the sort keys. These three variables uniquely identify people in the longitudinal research files. If the full panel extract is in the person-month format, include WAVE and REFMTH as the final sort keys.
4. Create an extract from the topical module file of interest. Sort the extract; use ID (the variable name for the sample unit ID in the topical module files), ENTRY, and PNUM as the sort keys.
5. Merge the core wave extract with the topical module extract based on the sort keys described here and shown in Table 13-5.

Table 13-5. Variables Identifying People in the Topical Module and Longitudinal Research Files Prior to the 1996 Panel

Variable	Topical Module Files		Longitudinal Research Files
Sample Unit ID	ID	is matched to	PP-ID
Entry Address ID	ENTRY	is matched to	PP-ENTRY
Person Number	PNUM	is matched to	PP-PNUM

Because the longitudinal research files contain a record for every person who was ever a member of a SIPP household, every person with a record in a topical module file should have a record in the longitudinal research file. However, analysts working with a person-month-format file containing records only for months when PP-MIS = 1 may find nonmatches.

Nonmatches When Merging Files

SIPP is designed to follow a group of people over an extended period of time. This group includes only those who were interviewed in the first wave of the panel and the children subsequently born to or adopted by them.[21] Over the course of the panel, these original sample members are followed and interviewed every 4 months. Secondary sample members, on the

[21] In the 1993 Panel all original sample members were followed no matter what their ages. In all other panels, only original sample members aged 15 years or older are followed when they move to new addresses. In all cases, however, the SIPP data files contain a record for all people, including children, who reside in a household with at least one original panel member present.

other hand, are part of the SIPP sample only for as long as they continue to reside with at least one original sample member. As long as they are part of the SIPP sample, the secondary sample members are interviewed and included in the SIPP data files.

The problem of nonmatches occurs only when users merge across waves for any types of files. There is no matching problem when the same or different types of files are merged within the same wave.

As shown in Table 13-6, there are a variety of reasons why a person may be in one SIPP data file but not in another. All but one of the reasons are associated with people entering and leaving the SIPP sample:[22]

1. The original sample person may have left the SIPP sample universe (e.g., died, moved abroad, moved into military barracks, or moved into an institution);
2. The original sample person may have left the sample but is still in the sample universe (sample attrition);
3. The original sample person may have just reentered the SIPP sample universe (after living abroad, etc.);
4. The person is a newborn (a special case of a person joining the sample universe);
5. The secondary sample member has just begun living with an original sample person;
6. The secondary sample member no longer lives with an original sample member;
7. The person had data for a "missing wave" imputed in the longitudinal research file and has no records in the core wave or topical module files for that wave; and
8. Prior to the 1996 Panel, the Census Bureau may have intentionally altered the identification information of the person, thereby making it difficult to find a match for this person (in rare situations referred to as *merged households*).

A person's reason for leaving the SIPP sample is identified in the core wave and longitudinal research files. In the former, the variable name is ULFTMAIN (REALFT). In the longitudinal research files, the name is REASLEFT, and it has a value for each wave rather than each month. Figure 13-3 shows the variable values and corresponding descriptions.

Procedures for dealing with nonmatches vary, depending largely on the reasons the person entered or left the SIPP sample. A number of common scenarios are presented below.

[22] The SIPP following rules are described in greater detail in Chapter 2.

Table 13-6. Reasons for Nonmatches

Reasons	File #1 (earlier time period)	File #2 (later time period)
People Exiting the Sample		
Original sample people left the SIPP sample universe (left the population of inference) Person died Moved abroad—left sample universe Moved into military barracks—left sample universe Moved into an institution—left sample universe	Present	Not present
Original sample person exited from the sample (still in the sample universe but no longer in the sample) Refused to be interviewed	Present	Not present
Secondary sample person no longer lives with an original sample member	Present	Not present
People Entering the Sample		
Newborn	Not present	Present
Original sample person returns to SIPP sample universe (returns to the population of inference) Moved from abroad—entered sample universe Moved from military barracks—entered sample universe Moved from an institution—entered sample universe	Not present	Present
Original sample member returns to sample Original sample member agrees to be interviewed and returns to sample	Not present	Present
Secondary sample person now lives with an original sample member	Not present	Present
Missing Wave Imputation in the Longitudinal Research File (Beginning with the 1991 Panel)		
Person has data in the longitudinal research file but no data in the corresponding wave in the core wave or topical module files.		
Merged Households—Special Case		
"Old" version of the ID information	Present	Not present
"New" version of the ID information	Not present	Present

Exiting or Entering the Population

There is a fundamental distinction between situations in which people leave the sample because they leave the SIPP sample universe and situations in which they leave the sample despite the fact that they are still part of that population. The SIPP sample universe (the population that the SIPP sample represents) is the noninstitutionalized, resident population of the United States. It includes both civilian and military people; it includes adults and children who reside in the United States and outside of institutions.

People who leave this population because they die, move abroad, or move into institutions exit the SIPP sample because they are no longer a part of the population that SIPP represents. *In general, when nonmatches occur because people have entered or exited the population represented by the SIPP sample, data should not be imputed and weights should not be adjusted for the period when these people are outside of that population.* From the perspective of SIPP, these people do not exist when they are outside of the population represented by the sample.

Figure 13-3. Data Dictionary Entries for Variables Identifying the Reason a Person Left the SIPP Sample

Wave 2, 1996 Panel Core Wave File

```
D ULFTMAIN    2    606
T PE: UNEDITED VARIABLE - Main reason left Household
     What is the main reason ... left the household?
U Movers from households which contain sample persons at the time of
   interview, movers from a household which splits into multiple
   households. Note: This is an unedited field and the universe is not
   exact.<BR>
V          0 .Not answered
V          1 .Deceased
V          2 .Institutionalized
V          3 .On active duty in the Armed Forces
V          4 .Moved outside of U.S.
V          5 .Separation or divorce
V          6 .Marriage
V          7 .Became employed/unemployed
V          8 .Due to job change - other
V          9 .Listed in error in prior wave
V         10 .Other
V         11 .Moved to type C household
```

1993 Full Panel Files

```
D REASLEFT    9    143  9   1
     Range = (0:9)
     Preedited reason for leaving the Household Control Card item 23
U Persons who left at any time during the reference period
  Subscript 1: not applicable for Observation 1
  Subscript 2 - 8: reason left in Observations 2 - 8
V   0 .Not applicable or not answered or nonmatch
V   1 .Left - deceased
V   2 .Left - institutionalized
V   3 .Left - living in armed forces barracks
V   4 .Left - moved outside of country
V   5 .Left - separation or divorce
V   6 .Left - person #201 or greater no longer living with sample person
V   7 .Left - other
V   8 .Entered merged household
V   9 .Interviewed in previous wave but not in sample
```

(figure continues)

Figure 13-3. Data Dictionary Entries for Variables Identifying the Reason a Person Left the SIPP Sample (*continued*)

1993 Core Wave Files

```
D REALFT      2    521
     Reason for leaving the household
     Applicable when previous wave address ID is not equal to control
     card address ID
     Range=(00:00,05:12,25:31,99:99)
U All persons, including children, no longer in the household
V   00 .Not applicable or not answered
V   05 .Left – deceased
V   06 .Left – institutionalized
V   07 .Left – living in Armed Forces barracks
V   08 .Left – moved outside of country
V   09 .Left – separation or divorce
V   10 .Left – person #201+  no longer living with sample person
V   11 .Left – other
V   12 .Left – entered merged household
* Should have been deleted in a previous wave:
V   25 .Left – deceased
V   26 .Left – institutionalized
V   27 .Left – living in Armed Forces barracks
V   28 .Left – moved outside of country
V   29 .Left – separation or divorce
V   30 .Left - 201+ person no longer living with sample person
V   31 .Left – other
V   99 .Listed in error
```

The following examples help explain why weighting adjustments and imputation are problematic in these situations:

- *A person is in the SIPP sample at Time 1 but dies before Time 2.* In this case, the person is not part of the population at Time 2. In computing the aggregate (total) income of the population at Time 1, this person's income would be included. To impute income to this person for the Time 2 observation, analysts would compute an aggregate income that is too high: The person had no income at Time 2, and so none should be imputed.[23] If this case is dropped from the analysis file and the weights are inflated for the remaining sample, the estimate of the total population at Time 2 would be too high. Because this person was not a part of the population at Time 2, the weights for the remaining sample members should not be inflated to represent this individual.

[23] If the person had been alive with income that she or he did not report to the Census Bureau, an estimate of his or her unreported income would be imputed to the individual. Failing to impute that unreported income would mean that the income received by a member of the population is not represented anywhere in the sample. That value would result in a sample estimate of aggregate income in the population that was lower than the actual value in the population.

- *A person is overseas at Time 1 but at Time 2 is living with an original sample member in the United States.* At Time 1, this person was not part of the population represented by the SIPP sample. Because this person was not a part of that population, the SIPP sample should not be adjusted in any way to represent this individual.

A number of strategies are possible for dealing with cases in which nonmatches result from people entering or leaving the population represented by the SIPP sample. One approach is to drop those people from the analysis sample entirely. No adjustment would be made to the weights of the remaining cases. However, the definition of the population represented by the remaining sample would change. The remaining sample represents the population that existed at both Time 1 and Time 2. It does not represent anyone who either entered or left the population.

That approach has the advantage of being simple to implement. It also results in a clearly defined population of inference. Caution is necessary, however, to the extent that people entering and leaving the population are systematically different from those who are present throughout the period being studied: the remaining sample *cannot* be used to draw inferences about this other part of the population. People entering and leaving prisons and nursing homes, for example, likely have very different income profiles than the population that remains outside of these institutions over the period under study.

If event-history models are used to analyze the data, another approach is possible.[24] With these models, exits from the population can be treated as competing outcomes. For example, in a study of unemployment dynamics, a competing risks model might allow for three possible outcomes: spells of unemployment can end because (1) a person becomes employed, (2) a person exits the labor force, or (3) a person exits the population.[25]

Exiting the Sample but Remaining in the Population (Sample Attrition)

Sample attrition occurs when people leave the SIPP sample but remain a part of the population represented by that sample. In these instances the remaining sample generally should be adjusted to represent the full population, including the part of the population represented by those who leave the sample.

There are several options for handling such cases:

- *Impute the missing data and proceed.* This option is appropriate for researchers familiar with the statistical literature on imputation for missing data. A full discussion of this topic is well beyond the scope of this manual. Analysts are cautioned, however, against using the common practice of "substituting the mean" for missing data. That practice can yield biased estimates

[24] For a description of these methods, see, for example, Tuma and Hannan (1984).

[25] In actual applications, more than three outcomes would likely be modeled. The determinants of entering a nursing home, for example, are likely quite different from the determinants of entering a prison.

of multivariate statistics (such as regression coefficients) and generally leads to downward-biased estimates of standard errors.

- *Drop cases with missing data, adjust (poststratify) the weights for the retained cases, and proceed.* This poststratification involves several steps.

  1. Tabulate the weighted number of cases by various socioeconomic categories before dropping any cases.
  2. Repeat the tabulation after dropping the nonmatches.
  3. Compute adjustment factors by dividing the weighted numbers from step 1 (before dropping any cases) by the weighted numbers from step 2 (after dropping cases).
  4. Create a new weight variable by multiplying the original weight variable by the appropriate poststratification factor computed in step 3.

This situation requires caution. A user who drops records may introduce selection biases because those in the retained sample may be more stable than those who leave. For example, the fact that a (former) sample member has left may be associated with other changes in that person's life, such as giving birth, getting married, or getting a new job. Because the person left the sample, it is not possible to know from the available data what changes actually did occur in each case. Also, when records are dropped, the procedures for computing standard errors as described in the source and accuracy statements provided with the data will no longer apply. The procedures described in Chapter 7 for the direct estimation of standard errors should, however, work without any modification. If the number of cases lacking complete information is small relative to the full analysis sample (the full sample with positive weights), the biases introduced by dropping those cases also are likely to be small and this procedure may be a viable alternative.

- *If the longitudinal research file is available, use a subset of the cases with complete data for which Census Bureau–provided weights are available and proceed.* At the extreme, this procedure entails retaining only cases with positive full panel weights and using those weights for any analyses performed.[26] This is a conservative approach, but one that is relatively easy to implement because the weights already exist, they have already been adjusted for the observed sample attrition, and the population of inference is clearly defined.

- *Use other missing data methods to provide estimates and their standard errors.* A full discussion of these methods is beyond the scope of this manual. The methods are designed to make use of all available information from the cases with complete data without (directly) imputing data to cases with incomplete information. Interested users can consult the literature on the E-M algorithm for one example of how this can be done.[27] Also, Skinner et al. (1989) discuss model-based approaches to the analysis of complex surveys with missing data.

[26] The calendar year weights on the longitudinal research files are also options worth exploring. Chapter 8 provides a detailed discussion of the SIPP sample weights, their derivation, and use.

[27] For example, see Little and Rubin (1987). Users should also note that some statistical packages (e.g., SPSS) have incorporated more sophisticated options for handling missing data than have generally been available in the past.

Missing Wave Imputation in the Longitudinal Research Files Prior to 1996

Beginning with the 1991 Panel, a missing wave imputation procedure has been applied to the longitudinal research files: persons who had missing data from one wave but complete data from the two adjacent waves had data imputed for the missing wave in the longitudinal research files.[28] Some of those cases are Type Z nonrespondents and will have records with different data in the core wave files.[29] Other people will have data in the longitudinal research files for months when they have no records in the associated core wave or topical module files.

The correct procedure for dealing with the resulting nonmatches depends on which weight variables will be used. If the weights are coming from the core wave or topical module files, observations from the longitudinal research files not present in the cross-sectional files should be dropped. That is because the weights on the core wave and topical module files are computed for the samples in those files, samples that do not include the people who have had that wave imputed in the longitudinal research files.

If the weights are coming from the longitudinal research file, then other procedures must be used to deal with the missing data from the core wave and topical module files. In those instances, the procedures described for dealing with sample attrition should be considered.

Merged Households in Panels Prior to 1996

Finally, nonmatches can occur when the Census Bureau changes the ID numbers for sample members.[30] Prior to the 1996 Panel, there were two very rare occasions when this happened. The first occurred when two separate sampling units, each containing original sample members, were merged together, perhaps because of a marriage. In this situation, the people in one of the sampling units retained their identification information, while the people in the other sampling unit had their identification information changed to agree with the retained set. The person numbers of the changed set were modified to be between 180 and 199.

The second instance occurred when a SIPP household split into two new households (in which each new household gained a new sample person), which later recombined. For example, a

[28] Imputed waves can be identified on the longitudinal research files by using the WAVFLG variable.

[29] The data are different because different imputation procedures are used.

[30] Because the Census Bureau is using new procedures in the 1996 Panel, merged households will not be an identifiable source of nonmatches when files from the 1996 Panel are merged. Rather, they will appear no different from other situations where people enter and leave the SIPP sample, such as through marriages, divorces, deaths, and sample attrition. For example, in the 1996 Panel, there will be no way to identify which (if any) of the people who appear to have entered the sample in Wave 3 were also sample members who appear to have left the sample following Wave 2. The "new" sample members will be given person numbers in the same range as others who enter the sample in Wave 3, and no previous wave information will be attached to them. The new procedures greatly simplify the handling of these rare cases for both the Census Bureau and outside data users.

married couple separated in Wave 3, each moving in with a sibling. Both siblings were assigned a person number of 301, because they entered the sample in Wave 3 at different addresses. If the husband and wife reunited in Wave 6, bringing the siblings with them, one sibling's person number was changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699 because the households recombined in Wave 6).

Different file types (i.e., core wave, topical, and full panel) keep track of the changed ID values differently. If the move occurred after the first month of a reference period, the core wave file contains two records for the person whose identification information changed. The first record contains the original identification information of the person before the move and identifies the person as having exited the sample at the time of the move. The second record contains the new identification information after the move and identifies the person as having entered the sample at the time of the move. When the move occurs at the start of a reference period, only the second record is retained in the core wave file. The topical module file, however, contains only the second record, no matter when the move took place. The longitudinal research file contains both records, no matter when the move took place.

The easiest way to find these people is to search the core wave file for people with a previous wave identified as present, that is, PWSUID > 0 or PWENTRY > 0 or PWPNUM > 0. Users then need to decide how they want to handle these special cases. There are several possibilities:

- Change the identification information used in the waves before the move to the new values seen in the wave(s) after the move, and then merge the records using these ID values. This option is useful when working primarily with the person's core wave data after the move.
- Change the identification information in the waves after the move to the original values, and then use those ID values to merge records. This option is useful when working primarily with the person's core wave data before the move.
- Duplicate the person's record, and use the initial identification information with one record and the new identification information with the other record; then merge those records. With this approach, the weights for the duplicated records will need to be adjusted so that the duplicated weights sum to the original (unduplicated) weights.
- Treat this person as two people: once as someone who exits the sample at the time of the move and once as someone who enters the sample at the time of the move. That is how these cases are treated in the longitudinal research files. The weighting implications of this approach depend on the planned analysis.

Appendixes

A. *SIPP Users' Guide* Variable Crosswalk: 1993 to 1996

This appendix contains four sections showing the correspondences between the core wave file variables in 1993 and those in 1996. The sections differ by order as follows:

1. By 1993 Variable Name
2. By 1996 Variable Name
3. By 1993 File Position
4. By 1996 File Position

Ordered by 1993 Variable Name	
1993	**1996**
ADDID	SHHADID
AFDC	RCUTYP20
AFDCPNUM	RCUOWN20
AFDPCT	n/a
AFDSAB	n/a
AFTIME	n/a
AGE	TAGE
BFFREE	EFRERDBK
BFTOT	n/a
BREAKF	EBRKFST
BRTHMN	EBMNTH
BRTHYR	TBYEAR
CAIDCOV	RCUTYP57
CARECOV	ECRMTH
CHAMP	RCHAMPM
CHPNUM	n/a
CJ10003	ASVJTINT
CJ10407	AMDJTINT
CO10003	ASVOINT
CO10407	AMDOINT
CWORK	ER55
DAYENT	n/a
DAYLFT	n/a
DESGPNPT	RDESGPNT
DISAB	EDISABL
DISAGE	TAGESS
EARN	TPEARN
EASTAMT	EEGYAMT
EDASST	EEDFUND
EMPLED	n/a
EMPLYR	EASST10
ENROLD	RENROLL, EENRLM, RENRLMA
ENTRY	EENTAID
ESR	RMESR
ETHNCTY	EORIGIN
EWID	UEVRWID
FAFDC	TFAFDC
FAMREL	ERRP
FAMTYP	ESFT
FCHANGE	RFCHANGE
FEARN	TFEARN
FFDSTP	TFFDSTP
FID	RFID
FID2	RFID2

Ordered by 1993 Variable Name	
1993	**1996**
FKIND	EFKIND
FKPNUM	RCUOWN23
FNKIDS	RFNKIDS
FNLWGT	WPFINWGT
FNP	EFNP
FNSSR	RFNSSR
FOKLT18	RFOKLT18
FOODSTMP	RCUTYP27
FOSTKID	RCUTYP23
FOTHER	TFOTHINC
FOWNKID	RFOWNKID
FPOV	TFPOV
FPROP	THPRPINC
FREFPER	EFREFPER
FSOCSEC	TFSOCSEC
FSPNUM	RCUOWN27
FSPOUSE	EFSPOUSE
FSSHIP	EASST06, EASST08, EASST09
FSSI	TFSSI
FTOTINC	TFTOTINC
FTRAN	TFTRNINC
FTYPE	EFTYPE
FUNEMP	TFUNEMP
FVETS	TFVETS
FWGT	WFFINWGT
GAPNUM	RCUOW21A
GENASST	RCUTYP21
GIBILL	ER40
GRDCMPL	n/a
H5ADDID	n/a
H5MIS	EOUTCOME
H5NP	EHHNUMPP
H5REF	EHREFPER
H5WGT	WHFNWGT
HACCESS	EACCESS
HAFDC	THAFDC
HCASH	RHCBRF
HCHANGE	RHCHANGE
HEARN	THEARN
HENRGY	EEGYPMT1, EEGYPMT2, EEGYPMT3
HFDSTP	THFDSTP
HHSC	GHLFSAM
HIFAM	n/a
HIGRADE	EEDUCATE

Ordered by 1993 Variable Name	
1993	1996
HIIND	RCUTYP58
HINONH	EHIOWNER
HIOWN	EHIOWNER
HIPAY	EHICOST
HIPNUM	RCUOW58A, RCUOW58B
HISRC	EHEMPLY
HITM36B	n/a
HITYPE	EHIOWNER
HLORNT	EGVTRNT
HLVQTR	ELIVQRT
HMEANS	RHMTRF
HMETRO	TMETRO
HMSA	TMSA
HNCASH	RHNBRF
HNF	RHNF
HNFAM	RHNFAM
HNONCSH	THNONCSH
HNP	EHHNUMPP
HNSF	RHNSF
HNSSR	RHNSSR
HOTHER	THOTHINC
HPOV	THPOV
HPROP	THPRPINC
HPUBHS	EPUBHSE
HREFPER	EHREFPER
HSOCSEC	THSOCSEC
HSSI	THSSI
HSTATE	TFIPSST
HSTRAT	GVARSTR
HTENURE	ETENURE
HTOTINC	THTOTINC
HTRAN	THTRNINC
HTYPE	RHTYPE
HUNEMP	THUNEMP
HUNITS	EUNITS
HVETS	THVETS
HWGT	WHFNWGT
IBFFREE	AFRERDBK
IBFTOT	n/a
IBREAKF	ABRKFST
ICAIDCOV	n/a
ICARECOV	ACRMTH
ICWORK	AR55
IDISAB	ADISABL

Ordered by 1993 Variable Name	
1993	1996
IDISAGE	AAGESS
IEASTAMT	AEGYAMT
IEDASST	AEDFUND
IEMPLYR	AEDASST
IENROLD	ARENROLL, AENRLM, EENLEVEL
IETHNCTY	AORIGIN
IEWID	n/a
IFSSHIP	AEDASST
IGIBILL	AR40
IGRDCMPL	n/a
IHENRGY	AEGYPMT
IHIGRADE	AEDUCATE
IHIIND	n/a
IHIOWN	AHIOWNER
IHIPAY	AHICOST
IHISRC	AHEMPLY
IHITYPE	AHIOWNER
IINAF	AAFNOW
IJ10003	ASVJTINT
IJ10407	AMDJTINT
IJ110	ASJNTDIV
IJ110RI	AMJADIV
IJ120OT	AJACLR2
IJ130	AMIJNT
IJGRENT	AJARNT
IJNRENT	AJACLR
IJO110	AMOWNDIV
IJO110RI	AMOTHDIV
ILCHCOST	n/a
ILCHFREE	AFRERDLN
ILCHPT	AFREELUN
ILCHTOT	n/a
ILEVEL	AENLEVEL
ILUNCH	AHOTLUNC
IMCOPT	n/a
INAF	EAFNOW
INDSL	AEDASST
INKIDSBF	n/a
INKIDSHL	n/a
INONHHI	AHIOTHER
INTVW	EPPINTVW
IO10003	ASVOINT
IO10407	AMDOINT
IO110	ASOWNDIV

Ordered by 1993 Variable Name	
1993	**1996**
IO110RI	AMOWNADV
IO130	AMIOWN
IO14050	ARNDUP1
IOGRENT	AOARNT
IONRENT	AOACLR
IOTHAID	AEDASST
IOTHVET	AEDASST
IPELL	AEDASST
IPHRENT	AGVTRNT
IPLUS	AEDASST
IR01A	AR01A
IR01K	AR01K
IR02A	AR02
IR03	AR03A, AR03K
IR05	AR05
IR06	AR06
IR07	AR07
IR08	AR08
IR10	AR10
IR100	AAST2B
IR101	AAST2C
IR102	AAST2D
IR103	AAST2A
IR104	AMDJT, AMDOAST
IR105	AAST3D
IR106	AAST3C
IR107	AAST4C
IR110	AMANYCHK
IR12	AR12
IR120	AAST4A
IR13	AR13
IR130	AAST3E
IR140	AAST4B
IR150	EOTHPROP
IR20	AR20
IR21	AR21
IR23	AR23
IR24	AR24
IR25	AR25
IR27	AR27
IR28	AR28
IR29	AR29
IR30	AR30
IR31	AR31

Ordered by 1993 Variable Name	
1993	**1996**
IR32	AR32
IR34	AR34
IR35	AR35
IR36	AR36
IR37	AR37
IR38	AR38
IR40	AR40
IR41	AR41
IR50	AR50
IR51	AR51
IR52	AR52
IR53	AR53
IR54	AR54
IR55	AR55
IR56	AR56
IRACE	ARACE
IREASAB	AABRE
IRETIRD	AEVERET
IRHCDIS	n/a
IRJ10003	ASVJT
IRJ10407	n/a
IRJ120	AJNTRNT
IRJ120OT	AJRNT2
IRJ130	AMRTJNT
IRO10003	ASVOAST
IRO10407	n/a
IRO120	AOWNRNT
IRO130	AMRTOWN
IS01A	A01AMTA
IS01K	A01AMTK
IS02A	A02AMT
IS02K	n/a
IS03	A03AMTA, A03AMTK
IS05	A05AMT
IS06	A06AMT
IS07	A07AMT
IS08	A08AMT
IS10	A10AMT
IS12	A12AMT
IS13	A13AMT
IS20	A20AMT
IS21	A21AMT
IS23	A23AMT
IS24	A24AMT

Ordered by 1993 Variable Name	
1993	1996
IS27	A27AMT
IS28	A28AMT
IS29	A29AMT
IS30	A30AMT
IS31	A31AMT
IS32	A32AMT
IS34	A34AMT
IS35	A35AMT
IS36	A36AMT
IS37	A37AMT
IS38	A38AMT
IS40	A40AMT
IS41	n/a
IS50	A50AMT
IS51	A51AMT
IS52	A52AMT
IS53	A53AMT
IS54	A54AMT
IS55	A55AMT
IS56	A56AMT
IS75	A75AMT
ISE12214	AGROSB1
ISE12218	AEMPB1
ISE12220	AINCPB1
ISE12222	APROPB1
ISE12232	ASLRYB1
ISE12234	AOINCB1
ISE12254	APRFTB1
ISE12256	APRFTB1
ISE12260	ABMSUM1
ISE1AMT	ABMSUM1
ISE1IND	ABSIND1
ISE1OCC	ABSOCC1
ISE22314	AGROSB2
ISE22318	AEMPB2
ISE22320	AINCPB2
ISE22322	APROPB2
ISE22332	ASLRYB2
ISE22334	AOINCB2
ISE22354	APRFTB2
ISE22356	APRFTB2
ISE22360	ABMSUM2
ISE2AMT	ABMSUM2
ISE2IND	ABSIND2

Ordered by 1993 Variable Name	
1993	1996
ISE2OCC	ABSOCC2
ISEX	ASEX
ISPDAF	AAFSRVDI
ISPINAF	n/a
ISTLOAN	AEDASST
ISUPPED	AEDASST
ITAKJOB	n/a
ITAKJOBN	n/a
IUHOURS	AJBHRS1
IUTILS	AUTILYN
IVETSTAT	AAFEVER
IVETTYP	AVETTYP
IWKSJOB	n/a
IWKSLOK	AWKLKG
IWKSPT	APTWRK
IWKSPTR	APTRESN
IWKSTDY	AEDASST
IWKSWOP	AWKSAB
IWS12012	ACLWRK1
IWS12024	ARSEND1
IWS12026	APAYHR1
IWS12028	APYRATE1
IWS12029	n/a
IWS12030	n/a
IWS12031	n/a
IWS12044	AUNION1
IWS12046	ACNTRC1
IWS1IND	AJBIND1
IWS1OCC	AJBOCC1
IWS22112	AEJDATE2
IWS22124	ARSEND2
IWS22126	APAYHR2
IWS22128	APYRATE2
IWS22129	n/a
IWS22130	n/a
IWS22131	n/a
IWS22144	AUNION2
IWS22146	ACNTRC2
IWS2IND	AJBIND2
IWS2OCC	AJBOCC2
J10003	TSVJTINT
J10407	TMDJTINT
J110	TSJNTDIV
J110RI	TMJADIV

Ordered by 1993 Variable Name	
1993	**1996**
J120OT	TJACLR2
J130	TMIJNT
JGRENT	TJARNT
JNRENT	TJACLR
LCHCOST	n/a
LCHFREE	EFRERDLN
LCHPT	EFREELUN
LCHTOT	n/a
LEVEL	EENLEVEL
LUNCH	EHOTLUNC
MCDPNUM	RCUOWN57
MCOPT	n/a
MEDCODE	RMEDCODE
MIS5	n/a
MONENT	n/a
MONLFT	n/a
MONTH	RHCALMN
MS	EMS
NDSL	EASST05
NJOBS	EJOBCNTR
NKIDSBF	RNKBRK
NKIDSHL	RNKLUN
NOINC	n/a
NONHHI	EHIOTHER
O10003	TSVOINT
O10407	TMDOINT
O110	TSOWNDIV
O110RI	TMOWNADV
O130	TMIOWN
O14050	TRNDUP1
OGRENT	TOARNT
ONRENT	TOACLR
OTHAID	EASST11, EASST07
OTHER	TPOTHINC
OTHINC	ER56
OTHVET	EASST02
OTHWELF	RCUTYP24
OWPNUM	RCUOW24A
P5WGT	WPFINWGT
PANEL	SPANEL
PELL	EASST01
PHRENT	TMTHRNT
PLUS	EASST05
PNGDU	EPNGUARD

Ordered by 1993 Variable Name	
1993	**1996**
PNPT	EPNMOM, EPNDAD
PNSP	EPNSPOUS
PNUM	EPPPNUM
POPSTAT	EPOPSTAT
PROP	TPPRPINC
PWADDID	n/a
PWENTRY	n/a
PWPNUM	n/a
PWRRP	n/a
PWSUID	n/a
R01A	ER01A
R01K	ER01K
R02A	ER02
R02K	n/a
R03	ER03A, ER03K
R05	ER05
R06	ER06
R07	ER07
R08	ER08
R10	ER10
R100	EAST2B
R101	EAST2C
R102	EAST2D
R103	EAST2A
R104	EMDJT, EMDOAST
R105	EAST3D
R106	EAST3C
R107	EAST4C
R110	EAST3A, EAST3B
R12	AR12
R120	EAST4A
R13	ER13
R130	EAST3E
R140	EAST4B
R150	ERNDUP2
R20	ER20
R21	ER21
R23	ER23
R24	ER24
R25	ER25
R27	ER27
R28	ER28
R29	ER29
R30	ER30

Ordered by 1993 Variable Name	
1993	**1996**
R31	ER31
R32	ER32
R34	ER34
R35	ER35
R36	ER36
R37	ER37
R38	ER38
R40	ER40
R41	ER41
R50	ER50
R51	ER51
R52	ER52
R53	ER53
R54	ER54
R55	ER55
R56	ER56
R75	ER75, ER09, ER33
RACE	ERACE
RAILRD	n/a
REAENT	n/a
REALFT	n/a
REASAB	EABRE
REFMTH	SREFMON
RENVELOP	n/a
RETIRD	EEVERET
RHCDIS	n/a
RJ10003	ESVJT
RJ10407	n/a
RJ110	ESANYCHK
RJ110RI	EMOTHDIV
RJ120	EJNTRNT
RJ120OT	EJRNT2
RJ130	EMRTJNT
RO10003	ESVOAST
RO10407	n/a
RO110	EMANYCHK
RO110RI	EMOTHDIV
RO120	EOWNRNT
RO130	EMRTOWN
RO14050	n/a
ROT	SROTATON
RRDAY	n/a
RRP	ERRP
RRPNUM	n/a

Ordered by 1993 Variable Name	
1993	**1996**
RRPU	n/a
S01AMTA	T01AMTA
S01AMTK	T01AMTK
S02AMTA	T02AMT
S02AMTK	n/a
S03AMT	T03AMTA, T03AMTK
S05AMT	T05AMT
S06AMT	n/a
S07AMT	T07AMT
S08AMT	T08AMT
S10AMT	T10AMT
S12AMT	T12AMT
S13AMT	T13AMT
S20AMT	T20AMT
S21AMT	A20AMT
S23AMT	T23AMT
S24AMT	T24AMT
S27AMT	T27AMT
S28AMT	T28AMT
S29AMT	T29AMT
S30AMT	T30AMT
S31AMT	T31AMT
S32AMT	T32AMT
S34AMT	T34AMT
S35AMT	T35AMT
S36AMT	T36AMT
S37AMT	T37AMT
S38AMT	T38AMT
S40AMT	T39AMT
S41AMT	n/a
S50AMT	T50AMT
S51AMT	T51AMT
S52AMT	T52AMT
S53AMT	T53AMT
S54AMT	n/a
S55AMT	T55AMT
S56AMT	T56AMT
S75AMT	T75AMT
SAFDC	TSAFDC
SC1000	EPDJBTHN
SCHANGE	RSCHANGE
SE12201	EBNO1
SE12202	EBIZNOW1
SE12203	n/a

Ordered by 1993 Variable Name	
1993	1996
SE12212	EHRSBS1
SE12214	EGROSB1
SE12218	TEMPB1
SE12220	EINCPB1
SE12222	EPROPB1
SE12224	EHPRTB1
SE12226	EPARTB11
SE12228	EPARTB21
SE12230	EPARTB31
SE12232	ESLRYB1
SE12234	EOINCB1
SE12252	n/a
SE12254	TPRFTB1
SE12256	TPRFTB1
SE12260	TBMSUM1
SE1AMT	TBMSUM1
SE1IND	TBSIND1
SE1OCC	TBSOCC1
SE1WKS	n/a
SE22301	EBNO2
SE22302	EBIZNOW2
SE22303	n/a
SE22312	EHRSBS2
SE22314	EGROSB2
SE22318	TEMPB2
SE22320	EINCPB2
SE22322	EPROPB2
SE22324	EHPRTB2
SE22326	EPARTB12
SE22328	EPARTB22
SE22330	EPARTB32
SE22332	ESLRYB2
SE22334	EOINCB2
SE22352	n/a
SE22354	TPRFTB2
SE22356	TPRFTB2
SE22360	TBMSUM2
SE2AMT	TBMSUM2
SE2IND	TBSIND2
SE2OCC	TBSOCC2
SE2WKS	n/a
SEARN	TSFEARN
SENVELOP	n/a
SEX	ESEX

Ordered by 1993 Variable Name	
1993	1996
SFDSTP	TSFDSTP
SID	RSID
SKIND	ESFKIND
SNP	ESFNP
SOCSEC	RCUTYP01
SOCSR1	ERESNSS1
SOCSR2	ERESNSS2
SOKLT18	ESOKLT18
SOTHER	TSOTHINC
SOWNKID	ESOWNKID
SPDAF	EAFSRVDI
SPINAF	n/a
SPOV	TSFPOV
SPROP	TSPRPINC
SREFPER	ESFRFPER
SSDAY	n/a
SSICOVRG	ESSICHLD, ESSISELF
SSOCSEC	TSSOCSEC
SSPNUM	RCUOWN01
SSPOUSE	ESFSPSE
SSSI	TSSSI
SSUNIT	n/a
STLOAN	EASST05
STOTINC	TSTOTINC
STRAN	TSTRNINC
STYPE	ESFTYPE
SUID	SSUID
SUNEMP	TSUNEMP
SUPPED	EASST04
SURGC	GRGC
SUSEQNUM	SSUSEQ
SUSTATE	TFIPSST
SVETS	TSVETS
SWGT	WSFINWGT
TAKJOB	RTAKJOB
TAKJOBN	RNOTAKE
TOTINC	TPTOTINC
TRAN	TPTRNINC
UHOURS	EJBHRS1
USRVDT1	UAF1
USRVDT2	UAF2
USRVDT3	UAF3
UTILS	EUTILYN
VETNUM	RCUOWN8A, RCUOWN8B

Ordered by 1993 Variable Name	
1993	**1996**
VETS	RCUTYP08
VETSMT	EVAQUES
VETSTAT	EAFEVER
VETTYP	EVETTYP
WAVE	SWAVE
WEEKS	EMAX
WESR1	RWKESR1
WESR2	RWKESR2
WESR3	RWKESR3
WESR4	RWKESR4
WESR5	RWKESR5
WICCOV	RCUTYP25
WICPNUM	RCUOWN25
WICVAL	EMTHAM25
WKSJOB	RMWKWJB
WKSLOK	RMWKLKG
WKSPT	EPTWRK
WKSPTR	EPTRESN
WKSTDY	EASST03
WKSWOP	RMWKSAB
WS12002	EENO1
WS12003	ESTLEMP1
WS12004	n/a
WS12012	ECLWRK1
WS12016	TSJDATE1
WS12018	TSJDATE1
WS12020	TEJDATE1
WS12022	TEJDATE1
WS12023	TEJDATE1
WS12024	ERSEND1
WS12025	EJBHRS1
WS12026	EPAYHR1
WS12028	TPYRATE1
WS12029	RPYPER1
WS12030	n/a
WS12031	n/a
WS12044	EUNION1
WS12046	ECNTRC1
WS1AMT	TPMSUM1
WS1CALC	APAYHR1, APYRATE1
WS1CHG	n/a
WS1IND	EJBIND1
WS1OCC	TJBOCC1
WS1WKS	n/a

Ordered by 1993 Variable Name	
1993	**1996**
WS22102	EENO2
WS22103	ESTLEMP2
WS22104	n/a
WS22112	ECLWRK2
WS22116	TSJDATE2
WS22118	TSJDATE2
WS22120	TEJDATE2
WS22122	TEJDATE2
WS22123	TEJDATE2
WS22124	ERSEND2
WS22125	EJBHRS2
WS22126	EPAYHR2
WS22128	TPYRATE2
WS22129	RPYPER2
WS22130	n/a
WS22131	n/a
WS22144	EUNION2
WS22146	ECNTRC2
WS2AMT	TPMSUM2
WS2CALC	APAYHR2, APYRATE2
WS2CHG	n/a
WS2IND	EJBIND2
WS2OCC	TJBOCC2
WS2WKS	n/a
YEAR	RHCALYR

Ordered by 1996 Variable Name	
1993	**1996**
IS01A	A01AMTA
IS01K	A01AMTK
IS02A	A02AMT
IS03	A03AMTA, A03AMTK
IS05	A05AMT
IS06	A06AMT
IS07	A07AMT
IS08	A08AMT
IS10	A10AMT
IS12	A12AMT
IS13	A13AMT
S21AMT	A20AMT
IS20	A20AMT
IS21	A21AMT
IS23	A23AMT
IS24	A24AMT
IS27	A27AMT
IS28	A28AMT
IS29	A29AMT
IS30	A30AMT
IS31	A31AMT
IS32	A32AMT
IS34	A34AMT
IS35	A35AMT
IS36	A36AMT
IS37	A37AMT
IS38	A38AMT
IS40	A40AMT
IS50	A50AMT
IS51	A51AMT
IS52	A52AMT
IS53	A53AMT
IS54	A54AMT
IS55	A55AMT
IS56	A56AMT
IS75	A75AMT
IREASAB	AABRE
IVETSTAT	AAFEVER
IINAF	AAFNOW
ISPDAF	AAFSRVDI
IDISAGE	AAGESS
IR103	AAST2A
IR100	AAST2B
IR101	AAST2C

Ordered by 1996 Variable Name	
1993	**1996**
IR102	AAST2D
IR106	AAST3C
IR105	AAST3D
IR130	AAST3E
IR120	AAST4A
IR140	AAST4B
IR107	AAST4C
ISE12260	ABMSUM1
ISE1AMT	ABMSUM1
ISE22360	ABMSUM2
ISE2AMT	ABMSUM2
IBREAKF	ABRKFST
ISE1IND	ABSIND1
ISE2IND	ABSIND2
ISE1OCC	ABSOCC1
ISE2OCC	ABSOCC2
IWS12012	ACLWRK1
IWS12046	ACNTRC1
IWS22146	ACNTRC2
ICARECOV	ACRMTH
IDISAB	ADISABL
ISTLOAN	AEDASST
IOTHVET	AEDASST
IWKSTDY	AEDASST
IPELL	AEDASST
INDSL	AEDASST
IPLUS	AEDASST
IEMPLYR	AEDASST
IOTHAID	AEDASST
IFSSHIP	AEDASST
ISUPPED	AEDASST
IEDASST	AEDFUND
IHIGRADE	AEDUCATE
IEASTAMT	AEGYAMT
IHENRGY	AEGYPMT
IWS22112	AEJDATE2
ISE12218	AEMPB1
ISE22318	AEMPB2
ILEVEL	AENLEVEL
IRETIRD	AEVERET
ILCHPT	AFREELUN
IBFFREE	AFRERDBK
ILCHFREE	AFRERDLN
ISE12214	AGROSB1

Ordered by 1996 Variable Name	
1993	1996
ISE22314	AGROSB2
IPHRENT	AGVTRNT
IHISRC	AHEMPLY
IHIPAY	AHICOST
INONHHI	AHIOTHER
IHIOWN	AHIOWNER
IHITYPE	AHIOWNER
ILUNCH	AHOTLUNC
ISE12220	AINCPB1
ISE22320	AINCPB2
IJNRENT	AJACLR
IJ120OT	AJACLR2
IJGRENT	AJARNT
IUHOURS	AJBHRS1
IWS1IND	AJBIND1
IWS2IND	AJBIND2
IWS1OCC	AJBOCC1
IWS2OCC	AJBOCC2
IRJ120	AJNTRNT
IRJ120OT	AJRNT2
IR110	AMANYCHK
IR104	AMDJT, AMDOAST
IJ10407	AMDJTINT
CJ10407	AMDJTINT
CO10407	AMDOINT
IO10407	AMDOINT
IJ130	AMIJNT
IO130	AMIOWN
IJ110RI	AMJADIV
IJO110RI	AMOTHDIV
IO110RI	AMOWNADV
IJO110	AMOWNDIV
IRJ130	AMRTJNT
IRO130	AMRTOWN
IONRENT	AOACLR
IOGRENT	AOARNT
ISE12234	AOINCB1
ISE22334	AOINCB2
IETHNCTY	AORIGIN
IRO120	AOWNRNT
WS1CALC	APAYHR1, APYRATE1
IWS12026	APAYHR1
IWS22126	APAYHR2
WS2CALC	APAYHR2, APYRATE2

Ordered by 1996 Variable Name	
1993	1996
ISE12256	APRFTB1
ISE12254	APRFTB1
ISE22356	APRFTB2
ISE22354	APRFTB2
ISE12222	APROPB1
ISE22322	APROPB2
IWKSPTR	APTRESN
IWKSPT	APTWRK
IWS12028	APYRATE1
IWS22128	APYRATE2
IR01A	AR01A
IR01K	AR01K
IR02A	AR02
IR03	AR03A, AR03K
IR05	AR05
IR06	AR06
IR07	AR07
IR08	AR08
IR10	AR10
IR12	AR12
R12	AR12
IR13	AR13
IR20	AR20
IR21	AR21
IR23	AR23
IR24	AR24
IR25	AR25
IR27	AR27
IR28	AR28
IR29	AR29
IR30	AR30
IR31	AR31
IR32	AR32
IR34	AR34
IR35	AR35
IR36	AR36
IR37	AR37
IR38	AR38
IR40	AR40
IGIBILL	AR40
IR41	AR41
IR50	AR50
IR51	AR51
IR52	AR52

Ordered by 1996 Variable Name	
1993	**1996**
IR53	AR53
IR54	AR54
IR55	AR55
ICWORK	AR55
IR56	AR56
IRACE	ARACE
IENROLD	ARENROLL, AENRLM, EENLEVEL
IO14050	ARNDUP1
IWS12024	ARSEND1
IWS22124	ARSEND2
ISEX	ASEX
IJ110	ASJNTDIV
ISE12232	ASLRYB1
ISE22332	ASLRYB2
IO110	ASOWNDIV
IRJ10003	ASVJT
CJ10003	ASVJTINT
IJ10003	ASVJTINT
IRO10003	ASVOAST
CO10003	ASVOINT
IO10003	ASVOINT
IWS12044	AUNION1
IWS22144	AUNION2
IUTILS	AUTILYN
IVETTYP	AVETTYP
IWKSLOK	AWKLKG
IWKSWOP	AWKSAB
REASAB	EABRE
HACCESS	EACCESS
VETSTAT	EAFEVER
INAF	EAFNOW
SPDAF	EAFSRVDI
PELL	EASST01
OTHVET	EASST02
WKSTDY	EASST03
SUPPED	EASST04
PLUS	EASST05
NDSL	EASST05
STLOAN	EASST05
FSSHIP	EASST06, EASST08, EASST09
EMPLYR	EASST10
OTHAID	EASST11, EASST07
R103	EAST2A
R100	EAST2B

Ordered by 1996 Variable Name	
1993	**1996**
R101	EAST2C
R102	EAST2D
R110	EAST3A, EAST3B
R106	EAST3C
R105	EAST3D
R130	EAST3E
R120	EAST4A
R140	EAST4B
R107	EAST4C
SE12202	EBIZNOW1
SE22302	EBIZNOW2
BRTHMN	EBMNTH
SE12201	EBNO1
SE22301	EBNO2
BREAKF	EBRKFST
WS12012	ECLWRK1
WS22112	ECLWRK2
WS12046	ECNTRC1
WS22146	ECNTRC2
CARECOV	ECRMTH
DISAB	EDISABL
EDASST	EEDFUND
HIGRADE	EEDUCATE
EASTAMT	EEGYAMT
HENRGY	EEGYPMT1, EEGYPMT2, EEGYPMT3
LEVEL	EENLEVEL
WS12002	EENO1
WS22102	EENO2
ENTRY	EENTAID
RETIRD	EEVERET
FKIND	EFKIND
FNP	EFNP
LCHPT	EFREELUN
FREFPER	EFREFPER
BFFREE	EFRERDBK
LCHFREE	EFRERDLN
FSPOUSE	EFSPOUSE
FTYPE	EFTYPE
SE12214	EGROSB1
SE22314	EGROSB2
HLORNT	EGVTRNT
HISRC	EHEMPLY
H5NP	EHHNUMPP
HNP	EHHNUMPP

Ordered by 1996 Variable Name	
1993	1996
HIPAY	EHICOST
NONHHI	EHIOTHER
HINONH	EHIOWNER
HITYPE	EHIOWNER
HIOWN	EHIOWNER
LUNCH	EHOTLUNC
SE12224	EHPRTB1
SE22324	EHPRTB2
H5REF	EHREFPER
HREFPER	EHREFPER
SE12212	EHRSBS1
SE22312	EHRSBS2
SE12220	EINCPB1
SE22320	EINCPB2
UHOURS	EJBHRS1
WS12025	EJBHRS1
WS22125	EJBHRS2
WS1IND	EJBIND1
WS2IND	EJBIND2
RJ120	EJNTRNT
NJOBS	EJOBCNTR
RJ120OT	EJRNT2
HLVQTR	ELIVQRT
RO110	EMANYCHK
WEEKS	EMAX
R104	EMDJT, EMDOAST
RJ110RI	EMOTHDIV
RO110RI	EMOTHDIV
RJ130	EMRTJNT
RO130	EMRTOWN
MS	EMS
WICVAL	EMTHAM25
SE12234	EOINCB1
SE22334	EOINCB2
ETHNCTY	EORIGIN
IR150	EOTHPROP
H5MIS	EOUTCOME
RO120	EOWNRNT
SE12226	EPARTB11
SE22326	EPARTB12
SE12228	EPARTB21
SE22328	EPARTB22
SE12230	EPARTB31
SE22330	EPARTB32

Ordered by 1996 Variable Name	
1993	1996
WS12026	EPAYHR1
WS22126	EPAYHR2
SC1000	EPDJBTHN
PNGDU	EPNGUARD
PNPT	EPNMOM, EPNDAD
PNSP	EPNSPOUS
POPSTAT	EPOPSTAT
INTVW	EPPINTVW
PNUM	EPPPNUM
SE12222	EPROPB1
SE22322	EPROPB2
WKSPTR	EPTRESN
WKSPT	EPTWRK
HPUBHS	EPUBHSE
R01A	ER01A
R01K	ER01K
R02A	ER02
R03	ER03A, ER03K
R05	ER05
R06	ER06
R07	ER07
R08	ER08
R10	ER10
R13	ER13
R20	ER20
R21	ER21
R23	ER23
R24	ER24
R25	ER25
R27	ER27
R28	ER28
R29	ER29
R30	ER30
R31	ER31
R32	ER32
R34	ER34
R35	ER35
R36	ER36
R37	ER37
R38	ER38
R40	ER40
GIBILL	ER40
R41	ER41
R50	ER50

Ordered by 1996 Variable Name	
1993	**1996**
R51	ER51
R52	ER52
R53	ER53
R54	ER54
R55	ER55
CWORK	ER55
R56	ER56
OTHINC	ER56
R75	ER75, ER09, ER33
RACE	ERACE
SOCSR1	ERESNSS1
SOCSR2	ERESNSS2
R150	ERNDUP2
RRP	ERRP
FAMREL	ERRP
WS12024	ERSEND1
WS22124	ERSEND2
RJ110	ESANYCHK
SEX	ESEX
SKIND	ESFKIND
SNP	ESFNP
SREFPER	ESFRFPER
SSPOUSE	ESFSPSE
FAMTYP	ESFT
STYPE	ESFTYPE
SE12232	ESLRYB1
SE22332	ESLRYB2
SOKLT18	ESOKLT18
SOWNKID	ESOWNKID
SSICOVRG	ESSICHLD, ESSISELF
WS12003	ESTLEMP1
WS22103	ESTLEMP2
RJ10003	ESVJT
RO10003	ESVOAST
HTENURE	ETENURE
WS12044	EUNION1
WS22144	EUNION2
HUNITS	EUNITS
UTILS	EUTILYN
VETSMT	EVAQUES
VETTYP	EVETTYP
HHSC	GHLFSAM
SURGC	GRGC
HSTRAT	GVARSTR

Ordered by 1996 Variable Name	
1993	**1996**
CHAMP	RCHAMPM
GAPNUM	RCUOW21A
OWPNUM	RCUOW24A
HIPNUM	RCUOW58A, RCUOW58B
SSPNUM	RCUOWN01
AFDCPNUM	RCUOWN20
FKPNUM	RCUOWN23
WICPNUM	RCUOWN25
FSPNUM	RCUOWN27
MCDPNUM	RCUOWN57
VETNUM	RCUOWN8A, RCUOWN8B
SOCSEC	RCUTYP01
VETS	RCUTYP08
AFDC	RCUTYP20
GENASST	RCUTYP21
FOSTKID	RCUTYP23
OTHWELF	RCUTYP24
WICCOV	RCUTYP25
FOODSTMP	RCUTYP27
CAIDCOV	RCUTYP57
HIIND	RCUTYP58
DESGPNPT	RDESGPNT
ENROLD	RENROLL, EENRLM, RENRLMA
FCHANGE	RFCHANGE
FID	RFID
FID2	RFID2
FNKIDS	RFNKIDS
FNSSR	RFNSSR
FOKLT18	RFOKLT18
FOWNKID	RFOWNKID
MONTH	RHCALMN
YEAR	RHCALYR
HCASH	RHCBRF
HCHANGE	RHCHANGE
HMEANS	RHMTRF
HNCASH	RHNBRF
HNF	RHNF
HNFAM	RHNFAM
HNSF	RHNSF
HNSSR	RHNSSR
HTYPE	RHTYPE
MEDCODE	RMEDCODE
ESR	RMESR
WKSLOK	RMWKLKG

Ordered by 1996 Variable Name	
1993	1996
WKSWOP	RMWKSAB
WKSJOB	RMWKWJB
NKIDSBF	RNKBRK
NKIDSHL	RNKLUN
TAKJOBN	RNOTAKE
WS12029	RPYPER1
WS22129	RPYPER2
SCHANGE	RSCHANGE
SID	RSID
TAKJOB	RTAKJOB
WESR1	RWKESR1
WESR2	RWKESR2
WESR3	RWKESR3
WESR4	RWKESR4
WESR5	RWKESR5
ADDID	SHHADID
PANEL	SPANEL
REFMTH	SREFMON
ROT	SROTATON
SUID	SSUID
SUSEQNUM	SSUSEQ
WAVE	SWAVE
S01AMTA	T01AMTA
S01AMTK	T01AMTK
S02AMTA	T02AMT
S03AMT	T03AMTA, T03AMTK
S05AMT	T05AMT
S07AMT	T07AMT
S08AMT	T08AMT
S10AMT	T10AMT
S12AMT	T12AMT
S13AMT	T13AMT
S20AMT	T20AMT
S23AMT	T23AMT
S24AMT	T24AMT
S27AMT	T27AMT
S28AMT	T28AMT
S29AMT	T29AMT
S30AMT	T30AMT
S31AMT	T31AMT
S32AMT	T32AMT
S34AMT	T34AMT
S35AMT	T35AMT
S36AMT	T36AMT

Ordered by 1996 Variable Name	
1993	1996
S37AMT	T37AMT
S38AMT	T38AMT
S40AMT	T39AMT
S50AMT	T50AMT
S51AMT	T51AMT
S52AMT	T52AMT
S53AMT	T53AMT
S55AMT	T55AMT
S56AMT	T56AMT
S75AMT	T75AMT
AGE	TAGE
DISAGE	TAGESS
SE1AMT	TBMSUM1
SE12260	TBMSUM1
SE2AMT	TBMSUM2
SE22360	TBMSUM2
SE1IND	TBSIND1
SE2IND	TBSIND2
SE1OCC	TBSOCC1
SE2OCC	TBSOCC2
BRTHYR	TBYEAR
WS12023	TEJDATE1
WS12022	TEJDATE1
WS12020	TEJDATE1
WS22122	TEJDATE2
WS22120	TEJDATE2
WS22123	TEJDATE2
SE12218	TEMPB1
SE22318	TEMPB2
FAFDC	TFAFDC
FEARN	TFEARN
FFDSTP	TFFDSTP
SUSTATE	TFIPSST
HSTATE	TFIPSST
FOTHER	TFOTHINC
FPOV	TFPOV
FSOCSEC	TFSOCSEC
FSSI	TFSSI
FTOTINC	TFTOTINC
FTRAN	TFTRNINC
FUNEMP	TFUNEMP
FVETS	TFVETS
HAFDC	THAFDC
HEARN	THEARN

Ordered by 1996 Variable Name	
1993	**1996**
HFDSTP	THFDSTP
HNONCSH	THNONCSH
HOTHER	THOTHINC
HPOV	THPOV
HPROP	THPRPINC
FPROP	THPRPINC
HSOCSEC	THSOCSEC
HSSI	THSSI
HTOTINC	THTOTINC
HTRAN	THTRNINC
HUNEMP	THUNEMP
HVETS	THVETS
JNRENT	TJACLR
J120OT	TJACLR2
JGRENT	TJARNT
WS1OCC	TJBOCC1
WS2OCC	TJBOCC2
J10407	TMDJTINT
O10407	TMDOINT
HMETRO	TMETRO
J130	TMIJNT
O130	TMIOWN
J110RI	TMJADIV
O110RI	TMOWNADV
HMSA	TMSA
PHRENT	TMTHRNT
ONRENT	TOACLR
OGRENT	TOARNT
EARN	TPEARN
WS1AMT	TPMSUM1
WS2AMT	TPMSUM2
OTHER	TPOTHINC
PROP	TPPRPINC
SE12254	TPRFTB1
SE12256	TPRFTB1
SE22356	TPRFTB2
SE22354	TPRFTB2
TOTINC	TPTOTINC
TRAN	TPTRNINC
WS12028	TPYRATE1
WS22128	TPYRATE2
O14050	TRNDUP1
SAFDC	TSAFDC
SFDSTP	TSFDSTP

Ordered by 1996 Variable Name	
1993	**1996**
SEARN	TSFEARN
SPOV	TSFPOV
WS12018	TSJDATE1
WS12016	TSJDATE1
WS22118	TSJDATE2
WS22116	TSJDATE2
J110	TSJNTDIV
SOTHER	TSOTHINC
O110	TSOWNDIV
SPROP	TSPRPINC
SSOCSEC	TSSOCSEC
SSSI	TSSSI
STOTINC	TSTOTINC
STRAN	TSTRNINC
SUNEMP	TSUNEMP
SVETS	TSVETS
J10003	TSVJTINT
O10003	TSVOINT
USRVDT1	UAF1
USRVDT2	UAF2
USRVDT3	UAF3
EWID	UEVRWID
FWGT	WFFINWGT
H5WGT	WHFNWGT
HWGT	WHFNWGT
P5WGT	WPFINWGT
FNLWGT	WPFINWGT
SWGT	WSFINWGT

Ordered by 1993 File Position	
1993	**1996**
SUSEQNUM	SSUSEQ
SUID	SSUID
ADDID	SHHADID
PANEL	SPANEL
WAVE	SWAVE
MONTH	RHCALMN
YEAR	RHCALYR
ROT	SROTATON
REFMTH	SREFMON
SUSTATE	TFIPSST
SURGC	GRGC
HHSC	GHLFSAM
HSTRAT	GVARSTR
HNF	RHNF
HNFAM	RHNFAM
HNSF	RHNSF
HREFPER	EHREFPER
HNP	EHHNUMPP
HTYPE	RHTYPE
HWGT	WHFNWGT
HSTATE	TFIPSST
HMETRO	TMETRO
HMSA	TMSA
HNSSR	RHNSSR
HACCESS	EACCESS
HLVQTR	ELIVQRT
HUNITS	EUNITS
HTENURE	ETENURE
HPUBHS	EPUBHSE
HLORNT	EGVTRNT
HITM36B	n/a
HMEANS	RHMTRF
HCASH	RHCBRF
HNCASH	RHNBRF
HPOV	THPOV
HTOTINC	THTOTINC
HEARN	THEARN
HPROP	THPRPINC
HTRAN	THTRNINC
HOTHER	THOTHINC
HNONCSH	THNONCSH
HSOCSEC	THSOCSEC
HSSI	THSSI
HUNEMP	THUNEMP

Ordered by 1993 File Position	
1993	**1996**
HVETS	THVETS
HAFDC	THAFDC
HFDSTP	THFDSTP
PHRENT	TMTHRNT
UTILS	EUTILYN
HENRGY	EEGYPMT1, EEGYPMT2, EEGYPMT3
EASTAMT	EEGYAMT
LUNCH	EHOTLUNC
NKIDSHL	RNKLUN
LCHTOT	n/a
LCHPT	EFREELUN
LCHFREE	EFRERDLN
LCHCOST	n/a
BREAKF	EBRKFST
NKIDSBF	RNKBRK
BFTOT	n/a
BFFREE	EFRERDBK
IPHRENT	AGVTRNT
IUTILS	AUTILYN
IHENRGY	AEGYPMT
IEASTAMT	AEGYAMT
ILUNCH	AHOTLUNC
INKIDSHL	n/a
ILCHTOT	n/a
ILCHPT	AFREELUN
ILCHFREE	AFRERDLN
ILCHCOST	n/a
IBREAKF	ABRKFST
INKIDSBF	n/a
IBFTOT	n/a
IBFFREE	AFRERDBK
H5REF	EHREFPER
H5NP	EHHNUMPP
H5MIS	EOUTCOME
H5ADDID	n/a
H5WGT	WHFNWGT
FID	RFID
FID2	RFID2
FNP	EFNP
FREFPER	EFREFPER
FSPOUSE	EFSPOUSE
FTYPE	EFTYPE
FKIND	EFKIND
FNKIDS	RFNKIDS

Ordered by 1993 File Position	
1993	**1996**
FOWNKID	RFOWNKID
FOKLT18	RFOKLT18
FNSSR	RFNSSR
FWGT	WFFINWGT
FPOV	TFPOV
FTOTINC	TFTOTINC
FEARN	TFEARN
FPROP	THPRPINC
FTRAN	TFTRNINC
FOTHER	TFOTHINC
FSOCSEC	TFSOCSEC
FSSI	TFSSI
FUNEMP	TFUNEMP
FVETS	TFVETS
FAFDC	TFAFDC
FFDSTP	TFFDSTP
SID	RSID
SNP	ESFNP
SREFPER	ESFRFPER
SSPOUSE	ESFSPSE
STYPE	ESFTYPE
SKIND	ESFKIND
SOWNKID	ESOWNKID
SOKLT18	ESOKLT18
SWGT	WSFINWGT
SPOV	TSFPOV
STOTINC	TSTOTINC
SEARN	TSFEARN
SPROP	TSPRPINC
STRAN	TSTRNINC
SOTHER	TSOTHINC
SSOCSEC	TSSOCSEC
SSSI	TSSSI
SUNEMP	TSUNEMP
SVETS	TSVETS
SAFDC	TSAFDC
SFDSTP	TSFDSTP
ENTRY	EENTAID
PNUM	EPPPNUM
INTVW	EPPINTVW
MIS5	n/a
FNLWGT	WPFINWGT
P5WGT	WPFINWGT
RRP	ERRP

Ordered by 1993 File Position	
1993	**1996**
RRPU	n/a
AGE	TAGE
BRTHMN	EBMNTH
BRTHYR	TBYEAR
POPSTAT	EPOPSTAT
SEX	ESEX
RACE	ERACE
ETHNCTY	EORIGIN
MS	EMS
EWID	UEVRWID
FAMTYP	ESFT
FAMREL	ERRP
PNSP	EPNSPOUS
PNPT	EPNMOM, EPNDAD
PNGDU	EPNGUARD
DESGPNPT	RDESGPNT
REALFT	n/a
REAENT	n/a
DAYLFT	n/a
MONLFT	n/a
YRLFT	n/a
DAYENT	n/a
MONENT	n/a
YRENT	n/a
HCHANGE	RHCHANGE
FCHANGE	RFCHANGE
SCHANGE	RSCHANGE
TOTINC	TPTOTINC
EARN	TPEARN
PROP	TPPRPINC
TRAN	TPTRNINC
OTHER	TPOTHINC
SC1000	EPDJBTHN
ESR	RMESR
WEEKS	EMAX
WESR1	RWKESR1
WESR2	RWKESR2
WESR3	RWKESR3
WESR4	RWKESR4
WESR5	RWKESR5
WKSJOB	RMWKWJB
WKSWOP	RMWKSAB
WKSLOK	RMWKLKG
REASAB	EABRE

Ordered by 1993 File Position	
1993	**1996**
TAKJOB	RTAKJOB
TAKJOBN	RNOTAKE
CWORK	ER55
UHOURS	EJBHRS1
WKSPT	EPTWRK
WKSPTR	EPTRESN
EMPLED	n/a
DISAB	EDISABL
RHCDIS	n/a
VETSTAT	EAFEVER
INAF	EAFNOW
SPINAF	n/a
USRVDT1	UAF1
USRVDT2	UAF2
USRVDT3	UAF3
AFTIME	n/a
AFDSAB	n/a
AFDPCT	n/a
SPDAF	EAFSRVDI
VETS	RCUTYP08
VETSMT	EVAQUES
VETNUM	RCUOWN8A, RCUOWN8B
RETIRD	EEVERET
SOCSEC	RCUTYP01
SSPNUM	RCUOWN01
SOCSR1	ERESNSS1
SOCSR2	ERESNSS2
DISAGE	TAGESS
RAILRD	n/a
RRPNUM	n/a
CARECOV	ECRMTH
MEDCODE	RMEDCODE
MCOPT	n/a
FOODSTMP	RCUTYP27
FSPNUM	RCUOWN27
AFDC	RCUTYP20
AFDCPNUM	RCUOWN20
GENASST	RCUTYP21
GAPNUM	RCUOW21A
FOSTKID	RCUTYP23
FKPNUM	RCUOWN23
OTHWELF	RCUTYP24
OWPNUM	RCUOW24A
WICCOV	RCUTYP25

Ordered by 1993 File Position	
1993	**1996**
WICVAL	EMTHAM25
WICPNUM	RCUOWN25
CAIDCOV	RCUTYP57
MCDPNUM	RCUOWN57
HIIND	RCUTYP58
HIPNUM	RCUOW58A, RCUOW58B
HINONH	EHIOWNER
CHAMP	RCHAMPM
CHPNUM	n/a
HIOWN	EHIOWNER
HISRC	EHEMPLY
HIPAY	EHICOST
HITYPE	EHIOWNER
HIFAM	n/a
NONHHI	EHIOTHER
HIGRADE	EEDUCATE
GRDCMPL	n/a
ENROLD	RENROLL, EENRLM, RENRLMA
LEVEL	EENLEVEL
EDASST	EEDFUND
GIBILL	ER40
OTHVET	EASST02
WKSTDY	EASST03
PELL	EASST01
SUPPED	EASST04
NDSL	EASST05
STLOAN	EASST05
PLUS	EASST05
EMPLYR	EASST10
FSSHIP	EASST06, EASST08, EASST09
OTHAID	EASST11, EASST07
OTHINC	ER56
NOINC	n/a
PWSUID	n/a
PWENTRY	n/a
PWPNUM	n/a
PWRRP	n/a
PWADDID	n/a
ISEX	ASEX
IRACE	ARACE
IETHNCTY	AORIGIN
IHIGRADE	AEDUCATE
IGRDCMPL	n/a
IEWID	n/a

Ordered by 1993 File Position	
1993	**1996**
IWKSJOB	n/a
IWKSWOP	AWKSAB
IWKSLOK	AWKLKG
IREASAB	AABRE
ITAKJOB	n/a
ITAKJOBN	n/a
ICWORK	AR55
IUHOURS	AJBHRS1
IWKSPT	APTWRK
IWKSPTR	APTRESN
IDISAB	ADISABL
IDISAGE	AAGESS
IRHCDIS	n/a
IVETSTAT	AAFEVER
IINAF	AAFNOW
ISPINAF	n/a
ISPDAF	AAFSRVDI
IRETIRD	AEVERET
ICARECOV	ACRMTH
IMCOPT	n/a
ICAIDCOV	n/a
IHIIND	n/a
IHIOWN	AHIOWNER
IHISRC	AHEMPLY
IHIPAY	AHICOST
IHITYPE	AHIOWNER
INONHHI	AHIOTHER
IENROLD	ARENROLL, AENRLM, EENLEVEL
ILEVEL	AENLEVEL
IEDASST	AEDFUND
IGIBILL	AR40
IOTHVET	AEDASST
IWKSTDY	AEDASST
IPELL	AEDASST
ISUPPED	AEDASST
INDSL	AEDASST
ISTLOAN	AEDASST
IPLUS	AEDASST
IEMPLYR	AEDASST
IFSSHIP	AEDASST
IOTHAID	AEDASST
NJOBS	EJOBCNTR
WS12003	ESTLEMP1
WS12004	n/a

Ordered by 1993 File Position	
1993	**1996**
WS1OCC	TJBOCC1
WS1IND	EJBIND1
WS1WKS	n/a
WS1AMT	TPMSUM1
WS12002	EENO1
WS12012	ECLWRK1
WS1CHG	n/a
WS12018	TSJDATE1
WS12016	TSJDATE1
WS12022	TEJDATE1
WS12020	TEJDATE1
WS12023	TEJDATE1
WS12024	ERSEND1
WS12025	EJBHRS1
WS12026	EPAYHR1
WS12028	TPYRATE1
WS12029	RPYPER1
WS12031	n/a
WS12030	n/a
WS12044	EUNION1
WS12046	ECNTRC1
IWS1OCC	AJBOCC1
IWS1IND	AJBIND1
IWS12012	ACLWRK1
IWS12024	ARSEND1
IWS12026	APAYHR1
IWS12028	APYRATE1
IWS12029	n/a
IWS12031	n/a
IWS12030	n/a
IWS12044	AUNION1
IWS12046	ACNTRC1
WS1CALC	APAYHR1, APYRATE1
WS22103	ESTLEMP2
WS22104	n/a
WS2OCC	TJBOCC2
WS2IND	EJBIND2
WS2WKS	n/a
WS2AMT	TPMSUM2
WS22102	EENO2
WS22112	ECLWRK2
WS2CHG	n/a
WS22118	TSJDATE2
WS22116	TSJDATE2

Ordered by 1993 File Position	
1993	1996
WS22122	TEJDATE2
WS22120	TEJDATE2
WS22123	TEJDATE2
WS22124	ERSEND2
WS22125	EJBHRS2
WS22126	EPAYHR2
WS22128	TPYRATE2
WS22129	RPYPER2
WS22131	n/a
WS22130	n/a
WS22144	EUNION2
WS22146	ECNTRC2
IWS2OCC	AJBOCC2
IWS2IND	AJBIND2
IWS22112	AEJDATE2
IWS22124	ARSEND2
IWS22126	APAYHR2
IWS22128	APYRATE2
IWS22129	n/a
IWS22131	n/a
IWS22130	n/a
IWS22144	AUNION2
IWS22146	ACNTRC2
WS2CALC	APAYHR2, APYRATE2
SE12202	EBIZNOW1
SE12203	n/a
SE1IND	TBSIND1
SE1OCC	TBSOCC1
SE1WKS	n/a
SE1AMT	TBMSUM1
SE12201	EBNO1
SE12212	EHRSBS1
SE12214	EGROSB1
SE12218	TEMPB1
SE12220	EINCPB1
SE12222	EPROPB1
SE12224	EHPRTB1
SE12226	EPARTB11
SE12228	EPARTB21
SE12230	EPARTB31
SE12232	ESLRYB1
SE12234	EOINCB1
SE12252	n/a
SE12254	TPRFTB1

Ordered by 1993 File Position	
1993	1996
SE12256	TPRFTB1
SE12260	TBMSUM1
ISE1OCC	ABSOCC1
ISE1IND	ABSIND1
ISE12214	AGROSB1
ISE12218	AEMPB1
ISE12220	AINCPB1
ISE12222	APROPB1
ISE12232	ASLRYB1
ISE12234	AOINCB1
ISE12254	APRFTB1
ISE12256	APRFTB1
ISE12260	ABMSUM1
ISE1AMT	ABMSUM1
SE22302	EBIZNOW2
SE22303	n/a
SE2IND	TBSIND2
SE2OCC	TBSOCC2
SE2WKS	n/a
SE2AMT	TBMSUM2
SE22301	EBNO2
SE22312	EHRSBS2
SE22314	EGROSB2
SE22318	TEMPB2
SE22320	EINCPB2
SE22322	EPROPB2
SE22324	EHPRTB2
SE22326	EPARTB12
SE22328	EPARTB22
SE22330	EPARTB32
SE22332	ESLRYB2
SE22334	EOINCB2
SE22352	n/a
SE22354	TPRFTB2
SE22356	TPRFTB2
SE22360	TBMSUM2
ISE2OCC	ABSOCC2
ISE2IND	ABSIND2
ISE22314	AGROSB2
ISE22318	AEMPB2
ISE22320	AINCPB2
ISE22322	APROPB2
ISE22332	ASLRYB2
ISE22334	AOINCB2

Ordered by 1993 File Position	
1993	**1996**
ISE22354	APRFTB2
ISE22356	APRFTB2
ISE22360	ABMSUM2
ISE2AMT	ABMSUM2
R01A	ER01A
R01K	ER01K
R02A	ER02
R02K	n/a
R03	ER03A, ER03K
R05	ER05
R06	ER06
R07	ER07
R08	ER08
R10	ER10
R12	AR12
R13	ER13
R20	ER20
R21	ER21
R23	ER23
R24	ER24
R25	ER25
R27	ER27
R28	ER28
R29	ER29
R30	ER30
R31	ER31
R32	ER32
R34	ER34
R35	ER35
R36	ER36
R37	ER37
R38	ER38
R40	ER40
R41	ER41
R50	ER50
R51	ER51
R52	ER52
R53	ER53
R54	ER54
R55	ER55
R56	ER56
R75	ER75, ER09, ER33
S01AMTA	T01AMTA
S01AMTK	T01AMTK

Ordered by 1993 File Position	
1993	**1996**
S02AMTA	T02AMT
S02AMTK	n/a
S03AMT	T03AMTA, T03AMTK
S05AMT	T05AMT
S06AMT	n/a
S07AMT	T07AMT
S08AMT	T08AMT
S10AMT	T10AMT
S12AMT	T12AMT
S13AMT	T13AMT
S20AMT	T20AMT
S21AMT	A20AMT
S23AMT	T23AMT
S24AMT	T24AMT
S27AMT	T27AMT
S28AMT	T28AMT
S29AMT	T29AMT
S30AMT	T30AMT
S31AMT	T31AMT
S32AMT	T32AMT
S34AMT	T34AMT
S35AMT	T35AMT
S36AMT	T36AMT
S37AMT	T37AMT
S38AMT	T38AMT
S40AMT	T39AMT
S41AMT	n/a
S50AMT	T50AMT
S51AMT	T51AMT
S52AMT	T52AMT
S53AMT	T53AMT
S54AMT	n/a
S55AMT	T55AMT
S56AMT	T56AMT
S75AMT	T75AMT
IR01A	AR01A
IR01K	AR01K
IR02A	AR02
IR03	AR03A, AR03K
IR05	AR05
IR06	AR06
IR07	AR07
IR08	AR08
IR10	AR10

Ordered by 1993 File Position	
1993	**1996**
IR12	AR12
IR13	AR13
IR20	AR20
IR21	AR21
IR23	AR23
IR24	AR24
IR25	AR25
IR27	AR27
IR28	AR28
IR29	AR29
IR30	AR30
IR31	AR31
IR32	AR32
IR34	AR34
IR35	AR35
IR36	AR36
IR37	AR37
IR38	AR38
IR40	AR40
IR41	AR41
IR50	AR50
IR51	AR51
IR52	AR52
IR53	AR53
IR54	AR54
IR55	AR55
IR56	AR56
IS01A	A01AMTA
IS01K	A01AMTK
IS02A	A02AMT
IS02K	n/a
IS03	A03AMTA, A03AMTK
IS05	A05AMT
IS06	A06AMT
IS07	A07AMT
IS08	A08AMT
IS10	A10AMT
IS12	A12AMT
IS13	A13AMT
IS20	A20AMT
IS21	A21AMT
IS23	A23AMT
IS24	A24AMT
IS27	A27AMT

Ordered by 1993 File Position	
1993	**1996**
IS28	A28AMT
IS29	A29AMT
IS30	A30AMT
IS31	A31AMT
IS32	A32AMT
IS34	A34AMT
IS35	A35AMT
IS36	A36AMT
IS37	A37AMT
IS38	A38AMT
IS40	A40AMT
IS41	n/a
IS50	A50AMT
IS51	A51AMT
IS52	A52AMT
IS53	A53AMT
IS54	A54AMT
IS55	A55AMT
IS56	A56AMT
IS75	A75AMT
R100	EAST2B
R101	EAST2C
R102	EAST2D
R103	EAST2A
RJ10003	ESVJT
RO10003	ESVOAST
R104	EMDJT, EMDOAST
R105	EAST3D
R106	EAST3C
R107	EAST4C
RJ10407	n/a
RO10407	n/a
R110	EAST3A, EAST3B
RJ110	ESANYCHK
RO110	EMANYCHK
RJ110RI	EMOTHDIV
RO110RI	EMOTHDIV
R120	EAST4A
RJ120	EJNTRNT
RO120	EOWNRNT
RJ120OT	EJRNT2
R130	EAST3E
RJ130	EMRTJNT
RO130	EMRTOWN

Ordered by 1993 File Position	
1993	**1996**
R140	EAST4B
R150	ERNDUP2
RO14050	n/a
J10003	TSVJTINT
O10003	TSVOINT
J10407	TMDJTINT
O10407	TMDOINT
J110	TSJNTDIV
O110	TSOWNDIV
J110RI	TMJADIV
O110RI	TMOWNADV
JGRENT	TJARNT
JNRENT	TJACLR
OGRENT	TOARNT
ONRENT	TOACLR
J120OT	TJACLR2
J130	TMIJNT
O130	TMIOWN
O14050	TRNDUP1
CJ10003	ASVJTINT
CO10003	ASVOINT
CJ10407	AMDJTINT
CO10407	AMDOINT
IR100	AAST2B
IR101	AAST2C
IR102	AAST2D
IR103	AAST2A
IRJ10003	ASVJT
IRO10003	ASVOAST
IR104	AMDJT, AMDOAST
IR105	AAST3D
IR106	AAST3C
IR107	AAST4C
IRJ10407	n/a
IRO10407	n/a
IR110	AMANYCHK
IJO110	AMOWNDIV
IJO110RI	AMOTHDIV
IR120	AAST4A
IRJ120	AJNTRNT
IRO120	AOWNRNT
IRJ120OT	AJRNT2
IR130	AAST3E
IRJ130	AMRTJNT

Ordered by 1993 File Position	
1993	**1996**
IRO130	AMRTOWN
IR140	AAST4B
IR150	EOTHPROP
IJ10003	ASVJTINT
IO10003	ASVOINT
IJ10407	AMDJTINT
IO10407	AMDOINT
IJ110	ASJNTDIV
IO110	ASOWNDIV
IJ110RI	AMJADIV
IO110RI	AMOWNADV
IJGRENT	AJARNT
IJNRENT	AJACLR
IOGRENT	AOARNT
IONRENT	AOACLR
IJ120OT	AJACLR2
IJ130	AMIJNT
IO130	AMIOWN
IO14050	ARNDUP1
VETTYP	EVETTYP
IVETTYP	AVETTYP
SSUNIT	n/a
SENVELOP	n/a
SSDAY	n/a
RENVELOP	n/a
RRDAY	n/a
SSICOVRG	ESSICHLD, ESSISELF

Ordered by 1996 File Position	
1993	**1996**
SUSEQNUM	SSUSEQ
SUID	SSUID
PANEL	SPANEL
WAVE	SWAVE
ROT	SROTATON
REFMTH	SREFMON
MONTH	RHCALMN
YEAR	RHCALYR
ADDID	SHHADID
HSTRAT	GVARSTR
HHSC	GHLFSAM
SURGC	GRGC
SUSTATE	TFIPSST
HSTATE	TFIPSST
H5MIS	EOUTCOME
HNF	RHNF
HNFAM	RHNFAM
HNSF	RHNSF
H5REF	EHREFPER
HREFPER	EHREFPER
H5NP	EHHNUMPP
HNP	EHHNUMPP
HTYPE	RHTYPE
HWGT	WHFNWGT
H5WGT	WHFNWGT
HMETRO	TMETRO
HMSA	TMSA
HCHANGE	RHCHANGE
HNSSR	RHNSSR
HACCESS	EACCESS
HUNITS	EUNITS
HLVQTR	ELIVQRT
HTENURE	ETENURE
HPUBHS	EPUBHSE
HLORNT	EGVTRNT
IPHRENT	AGVTRNT
PHRENT	TMTHRNT
UTILS	EUTILYN
IUTILS	AUTILYN
HENRGY	EEGYPMT1, EEGYPMT2, EEGYPMT3
IHENRGY	AEGYPMT
EASTAMT	EEGYAMT
IEASTAMT	AEGYAMT
LUNCH	EHOTLUNC

Ordered by 1996 File Position	
1993	**1996**
ILUNCH	AHOTLUNC
NKIDSHL	RNKLUN
LCHPT	EFREELUN
ILCHPT	AFREELUN
LCHFREE	EFRERDLN
ILCHFREE	AFRERDLN
BREAKF	EBRKFST
IBREAKF	ABRKFST
NKIDSBF	RNKBRK
BFFREE	EFRERDBK
IBFFREE	AFRERDBK
HEARN	THEARN
FPROP	THPRPINC
HPROP	THPRPINC
HTRAN	THTRNINC
HOTHER	THOTHINC
HTOTINC	THTOTINC
HNCASH	RHNBRF
HCASH	RHCBRF
HMEANS	RHMTRF
HPOV	THPOV
HNONCSH	THNONCSH
HSOCSEC	THSOCSEC
HSSI	THSSI
HUNEMP	THUNEMP
HVETS	THVETS
HAFDC	THAFDC
HFDSTP	THFDSTP
FID	RFID
FID2	RFID2
FNP	EFNP
FREFPER	EFREFPER
FSPOUSE	EFSPOUSE
FTYPE	EFTYPE
FCHANGE	RFCHANGE
FKIND	EFKIND
FNKIDS	RFNKIDS
FOWNKID	RFOWNKID
FOKLT18	RFOKLT18
FNSSR	RFNSSR
FWGT	WFFINWGT
FEARN	TFEARN
FTRAN	TFTRNINC
FOTHER	TFOTHINC

Ordered by 1996 File Position	
1993	**1996**
FTOTINC	TFTOTINC
FPOV	TFPOV
FSOCSEC	TFSOCSEC
FSSI	TFSSI
FUNEMP	TFUNEMP
FVETS	TFVETS
FAFDC	TFAFDC
FFDSTP	TFFDSTP
SID	RSID
SNP	ESFNP
SREFPER	ESFRFPER
SSPOUSE	ESFSPSE
STYPE	ESFTYPE
SKIND	ESFKIND
SCHANGE	RSCHANGE
SOWNKID	ESOWNKID
SOKLT18	ESOKLT18
SWGT	WSFINWGT
SEARN	TSFEARN
SPROP	TSPRPINC
STRAN	TSTRNINC
SOTHER	TSOTHINC
STOTINC	TSTOTINC
SPOV	TSFPOV
SSOCSEC	TSSOCSEC
SSSI	TSSSI
SVETS	TSVETS
SUNEMP	TSUNEMP
SAFDC	TSAFDC
SFDSTP	TSFDSTP
ENTRY	EENTAID
PNUM	EPPPNUM
INTVW	EPPINTVW
POPSTAT	EPOPSTAT
BRTHMN	EBMNTH
BRTHYR	TBYEAR
SEX	ESEX
ISEX	ASEX
RACE	ERACE
IRACE	ARACE
ETHNCTY	EORIGIN
IETHNCTY	AORIGIN
EWID	UEVRWID
INAF	EAFNOW

Ordered by 1996 File Position	
1993	**1996**
IINAF	AAFNOW
VETSTAT	EAFEVER
IVETSTAT	AAFEVER
USRVDT1	UAF1
USRVDT2	UAF2
USRVDT3	UAF3
VETTYP	EVETTYP
IVETTYP	AVETTYP
VETSMT	EVAQUES
SPDAF	EAFSRVDI
ISPDAF	AAFSRVDI
FNLWGT	WPFINWGT
P5WGT	WPFINWGT
FAMTYP	ESFT
AGE	TAGE
FAMREL	ERRP
RRP	ERRP
MS	EMS
PNSP	EPNSPOUS
PNPT	EPNMOM, EPNDAD
PNGDU	EPNGUARD
DESGPNPT	RDESGPNT
EARN	TPEARN
PROP	TPPRPINC
TRAN	TPTRNINC
OTHER	TPOTHINC
TOTINC	TPTOTINC
SOCSEC	RCUTYP01
SSPNUM	RCUOWN01
VETS	RCUTYP08
VETNUM	RCUOWN8A, RCUOWN8B
AFDC	RCUTYP20
AFDCPNUM	RCUOWN20
GENASST	RCUTYP21
GAPNUM	RCUOW21A
FOSTKID	RCUTYP23
FKPNUM	RCUOWN23
OTHWELF	RCUTYP24
OWPNUM	RCUOW24A
WICCOV	RCUTYP25
WICPNUM	RCUOWN25
FOODSTMP	RCUTYP27
FSPNUM	RCUOWN27
CAIDCOV	RCUTYP57

Ordered by 1996 File Position	
1993	1996
MCDPNUM	RCUOWN57
HIIND	RCUTYP58
HIPNUM	RCUOW58A, RCUOW58B
ENROLD	RENROLL, EENRLM, RENRLMA
IENROLD	ARENROLL, AENRLM, EENLEVEL
LEVEL	EENLEVEL
ILEVEL	AENLEVEL
EDASST	EEDFUND
IEDASST	AEDFUND
PELL	EASST01
WKSTDY	EASST03
SUPPED	EASST04
NDSL	EASST05
STLOAN	EASST05
PLUS	EASST05
FSSHIP	EASST06, EASST08, EASST09
EMPLYR	EASST10
OTHAID	EASST11, EASST07
IOTHVET	AEDASST
IWKSTDY	AEDASST
IPELL	AEDASST
ISUPPED	AEDASST
INDSL	AEDASST
IPLUS	AEDASST
IEMPLYR	AEDASST
IOTHAID	AEDASST
IFSSHIP	AEDASST
ISTLOAN	AEDASST
HIGRADE	EEDUCATE
IHIGRADE	AEDUCATE
SC1000	EPDJBTHN
WEEKS	EMAX
NJOBS	EJOBCNTR
RETIRD	EEVERET
IRETIRD	AEVERET
DISAB	EDISABL
IDISAB	ADISABL
REASAB	EABRE
IREASAB	AABRE
WKSPT	EPTWRK
IWKSPT	APTWRK
WKSPTR	EPTRESN
IWKSPTR	APTRESN
TAKJOB	RTAKJOB

Ordered by 1996 File Position	
1993	1996
TAKJOBN	RNOTAKE
ESR	RMESR
WESR1	RWKESR1
WESR2	RWKESR2
WESR3	RWKESR3
WESR4	RWKESR4
WESR5	RWKESR5
WKSJOB	RMWKWJB
WKSWOP	RMWKSAB
IWKSWOP	AWKSAB
WKSLOK	RMWKLKG
IWKSLOK	AWKLKG
WS12002	EENO1
WS12003	ESTLEMP1
WS12016	TSJDATE1
WS12018	TSJDATE1
WS12023	TEJDATE1
WS12020	TEJDATE1
WS12022	TEJDATE1
WS12024	ERSEND1
IWS12024	ARSEND1
WS12025	EJBHRS1
UHOURS	EJBHRS1
IUHOURS	AJBHRS1
WS12012	ECLWRK1
IWS12012	ACLWRK1
WS12044	EUNION1
IWS12044	AUNION1
WS12046	ECNTRC1
IWS12046	ACNTRC1
WS1AMT	TPMSUM1
WS12026	EPAYHR1
IWS12026	APAYHR1
WS1CALC	APAYHR1, APYRATE1
WS12028	TPYRATE1
IWS12028	APYRATE1
WS12029	RPYPER1
WS1IND	EJBIND1
IWS1IND	AJBIND1
WS1OCC	TJBOCC1
IWS1OCC	AJBOCC1
WS22102	EENO2
WS22103	ESTLEMP2
WS22118	TSJDATE2

Ordered by 1996 File Position	
1993	1996
WS22116	TSJDATE2
WS22122	TEJDATE2
WS22120	TEJDATE2
WS22123	TEJDATE2
IWS22112	AEJDATE2
WS22124	ERSEND2
IWS22124	ARSEND2
WS22125	EJBHRS2
WS22112	ECLWRK2
WS22144	EUNION2
IWS22144	AUNION2
WS22146	ECNTRC2
IWS22146	ACNTRC2
WS2AMT	TPMSUM2
WS22126	EPAYHR2
WS2CALC	APAYHR2, APYRATE2
IWS22126	APAYHR2
WS22128	TPYRATE2
IWS22128	APYRATE2
WS22129	RPYPER2
WS2IND	EJBIND2
IWS2IND	AJBIND2
WS2OCC	TJBOCC2
IWS2OCC	AJBOCC2
SE12201	EBNO1
SE12202	EBIZNOW1
SE12212	EHRSBS1
SE12214	EGROSB1
ISE12214	AGROSB1
SE12218	TEMPB1
ISE12218	AEMPB1
SE12220	EINCPB1
ISE12220	AINCPB1
SE12222	EPROPB1
ISE12222	APROPB1
SE12224	EHPRTB1
SE12232	ESLRYB1
ISE12232	ASLRYB1
SE12234	EOINCB1
ISE12234	AOINCB1
SE12254	TPRFTB1
SE12256	TPRFTB1
ISE12256	APRFTB1
ISE12254	APRFTB1

Ordered by 1996 File Position	
1993	1996
SE12260	TBMSUM1
SE1AMT	TBMSUM1
ISE1AMT	ABMSUM1
ISE12260	ABMSUM1
SE12226	EPARTB11
SE12228	EPARTB21
SE12230	EPARTB31
SE1IND	TBSIND1
ISE1IND	ABSIND1
SE1OCC	TBSOCC1
ISE1OCC	ABSOCC1
SE22301	EBNO2
SE22302	EBIZNOW2
SE22312	EHRSBS2
SE22314	EGROSB2
ISE22314	AGROSB2
SE22318	TEMPB2
ISE22318	AEMPB2
SE22320	EINCPB2
ISE22320	AINCPB2
SE22322	EPROPB2
ISE22322	APROPB2
SE22324	EHPRTB2
SE22332	ESLRYB2
ISE22332	ASLRYB2
SE22334	EOINCB2
ISE22334	AOINCB2
SE22354	TPRFTB2
SE22356	TPRFTB2
ISE22354	APRFTB2
ISE22356	APRFTB2
SE22360	TBMSUM2
SE2AMT	TBMSUM2
ISE2AMT	ABMSUM2
ISE22360	ABMSUM2
SE22326	EPARTB12
SE22328	EPARTB22
SE22330	EPARTB32
SE2IND	TBSIND2
ISE2IND	ABSIND2
SE2OCC	TBSOCC2
ISE2OCC	ABSOCC2
SSICOVRG	ESSICHLD, ESSISELF
SOCSR1	ERESNSS1

Ordered by 1996 File Position		Ordered by 1996 File Position	
1993	**1996**	**1993**	**1996**
SOCSR2	ERESNSS2	IR32	AR32
DISAGE	TAGESS	R34	ER34
IDISAGE	AAGESS	IR34	AR34
R01A	ER01A	R35	ER35
IR01A	AR01A	IR35	AR35
R01K	ER01K	R36	ER36
IR01K	AR01K	IR36	AR36
R02A	ER02	R37	ER37
IR02A	AR02	IR37	AR37
R03	ER03A, ER03K	R38	ER38
IR03	AR03A, AR03K	IR38	AR38
R05	ER05	R50	ER50
IR05	AR05	IR50	AR50
R07	ER07	R51	ER51
IR07	AR07	IR51	AR51
R08	ER08	R52	ER52
IR08	AR08	IR52	AR52
R10	ER10	R53	ER53
IR10	AR10	IR53	AR53
IR12	AR12	CWORK	ER55
R12	AR12	R55	ER55
R13	ER13	ICWORK	AR55
IR13	AR13	IR55	AR55
R20	ER20	OTHINC	ER56
IR20	AR20	R56	ER56
R21	ER21	IR56	AR56
IR21	AR21	R75	ER75, ER09, ER33
R23	ER23	S01AMTA	T01AMTA
IR23	AR23	IS01A	A01AMTA
R24	ER24	S01AMTK	T01AMTK
IR24	AR24	IS01K	A01AMTK
R25	ER25	S02AMTA	T02AMT
IR25	AR25	IS02A	A02AMT
R27	ER27	S03AMT	T03AMTA, T03AMTK
IR27	AR27	IS03	A03AMTA, A03AMTK
R28	ER28	S05AMT	T05AMT
IR28	AR28	IS05	A05AMT
R29	ER29	S07AMT	T07AMT
IR29	AR29	IS07	A07AMT
R30	ER30	S08AMT	T08AMT
IR30	AR30	IS08	A08AMT
R31	ER31	S10AMT	T10AMT
IR31	AR31	IS10	A10AMT
R32	ER32	S12AMT	T12AMT

Ordered by 1996 File Position		Ordered by 1996 File Position	
1993	**1996**	**1993**	**1996**
IS12	A12AMT	S56AMT	T56AMT
S13AMT	T13AMT	IS56	A56AMT
IS13	A13AMT	S75AMT	T75AMT
S20AMT	T20AMT	IS75	A75AMT
S21AMT	A20AMT	R103	EAST2A
IS20	A20AMT	IR103	AAST2A
IS21	A21AMT	R100	EAST2B
S23AMT	T23AMT	IR100	AAST2B
IS23	A23AMT	R101	EAST2C
S24AMT	T24AMT	IR101	AAST2C
IS24	A24AMT	R102	EAST2D
S27AMT	T27AMT	IR102	AAST2D
IS27	A27AMT	R110	EAST3A, EAST3B
S28AMT	T28AMT	R106	EAST3C
IS28	A28AMT	IR106	AAST3C
S29AMT	T29AMT	R105	EAST3D
IS29	A29AMT	IR105	AAST3D
S30AMT	T30AMT	R130	EAST3E
IS30	A30AMT	IR130	AAST3E
S31AMT	T31AMT	R120	EAST4A
IS31	A31AMT	IR120	AAST4A
S32AMT	T32AMT	R140	EAST4B
IS32	A32AMT	IR140	AAST4B
S34AMT	T34AMT	R107	EAST4C
IS34	A34AMT	IR107	AAST4C
S35AMT	T35AMT	RJ120	EJNTRNT
IS35	A35AMT	IRJ120	AJNTRNT
S36AMT	T36AMT	JGRENT	TJARNT
IS36	A36AMT	IJGRENT	AJARNT
S37AMT	T37AMT	JNRENT	TJACLR
IS37	A37AMT	IJNRENT	AJACLR
S38AMT	T38AMT	RO120	EOWNRNT
IS38	A38AMT	IRO120	AOWNRNT
S40AMT	T39AMT	OGRENT	TOARNT
S50AMT	T50AMT	IOGRENT	AOARNT
IS50	A50AMT	ONRENT	TOACLR
S51AMT	T51AMT	IONRENT	AOACLR
IS51	A51AMT	RJ120OT	EJRNT2
S52AMT	T52AMT	IRJ120OT	AJRNT2
IS52	A52AMT	J120OT	TJACLR2
S53AMT	T53AMT	IJ120OT	AJACLR2
IS53	A53AMT	RJ130	EMRTJNT
S55AMT	T55AMT	IRJ130	AMRTJNT
IS55	A55AMT	J130	TMIJNT

Ordered by 1996 File Position	
1993	**1996**
IJ130	AMIJNT
RO130	EMRTOWN
IRO130	AMRTOWN
O130	TMIOWN
IO130	AMIOWN
O14050	TRNDUP1
IO14050	ARNDUP1
RJ10003	ESVJT
IRJ10003	ASVJT
J10003	TSVJTINT
IJ10003	ASVJTINT
CJ10003	ASVJTINT
RO10003	ESVOAST
IRO10003	ASVOAST
O10003	TSVOINT
CO10003	ASVOINT
IO10003	ASVOINT
R104	EMDJT, EMDOAST
IR104	AMDJT, AMDOAST
J10407	TMDJTINT
IJ10407	AMDJTINT
CJ10407	AMDJTINT
O10407	TMDOINT
IO10407	AMDOINT
CO10407	AMDOINT
RO110	EMANYCHK
IR110	AMANYCHK
IJO110	AMOWNDIV
RJ110RI	EMOTHDIV
RO110RI	EMOTHDIV
IJO110RI	AMOTHDIV
J110RI	TMJADIV
IJ110RI	AMJADIV
O110RI	TMOWNADV
IO110RI	AMOWNADV
RJ110	ESANYCHK
J110	TSJNTDIV
IJ110	ASJNTDIV
O110	TSOWNDIV
IO110	ASOWNDIV
CARECOV	ECRMTH
ICARECOV	ACRMTH
MEDCODE	RMEDCODE
HINONH	EHIOWNER

Ordered by 1996 File Position	
1993	**1996**
HITYPE	EHIOWNER
HIOWN	EHIOWNER
IHITYPE	AHIOWNER
IHIOWN	AHIOWNER
CHAMP	RCHAMPM
HISRC	EHEMPLY
IHISRC	AHEMPLY
HIPAY	EHICOST
IHIPAY	AHICOST
NONHHI	EHIOTHER
INONHHI	AHIOTHER
OTHVET	EASST02
R06	ER06
IR06	AR06
GIBILL	ER40
R40	ER40
IGIBILL	AR40
IR40	AR40
R41	ER41
IR41	AR41
R54	ER54
IR54	AR54
WICVAL	EMTHAM25
IS06	A06AMT
IS40	A40AMT
IS54	A54AMT
R150	ERNDUP2
IR150	EOTHPROP

B. SIPP Topcoding Specifications

Earnings

The topcoding of earnings amounts is based on the procedure used by the Current Population Survey (CPS). Monthly amounts are topcoded if the wave amount is greater than one-third of the annual earnings benchmark of $150,000. The Survey of Income and Program Participation (SIPP) uses the benchmark of $150,000 set by CPS to "annualize" the topcoding procedure. SIPP topcodes on a monthly basis (reporting level) for amounts exceeding $12,500 (1/12 of $150,000) if the wave amount is greater than $50,000 (1/3 of $150,000). The topcoded amounts are defined once for the Panel based on Wave 1 edited data.

Three variables require topcoding:

- EPM(1-4)SUM—wage and salary earnings,
- EBM(1-4)SUM—self-employed earnings,
- EMLM(1-4)SUM—earnings from additional jobs and moonlighting.

To compute the topcodes, the Census Bureau tallies all amounts that require topcoding based on the above criteria into a 12-cell matrix. The cells are based on sex, race/ethnic origin, and full-time/part-time worker definition. When all values have been tallied, a mean is computed for each cell based on the total amount divided by total number of occurrences. Those means will be used for the entire 1996 Panel with an adjustment for inflation and real growth in earned income of 1.019% per wave for all remaining waves in the panel.

Topcoding Earnings for the 1996 SIPP Panel

If the sum of the monthly earnings amounts for a job for the wave is greater than $50,000, then those monthly amounts that are greater than $12,500 are topcoded. After matching on sex, race/ethnic origin, and labor force status, the Census Bureau uses the topcode amounts from the topcoding matrix for earnings. See Table B-1 for examples of income amounts that need to be topcoded.

Table B-1. Examples of Income Amounts That Need to Be Topcoded

Example	Monthly Income Amounts				Sum for the Wave	Is the Sum Greater Than $50,000?	Topcoding Procedure
	Month 1	Month 2	Month 3	Month 4			
1	$3,000	$4,000	$5,000	$5,000	$17,000	No	None
2	$0	$0	$0	$55,000	$55,000	Yes	Topcode month 4 with the mean
3	$15,000	$15,000	$10,000	$12,000	$52,000	Yes	Topcode months 1 and 2 with the mean
4	$12,000	$12,000	$12,000	$15,000	$51,000	Yes	Topcode month 4 with the mean
5	$0	$0	$0	$49,000	$49,000	No	None
6	$15,000	$15,000	$15,000	$15,000	$60,000	Yes	Topcode all 4 months with the mean

Specification of the Matrix for Calculating the Means for Earnings

The mean values are created by summing the reported monthly amounts that are greater than $12,500 and dividing by the total number of inputs to the cell.

For cells with fewer than six amounts, create a mean value by summing all values for those cells with fewer than six amounts and dividing by the total number of inputs to the cells. Matrix definition: $2 \times 3 \times 2$ matrix for sex, race, and labor force status

Sex

Use the edited variable ESEX with the following values:

ESEX: 1 = Male
2 = Female

Race

Set the index RACORIG, using the edited ERACE and EORIGIN, as described below:

Create the index variable RACORIG, defined as follows:

RACORIG: 1 = Nonblack, non-Hispanic
2 = Black, non-Hispanic
3 = Hispanic, any race

IF (EORIGIN = 20 - 28) THEN RACORIG = 3
ELSE IF (ERACE = 2) THEN RACORIG = 2
ELSE THEN RACORIG = 1

Labor Force Status

Set the index FTFULYR, which will define a worker as a full-time, full-year or a full-time, not full-year worker.

FTFULYR:
1 = Yes, full-time, full-year worker
2 = No, not full-time, full-year worker

IF (RM1ESR = 1 AND RM2ESR = 1 AND RM3ESR = 1 AND RM4ESR = 1) **AND**

(the number of variables in the EHRSWK01 - EHRSWK(EMAX) array that equal 1 is greater than EMAX/2)

THEN FTFULYR = 1 (YES)
ELSE FTFULYR = 2 (NO)

Filling the Matrix to Create the Means for Topcoding

Perform the following calculations in the order shown:

- Sum the four monthly amounts reported for EPM1SUM, EPM2SUM, EPM3SUM, and EPM4SUM. If the sum is greater than $50,000, then store the amounts greater than $12,500 in the appropriate cell in the matrix (matched on ESEX, RACORIG, FTFULYR).
- Sum the four monthly amounts reported for EBM1SUM, EBM2SUM, EBM3SUM, and EBM4SUM. If the sum is greater than $50,000, then store the amounts greater than $12,500 in the appropriate cell in the matrix (matched on ESEX, RACORIG, FTFULYR).

- Sum the four monthly amounts reported for EMLM1SUM, EMLM2SUM, EMLM3SUM, and EMLM4SUM. If the sum is greater than $50,000, then store the amounts greater than $12,500 in the appropriate cell in the matrix (matched on ESEX, RACORIG, FTFULYR).
- Sum the values in each cell and divide by the number of inputs to the cell for the mean amount for the cell.
- For cells with fewer than six inputs, create the mean by combining all of the amounts from each of the cells and dividing by the total number of inputs to the cells. Use this mean for all cells with zero to six entries.

Table B-2. Earnings Topcodes

Sex	Race	Worker Status	Topcode
Sex = 1 (Male)	Nonblack, non-Hispanic	Full year, full time	$29,660
Sex = 1 (Male)	Nonblack, non-Hispanic	Not full year, full time	$38,270
Sex = 1 (Male)	Black, non-Hispanic	Full year, full time	$17,530
Sex = 1 (Male)	Black, non-Hispanic	Not full year, full time	$24,015
Sex = 1 (Male)	Hispanic, any race	Full year, full time	$26,250
Sex = 1 (Male)	Hispanic, any race	Not full year, full time	$24,015
Sex = 2 (Female)	Nonblack, non-Hispanic	Full year, full time	$21,990
Sex = 2 (Female)	Nonblack, non-Hispanic	Not full year, full time	$49,450
Sex = 2 (Female)	Black, non-Hispanic	Full year, full time	$24,015
Sex = 2 (Female)	Black, non-Hispanic	Not full year, full time	$24,015
Sex = 2 (Female)	Hispanic, any race	Full year, full time	$24,015
Sex = 2 (Female)	Hispanic, any race	Not full year, full time	$24,015

Note: The topcodes listed above for each cell are greater than the monthly value that is tested, $12,500. This topcode is the mean of all amounts greater than $12,500. The intention is to reveal as much information as possible by using the mean value.

Year of Birth (TBYEAR)

Year of birth is bottomcoded to 1912 to ensure that age does not exceed 88 during the panel. If year of birth (EBYEAR) is earlier than 1912, set year of birth to 1912. Age must be recalculated based on the new year of birth.

Age (TAGE)

Age is topcoded to 88 for the entire panel. TAGE is topcoded through birth year (EBYEAR), which is bottomcoded to 1912, and then age is recalculated.

Age at Receipt of Social Security Disability Benefits (TAGESS)

EAGESS is age at which person began receiving Social Security Disability benefits.

If EAGESS is greater than TAGE, set TAGESS equal to the topcoded value for age (88).

If EAGESS GT TAGE THEN TAGESS = TAGE

Age Respondent Started Job or Business (TSJDATE, TEJDATE, TSBDATE, TEBDATE)

ESJDATE is date respondent started job.

EEJDATE is date respondent ended job.

ESBDATE is date respondent started business.

EEBDATE is date respondent ended business

A respondent cannot be over 88 years old during the life of the panel. Therefore, year of birth is bottomcoded to 1912. A respondent cannot have "worked" or "owned a business" before age 14 years. The earliest a respondent can be shown beginning or ending a job or business is 1926 (1912 + 14). If the date in ESJDATE, EEJDATE, ESBDATE, or EEBDATE is earlier than 1926, set the date to 1926 (exclude values equal to –1).

After bottomcoding the year to 1926, check the month and day fields to ensure that the end date is after the start date for the job or business and then **switch** the dates as follows:

For Jobs:

If EEJDATE is less than ESJDATE

Then ESJDATE = EEJDATE

EEJDATE = ESJDATE

For Businesses:

If EEBDATE is less than ESBDATE

Then ESBDATE = EEBDATE

EEBDATE = ESBDATE

Table B-3. 1996 Panel Topcoding Specifications

	PUF Variable	MONTHLY Topcode at:	Bottom-code	Short Description
1	TBDJTINT	$2,500	NA	Assets: Amount of monthly interest on joint municipal-corporate bonds
2	TBDOINT	$3,200	NA	Assets: Amount of monthly interest on self-owned municipal-corporate bonds
3	TCDJTINT	$450	NA	Assets: Amount of monthly interest on joint certificates of deposit
4	TCDOINT	$825	NA	Assets: Amount of monthly interest on solely owned certificates of deposit
5	TCKJTINT	$55	NA	Assets: Amount of monthly interest from joint checking account
6	TCKOINT	$110	NA	Assets: Amount of monthly interest on solely owned checking account
7	TGVJTINT	$550	NA	Assets: Amount of monthly interest on joint U.S. government securities
8	TGVOINT	$1,725	NA	Assets: Amount of monthly interest on self-owned U.S. government securities
9	TJACLR	$1,375	($1,000)	Assets: Amount of net rent from property owned jointly with spouse
10	TJACLR2	$6,000	($1,000)	Assets: Amount of net income from rental property with others
11	TJARNT	$2,725	NA	Assets: Amount of gross rent from property owned jointly with spouse
12	TMDJTINT	$275	NA	Assets: Amount of monthly interest on joint money market account
13	TMDOINT	$550	NA	Assets: Amount of monthly interest on self-owned money market deposit account
14	TMIJNT	$1,775	NA	Assets: Amount of interest on mortgage owned with spouse
15	TMIOWN	$1,650	NA	Assets: Amount of interest on own mortgage
16	TMJADIV	$700	NA	Assets: Amount of dividend credited to joint margin account/reinvestment in mutual funds
17	TMJNTDIV	$1,100	NA	Assets: Amount of check for jointly own mutual funds
18	TMOWNADV	$1,825	NA	Assets: Amount of dividend credited to sole margin account/reinvestment in mutual funds
19	TMOWNDIV	$1,375	NA	Assets: Amount of check for solely owned mutual funds
20	TOACLR	$2,450	($1,250)	Assets: Amount of net income from own rental property
21	TOARNT	$4,350	NA	Assets: Amount of gross rent from own property
22	TRNDUP1	$3,300	NA	Assets: Amount of income from royalties
23	TRNDUP2	$4,750	($1,250)	Assets: Amount of other income from financial investments
24	TSJADIV	$825	NA	Assets: Amount of dividend credited to margin account/reinvestment in stocks owned jointly
25	TSJNTDIV	$775	NA	Assets: Amount of dividend check for jointly owned stocks
26	TSOWNADV	$1,375	NA	Assets: Amount of monthly dividend credited margin account/reinvestment in stock
27	TSOWNDIV	$1,150	NA	Assets: Amount of dividend check for solely owned stocks
28	TSVJTINT	$150	NA	Assets: Amount of monthly interest on joint savings account.

(table continues)

Table B-3. 1996 Panel Topcoding Specifications (*continued*)

	PUF Variable	MONTHLY Topcode at:	Bottom-code	Short Description
29	TSVOINT	$175	NA	Assets: Amount of monthly interest on self-only savings account
30	TCSAGY(M)	NA	NA	GenInc: Amount received by agency on your behalf
31	T28AMT	$1,200	NA	GenInc: Amount of child support payments
32	T29AMT	$3,275	NA	GenInc: Amount of alimony payments
33	T30AMT	$2,500	NA	GenInc: Amount of pension from a company or union
34	T31AMT	$3,925	NA	GenInc: Amount from federal civil service or other federal civilian employee pension
35	T32AMT	$3,825	NA	GenInc: Amount of U.S. military retirement pay
36	T34AMT	$3,270	NA	GenInc: Amount of state government pension
37	T35AMT	$3,600	NA	GenInc: Amount of local government pension
38	T36AMT	$2,200	NA	GenInc: Amount of income from a paid-up life insurance policy or annuity
39	T37AMT	$5,000	NA	GenInc: Amount from estates or trusts
40	T38AMT	$2,600	NA	GenInc: Amount of payments for retirement, disability, or as a survivor benefit
41	T39AMT	$110,000	NA	GenInc: Amount of payments for pension/retirement lump sums
42	T42AMT	$13,625	NA	GenInc: Amount of draw from an IRA/Keough/401k or Thrift Plan
43	T50AMT	$75	NA	GenInc: Amount of income assistance from a charitable group
44	T51AMT	$10,900	NA	GenInc: Amount of money from relatives or friends
45	T52AMT	$325	NA	GenInc: Amount of lump-sum payments
46	T53AMT	$1,960	NA	GenInc: Amount of income from roomers or boarders
47	T55AMT	$3,500	NA	GenInc: Amount of incidental or casual earnings
48	T56AMT	$21,800	NA	GenInc: Amount of miscellaneous cash income
49	TBM(M)SUM1/2	See Spec No. 1	NA	Business: Income received this month
50	TPM(M)SUM1/2	See Spec No. 1	NA	Job: Earnings from job received in MONTH1
51	TMLM(M)SUM	See Spec No. 1	NA	LabFor: Amount of income from this work (moonlighting) this month
52	TBYEAR	See Spec No. 2	NA	Person: Birth year
53	TAGE	See Spec No. 3	NA	Person: Age as of last birthday
54	TAGESS	See Spec No. 4	NA	GenInc: Age Social Security Disability receipt began
55	TSJDATE	See Spec No. 5	NA	Job: Date started this job
56	TEJDATE	See Spec No. 5	NA	Job: Date ended this job
57	TSBDATE	See Spec No. 5	NA	Business: Date started operating this business
58	TEBDATE	See Spec No. 5	NA	Business: Date ended operating this business
59	TPYRATE	$30	NA	Job: Regular hourly pay rate
60	TPRFTB	$17,450	($2,500)	Business: Net profit or loss
61	TROLLAMT	$999,000	NA	GenInc: Amount rolled over into a retirement account during the reference period
62	TMTHRNT(M)	$650	NA	Household: Amount of monthly rent

C. Computing the SIPP Sampling Weights

This appendix supplements the discussion in Chapter 8 (Using Sampling Weights on SIPP Files) with more detailed information about how the core wave file person-level weight FNLWGT and the full panel file person-level weights FNLWGT_x and PNLWGT are computed;[1] it is intended as a reference for users who require a comprehensive description of how the sampling weights are computed.

Sections 1 and 2 of this appendix discuss the algorithms that are used to compute the final core wave file person-level weights FNLWGT, with the first section discussing the Wave 1 weights and the second section discussing the Wave 2+ weights. The third section discusses the algorithm that computes the final full panel weights FNLWGT_x (the calendar year weight for year *x*) and PNLWGT (the panel weight).

Wave 1 Weights

For the 1996 Panel, the final weights used in deriving estimates consist of the product of four factors: the base weight, the duplication control factor, the household noninterview adjustment factor, and the second-stage adjustment factor. For panels prior to 1996, these four factors may have been multiplied by two other factors—the first-stage ratio estimate factor and the new construction noninterview adjustment factor—which are discussed later in this chapter.

Base Weight (BW)

The primary component of the sampling weight is the base weight. The base weight for any sampled person or sampled household is the reciprocal of the probability under the sample design of that person or household being selected. If there was full response and if there were no calibration adjustments, then the summation of base weights for a particular subgroup (e.g., Hispanics in the Southwest) is an unbiased estimator of the total U.S. population within that subgroup. In simplified terms, a base weight of 1,000 assigned to a sampled person means that the sampled person "represents" 1,000 people in the U.S. population. The base weight for a

[1] The remaining weights given in Table 12-2 (HWGT, FWGT, SWGT, P5WGT, H5WGT, and FINALWGT) are derived directly from the basic person-level weight FNLWGT. This derivation is discussed in the "How Weights Are Constructed" subsection of Chapter 8.

household and the base weight for a person within a household are the same, since every person within a sampled household is automatically selected (i.e., selected with a conditional probability of 1, given household selection).

Duplication Control Factor (DCF)

The duplication control factor, an integer value between 1 and 4 inclusive, is applied to the base weights of specified households to account for subsampling done in clusters of housing units selected at the last stage of sample selection. These clusters typically contain an unmanageable number of housing units. When this occurs, a sampling fraction, $1/N$, is determined by selecting a value of N such that the number of sample households in the cluster is reduced to a manageable size. After this is done, a duplication control factor of N or 4, whichever is smaller, is included as a weighting factor for sampled housing units in the cluster.

Household Noninterview Adjustment Factor (NAF)

The noninterview adjustment factor is intended to adjust for the presence of Type A noninterview households (households that are not interviewed because the occupants were temporarily absent, no one was home, the occupants refused participation, or the occupants could not be located). Noninterview adjustment factors are computed for each of a set of noninterview cells. These cells are based on 512 cells generated from all possible cross-classifications of the following household characteristics (256 cells for panels prior to 1996):

- Within-PSU oversampling strata: poverty stratum and nonpoverty stratum (only for 1996 and later panels);
- Census region;
- Race of reference person: black or nonblack;
- Tenure: owner or renter;
- Residence status: MSA urban, MSA nonurban, NonMSA Census place, or NonMSA not Census place; and
- Household size: one, two, three, or four or more persons.

Any cells with fewer than 30 interviewed households or with noninterview adjustment factors exceeding 2.0 are collapsed with a neighboring cell. To define cells as neighboring, the Census Bureau uses a sort order and scale values based on estimates of the 1979 poverty rate within the cell. The total number of noninterview cells is less than or equal to 512 for the 1996 Panel (256 or fewer for the earlier panels). In pre-1996 Panels, no cells were collapsed across the four cells defined by the cross-classification of race of reference person and tenure. For the 1996 Panel, no

cells are collapsed over the cross-cells defined by race of reference person, tenure, within-PSU oversampling strata, and Census region.

Within each final noninterview cell *c*, the formula for the noninterview adjustment factor (NAF*c*) is

$$\mathrm{NAF}_c = \frac{\text{sum of BW * DCF over all sampled households in cell } c}{\text{sum of BW * DCF over all interviewed households in cell } c} \quad . \qquad \text{(C-1)}$$

This factor is applied to the weight of each interviewed household in the cell; with these noninterview-adjusted weights, the interviewed households in each cell can be seen to "represent" themselves and also the Type A noninterviewed households in the cell.[2]

Wave 1 Second-Stage Calibration Adjustment (SSCA)

For the second-stage calibration adjustments, the Census Bureau uses tallies of Current Population Survey (CPS) weights for independent population controls. The CPS weights are calibrated to match population controls provided by the population division of the Census Bureau and then a "March type" adjustment is done to equalize the weights of husbands and wives. Because the population division does not produce family-type controls, SIPP family-type controls are in fact CPS sample estimates. SIPP controls for age, sex, and race, on the other hand, should not differ appreciably from the original population division controls.

The primary steps in the calibration (or ratio estimation) process are the attaching of second-stage calibration adjustment factors to the pre-second-stage weights (BW*DCF*NAF) within particular cells (e.g., male Hispanic 14-year-olds) so that the resulting adjusted weights (BW*DCF*NAF*SSCA) aggregate to independent CPS-derived population estimates within the cell. The summation of the pre-second-stage weights within any cell are unbiased estimates (assuming the nonresponse adjustment successfully adjusts for all effects of nonresponse) of the population totals (e.g., the summation of BW*DCF*NAF over all male Hispanic 14-year-olds in the panel is an unbiased estimate of the total number of male Hispanic 14-years-olds in the U.S. population).

For SIPP, the monthly CPS estimates of the population totals in these cells are generally superior to the aggregations of nonresponse-adjusted SIPP weights (superior in the sense of having lower sampling and/or nonsampling error). The adjusted weights (BW*DCF*NAF*SSCA) give estimates then for these cells that are equal to the independent estimates. This adjustment generally improves the overall precision of all estimates of these cells or any other related survey characteristics that are prevalent in these cells.

[2] In pre-1996 Panels, group quarters housing units were not included in the nonresponse computations, and received nonresponse adjustments equal to 1. Group quarters housing units are treated as other households in the 1996 Panel.

The population cells for which adjustments are made to independent estimates are given in Figures C-1, C-2, and C-3 (see pages C-6–C-11). The cells include (as can be seen in the figures) age, race, sex, Spanish origin, family relationship, and household type. As noted earlier, the independently derived estimates for these cells are based on CPS March supplement-type estimates, except the estimates for family type. (The CPS estimates are not the usual CPS monthly estimates. [See U.S. Census Bureau (1998) for more details.] The estimates are specially computed for this purpose by summing the CPS weights within a given cell for all sample units in the relevant CPS sample [there are some extra steps also, such as the equalization of husbands' and wives' CPS weights, which are not generally part of the CPS estimation process]).

Outline of the Second-Stage Calibration Algorithm

The second-stage calibration algorithm uses as its inputs the pre-second-stage weights BW*DCF*NAF computed for each sampled person represented on a completed questionnaire in a SIPP panel.[3] These weights are run through a series of adjustments, which result in a final weight (FNLWGT).[4] This final weight can be written as FNLWGT = SSCA*BW*DCF*NAF, with SSCA (the second-stage calibration adjustment) equal to the ratio of the pre-second-stage weight and the final weight after the calibration process is completed.

This algorithm can be segmented into five major steps[5]:

1. Calibration of Hispanic children weights;
2. Calibration of non-Hispanic children weights;
3. Initial calibration steps for all adults;
4. Calibration of Hispanic adults; and
5. Calibration of non-Hispanic adults.

Each of these steps consists of numerous substeps. The next two sections describe certain steps that are common to all of the steps in the algorithm (the ratio adjustment step, the raking step, the cell-collapsing step, and the computation of control totals), the third section discusses details of

[3] Children do not answer any SIPP questionnaires, but any children who are indicated as dependents by a sampled household receive weights in this process.

[4] In pre-1996 Panels, households with all adults categorized as military personnel were interviewed and assigned weights (except for households in barracks, which are ineligible for SIPP). These households were not included in the second-stage calibration process (as they are not eligible for CPS and are not included in the CPS-derived control totals), and they received final weights equal to their pre-second-stage weights. For the 1996 Panel, these households are assigned as ineligible households and are not included in the weighting at all.

[5] Separate runs of the calibration algorithm are made for each reference month and each rotation group (a total of 16 calibration runs for each panel wave).

particular calibration steps, and the last section describes steps that were carried out only for pre-1996 Panels.

Ratio Adjustments, Raking, and Cell Collapsing

The most important steps in the algorithm are the ratio adjustment and raking steps. Each ratio adjustment step takes all of the person weights (as they are at that point in the algorithm) within particular second-stage cells and multiplies them by a common ratio adjustment factor. The common factor is chosen for the second-stage cell so that the summation of the adjusted person weights within the cell equals the control total for that second-stage cell. The common ratio adjustment factor for each cell is equal to the control total divided by the summation of the current person weights for all sample persons in the cell.

The raking step is similar to the ratio adjustment step except that there are two sets of second-stage cells, with separate control totals (one set of second-stage cells is called the "row dimension," and the other set is called the "column dimension"). At the end of the raking process (also called iterative proportional fitting), each person weight (as it is at that point in the algorithm) has been adjusted so that all person weights aggregate to the appropriate control totals for both the row cells and the column cells. The adjusted person weights have the property of aggregating within the second-stage cells to each control total while remaining as "close as possible" (in terms of a particular algebraic distance function) to the person weight values at the beginning of the raking step. Thus, the new person weights are consistent with both sets of independent control totals and have been altered as little as possible from the person weights before the step.

Most of the ratio adjustment and raking steps are preceded by a cell-collapsing step. This step is designed to prevent extreme alterations in the person weights (which will increase variability of the estimators) in any of the ratio adjustment and raking steps. Each second-stage cell is checked in its sample size: if the sample size is less than 35, then the cell is collapsed with a neighboring cell. The second-stage cells are also checked by computing the ratio adjustment for that cell. If that adjustment is less than 0.67 or greater than 2.0, then the cell is collapsed with a neighboring cell.

Ratio adjustments are computed for each set of second-stage cells before the raking process is performed. Ratio adjustments are computed for the row cells and the column cells as if only a ratio adjustment were being done for the row cells alone or the column cells alone, rather than a full raking step. If the computed ratio adjustments for any of the row cells are less than 0.67 or greater than 2.0, or the sample size for any row cell is less than 35, then the row cell is collapsed with a neighboring row cell. The same process is carried out for the column cells. All collapsing of this kind is completed before the raking step is executed.

When a second-stage cell is designated as requiring collapsing during the cell-collapsing step, the neighboring cell is chosen through a predetermined mechanism. Hispanic second-stage cells (see Figure C-1) are collapsed by sex (e.g., Hispanic males 15–24 are collapsed with Hispanic

females 15–24). The same is true for the household status second-stage cells for non-Hispanic children (the column dimension for non-Hispanic children; see Figure C-2). For the household status second-stage cells for adults (the column dimension for adults; see Figure C-3, pp. C-8 through C-11), the following pairs are collapsed when collapsing is necessary (the numbers in parentheses are the column numbers in the Figure C-3 tables):[6]

- Spouse in primary family (1); spouse in subfamily (3).
- Householder, no spouse present, in household with family (2); householder in household without a family (5).
- Not a spouse in household with family (4); not a householder in household without family (6).

For the age status second stage for adults (the row dimension for adults: see Figure C-3), neighboring cells are found on the basis of the scale value (which is given for the 1996 Panel in Figure C-3). The cell with the scale value closest to that of the cell that requires collapsing becomes the neighboring cell used in collapsing.

Figure C-1. Second-Stage Cells for Hispanics

Second-stage cells for Hispanic children

Male	Female

Second-stage cells for Hispanic adults[7]

Male			Female		
15–24	25–44	45+	15–24	25–44	45+

Second-stage cells for unmarried Hispanic adults

Male	Female

[6] Collapsing is never done across black and nonblack status, or across sex, but only within the four primary groups: black males and females, and nonblack males and females (see Figure C-3).

[7] Hispanic adults in the military are not defined as Hispanics in the computation of control totals or in the calculation of second-stage adjustments.

Figure C-2. Second-Stage Cells for Non-Hispanic Children

Second-Stage Cells for Black Children (14 years of age and younger)

MALES Age (years)	Children in Family Households	Children Not in Family Households	SCALE
Under 2			15
2 to 3			17
4 to 5			25
6 to 7			27
8 to 9			45
10 to 11			47
12 to 13			55
14			57

FEMALES Age (years)	Children in Family Households	Children Not in Family Households	SCALE
Under 2			15
2 to 3			17
4 to 5			25
6 to 7			27
8 to 9			45
10 to 11			47
12 to 13			55
14			57

Second-Stage Cells for Nonblack Children (14 years of age and under)

MALES Age (years)	Children in Family Households	Children Not in Family Households	SCALE
Under 1			15
1			17
2			25
3			27
4			45
5			47
6			55
7			57
8			75
9			77
10 to 11			85
12 to 13			105
14			107

FEMALES Age (years)	Children in Family Households	Children Not in Family Households	SCALE
Under 1			15
1			17
2			25
3			27
4			45
5			47
6			55
7			57
8			75
9			77
10 to 11			85
12 to 13			105
14			107

Figure C-3. Second-Stage Cells for Non-Hispanic Adults

Second-Stage Cells for Black Males (15+ years of age)

Age (years)	Persons in Households That Contain a Primary Family or Subfamily				Persons Not in Households Containing a Primary Family or Subfamily		SCALE VALUE
	Husband of Primary Family	Male House-holder, No Spouse Present	Other Household Members		House-holder	Not a Householder or Person in Group Quarters	
			Husband of Subfamily	Not a Husband			
15							15
16–17							16
18–19							18
20–21							27
22–24							29
25–29							47
30–34							49
35–39							57
40–44							59
45–49							63
50–54							65
55–59							83
60–64							85
65–69							93
70+							95

(figure continues)

The cell-collapsing procedure in some cases requires more than one iteration if cells after collapsing to the nearest neighbor are still too small or show extreme ratio adjustments (this generally occurs only in row-dimension collapsing for adults). New scale values are computed for the collapsed cells and are used to designate neighboring cells for any further collapsing that is necessary.

Computation of Control Totals

The control totals are equal to the CPS March-type estimates within each second-stage cell for some of the earlier ratio adjustment and raking steps in the algorithm.[8] For the remaining ratio adjustment and raking steps, the control totals are derived by taking the CPS March-type estimate within the second-stage cell and subtracting from this the adjusted weights of any

[8] For the 1984 and 1985 Panels, the control totals excluded people illegally residing in the United States. For the 1986 Panel and all panels following, the people are included in the control totals.

Figure C-3. Second-Stage Cells for Non-Hispanic Adults (*continued*)

Second-Stage Cells for Black Females (15+ years of age)

Age (years)	Persons in Households That Contain a Primary Family or Subfamily				Persons Not in Households Containing a Primary Family or Subfamily		SCALE VALUE
	Wife of Primary Family	Female House-holder, No Spouse Present	Other Household Members		House-holder	Not a Householder or Person in Group Quarters	
			Wife of Subfamily	Not a Wife			
15							15
16-17							16
18-19							18
20-21							27
22-24							29
25-29							47
30-34							49
35-39							57
40-44							59
45-49							63
50-54							65
55-59							83
60-64							85
65-69							93
70-74							94
75+							96

(figure continues)

subgroups whose weights have been completed. For example, control totals are derived for non-Hispanic children by taking the CPS March-type estimates for all children in each row cell and column cell (see Figure C-2) and subtracting the adjusted weights of all SIPP panel-rotation-group Hispanic children within that cell.

Details of the Calibration Steps

The first step (for Hispanic children) is a direct ratio adjustment to CPS control totals (using only two cells defined by sex). The second step (for non-Hispanic children) is a raking adjustment to derived controls; for row cells and column cells, the second-stage cells given in Figure C-2 are used. The derived control totals for each second-stage cell are equal to CPS control totals for all children in the cell minus the adjusted weights of all sampled Hispanic children in the cell.

Figure C-3. Second-Stage Cells for Non-Hispanic Adults (*continued*)

Second-Stage Cells for Nonblack Males (15+ years of age)

Age (years)	Persons in Households That Contain a Primary Family or Subfamily				Persons Not in Households Containing a Primary Family or Subfamily		SCALE VALUE
	Husband of Primary Family	Male Householder, No Spouse Present	Other Household Members		Householder	Not a Householder or Person in Group Quarters	
			Husband of Subfamily	Not a Husband			
15							15
16–17							16
18–19							18
20–21							27
22–24							29
25–29							47
30–34							49
35–39							57
40–44							59
45–49							63
50–54							65
55–59							83
60–64							85
65–69							93
70–74							95
75–79							103
80–84							104
85+							106

(figure continues)

Following the steps for children (which complete all second-stage adjustments for the children's weights) are the initial calibration steps for adults. Those steps are as follows:

1. A raking adjustment to CPS control totals that uses the Figure C-3 second-stage cells (the input weights are the pre-second-stage weights of all sampled adults);
2. A direct ratio adjustment to CPS control totals for sampled Hispanic adults; the input weights are the adjusted weights from step 1, and the second-stage cells are the cells given in Figure C-3 (for adults);
3. An equalization of all husbands' weights to their wives' weights (so that spouses in one family have equal weights);
4. A second raking adjustment identical to step 1 except that the input weights are the adjusted weights after steps 1 through 3 are completed;
5. A second Hispanic adult ratio adjustment identical to step 2 except that the input weights are the Hispanic adult adjusted weights from step 4.

Figure C-3. Second-Stage Cells for Non-Hispanic Adults (*continued*)

Second-Stage Cells for Nonblack Females (15+ years of age)

Age (years)	Persons in Households That Contain a Primary Family or Subfamily				Persons Not in Households Containing a Primary Family or Subfamily		SCALE VALUE
	Wife of Primary Family	Female House-holder, No Spouse Present	Other Household Members		House-holder	Not a Householder or Person in Group Quarters	
			Wife of Subfamily	Not a Wife			
15							15
16–17							16
18–19							18
20–21							27
22–24							29
25–29							47
30–34							49
35–39							57
40–44							59
45–49							63
50–54							65
55–59							83
60–64							85
65–69							93
70–74							95
75–79							103
80–84							104
85+							106

The next two steps complete the weights for Hispanic adults. The first step is an equalization of all husbands' weights in married couples, including at least one Hispanic, to their wives' weights. The exception to this is when the wife is not Hispanic, in which case the wife's weight is set equal to the husband's weight. At this point, all married couples including at least one Hispanic have their final weights. The second step is a ratio adjustment for sampled unmarried Hispanics (only males and females are used as second-stage cells) to derived control totals, which are CPS control totals for all Hispanic adults minus the adjusted weights of the sampled married Hispanics.

The last steps complete the calibration process for sampled non-Hispanic adult weights. Those steps are as follows:

6. An equalization of wives' weights to their husbands' weights.
7. A raking adjustment to derived control totals that uses the Figure C-3 second-stage cells (the input weights are the current adjusted weights of all non-Hispanic adults). The control totals are the CPS control totals for all adults for the second-stage cells minus the adjusted weights of Hispanic adults within those cells.
8. An equalization of husbands' weights to their wives' weights. This step finalizes the weights for all non-Hispanic females and all non-Hispanic husbands.
9. A raking adjustment to derived control totals; the Figure C-3 second-stage cells for adult males (with the two husband columns deleted) are used, and the current adjusted weights of all non-Hispanic nonhusband males are used. The derived control totals are the CPS control totals minus the adjusted weights of all groups who have had their weights completed. This step produces the final weights for all non-Hispanic nonhusband male adults (the last group without completed weights).

Weighting Factors Used in Panels Prior to 1996

In all panels prior to the 1996 Panel, a first-stage ratio estimate factor (FSF) was applied to the base weight of each person in non-self-representing PSUs (i.e., PSUs not sampled with certainty). This first-stage factor was a ratio adjustment step that used as cells Census region, residence status, and race; it was designed to reduce the variance resulting from sampling of PSUs. Although this factor is no longer computed in the 1996 Panel, the cells are now used in the computation of noninterview adjustment factors.

Also, beginning with the 1985 Panel, a new construction noninterview adjustment factor (NCF) was applied to the base weight of new households in new construction housing-unit clusters. This factor was used to account for newly constructed housing units that were selected for the sample but were unavailable for interviewing. It was set equal to 1 in the 1986–1993 Panels (it was not used in the 1984 Panel), and eventually it was discontinued.

Thus, in the 1984 Panel, FNLWGT was equal to BW*DCF*HNF*FSF*SSCA (excludes NCF). FNLWGT was equal to BW*DCF*NCF*HNF*FSF*SSCA in the 1985–1993 Panels.

Wave 2+ Weights

The later wave cross-sectional weight is computed separately for each reference month of each wave. This Wave 2+ FNLWGT has the following factors for people in households whose residents have not changed from Wave 1: an initial weight (IW), a later wave noninterview

adjustment (LWNIA), and a second-stage calibration adjustment (SSCA). The initial weight is generally equal to the pre-second-stage weight for the Wave 1 household weight (with some exceptions). For households that have had people move into or out of the household after Wave 1, there is an adjustment to the initial weight called the mover's weight (MW). For these people, the cross-sectional weight has as factors the mover's weight, the later wave noninterview adjustment, and the second-stage calibration adjustment. In summary, people in households that do not need mover's adjustments receive the cross-sectional weight FNLWGT = IW*LWNIA*SSCA, and persons in households that do require a mover's adjustment receive the Wave 2+ final weight FNLWGT = MW*LWNIA*SSCA.

Wave 2+ Initial Weights

The initial weight is essentially the pre-second-stage Wave 1 weight, that is, IW = BW*DCF*NAF.[9] The second-stage calibration adjustment for the Wave 1 reference months is not included as a factor: the second-stage calibration adjustment is redone using control totals current for the later wave reference months. The initial weight allows the original sample person to represent unsampled persons in the population and persons in households who were not successfully interviewed in Wave 1. The initial weight does not generally change from wave to wave after Wave 1, unless special circumstances arise that cause an alteration in the panel sample (such as a cut in the sample for budgetary or other reasons).

Movers' Weights

People in any households that an original sample person enters during later waves, or any people who become part of a Wave 1 sample household during later waves, also become part of the sample for those waves. If the original sample person moves away from the household containing those people, the additional people immediately drop from the sample (their in-sample status in any given wave is entirely dependent on the presence of original sample persons in the household). Any of the additional people who were part of the SIPP population in Wave 1 (and therefore could have been sampled) and who become members of households with original sample persons are called associated sample persons. If any of these additional persons were not part of the SIPP population in Wave 1 (because they were out of the country, institutionalized, etc.), then they are called additional sample persons.

[9] The 1985 Panel had an initial weight that was computed differently. The initial weight for this panel included a new-construction noninterview adjustment factor and a first-stage ratio estimate factor. The Wave 1 noninterview adjustment factor was also recomputed in the 1985 Panel to account for sampled households mistakenly left off the sample roster during Wave 1, and sampled households that were noncooperative in Wave 1 but were converted during Wave 2. There was also an added "sample cut" factor, adjusting for sampled households that were deselected because of a reduction in the 1985 Panel sample. Pre-1996 Panels following 1985 had only one difference from the 1996 Panel initial weight described in the text: the presence of the first-stage ratio estimate factor.

Any household that consists of people who were in the SIPP universe who lived in separate households during the Wave 1 reference period (with at least one of the households sampled in Wave 1) is called an enhanced household. In most cases, an enhanced household consists of original sample persons from a Wave 1 sample household and associated sample persons from a household (or households) not sampled in Wave 1. In a few rare cases, an enhanced household will contain original sample persons from more than one Wave 1 sample household. Those households are rare because the probability of selection of any given household in SIPP is quite small, making the joint probability of a later wave merged household having two or more of its Wave 1 predecessor households selected in Wave 1 quite small (but the situation does occur in the SIPP panels).

Enhanced households require an adjustment of the Wave 1 base weight for each person in the household. These people in effect had multiple chances of being in the selected enhanced household: they could have been selected as original sample persons in the household they were in during Wave 1 (which then became an enhanced household), or they could become an associated sample person if their Wave 1 household was not selected but merged later with a sampled Wave 1 household. Their true probability of being included in the enhanced household is higher than their nominal Wave 1 probability of selection, and their assigned base weight should be the reciprocal of this true sample inclusion probability.

This true inclusion probability is not computed directly, for it requires the computation of joint probabilities of selection of multiple households, some of which were not in the original Wave 1 household sample. Instead, a "mover's weight" is assigned to each original and associated sample person in the enhanced household, which has as its expectation the inverse of the true sample inclusion probability. In other words, the movers' weights are unbiased weights, taking into account the complex realized sample design for enhanced households.

In the case in which an enhanced household is formed from only one Wave 1 sample household (with associated persons added to it), the mover's weight for each person in the household (original, associated, or additional) is computed as follows for reference month t, enhanced household i:

$$W_{ti} = \frac{W_{1i} S_{1ti}}{S_{ti} - S_{tai}}, \qquad \text{(C-2)}$$

where W_{1i} is the initial weight that is common to all original sample persons in the ith enhanced household, S_{1ti} is the number of original sample persons in the ith enhanced household in month t, S_{ti} is the size of the ith enhanced household in month t (all persons), and S_{tai} is the number of additional sample persons in the ith enhanced household in month t. The numerator of this expression is the sum of the initial weights over all original sample persons in the household during month t, and the denominator of this expression is the number of original and associated sample persons in the ith enhanced household in month t. For a discussion of why these are unbiased weights, see, for example, Kalton and Brick (1994).

When two Wave 1 sample households merge, the mover's weight for each sample person (original, associated, or additional) in the household is computed as follows:

$$W_{ti} = \frac{W_{1i}S_{1ti} + W'_{1i}S'_{1ti}}{S_{ti} - S_{tai}}. \qquad \text{(C-3)}$$

The two terms in the numerator are for the first and second Wave 1 sample households. The movers' weights for more than two merged Wave 1 sample households are computed analogously.

Wave 2+ Later Wave Noninterview Adjustments

The initial weights have an adjustment for noncooperation in Wave 1; that is, the sample households with nonzero initial weights represent households for which an interview was not completed in Wave 1. There are, however, further losses of sample households in later waves for several reasons:

- The household refuses to cooperate in some or all of the later waves.
- The people in the household have moved and cannot be found.
- The household has moved, and has been found, but is too far away for a personal interview and cannot be reached by telephone. [10]

The weights of households for which later wave interviews are completed are adjusted to "represent" sample households (who cooperated in Wave 1) whose interviews are not completed for any of the above reasons. Those adjustments are computed by assigning each sample household with a nonzero initial weight to one of 109 later wave noninterview cells.[11] The noninterview cells are based on the following household characteristics:

1. Reference person is a non-Hispanic white person, or other (two categories).
2. Reference person is a female householder without a spouse and with her own children, a householder 65 years of age or older, or other (three categories).
3. Household income includes welfare payments (AFDC, WIC, Food Stamps, Medicaid, or other welfare), or not (two categories).
4. Household size is 1, 2, 3, or 4 or more persons (four categories).
5. Household has some bond-type financial assets, or not (two categories).

[10] The SIPP sample is designed so that most of the field work takes place within the SIPP PSUs, to reduce traveling costs. If a household moves too far away from the field areas, a telephone interview is attempted.
[11] In pre-1996 Panels, 53 noninterview cells were used, based on the first 7 of the 10 listed household characteristics.

6. Reference person's education level is less than 8 years, 8 to 11 years, 12 to 15 years, or 16 or more years (four categories).
7. Household owns housing unit, is renter, or is living in a public housing project or receiving a rent subsidy from the government (three categories).
8. Census division (nine categories).
9. Number of imputations in household Wave 1 questionnaire is none, 1, or more than 1 (three categories).
10. Household income as a percentage of the household poverty threshold (with both averaged over 4 reference months): less than or equal to 175 percent, 176 through 450 percent, and more than 450 percent (three categories).

These categories have been found in empirical research to be consistently heterogeneous in later wave noninterview rates (i.e., the categories have divergent noninterview rates). The later wave noninterview adjustment for each noninterview cell is equal to the sum of the initial or mover's weights of all households that have had the later wave interview completed, divided by the sum of the initial or mover's weights of all Wave 1 sample households.[12] (The mover's weight is used whenever a mover's weight is computed for the household.) These adjustments are made separately for each reference month of each later wave of the panel.

Before the final noninterview adjustment is computed for each wave, each noninterview cell is checked. Any noninterview cell with fewer than 30 interviewed households, or with a noninterview adjustment greater than 2, is collapsed with a neighboring cell. Cells are defined as neighboring on the basis of a set of scale values assigned to each noninterview cell. This procedure prevents extreme noninterview adjustments from being made (which will increase sampling variability). The final noninterview adjustment (LWNIA) for the cell, or collapsed cell, is assigned to each household within the cell.

Table C-1 presents the major groupings of noninterview cells (the noninterview cells within these major groupings have similar scale values and would be collapsed together within these groupings before any collapsing was done across groupings).

Wave 2+ Second-Stage Calibration Adjustment (SSCA)

A second-stage calibration adjustment is carried out for each reference month in each later wave, for each rotation group of the panel separately. This adjustment uses the same algorithm as described for Wave 1 weights, with new CPS or CPS-derived control totals computed for each

[12] In pre-1996 Panels, general quarters households were not included in these calculations and receive noninterview adjustments equal to 1. In the 1996 Panel, these households are treated in the same way as family households in noninterview calculations, but households with only military adults were included.

Table C-1. Major Groupings of Later Wave Noninterview Cells

Household Characteristics	Number of Nonresponse Cells
Hispanic or nonwhite	
Minimal assets	15
Assets include bonds	9
White Non-Hispanic	
Single female householder	1
Householder 65 and older	14
Other householder	
No welfare income	
One person in household	20
Two people in household	14
Three people in household	7
Four or more in household	19
Has welfare income	10
Total	109

new reference month. The pre-second-stage weights in this case are IW*LWNIA, or MW*LWNIA if a mover's weight was computed for the household. The second-stage calibration adjustments reduce sampling variability by calibrating the final weights to agree with independent control totals. With the later wave cross-sectional weights, the second-stage calibration adjustments also have the effect of reducing biases from population undercoverage (arising from eligible people entering the U.S. population after the Wave 1 reference months).

Calendar Year and Panel Weights

The algorithm for generating the calendar year and panel weights is very similar to that used for computing Wave 2+ weights, with some differences. The most important differences are the following:

- A control date is associated with each calendar year and panel weight (rather than the weight being associated with a month, as with the Wave 1 and Wave 2+ weights).
- For a sample person to have a nonzero weight, data must be present for the sequence of months defined for the weight (12 months for the calendar year weights and all months of the panel for the panel weights). Months for which the sample person is ineligible are excluded from this check.

Calendar Year and Panel Initial Weights

The initial weight computed for each sample person for all calendar year and panel weights is IW = BW*DCF*NAF, that is, the same quantity that is used as the initial weight for all Wave 2+ weights. This initial weight allows each original sample person who has interviews for the months for which they are eligible in the calendar year (or panel) to represent unsampled people in the population and people in households that were not successfully interviewed in Wave 1.

Calendar Year and Panel Noninterview Adjustments

The noninterview adjustments for each calendar year and panel weight are computed by first assigning each sampled person with a nonzero initial weight to one of 149 noninterview cells.[13] These noninterview cells are based on the following person-level characteristics:

1. Person is a non-Hispanic white person, or other (two categories).
2. Person was self-employed, or not (two categories).
3. Family income was a percentage of the family poverty threshold (with both averaged over 4 reference months): less than or equal to 175 percent, 176 through 450 percent, and more than 450 percent (three categories).[14]
4. Person in household whose income includes welfare payments (SSI, AFDC, WIC, Food Stamps, Medicaid, or other welfare), person receiving unemployment compensation but not in household with welfare payments, or neither (three categories).
5. Person in household with some bond-type financial assets, or not (two categories).
6. Person's education level is less than 12 years, 12 to 15 years inclusive, or 16 or more years (three categories).
7. Person was in labor force at least 1 month of wave, or not (two categories).
8. Census division of household (nine categories).
9. Number of imputations in household Wave 1 questionnaire is none, 1, or more than 1 (three categories).
10. Within PSU, stratum code of household is poverty stratum or nonpoverty stratum (two categories).

[13] In pre-1996 Panels, 126 noninterview cells were used, based on the first 7 of the 10 listed person characteristics.
[14] In pre-1996 Panels, household income (averaged over 4 reference months) was used instead: less than $1,200 a month, between $1,200 and $4,000 a month, and greater than or equal to $4,000 a month.

These categories have been found in empirical research to be consistently heterogeneous in later wave noninterview rates. The noninterview adjustment for the noninterview cell (for the particular calendar year [panel] weight) is equal to the sum of the initial weights of all sampled persons whose households were interviewed in Wave 1,[15] divided by the sum of the initial weights of all sampled persons who have interviews for every month of the calendar year (panel) in which they are eligible.[16]

As with other noninterview adjustments discussed in this appendix, each noninterview cell is checked for small sample sizes and extreme noninterview adjustments. Any noninterview cell with fewer than 30 sampled persons with complete interview strings, or with a calendar year (panel) noninterview adjustment greater than 2, is collapsed with a neighboring cell for that calendar year and panel weight. If necessary, this process can be iterative: a cell may be collapsed into another cell, and then the combined cell may be collapsed further with other cells. A set of scale values determines how cells are collapsed when collapsing is necessary. Table C-2 presents the major groupings of noninterview cells (i.e., the noninterview cells with similar scale values). The noninterview cells within these groupings would be collapsed together among themselves before any collapsing would be done outside of these groupings.

Table C-2. Major Groupings of Calendar Year (Panel) Noninterview Cells

Person Characteristics	Number of Nonresponse Cells
Hispanic or nonwhite	50
White Non-Hispanic	
Less than 12 years of education	25
12 to 15 years of education	
In labor force	32
Not in labor force	18
16 or more years of education	24
Total	149

[15] People who entered the sample during or after the calendar year (panel) period (by entering a sampled household) are excluded from these calculations (and receive calendar year [panel] weights of zero). Children who move without their parents (into nonsampled households) during the period are also excluded from these computations and receive calendar year (panel) weights of zero.

[16] In pre-1996 Panels, sample persons living in group quarters are not included in these noninterview adjustments, and those people are given noninterview adjustments equal to 1 (when their calendar year and panel weights are nonzero). In the 1996 Panel, sample persons living in group quarters are treated in the same way as other sample persons.

Calendar Year and Panel Second-Stage Adjustments

The calendar year and panel weights that have been computed up to this point (called the pre-second-stage weights) for each sampled person (with a complete set of interviews for their eligible months) are equal to BW*DCF*NAF*LWNIA. The formula for the final calendar year weights (FNLWGT) is BW*DCF*NAF*LWNIA*SSCA, where SSCA is the second-stage calibration adjustment. The final panel weight follows the same formula: PNLWGT = BW*DCF*NAF*LWNIA*SSCA, though LWNIA and SSCA are computed differently here. The final weight is computed in both cases from the pre-second-stage weights BW*DCF*NAF*LWNIA in accordance with the algorithm described below. As with the Wave 1 and Wave 2+ weights, the algorithm for second-stage adjustment for calendar year and panel weights can be segmented into the following five major steps:

1. Calibration of Hispanic children weights;
2. Calibration of non-Hispanic children weights;
3. Initial calibration steps for all adults;
4. Calibration of Hispanic adults; and
5. Calibration of non-Hispanic adults.

However, the actual steps within these five major steps are different in their details for calendar year (panel) weights. The primary difference between the calendar year (panel) weights second-stage calibration algorithm and the Wave 2+ weights second-stage calibration algorithm is that a married couple weighting equalization is not done for the calendar year (panel) weights, and married and unmarried persons are not separated out for separate calibration steps in the calendar year (panel) weights algorithm.

The independent estimates for the control month are the same CPS March supplement-type estimates that were used for the Wave 2+ weights, except they are computed for different second-stage cells when used for calendar year (panel) weights. The second-stage cells for calendar year (panel) weights are given in Figures C-4, C-5, and C-6. The second-stage calibration algorithm is run separately for each rotation group, with the control totals for each rotation group equal to one-quarter of the CPS control totals.

Figure C-4. Calendar Year and Panel Weight Second-Stage Cells for Hispanics

Second-Stage Cells for Hispanics (14 years and younger)

Male	Female

Second-Stage Cells for Hispanics (15+ years of age)[17]

Male			Female		
15–24	25–44	45+	15–24	25–44	45+

Figure C-5. Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Children

Cells for Children (14 years and younger)

Age	Nonblack Males	Nonblack Females	Black Males	Black Females	SCALE
Under 2					15
2 to 3					17
4 to 5					25
6 to 7					27
8 to 9					45
10 to 11					47
12 to 13					55
14					57

[17] Hispanic adults in the military are not defined as Hispanics in the computation of control totals or in the calculation of second-stage adjustments.

Figure C-6. Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Adults

1996 Panel Second-Stage Cells for Nonblack Females (15+ years of age)

	Householder				Not Householder			
Age (years)	1. Female Householder No Spouse Present with Own Children	2. Other Female Householder No Spouse Present	3. Other Female Householder Living with Relative	4. Female Householder Not Living with Relative	6. Spouse of Householder or Spouse of Related Subfamily	7. Other Female Related to Householder	9. Other Female Not Related to Householder	SCALE VALUE
15								15
16–17								16
18–19								18
20–21								27
22–24								29
25–29								47
30–34								49
35–39								57
40–44								59
45–49								63
50–54								65
55–59								73
60–61								74
62–64								76
65–69								93
70–74								95
75–79								103
80–84								104
85+								106

(figure continues)

Details of the Calendar Year and Panel Second-Stage Calibration Steps

The individual steps in the calendar year (panel) second-stage calibration algorithm are generally the same as the corresponding steps in the Wave 1 and Wave 2+ second-stage calibration

Figure C-6. Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Adults (*continued*)

1996 Panel Second-Stage Cells for Black Females (15+ years of age)

	Householder			Not Householder			
Age (years)	2. Female Householder No Spouse Present	3. Other Female Householder Living with Relative	4. Female Householder Not Living with Relative	6. Spouse of Householder or Spouse of Related Subfamily	7. Other Female Related to Householder	9. Other Female Not Related to Householder	SCALE VALUE
15							15
16–17							16
18–19							18
20–21							27
22–24							29
25–29							47
30–34							49
35–39							57
40–44							59
45–49							63
50–54							65
55–59							73
60–61							74
62–64							76
65–69							93
70–74							94
75+							96

(figure continues)

algorithm.[18] The differences in the two calibration algorithms are primarily the second-stage cells, with some other minor differences, as described in this section.

The first step (for Hispanic children) is a ratio adjustment to CPS control totals that uses only the two cells defined by sex (this step is identical to the Wave 1 and Wave 2+ algorithm step for Hispanic children). The second step (for non-Hispanic children) is a ratio adjustment step to derived controls that uses as cells the second-stage cells given in Figure C-5.

[18] The cell-collapsing procedures described for the Wave 1 and Wave 2+ weights are also used as stated in that section for the calendar year and panel weights, except for the column dimension collapsing for non-Hispanic adults. For calendar year and panel weights, and for any of the four race/sex groups given in Figure C-6, columns 1 and 2 (see Figure C-6 for the numbering of the columns) are collapsed if either does not meet the criterion (which is the same as described in the earlier section on ratio adjustment, raking, and cell collapsing), column 4 is collapsed with column 2 if it does not meet the criterion, column 7 is collapsed with column 9 if either does not meet the criterion, and column 8 is collapsed with column 10. Collapsing of columns 3, 5, and 6 and further collapsing of the other columns should never be necessary in practice.

Figure C-6. Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Adults (*continued*)

1996 Panel Second-Stage Cells for Nonblack Males (15+ years of age)

	Householder		Not Householder			
Age (years)	3. Male Householder Living with Relative	5. Male Householder Not Living with Relative	6. Spouse of Householder or Spouse of Related Subfamily	8. Other Male Related to Householder	10. Other Male Not Related to Householder	SCALE VALUE
15						215
16–17						216
18–19						218
20–21						227
22–24						229
25–29						247
30–34						249
35–39						257
40–44						259
45–49						263
50–54						265
55–59						273
60–61						274
62–64						276
65–69						293
70–74						295
75–79						303
80–84						304
85+						306

(figure continues)

Following these steps for children (which complete all second-stage adjustments for the children's weights) are the initial calibration steps for adults. Those steps are as follows:

1. A raking adjustment to CPS control totals that uses the Figure C-6 second-stage cells; the input weights are the pre-second-stage weights of all sampled adults.
2. A direct ratio adjustment to CPS control totals for sampled Hispanic adults; the input weights are the adjusted weights from step 1, and the second-stage cells are the cells given in Figure C-4 (for adults).
3. A second raking adjustment identical to step 1 except that the input weights are the adjusted weights after steps 1 and 2 are completed.

Figure C-6. Calendar Year and Panel Weight Second-Stage Cells for Non-Hispanic Adults (*continued*)

1996 Panel Second-Stage Cells for Black Males (15+ years of age)

	Householder		Not Householder			
Age (years)	3. Male Householder Living with Relative	5. Male Householder Not Living with Relative	6. Spouse of Householder or Spouse of Related Subfamily	8. Other Male Related to Householder	10. Other Male Not Related to Householder	SCALE VALUE
15						215
16–17						216
18–19						218
20–21						227
22–24						229
25–29						247
30–34						249
35–39						257
40–44						259
45–49						263
50–54						265
55–59						273
60–61						274
62–64						276
65–69						293
70+						295

4. A second Hispanic adult ratio adjustment identical to step 2 except that the input weights are the Hispanic adult adjusted weights from step 3.

At this point, the weights are completed for Hispanic adults. The final step is a raking adjustment to derived control totals that uses the Figure C-6 second-stage cells. The derived control totals are the CPS control totals for all adults for the second-stage cells minus the adjusted weights of Hispanic adults within those cells. The input weights are the current adjusted weights for non-Hispanic adults.

D. Acronyms

ADL	=	Activities of Daily Living
AFDC	=	Aid to Families with Dependent Children
ASA	=	American Statistical Association
BLS	=	Bureau of Labor Statistics
BW	=	base weight
CAI	=	computer-assisted interviewing
CAPI	=	computer-assisted personal interviewing
CMSA	=	Consolidated Metropolitan Statistical Area
CPS	=	Current Population Survey
DADS	=	Data Access and Dissemination System
DCF	=	duplication control factor
DES	=	Data Extraction System
EDs	=	enumeration districts
FERRET	=	Federal Electronic Research Review and Extraction Tool
FHNSP	=	female with no spouse present living with relatives
GA	=	General Assistance
GVFs	=	generalized variance functions
ICPSR	=	Inter-university Consortium for Political and Social Research
ISDP	=	Income Survey Development Program
MSA	=	Metropolitan Statistical Area
NAF	=	noninterview adjustment factor

NCF	=	new-construction noninterview adjustment factor
NCHS	=	National Center for Health Statistics
NLS	=	National Longitudinal Surveys
NSR PSUs	=	non-self-representing PSUs
OASDI	=	Old-Age, Survivors, and Disability Insurance
OMB	=	Office of Management and Budget
PRWORA	=	Personal Responsibility and Work Opportunity Reconciliation Act
PSID	=	Panel Study of Income Dynamics
PSU	=	primary sampling units
SIPP	=	Survey of Income and Program Participation
SPD	=	Survey of Program Dynamics
SRS	=	simple random sample
SSCA	=	second-stage calibration adjustment
SSI	=	Supplemental Security Income
TANF	=	Temporary Assistance for Needy Families
WIC	=	Women, Infants, and Children nutrition program

E. Glossary

A

address unit

This collection unit is a person or group of persons living at the same address at the time of the interview. The address unit may consist of one person living by himself or herself, a group of unrelated individuals, or one or more families.

allocation flag

See *imputation flag.*

B

C

CAI (computer-assisted interviewing)

A method of interviewing in which a computer is used as the data collection instrument.

CAPI (computer-assisted personal interviewing)

A method of interviewing in which field representatives use a laptop computer to collect data during in-person interviews. In SIPP, the field representatives also periodically use the laptop computers during telephone interviews conducted from their homes.

cold-deck matrix

The matrix of starting values that constitutes the first step in the hot-deck imputation procedure. The matrix values can be determined a priori from information external to the current file being processed or can be determined from reported information from the current file.

control card

In the paper instrument for SIPP, a mechanism for carrying demographic and case management information forward from one wave to the next for each sample member.

core content

Questions asked at every SIPP interview. They cover demographic characteristics, work experience, earnings, program participation, transfer income, and asset income.

core wave files

Files containing the core data from one wave of interviews.

cross-sectional

Pertaining to data collected for a single time period from a representative sample. In SIPP hot-deck imputation procedures, cross-sectional refers to current-wave data.

Current Population Survey (CPS)

A labor force survey sponsored jointly by the Census Bureau and the Bureau of Labor Statistics that is used to compute the government's official monthly unemployment statistics along with other estimates of labor force characteristics.

D

data dictionary

Contains information about the file structure and the names, locations, and contents of all variables in a microdata file.

data editing

The use of related information to replace missing or inconsistent data in the survey.

departure noninterview

This type of noninterview occurs when someone was a member of a SIPP interviewed household during the 4-month reference period but was no longer a household member on the date of the interview.

E

F

family

Two or more people who are living together and are related by blood, marriage, or adoption.

FERRET

An on-line data access tool available on the SIPP Web site. SIPP data are available on FERRET beginning with the 1992 longitudinal panel.

following rules

SIPP rules that guide which original sample members continue to be interviewed should they move.

full panel files

Files containing all data for every person who was a member of a SIPP panel at any time during the life of that panel.

G

general income

Any type of income except earnings and asset income.

geographic (GRIN) codes

Codes that identify where each sample household is located and permit linkage to a file that contains a full set of geographic codes for different kinds of areas. This level of geography is not available on the public use files.

group quarters

Noninstitutional living quarters, such as rooming and boarding houses, college dormitories, convents, and monasteries. These do not constitute households and are often treated differently from households.

H

hot-deck matrix

The matrix used in all but the first stage of hot-deck imputation. As cold-deck values are replaced with information from the current wave, the resulting array of cells constitutes the hot-deck matrix.

hot-deck procedure

The statistical method used to impute items missing from the core questionnaire and topical modules. This procedure replaces missing item data in a wave with nonmissing values from similar interviewed cases. The imputation method can be a purely cross-sectional procedure of locating donors from the current file on the basis of characteristics reported in this wave, or it can be a longitudinal procedure of locating donors from the prior wave on the basis of characteristics reported at that earlier time for items missing in the current wave.

household

People living in a housing unit at the time of the interview. SIPP infers households from the interviews conducted at each address.

household-level noninterviews

See *household nonresponse*.

household nonresponse

Nonresponse that occurs when the interviewer either cannot locate a household or cannot interview any of its adult members. See *Type A, Type B, Type C,* and *Type D noninterviews*.

household reference person

See *reference person.*

housing unit

Living quarters with its own entrance and cooking facilities.

I

imputation

The most common method for handling missing data in SIPP. Imputation replaces missing values with statistical estimates that are based on the best relevant information available.

imputation flag

An imputation flag is associated with each core questionnaire item subject to statistical imputation and indicates whether information has been imputed.

in-sample variables

See *monthly interview status variables.*

in scope

Being part of the survey universe.

interview month

The month during which the interview takes place.

item nonresponse

A source of missing data that occurs when a respondent does not answer one or more questions, even though most of the questionnaire is completed.

J

K

L

logical imputation

See *data editing.*

longitudinal

Pertaining to data collected at different times over an extended period from a representative sample. In SIPP hot-deck imputation procedures, longitudinal refers to previous-wave data.

M

merged households

Households created either when two separate sampling units, each containing original sample members, are merged together, perhaps because of a marriage, or when a household splits into two new households and later the households recombine.

microdata files

Data files containing information at the person, family, or household level. For SIPP, they include the core wave files, topical module files, and full panel files.

missing item data

Data that are missing for one or more individual questions or variables, but the observation has sufficient reported information to be classified as interviewed.

missing waves

Waves in which a respondent has no data, although data are present for other waves.

monthly interview status variables

Variables that indicate whether a person was in sample in a particular month, and whether a person was in sample in the interview month. They are known as the PP-MIS variables.

mover

An original sample person who moves during the life of the panel.

N

National Longitudinal Survey (NLS)

Collects data on current labor force and employment status, work history, and characteristics of the current or last job.

non-self-representing (NSR) primary sampling units (PSUs)

Smaller PSUs that must be grouped with similar PSUs from the same region in order to form strata for sampling. This level of geography is not available on the public use files.

O

original sample members

All people who were interviewed in the first wave of the panel and any children subsequently born to or adopted by them.

oversampling

Sampling that involves selecting certain groups or units with higher probabilities than others, resulting in the oversampled group having greater representation than occurs in the population from which it was drawn.

P

P-70 reports

Primary source for published estimates from the SIPP. These reports can be obtained from the SIPP Web site or from the Census Bureau.

panel

Refers both to a new sample that is introduced periodically in the SIPP and to the full collection of information for that sample. For example, the 1996 Panel refers to both the sample introduced in 1996 and the 12 waves of interviews conducted with that sample.

panel nonrespondents

Persons for whom an interview is missing for a wave.

Panel Study of Income Dynamics (PSID)

A nationally representative, longitudinal survey of the U.S. population, conducted by the University of Michigan. The focus of the survey is economics and demographics, especially income sources and amounts, employment, family composition changes, and residential location.

Partial panel files

Longitudinal files to be released by the Census Bureau prior to the conclusion of the 1996 Panel because of the 4-year duration of the 1996 Panel.

person-level noninterviews

This type of noninterview occurs when data are collected for at least one member of a household, but are missing for one or more other sample persons within that household.

person-month files

Microdata files containing a record for each person in a wave, for each month of the reference period the person was in the sample.

person nonresponse

Nonresponse that occurs when at least one person in the household is interviewed, while at least one other person is not. See *Type Z noninterview.*

primary family

Family containing the household reference person and related individuals.

primary individual

A household reference person who lives alone or lives with only nonrelatives.

primary sample members

See *original sample members.*

primary sampling units (PSUs)

Geographic units based on Census data and used in developing the SIPP sample. This level of geography is not available on the public use files.

program units

The group of individuals which constitutes one case, as defined by a particular benefit program. In SIPP, program units apply to health insurance and transfer programs and are identified for programs in which a case can consist of more than one person.

proxy interviews

Interviews taken on behalf of a sample member who is unable to answer.

public use microdata files

Data files that have been prepared by the Census Bureau for public use. These files have already been processed to impute missing data, to edit data for confidentiality, and to provide weights. Microdata files are available from the Census Bureau or on-line from the SIPP Web site.

Q

R

random carryover method

Longitudinal imputation procedure used to impute missing wave data.

1996 Redesign

A revamping of SIPP in order to improve the quality of estimates and to make the data more useful to analysts.

reference months

The months that constitute the reference period for a wave. The months vary for different rotation groups.

reference period

The 4 calendar months preceding the month of interview. The reference period is a different calendar period for each rotation group.

reference person

An owner or renter of record who can reasonably be expected to answer questions about the household in general and about other household members should they be unavailable for interview. All people in the household are listed according to their relationship to the reference person.

related subfamily

A married couple and dependents or parent-child family related to the reference person but not including him or her. An example would be the reference person's daughter and son-in-law.

rotation group

A subsample containing roughly one-quarter of the sample members. One rotation group is interviewed each month of a 4-month wave.

S

sample attrition

Loss of sample members. Sample attrition rates decline over time, but total attrition numbers increase.

seam effect

The tendency of respondents to report a disproportionate number of changes as occurring at the "seam" between the fourth month of one wave and the first month of the following wave.

secondary families

Two or more people living in the same household who are related to each other but not to the household reference person.

secondary individual

An individual who is neither a household reference person nor a relative of any other people in the household.

secondary sample members

People living with original sample members.

self-representing (SR) primary sampling units (PSUs)

Larger PSUs that do *not* have to be combined with other PSUs in order to form strata for sampling. This level of geography is not available on the public use files.

sequential hot-deck procedure

See *hot-deck procedure.*

short waves

Waves that contain three rotation groups instead of the standard four.

skip patterns

Mechanisms embedded in the survey that allow the interviewer to skip over irrelevant questions and call up the next relevant question.

source and accuracy statement

A statement included with the technical documentation that accompanies public use files; it contains detailed information about weights on the files, when and how to make adjustments to the weights, and how to use generalized variance procedures to compute standard errors for some common types of estimates. It also includes cautions for users about sources of nonsampling error.

Survey of Program Dynamics (SPD)

An offshoot of SIPP that began recontacting members of the 1992 and 1993 Panels, with data collection to continue through 2001 in order to collect 10 years of data.

Surveys-on-Call

An on-line data access tool available on the SIPP Web site. Surveys-on-Call allows users to define microdata extracts from SIPP public use files through the 1993 Panel.

T

technical documentation

Information that accompanies microdata files and that includes a description of file contents, a glossary, codes, a data dictionary, a source and accuracy statement, and a copy of the core questions for the panel in question.

time-in-sample effect

Tendency of sample members to "learn" the survey over time, possibly resulting in altered responses.

topcoding

Practice of recoding income variables to protect against the possibility that a user might recognize the identity of a SIPP respondent with very high income. Incomes exceeding a maximum value are recoded to that maximum value or to a mean of responses in excess of that value.

topical content

Questions that are not repeated in every wave. They cover a wide range of topics and can occur once or more than once in a panel. The questions are grouped into modules by topic.

topical module files

Files containing all topical module data from the wave in question.

topical modules

Collections of questions asked periodically, but not at every interview, about various topics that might be outside the range of the core content.

topical module imputation procedure

Missing data in topical modules are imputed using the same hot-deck procedure used to impute missing data in the core questionnaire.

Type A noninterview

Households that are occupied by people eligible for interview but for which no interview is obtained.

Type B noninterview

A household noninterview that occurs when the address unit is vacant or in some way unfit for residence.

Type C noninterview

In Wave 1, a household noninterview that occurs when the housing unit has been demolished or converted to some other use; in subsequent waves, a household noninterview that occurs when all sample members in a household are outside the scope of the survey, for example, deceased, living abroad, living in institutions, or living in armed forces barracks.

Type D noninterview

Households or people who have moved to an unknown address, or who have moved more than 100 miles from the nearest field representative and for whom no telephone interview is conducted. This type of noninterview applies only to Wave 2 and beyond.

Type Z imputation

Procedures used to impute missing data for Type Z noninterviews and for situations when a person was in sample early in the wave but not in sample by the month of interview.

Type Z noninterview

An eligible person in an interviewed household from whom the field representative could not get an interview or for whom the interviewer could not obtain a proxy interview. A noninterview also occurs when a person who was part of the household for a portion of the reference period moves and is no longer a household member on the date of the interview. If the person is an original sample member, an effort will be made to locate and follow the person.

U

undercoverage

Underrepresentation of demographic subgroups within the surveyed population.

unrelated subfamily

A family, that is, a group of two or more related individuals, living at a sample address unit that does not contain the reference person or anyone related to the reference person.

User Notes

Issued periodically by the Census Bureau, these contain updated information for specific microdata files.

usual place of residence

Place where a person normally lives and sleeps; specific living quarters held for the person, to which he or she is free to return at any time.

V

variable metadata

Provides a complete characterization of a variable's content. Variable metadata are available on the SIPP Web site.

W

wave

One round of interviewing, which takes 4 months to complete; one fourth of the sample (i.e., a rotation group) is interviewed each month.

wave files

See *core wave files.*

weights

Estimates of the number of units in the target population that a given unit represents.

X

Y

Z

References

Allen, T. M., Petroni, R. J., and Singh, R. P. (1993). The effectiveness of oversampling low-income households in the Survey of Income and Program Participation, U.S. Bureau of the Census, Washington, DC. *Proceedings of the American Statistical Association*. Alexandria, VA: American Statistical Association.

Brick, J. M., and Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research* 5, 215–238.

Bye, B., and Gallicchio, S. (1989). Two Notes on Sampling Variance Estimates from the 1984 SIPP Public-Use Files. SIPP Working Paper No. 8902. Washington, DC: U.S. Bureau of the Census.

Citro, C. F., Hernandez, D., and Herriot, R. (1986). Longitudinal household concepts in SIPP: Preliminary results. *Proceedings of the Bureau of the Census Second Annual Research Conference,* Washington, DC: U.S. Department of Commerce, pp. 598-619. (Also available as SIPP Working Paper No. 8611, Washington, DC: U.S. Bureau of the Census.)

Citro, C. F., and Kalton, G. (1993). *The Future of the Survey of Income and Program Participation*. Washington, DC: National Academy Press.

Citro, C. F., Michael, R. T., and Maritano, N. (eds.) (1995). *Measuring Poverty: A New Approach*. Washington, DC: National Academy Press, Appendix B.

Coder, J., and Scoon-Rogers, L. S. (1996). Evaluating the Quality of Income Data Collection in the Annual Supplement to the March Current Population Survey and the Survey of Income and Program Participation. SIPP Working Paper No. 9604. Washington, DC: U.S. Census Bureau.

Doyle, P., and Dalrymple, R. (1987). The impact of imputation procedures on distribution characteristics of the low income population. *Proceedings of the Bureau of the Census Third Annual Research Conference.* Washington, DC: U.S. Department of Commerce, pp. 483–508. (Also available as SIPP Working Paper No. 8710, Washington, DC: U.S. Census Bureau)

Duncan, G., and Hill, M. (1985). Conceptions of longitudinal households: Fertile or futile? *Journal of Economic and Social Measurement* 13, 361–376.

Eargle, J. (1990). *Household Wealth and Asset Ownership: 1988.* Current Population Reports P70-22. Washington, DC: U.S. Census Bureau.

Guo, G. (1993). Event-history analysis for left-truncated data. *Sociological Methodology* 23, 217–243.

Huggins, V. J., and King, K. E. (1997). Evaluation of oversampling the low-income population in the 1996 Survey of Income and Program Participation (SIPP), U.S. Bureau of the Census, Washington, DC. *Proceedings of the American Statistical Association, Survey Research Methods Section*. Anaheim, CA: American Statistical Association.

Jabine, T., King, K., and Petroni, R. (1990). *SIPP Quality Profile,* 2nd Ed. Washington, DC: U.S. Census Bureau.

Jinn, J. H., and Sedransk, J. (1987). Effect on secondary data analysis of different imputation methods. *Proceedings of the Bureau of the Census Third Annual Research Conference.* Washington, DC: U.S. Department of Commerce, pp. 509–530.

Kalbfleisch, J. D., and Prentice, R. L. (1980). *The Analysis of Failure Time Data.* New York: John Wiley & Sons.

Kalton, G., and Brick, J. M. (1995). *Survey Methodology*, 21, 33-44.

Kalton, G., and Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology* 12(1), 1–16.

Kalton, G., Lepkowski, J., Heeringa, S., Lin, T., and Miller, M. E. (1987). The Treatment of Person-Wave Nonresponse in Longitudinal Surveys. SIPP Working Paper No. 8704. Washington, DC: U.S. Census Bureau.

Kalton, G., Miller, D. P., and Lepkowski, J. (1992*).* Analyzing Spells of Program Participation in the SIPP. SIPP Working Paper No. 9210 (171). Washington, DC: U.S. Census Bureau.

Kalton, G., Winglee, M., and Jabine, T. (1998). *SIPP Quality Profile,* 3rd Ed. Washington, DC: U.S. Census Bureau.

King, K., Petroni, R., and Singh, R.P. (1987). *SIPP Quality Profile.* Washington, DC: U.S. Census Bureau.

Lepkowski, J., and Bowles, J. (1996). Sampling error software for personal computers. *Survey Statistician* 35, 10–17.

Lepkowski, J. M., Landis, R. L., and Stehouwer, S. A. (1987). Strategies for the analysis of imputed data from a sample survey. *Medical Care* 25(8), 705–716.

Little, R. J. A. (1986). Missing data in Census Bureau surveys. *Proceedings of the Bureau of the Census Second Annual Research Conference.* Washington, DC: U.S. Department of Commerce, pp. 442–454.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, pp.129–139.

Marquis, K. H., and Moore, J. C. (1989a). Response errors in SIPP: Preliminary results. *Proceedings of the Bureau of the Census Fifth Annual Research Conference.* Washington, DC: U.S. Department of Commerce, pp. 515–536.

Marquis, K. H., and Moore, J. C. (1989b). Some response errors in SIPP—with thoughts about their effects and remedies. *Proceedings of the, American Statistical Association, Survey Research Methods Section.* Anaheim, CA: American Statistical Association, pp. 381–386.

Marquis, K. H., and Moore, J. C. (1990). Measurement errors in SIPP program reports. *Proceedings of the U.S. Bureau of the Census' 1990 Annual Research Conference.* Washington, DC: U.S. Department of Commerce, pp. 721–745.

Marquis, K. H., Moore, J. C., and Huggins, V. J. (1990). Implications of SIPP Record Check results for measurement principles and practice. *Proceedings of the American Statistical Association, Survey Research Methods Section*. Anaheim, CA: American Statistical Association, pp. 564–569.

McCormick, M. K., Butler, D. M., and Singh, R. P. (1992). Investigating time in sample effect for the Survey of Income and Program Participation. Paper prepared for the American Statistical Association Annual Meeting. Washington, DC: U.S. Census Bureau.

McMillen, D., and Herriot, R. (1985). Toward a longitudinal definition of households. *Journal of Economic and Social Measurement* 13, 504–509. (Also available as SIPP Working Paper No. 8402. Washington, DC: U.S. Census Bureau.)

McNeil, J. (1988). CPS and SIPP Estimates of Health Insurance Coverage Status. Census Bureau Internal Memorandum, May 3.

Moore, J.C. (1988). Self/proxy Response Status and Survey Response Quality—A Review of the Literature. *Journal of Official Statistics* 4, 155–172.

Pennell, S. G. (1993). *Cross-Sectional Imputation and Longitudinal Editing Procedures in the Survey of Income and Program Participation.* Prepared by the University of Michigan Survey Research Center, Ann Arbor. Washington, DC: U.S. Census Bureau.

Pennell, S. G., and Lepkowski, J. M. (1992). Panel Conditioning Effects in the Survey of Income and Program Participation. *Proceedings of the American Statistical Association, Survey Research Methods Section*. Alexandria, VA: American Statistical Association, pp. 566–571.

Ruggles, P., and Williams, R. (1989). Measuring the Duration of Poverty Spells. SIPP Working Paper No. 8909. Washington, DC: U.S. Census Bureau.

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics* 1, 381–397.

Sedransk, J. (1985) The objectives and practice of imputation. *Proceedings of the Bureau of the Census First Annual Research Conference.* Washington, DC: U.S. Census Bureau, pp. 445–452.

Shapiro, G. M., Diffendal, G., and Cantor, D. (1993). Survey Undercoverage: Major Causes and New Estimates of Magnitude. Census Bureau Internal Memorandum.

Shea, M. (1995a). *Dynamics of Economic Well-Being: Poverty 1990–1992*. Current Population Reports P70-112. Washington, DC: U.S. Census Bureau.

Shea, M. (1995b). *Dynamics of Economic Well-Being: Program Participation, 1990 to 1992* Current Population Reports P70-41. Washington, DC: U.S. Census Bureau.

Skinner, C. J., Holt, D., and Smith, T. M. F. (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons.

Tuma, N. B., and Hannan, M. T. (1984). *Social Dynamics, Models and Methods.* Orlando, FL: Academic Press.

U.S. Census Bureau (1991). *Survey of Income and Program Participation Users' Guide,* 2nd Ed. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau (1993). *Survey of Income and Program Participation Initial Training Guide.* Washington, DC: U.S. Census Bureau.

U.S. Census Bureau (1994). SIPP Information Booklet: 1990 and 1991 Panels. Form SIPP-7004A. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau (1998a). *Survey of Income and Program Participation Quality Profile,* 3rd Ed. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau (1998b). The Current Population Survey: Design and Methodology. Technical Paper 63. Washington, DC: U.S. Census Bureau.

Waite, P.J. (1996). SIPP (1996) Specifications for Interview Mode Flag. Internal Census Bureau Memorandum to Chester Bowie, May 17th.

Williams, T., and Bailey, L. (1996). Compensating for Missing Wave Data in the SIPP. SIPP Working Paper No. 9605. Washington, DC: U.S. Census Bureau.

Index

Accessing SIPP information. *See also* Information resources
published estimates, **1**-5–**1**-6, **5**-1, **5**-2–**5**-3
Activities of Daily Living (ADL) instrument, **3**-10, **3**-11
Additional household members. *See also* Household composition
births, **2**-14, **8**-5, **8**-7, **8**-17, **9**-5, **9**-8, **10**-25, **13**-16, **13**-17
defined, **C**-13
following rules, **1**-4, **2**-1, **2**-9, **C**-13
identification, **9**-3, **10**-8, **10**-25, **11**-13, **11**-14, **12**-14, **12**-24–**12**-25
imputation of records, **4**-6–**4**-7, **10**-36
interview procedures, **2**-16, **2**-17
movers, **4**-6–**4**-7, **8**-6, **10**-8, **10**-20, **11**-24, **12**-24–**12**-25
weighting adjustment, **8**-5, **8**-7, **8**-17, **9**-5, **9**-8
Address. *See also* Current Address IDs; Entry Address IDs
clusters, **2**-6, **8**-4, **8**-5, **10**-8, **11**-13, **C**-2
enumeration districts frame. *See* Unit frame
screening, **2**-6
subsampling, **2**-6
units, **2**-6, **2**-10, **2**-18, **12**-14, **E**-1
Adjustment cells, **4**-8–**4**-9, **4**-12
Administrative records, responses compared to, **6**-3–**6**-4
Age
core wave file structure, **13**-7
following rules, **2**-9, **2**-12, **10**-25, **11**-24, **12**-26, **13**-15
imputation, **10**-37
job or business started, **B**-5
population status based on, **11**-12
at receipt of Social Security Disability benefits, **B**-5
respondents, **1**-2, **2**-7, **2**-16, **3**-1, **3**-6, **3**-7, **3**-9, **3**-10, **11**-6, **11**-10
topcoding, **4**-17, **B**-4–**B**-5
variable name, **11**-11, **11**-12
weighting, **8**-5, **C**-3–**C**-4, **C**-6–**C**-8
Aging population, **5**-16
Aid to Families with Dependent Children (AFDC)
authorized recipient, **10**-28, **12**-30, **12**-31
coverage, **12**-30, **12**-31
history, **3**-15
ID variables, **9**-7, **9**-14, **10**-27, **10**-28, **10**-29, **10**-30–**10**-31, **12**-29, **12**-30, **12**-31
misinterpretation of questions on, **6**-3
replacement with TANF, **1**-3, **9**-7, **10**-27
weights, **8**-2
Algorithms
calendar-year and panel weight generation, **C**-17
family identification variables, **12**-17, **12**-18
monthly program income variables, **12**-30, **12**-36
reference months aligned to calendar months, **12**-9, **12**-10
second-stage calibration, **C**-4–**C**-12, **C**-16, **C**-23
topcoding, **10**-33–**10**-34
Alimony payments, **3**-3, **3**-6
Allocation flags, **4**-11, **4**-13–**4**-14, **4**-15, **10**-36–**10**-37, **11**-28, **12**-37, **13**-8, **13**-22
American Statistical Association (ASA), **1**-14, **5**-15
Area enumeration districts frame. *See* Area frame
Area frame, **2**-5–**2**-6
Asset ownership
comparison of surveys, **1**-9, **1**-10
core questions, **3**-3–**3**-4, **3**-5, **3**-6, **3**-8
errors in estimates, **6**-4, **13**-12
household, **C**-15
imputation, **4**-4, **4**-7, **4**-9
income, **3**-3–**3**-4, **3**-5, **3**-6, **3**-13, **10**-29, **10**-32
information resources, **5**-2, **5**-3, **5**-16, **13**-12
joint, **3**-4, **3**-8
municipal/corporate bonds, **10**-29
nonresponse, **6**-2, **C**-18
topcoding, **11**-28, **B**-6–**B**-7
topical modules, **3**-6, **3**-8, **3**-13, **3**-14
Associated sample persons, **C**-13, **C**-14
Attrition
bias, **1**-6, **1**-7, **2**-2, **6**-3
confounding with time-in-sample bias, **6**-3
defined, **E**-9
and merging files or data, **13**-16, **13**-17, **13**-20–**13**-21
by panel, **2**-19
spell construction, **8**-19
total sample, **2**-17–**2**-18
weighting adjustments, **8**-4, **8**-19, **13**-22

Balanced repeated replications, **7**-2, **7**-3
Basic needs information, **3**-8, **3**-10, **5**-3
Benefits
electronic transfer of, **3**-15
employer-provided, **3**-4, **3**-8, **3**-9–**3**-10
offered solely to children, **10**-27, **10**-28, **12**-29
topical modules, **3**-8
Bias
attrition, **1**-6, **1**-7, **2**-2, **6**-3, **13**-20–**13**-21
in imputation of missing data, **13**-20–**13**-21
linking families or households, **13**-1–**13**-2
multivariate statistics, **13**-20–**13**-21
nonmetropolitan samples, **10**-39
nonresponse, **2**-17, **4**-2, **6**-1
sampling error estimation, **1**-7, **2**-5
selection, **13**-21
standard error estimates, **2**-5, **13**-21
systematic, **6**-3
time-in-sample, **1**-7, **2**-2, **6**-3, **8**-19
undercoverage of subpopulations, **C**-17
unweighted analyses, **8**-1, **8**-2, **9**-8
Bibliography, online, **1**-13, **5**-15
Birth year, bottomcoding, **B**-4, **B**-7
Births
errors in estimates, **6**-4
ID variables, **10**-25, **11**-24, **12**-26
order of, **3**-10
to original sample members, **2**-14, **10**-25, **11**-24, **13**-16, **13**-17
to single mothers, **8**-19
weighting adjustments, **8**-5, **8**-7, **8**-17, **9**-5, **9**-8
Boarding houses, **2**-6, **10**-17, **12**-15
Bottomcoding, **4**-17, **B**-4
Building permits, **2**-6
Bureau of Labor Statistics (BLS), **1**-9, **5**-13
Business. *See also* Employers; Self-employment
characteristics, **4**-14
ownership, **3**-3, **3**-8

Calendar month
alignment of data by, **8**-19, **12**-7, **12**-9, **12**-10, **12**-11–**12**-12, **13**-4
estimates, **8**-12, **8**-14–**8**-16, **8**-19, **9**-8, **9**-9, **10**-7
format, **10**-7
interview month correspondence, **13**-13
topcodes, **10**-36, **12**-37
weights, **8**-12, **8**-14–**8**-15, **8**-19, **9**-8, **12**-7, **13**-1, **13**-8
Calendar year
estimates, **8**-18, **9**-8, **11**-21
weights, **8**-3, **8**-7–**8**-8, **8**-16–**8**-17, **8**-18, **9**-5, **9**-8, **12**-37–**12**-38, **13**-21, **C**-17–**C**-25
calendar year estimates, **8**-18, **C**-17–**C**-25
Callbacks, **2**-17, **2**-21
Census Region, **8**-5
Censuses of the Population
Decennial, **2**-6, **2**-8
CHAMPUS, **9**-14, **10**-27, **12**-29
CHAMPVA, **9**-14
Child care
foster care, **9**-14, **10**-27, **12**-29
ID variables, **9**-14, **10**-27
information resources, **5**-2, **5**-3, **5**-16
topical modules, **3**-7, **3**-8–**3**-9
Child support
agreements, **3**-9
income, **3**-3
paid, **1**-10, **3**-9, **3**-15, **12**-37
pass-through payments, **3**-5, **3**-9
topcoded payments, **12**-37
topical modules, **3**-7, **3**-9, **3**-15
Children. *See also* Births; Infants
benefits offered solely to, **10**-27, **10**-28, **12**-29
core wave file records, **10**-6
custodial arrangements, **3**-9, **3**-14
disability, **10**-28, **10**-29, **10**-30–**10**-31, **12**-30
following rules, **1**-4, **2**-9
foster, **9**-14, **10**-16, **10**-17, **10**-27, **11**-20
health status, **3**-11
imputation of program participation, **10**-28, **12**-28
income, **3**-6
interview procedures, **2**-17, **3**-1
living arrangements, **5**-2
moves without parents, **C**-19
of original sample members, **10**-6
P-70 publications, **5**-2
parents linked to, **10**-7, **11**-13, **11**-16, **12**-13
paternity establishment status, **3**-9
program units, coverage, and recipiency, **10**-29, **10**-30–**10**-31, **12**-29
relationship to reference person, **10**-16, **10**-17, **10**-18, **11**-20
special education services, **3**-11
topical modules, **3**-9, **3**-10–**3**-11
weighting adjustments, **8**-17, **C**-4, **C**-7, **C**-10, **C**-19, **C**-24–**C**-25
well-being, **3**-7, **3**-9, **5**-16, **11**-21
Clustering of addresses, **2**-6, **8**-4, **8**-5, **10**-8, **11**-13, **12**-14, **C**-2
Cold-deck values, **4**-8, **4**-11–**4**-12, **E**-1
College students, **2**-16
Computer-assisted interviewing (CAI)
advantages over paper instrument, **3**-1, **4**-15, **8**-6
case management features, **3**-1, **3**-2, **3**-3, **13**-13
data editing, **1**-3, **1**-5, **2**-17, **4**-6, **4**-15

defined, **E-**1
mode of interviewing, **6-**2
quality of data, **1-**3, **3-**1, **6-**2, **8-**16
questionnaire documentation, **5-**14, **11-**2, **12-**2
skip patterns, **2-**17, **3-**2, **10-**2, **10-**6, **11-**2
variable name changes, **10-**6

Computer-assisted personal interviewing (CAPI), **6-**2, **E-**1

Confidentiality. *See also* Topcoding
bottomcoding, **4-**17
core wave files, **10-**38–**10-**39
employment information, **4-**17
geographic information, **4-**17, **5-**1, **10-**8, **10-**38–**10-**39, **11-**13, **12-**14
procedures for public use files, **1-**5, **4-**4, **4-**5, **4-**17–**4-**18, **7-**2, **10-**6, **10-**8, **11-**13, **12-**14
telephone interviews, **2-**17

Consolidated Metropolitan Statistical Areas (CMSAs), **10-**39

Control cards, **3-**2, **4-**6, **8-**6, **E-**2

Control date, **8-**7, **8-**16

Control file, **4-**15

Core content
asset ownership, **3-**3–**3-**4, **3-**5, **3-**6, **3-**8
defined, **3-**1, **E-**2
earnings, **3-**3, **3-**4, **3-**5
income amounts, **1-**8, **3-**6
labor force status, **3-**3, **3-**4
1996 and subsequent panels, **3-**3–**3-**4
overview, **3-**2
pre-1996 panels, **3-**2, **3-**4–**3-**6
program participation, **1-**8, **3-**3, **3-**4, **3-**5, **3-**6
topics, **1-**4, **3-**3–**3-**6
unearned income, **3-**3–**3-**4

Core data, **2-**3, **4-**5, **9-**7, **9-**9, **11-**8

Core items
coverage, **1-**4
defined, **3-**1
full panel files, **1-**8, **12-**6, **13-**1
imputation, **4-**6–**4-**7, **4-**13, **11-**9
topical module files, **1-**8, **11-**10

Core questionnaire, **2-**3, **3-**1, **3-**2–**3-**6

Core wave files
allocation flags, **4-**13–**4-**14, **10-**36–**10-**37
calendar month estimation, **8-**12, **8-**14, **8-**19, **9-**8, **10-**7
confidentiality procedures, **10-**38–**10-**39
content, **1-**8, **5-**4
creation, **4-**3, **4-**4
cross-wave consistency, **4-**15
data dictionary, **9-**11, **10-**2–**10-**4, **10-**5, **10-**35, **12-**3, **13-**18, **13-**19
defined, **E-**2
edits, **4-**4, **4-**15, **8-**16, **10-**37, **12-37**, **13-**6–**13-**7, **13-**14
family characteristics, **9-**12
family composition variables, **9-**13, **9-**15, **10-**15–**10-**20
family identification, **9-**6, **9-**12, **10-**11–**10-**14, **10-**21, **12-**17
full panel files compared, **9-**11–**9-**15, **10-**37, **12-**6, **12-**10, **12-**17, **12-**30, **12-**37, **13-**1, **13-**14
household composition variables, **9-**11, **9-**12, **9-**13, **9-**15, **10-**8, **10-15**–**10-**20, **10-**23–**10-**24, **11-**19
household identification, **9-**11, **10-**9–**10-**11
ID variables, **9-**3, **9-**12, **10-**6–**10-**14, **10-**20–**10-**28, **10-**29–**10-**30, **11-**11–**11-**12, **11-**13, **11-**23, **13-**9, **13-**23
imputation procedures, **4-**2, **4-**4, **4-**6–**4-**7, **4-**13, **8-**16, **9-**15, **10-**6, **10-**25, **10-**36–**10-**37, **11-**9, **12-**10, **12-**17, **12-**37, **13-**6–**13-**7, **13-**14
income variables, **9-**12, **10-**19–**10-**20, **10-**21, **10-**27, **10-**37
linking between two or more, **4-**5, **5-**4, **13-**4, **13-**6–**13-**8
linking with full panel files, **1-**9, **12-**28, **13-**8–**13-**11
linking with topical module files, **1-**9, **13-**12–**13-**14
longitudinal analysis of data from, **13-**6–**13-**7, **13-**8
merging data within, **1-**9, **12-**13, **13-**3–**13-**4, **13-**5–**13-**6
merging with full panel files, **10-**6, **12-**1, **12-**6, **12-**17, **12-**20, **12-**28, **12-**30, **13-**1, **13-**3, **13-**4
merging with topical module files, **1-**8, **3-**10, **9-**6, **9-**9, **10-**6, **11-**1, **11-**7, **11-**8, **11-**10, **11-**11, **11-**13, **11-**17, **11-**19, **12-**6, **12-**13, **13-**1, **13-**3, **13-**4, **13-**12, **13-**13, **13-**14, **13-**15
merging two or more, **10-**1, **10-**6, **12-**13
metropolitan area identification, **9-**15, **10-**38–**10-**39
monthly interview status variable, **9-**4, **9-**5, **9-**11, **11-**9, **11-**12
mover identification, **10-**8, **10-**20, **10-**22–**10-**26, **11-**23, **13-**23
overview, **1-**8
person identification, **9-**11, **9-**15, **10-**6–**10-**9, **11-**11, **13-**9, **13-**23
person-month format, **1-**8, **5-**4, **5-**5, **8-**8, **9-**1, **9-**3, **9-**5, **9-**6, **9-**11, **10-**6, **10-**7, **10-**25, **11-**7, **13-**2, **13-**3–**13-**4, **13-**5–**13-**6, **13-**7, **13-**9, **13-**13, **13-**15
person nonresponse in, **4-**2, **13-**22
person-record format, **9-**4, **9-**5, **9-**7, **9-**11, **10-**6, **10-**7, **13-**3–**13-**4, **13-**5–**13-**6
previous wave variables, **11-**27, **13-**23
program unit identification, **9-**14, **10-**26–**10-**29, **10-**30–**10-**31

public use version, **4-**4, **9-**1–**9-**2, **9-**3, **10-**1–**10-**39
quarterly estimates, **8-**14–**8-**16
questionnaire correspondence to variables on, **10-**4–**10-**6
reference period, **9-**2, **10-**7, **11-**8, **13-**4, **13-**7
reformatting, **13-**3–**13-**4, **13-**5–**13-**6
sort order, **13-**3, **13-**4, **13-**6
state variable, **9-**15, **10-**38
structure, **5-**4, **5-**5, **8-**8, **9-**1–**9-**2, **9-**11, **10-**6, **10-**7, **11-**7, **12-**6, **13-**6–**13-**7
technical documentation, **10-**2–**10-**4
topcoding, **9-**15, **10-**6, **10-**29, **10-**32–**10-**36, **11-**28
topical module files compared, **9-**11–**9-**15, **11-**7, **11-**8, **11-**11–**11-**12, **13-**13
uses, **5-**4
variable names, **9-**1, **9-**13, **10-**1, **10-**4, **10-**8, **10-**11, **11-**11–**11-**12, **A-**1–**A-**34
variance estimation variables, **7-**3
weighting procedures, **5-**4, **8-**8–**8-**16, **10-**37
weights, **5-**4, **8-**3, **8-**4–**8-**5, **8-**7, **8-**8–**8-**13, **9-**8, **9-**15, **10-**1, **10-**2, **13-**8, **13-**22, **C-**1–**C-**25
wide-record format, **13-**7

Coverage
core items, **1-**4
CPS, **1-**9
housing units, **2-**6
improvement frame, **2-**6
ratio, **1-**6, **6-**1
transfer program unit, **4-**16, **9-**14, **10-**26–**10-**28, **10-**29, **10-**30–**10-**31, **12-**28, **12-**30–**12-**31

Cross-sectional analyses
core wave files, **5-**4
defined, **E-**2
editing and imputation, **4-**1, **4-**8, **4-**9
full panel files, **12-**7
quarterly estimates, **8-**16
sample size and, **2-**2
seam effect and, **6-**3
weights, **8-**3, **8-**4, **8-**16, **C-**12–**C-**13

Cross-walks
reference periods, **10-**2, **11-**2, **12-**2
variables names for core wave files, **A-**1–**A-**34

Current Address IDs
components, **9-**3–**9-**4, **10-**20, **11-**22
core wave files, **9-**3, **10-**7, **10-**10, **10-**13–**10-**14, **10-**20, **10-**22, **10-**23–**10-**24, **11-**11, **11-**23
family identification, **10-**11, **10-**13–**10-**14, **10-**21, **11-**17, **11-**18, **12-**18, **12-**20
family-level income, **12-**23
by file type, **9-**3
full panel files, **12-**15, **12-**16, **12-**18, **12-**20, **12-**23, **12-**24–**12-**25, **12-**26, **12-**27
household composition, **9-**6, **10-**10, **10-**23–**10-**24, **11-**14, **11-**16, **11-**25–**11-**26, **12-**15, **12-**16, **12-**27
movers, **10-**20, **10-**22, **10-**23–**10-**24, **10-**25, **11-**22, **11-**23, **11-**24, **12-**23, **12-**24–**12-**25, **12-**26, **12-**27
newborns, **10-**25, **12-**26
split households, **9-**3, **11-**22, **12-**28
topical module files, **9-**3, **11-**7, **11-**10, **11-**11, **11-**14, **11-**15, **11-**16, **11-**17, **11-**18, **11-**22, **11-**26
transfer program unit composition, **9-**8
variable names, **9-**3, **10-**10, **11-**11, **12-**15

Current Population Reports, **1-**13

Current Population Survey (CPS), **1-**1, **1-**9, **1-**10, **6-**4, **C-**3–**C-**4, **C-**8, **C-**9, **C-**16, **C-**20, **C-**24, **C-**25, **E-**2

Data Access and Dissemination System (DADS), **5-**12

Data collection procedures, **5-**16, **6-**2

Data dictionary
accuracy of definitions, **11-**6, **12-**3
contents, **4-**13, **5-**14, **10-**2, **11-**2, **12-**2–**12-**3
core wave files, **9-**11, **10-**2–**10-**4, **10-**5, **10-**35, **12-**3, **13-**18, **13-**19
corrections to, **5-**14
defined, **E-**2
differences by file types, **9-**11, **12-**3
excerpts from, **10-**3–**10-**4, **11-**3–**11-**4, **12-**4, **13-**18, **13-**19
exiting sample member variables, **13-**18–**13-**19
format, **10-**2–**10-**4, **11-**3–**11-**4, **12-**3–**12-**5
full panel files, **9-**11, **12-**2–**12-**5, **12-**31, **13-**19
machine-readable version, **10-**2, **11-**2, **12-**3
questionnaire correspondence to, **10-**4–**10-**6, **11-**6, **12-**5–**12-**6
SAS and FORTRAN syntax, **10-**4, **10-**5, **11-**4, **11-**5, **12-**3, **12-**5
topcodes, **10-**35, **12-**31
topical module files, **9-**11, **11-**2–**11-**5, **11-**6, **12-**3
universe definitions, **10-**3, **10-**6, **11-**4, **11-**6, **12-**3
variable metadata, **5-**15
variable name–content correspondence, **10-**6

Data editing
advantages over imputation, **4-**3
allocation flags, **4-**13, **10-**37
CAI, **1-**3, **1-**5, **2-**17, **4-**6, **4-**15
confidentiality-related, **4-**17
core wave files, **4-**4, **4-**15, **8-**16, **10-**37, **12-**37, **13-**1, **13-**6–**13-**7
cross-sectional, **4-**1
defined, **E-**2
effect on analyses, **4-**15, **8-**16, **13-**1, **13-**6–**13-**7, **13-**8, **13-**12
full panel files, **1-**5, **4-**3, **4-**5, **4-**14, **4-**15–**4-**16, **12-**7, **12-**37, **13-**1, **13-**8

geographic information, **4-**17–**4-**18
for internal consistency, **4-**4, **10-**37
item nonresponse from, **2-**21
longitudinal, **1-**5, **4-**1, **4-**4, **4-**5, **4-**14, **4-**15–**4-**16
paper questionnaires, **2-**17, **4-**6
procedures, **4-**1, **4-**4, **4-**8, **4-**15–**4-**16
topcoding, **1-**5, **4-**17
topical modules, **4-**4, **13-**12
uses, **2-**21, **4-**1, **4-**3

Data entry, **4-**2, **4-**6

Data Extraction System (DES), **5-**12

Data processing. *See also* Data editing; Imputation
overview, **4-**3–**4-**5
phase 1, **4-**3, **4-**4–**4-**5, **4-**6–**4-**14
phase 2, **4-**3, **4-**5, **4-**15–**4-**16

Deaths, **8-**4, **8-**5, **8-**7, **9-**5, **9-**8, **11-**11, **12-**13, **13-**16, **13-**17, **13-**19

Department of Health, Education, and Welfare, **1-**1

Dependent care, **3-**8

Design of SIPP. *See also* Redesign (1996) of SIPP; Sample design
comparison with other surveys, **1-**9–**1-**11
evolution, **1-**1–**1-**2
features, **1-**2–**1-**3
information resources, **5-**16
organizing principles, **2-**1–**2-**5
topics, **1-**4–**1-**5, **2-**1

Disability
children, **3-**11, **10-**28, **10-**29, **10-**30–**10-**31, **12-**30
functional limitations, **3-**10–**3-**11, **5-**2
history, **3-**15
income, **3-**3, **3-**5, **12-**30
long-term care needs, **3-**12
medical expenses, **3-**12
P-70 publications, **5-**2, **5-**3
topical modules, **3-**7, **3-**10, **3-**11
work-related, **3-**11, **3-**12, **3-**15

Divorces, **6-**4

Earnings. *See also* Income, earned; Wages and salaries
annual, **3-**8
core questions, **3-**3, **3-**4, **3-**5
information resources, **5-**16
misinterpretation of questions about, **6-**3
self-employed, **10-**32
topcoding, **10-**32–**10-**35, **12-**37, **B-**1–**B-**4, **B-**7
topical modules, **3-**8

Edits. *See* Data editing

Education and training
financial assistance, **3-**4, **3-**5, **3-**14, **5-**2
history, **3-**4, **3-**9, **3-**14, **11-**12, **11-**28
household characteristics, **8-**6
information resources, **5-**2, **5-**16
noninterview adjustments, **C-**18
topical modules, **3-**7, **3-**9, **3-**10, **3-**14, **11-**12

Eligibility, program, **3-**8, **3-**15, **10-**38, **11-**29, **12-**38

E-M algorithm, **13-**21

Emigration, **8-**5

Employers
characteristics, **3-**3, **10-**36, **10-**37
health benefits provided by, **3-**4, **3-**8, **3-**9–**3-**10
maternity leave policies, **3-**10
variables, **10-**5

Employment. *See also* Labor force status; Unemployment; Work
confidentiality procedures, **4-**17
core questions, **3-**3, **3-**4
gender differences, **5-**2
history, **3-**10
home-based, **3-**6, **3-**16
income, **10-**32–**10-**36
information resources, **5-**2, **5-**16
job offers for unemployed respondents, **3-**12
number in second business, **10-**6
pregnancy and, **3-**10
starting dates, **4-**17
topical modules, **3-**7, **3-**10, **3-**12, **3-**15–**3-**16
variables, **10-**5

Energy assistance, **3-**4, **3-**6

Energy usage, **3-**12

Entry Address IDs
changes in, **10-**26, **11-**13, **11-**27, **12-**14
components, **9-**4, **10-**8, **11-**14, **12-**14
core wave files, **9-**3, **10-**7, **10-**8, **10-**9, **10-**20, **10-**22, **10-**23–**10-**24, **11-**23, **13-**3, **13-**7
family-level income, **12-**23
full panel files, **9-**3, **12-**7, **12-**8, **12-**11–**12-**12, **12-**13, **12-**14, **12-**15, **12-**16, **12-**21, **12-**23–**12-**27
household identification, **12-**16
movers, **10-**8, **10-**20, **10-**22, **10-**23–**10-**24, **11-**14, **11-**22, **11-**23, **11-**24, **11-**25–**11-**26, **12-**23–**12-**27
newborns, **10-**25, **12-**26
purpose, **9-**3, **9-**4, **11-**14
redesign of 1996 and, **9-**4, **10-**7, **10-**8, **10-**9, **11-**13, **12-**13, **13-**3
sorting files for linking, **13-**3, **13-**4, **13-**9, **13-**14, **13-**15
spouses, parents, and guardians, **12-**21, **12-**22

topical module files, **9**-3, **11**-7, **11**-10, **11**-12, **11**-13, **11**-14, **11**-15, **11**-22, **11**-24, **11**-25–**11**-26, **11**-27
values, **10**-8
variable names, **9**-3, **11**-12
by wave, **10**-9

EPDJBTHN variable, **4**-14

EPPFLAG imputation, **4**-10, **4**-13, **4**-14, **10**-36–**10**-37

EPPINTVW field, **4**-13–**4**-14, **10**-36

Errors. *See also* Nonsampling errors; Sampling errors; Standard errors
imputation-related, **12**-7, **13**-7, **13**-8, **13**-12, **13**-14
information sources on, **1**-13
keying/recording, **4**-2
measurement, **6**-2–**6**-3, **13**-12
in microdata files, **5**-14
respondent recall, **2**-3, **6**-2

Evaluation studies, **6**-4

Event-history analysis, **8**-18, **13**-20

Expenditure data
comparison of surveys, **1**-10
medical, **3**-12
work-related, **3**-15

Family(ies). *See also* Subfamily
defined, **3**-11, **8**-11, **9**-6, **10**-11, **10**-12, **11**-16, **11**-17, **12**-16, **12**-17, **12**-18, **E**-3
disruption, **5**-2
grouping of, **10**-12
grouping people into, **12**-19
head of, **10**-15
identification, **3**-11, **9**-6, **9**-7, **9**-12, **10**-11–**10**-14, **10**-21, **11**-12, **11**-16–**11**-18, **12**-16–**12**-19, **12**-20, **12**-23
information resources, **5**-2, **5**-16
methods for distinguishing, **10**-12–**10**-14, **11**-17–**11**-18, **12**-17–**12**-18
number in household, **10**-15
primary, **3**-11, **8**-11, **8**-12, **9**-6, **9**-12, **10**-11, **10**-12, **10**-19, **10**-20, **10**-21, **11**-16, **11**-17–**11**-18, **12**-16, **12**-19, **12**-20, **12**-23, **E**-8
reference person, **3**-11, **8**-11–**8**-12, **9**-6, **10**-11, **10**-12, **10**-15, **10**-16
secondary, **9**-6, **10**-11, **11**-16, **12**-17, **12**-19, **E**-9
types, **8**-11, **9**-12, **10**-11, **10**-13–**10**-14, **10**-15, **11**-16–**11**-17, **12**-16–**12**-17, **12**-20, **12**-21, **C**-3
weights, **8**-4–**8**-5, **8**-6, **8**-8, **8**-11–**8**-12, **8**-13, **9**-15, **C**-3

Family characteristics
assigning to individuals, **13**-2
constructing, **9**-8, **12**-17, **12**-18
core wave files, **9**-12
income, **9**-12, **10**-19–**10**-20, **10**-21, **10**-35, **10**-36, **12**-23, **12**-37, **C**-18
merging files to obtain, **9**-6, **11**-13, **11**-17, **12**-17, **12**-20
support networks, **5**-2
topical modules, **3**-7, **3**-11, **9**-12
transfer program income recipient, **10**-7, **10**-27, **10**-28

Family composition
background information, **3**-10
core wave files, **9**-13, **9**-15, **10**-15–**10**-20
determining, **9**-6–**9**-7
excluding related subfamily members, **10**-12, **10**-13–**10**-14, **10**-15, **11**-12, **11**-17, **11**-18, **12**-19, **12**-20
full panel files, **9**-13, **9**-15, **12**-19–**12**-22
households, **8**-12, **8**-13
ID variables, **9**-6–**9**-7, **9**-12, **9**-13, **10**-11, **10**-12, **10**-19, **11**-17, **11**-18, **12**-18, **12**-20
including related subfamily, **10**-19–**10**-20, **10**-21, **10**-13–**10**-14, **11**-18, **12**-19, **12**-20, **12**-23
interrelationships, **10**-15, **10**-16, **12**-21, **C**-3–**C**-4, **C**-6–**C**-8
monthly, **9**-6–**9**-7, **9**-8, **12**-17–**12**-18, **12**-20
multigenerational household, **9**-7, **10**-12, **10**-18, **10**-19, **11**-21, **11**-22, **12**-19, **12**-22
one-person, **9**-6, **11**-17
restrictions on analyses, **12**-15, **12**-16
topical module files, **9**-6, **9**-12, **9**-13, **9**-15, **11**-16–**11**-18, **11**-19–**11**-21, **11**-22
variables, **9**-13, **9**-15, **10**-15–**10**-20, **11**-16–**11**-18, **11**-19–**11**-21, **11**-22, **12**-19–**12**-22

Fathers, **10**-15

Fay's method for variance estimation, **7**-3

Federal Reserve Board, **6**-4

FERRET, **1**-6, **5**-12, **5**-13, **7**-3, **E**-3

Fertility history, **3**-10, **5**-16

Financial data, topical modules, **3**-7

Following rules. *See also* Moves/movers
additional household members, **1**-4, **2**-1, **2**-9
age and, **2**-9, **2**-12, **10**-25, **11**-24, **12**-26, **13**-15
children, **1**-4, **2**-9
defined, **E**-3
example, **2**-10–**2**-14
excluded individuals, **2**-9
original sample members, **1**-4, **2**-7, **2**-9–**2**-15, **10**-25, **11**-24
temporarily absent members, **2**-15–**2**-16

Food stamps
history, **3**-15
ID variables, **9**-14, **10**-27, **10**-28, **12**-29, **12**-30, **12**-31
income, **3**-3, **4**-16, **10**-32, **12**-30, **12**-34–**12**-36
members of a common unit, **10**-28

program units, coverage, and recipiency, **9**-7, **10**-29, **10**-30–**10**-31, **12**-28, **12**-29, **12**-30, **12**-31
quarterly estimates, **8**-15–**8**-16
spell estimation, **8**-18
user-created monthly variables, **12**-30, **12**-34–**12**-36
weights, **8**-2

FORTRAN approach for file format change, **13**-3

FORTRAN syntax, **10**-4, **10**-5, **11**-4, **11**-5, **12**-5

Foster children, **9**-14, **10**-16, **10**-17, **10**-27, **11**-20, **12**-29

Frames, non-overlapping, **2**-6

Full panel files
allocation flags, **4**-14, **4**-15, **12**-37
attrition adjustments, **13**-22
calendar month alignment of data, **8**-19, **12**-7, **12**-9, **12**-10, **12**-11–**12**-12
calendar year estimates, **8**-18, **9**-8, **11**-21
content, **1**-8, **5**-12, **12**-6
core wave files compared, **9**-11–**9**-15, **10**-37, **12**-6, **12**-10, **12**-17, **12**-30, **12**-37, **13**-1, **13**-14
creation, **1**-5, **4**-3, **4**-4, **4**-5, **4**-15, **5**-12
data dictionary, **9**-11, **12**-2–**12**-5, **12**-31, **13**-19
data editing procedures, **1**-5, **4**-3, **4**-5, **4**-14, **4**-15–**4**-16, **12**-7, **12**-37, **13**-8, **13**-14
defined, **E**-3
family composition variables, **9**-13, **9**-15, **12**-19–**12**-22
family identification, **9**-6, **9**-7, **9**-12, **12**-16–**12**-19, **12**-20
format change, **5**-12, **13**-9–**13**-10
household composition variables, **9**-12, **9**-13, **9**-15, **12**-19, **12**-21–**12**-22, **12**-25, **12**-26
household identification, **9**-11, **12**-15–**12**-16
ID variables, **9**-3, **9**-12, **9**-14, **12**-6, **12**-23–**12**-28, **13**-9, **13**-15, **13**-23
imputation, **1**-8, **4**-3, **4**-5, **4**-14, **8**-17, **9**-15, **10**-37, **12**-7, **12**-10, **12**-17, **12**-37, **13**-8, **13**-11, **13**-14, **13**-22
income topcoding, **5**-1, **9**-15, **12**-31, **12**-36–**12**-37
income variables, **9**-12, **12**-23, **12**-30–**12**-31, **12**-32–**12**-36
linking with core wave files, **1**-9, **12**-28, **13**-8–**13**-11
linking with topical module files, **1**-9, **13**-14–**13**-15
metropolitan area identification, **12**-38
missing waves, **12**-10, **13**-22
merging with core wave files, **10**-6, **12**-1, **12**-6, **12**-17, **12**-20, **12**-28, **12**-30, **13**-1, **13**-3, **13**-4
monthly interview status variable, **1**-8, **9**-4, **9**-5, **9**-11, **11**-11, **12**-6, **12**-7, **12**-8, **12**-9–**12**-10, **12**-11–**12**-12, **12**-13, **12**-15, **12**-16, **12**-18, **12**-20, **12**-23, **12**-29
mover identification, **12**-23–**12**-27, **13**-23
1996, **4**-16, **9**-3, **9**-11–**9**-15, **13**-8, **13**-14
overview, **1**-8
person identification, **8**-17, **9**-11, **9**-15, **12**-13–**12**-15, **13**-23
person records, **8**-17, **9**-2, **9**-11, **9**-15, **13**-2
pre-1996, **4**-15–**4**-16, **7**-3, **9**-3, **9**-11–**9**-15, **12**-1–**12**-38
program unit identification, **9**-14, **12**-28–**12**-30
public use version, **4**-5, **5**-12, **9**-2, **9**-3, **12**-1–**12**-38
quarterly estimates, **8**-16
questionnaire correspondence with, **12**-5–**12**-6
release of, **9**-9
single files, **12**-1
spell estimations, **8**-18–**8**-19
state identification, **9**-15, **12**-38
structure, **5**-12, **9**-2, **9**-11, **11**-8, **12**-6–**12**-7, **12**-8, **12**-26, **12**-27, **13**-2
technical documentation, **12**-2–**12**-5, **12**-9
topical module files compared, **9**-11–**9**-15, **11**-8
variable name changes, **9**-3, **9**-15
variance estimation variables, **7**-3
weights, **8**-3, **8**-7–**8**-8, **8**-16–**8**-19, **9**-8, **9**-15, **12**-1, **12**-2, **12**-13, **12**-37–**12**-38, **13**-14, **13**-22, **C**-1–**C**-25

Functional limitations, **3**-10–**3**-11

Gender
imputation, **10**-37
and income topcoding, **10**-32, **10**-33, **B**-2, **B**-4
variable name, **11**-12
weighting adjustments, **C**-3–**C**-4, **C**-5, **C**-6–**C**-8

General Assistance (GA), **9**-7
ID variables, **9**-14, **10**-27, **12**-29
misinterpretation of questions on, **6**-3

General (G1) sources and amounts, **12**-30, **12**-31, **E**-3

General income questions, **3**-3

Generalized variance functions (GVFs), **5**-14, **7**-1
accuracy of estimates from, **7**-4
derivation, **7**-4
standard error of a mean, **7**-5–**7**-6
standard error of estimated number from, **7**-4–**7**-5

Geographic (GRIN) codes, **E**-3

Geographic information
sort variables for imputation, **4**-11
state-level, **4**-17–**4**-18, **10**-38, **11**-29
suppression, **4**-17, **5**-1, **10**-8, **10**-38–**10**-39, **11**-13

Group quarters, **8**-6, **8**-12, **9**-6, **10**-10, **11**-14, **11**-15, **11**-18, **12**-15, **12**-19, **12**-20, **C**-19, **E**-3

Group quarters frame, **2**-6
Guardians, **10**-15, **10**-19, **11**-12, **11**-19, **11**-21, **11**-22, **12**-21, **12**-22

Head of household, **2**-8, **8**-2
Health care
costs/expenditures, **3**-9, **3**-12
long-term, **3**-10, **3**-12
utilization, **3**-11, **3**-12
Health insurance coverage. *See also* Medicaid; Medicare
child support arrangements, **3**-15
characteristics of, **10**-26
data edits, **4**-16
errors in estimates, **6**-4
ID variables, **9**-14, **10**-27, **10**-29
information resources, **5**-2, **5**-3, **5**-16
time-specific data, **2**-4
topical modules, **3**-4, **3**-8, **3**-9–**3**-10, **3**-11, **3**-12, **3**-13
variables, **12**-29
Health status
children, **3**-11
disability, **3**-11, **3**-15
topical modules, **3**-7, **3**-9, **3**-11
Home-based employment, **3**-6
Home health care, **3**-11
Hospitalized persons, **2**-16
Hot-deck matrix, **4**-9–**4**-10, **4**-11, **4**-12, **E**-4
Hotel rooms, **2**-6
Household(s). *See also* Family
defined, **2**-6, **8**-10, **9**-6, **10**-9, **12**-15, **E**-4
enhanced, **C**-14
grouping of related primary families, **10**-12
identification, **9**-6, **9**-11, **10**-9–**10**-11, **11**-11, **11**-14, **11**-15, **12**-15–**12**-16
merged, **9**-11, **9**-12, **10**-25, **10**-26, **11**-27, **12**-28, **13**-16, **13**-22–**13**-23, **C**-14, **C**-15, **E**-6
number, by panel, **1**-2, **2**-2, **2**-8, **8**-20, **12**-7
recombined, **10**-26, **11**-27, **12**-28, **13**-22–**13**-23
split, **2**-11, **2**-12, **2**-14, **9**-3, **10**-12, **10**-13–**10**-14, **10**-20, **10**-26, **11**-18, **11**-22, **11**-24, **11**-27, **12**-23, **12**-24–**12**-25, **12**-28, **13**-22
types, **8**-12, **10**-15, **C**-3, **C**-6–**C**-8
weights, **8**-2, **8**-4–**8**-5, **8**-6, **8**-8, **8**-10–**8**-12, **8**-13, **9**-5, **9**-8, **9**-15
Household characteristics
assigning to individuals, **13**-2
caregiver members, **3**-11, **3**-12
constructing, **9**-8
economic, **3**-8, **5**-2, **5**-3, **7**-5, **8**-6, **9**-5, **10**-36, **10**-37, **11**-28, **12**-13, **12**-37, **13**-12, **B**-7, **C**-15, **C**-16
imputation, **10**-37, **C**-16
interview status of members, **9**-6, **11**-9, **12**-15, **12**-16
longitudinal analysis, **13**-2
merging files to obtain, **9**-6, **12**-28, **13**-22–**13**-23
program unit identification, **9**-7, **10**-28
reference person, **8**-10–**8**-11, **8**-12, **10**-11, **10**-12, **10**-15, **10**-16–**10**-19, **11**-6, **11**-12, **11**-16, **11**-17, **11**-19–**11**-21, **12**-17, **12**-21, **C**-15
size considerations, **8**-5, **8**-6, **9**-5, **12**-13, **C**-15
tenure, **8**-5, **8**-6, **C**-2, **C**-16
topical modules, **3**-7, **3**-12
weighting adjustments, **12**-13, **C**-2–**C**-3, **C**-15
Household composition. *See also* Additional household members; Family
calendar year weight and, **9**-5
changes in, **2**-10–**2**-14, **8**-5, **8**-10, **10**-11, **10**-20, **10**-23–**10**-24, **11**-14, **11**-22, **11**-24–**11**-27, **12**-16
core questions, **3**-11
core wave files, **9**-11, **9**-12, **9**-13, **9**-15, **10**-8, **10**-15–**10**-20, **10**-23–**10**-24
determining, **9**-6
full panel files, **9**-12, **9**-13, **9**-15, **12**-15, **12**-19, **12**-21–**12**-22, **12**-25, **12**-26
ID variables, **9**-6, **10**-23–**10**-24, **12**-15, **12**-16, **12**-25
identifying members, **2**-6–**2**-7, **9**-3, **9**-6, **10**-19, **11**-12
interrelationships, **3**-11–**3**-12, **9**-6, **10**-15, **10**-16
and linking topical module files, **13**-11–**13**-12
longitudinal edits, **4**-16
monthly, **9**-6, **9**-8
multigenerational family, **9**-7, **10**-12, **10**-18, **10**-19, **11**-21, **11**-22, **12**-22
number of families, **10**-15
reference period for, **11**-14
relationship to reference person, **11**-12, **12**-21
restrictions on analyses, **12**-15
rostering, **2**-7, **2**-16, **3**-2
temporarily absent members, **2**-15–**2**-16
topical modules, **9**-6, **3**-11, **10**-15
variables, **4**-16, **8**-10, **9**-11, **9**-12, **9**-13, **9**-15, **10**-8, **10**-10, **10**-15–**10**-20, **10**-23–**10**-24, **11**-19–**11**-21, **11**-22, **12**-15–**12**-16, **12**-19, **12**-21–**12**-22
weighting adjustments, **8**-10–**8**-11, **8**-18, **9**-5, **12**-13, **C**-6
Household Economic Studies, **1**-13–**1**-14
Household noninterview. *See* Household nonresponse
Household nonresponse
adjustment factors, **8**-5, **C**-2–**C**-3
defined, **E**-4
errors, **6**-1–**6**-2

interview attempts at subsequent waves, **2-**18
rate calculations, **2-**20
refusals, **11-**8, **C-**15
sources of, **2-**18, **C-**15
topical module files, **11-**8
Type A, **2-**18–**2-**20, **C-**2–**C-**3, **E-**13
Type B, **2-**18, **E-**13
Type C, **2-**18, **E-**13
Type D, **2-**18, **2-**19, **2-**20, **E-**12
by wave and panel, **2-**19
weights, **2-**20, **8-**5, **8-**6

Housemates/roommates, **10-**17, **11-**20

Housing
conditions, **3-**12
costs, **3-**7, **3-**8, **3-**12, **3-**14
subsidized, **3-**6
units, **1-**9, **2-**6, **2-**8–**2-**9, **2-**16, **2-**18, **9-**3, **10-**8, **10-**9–**10-**10, **11-**13, **12-**15, **E-**4

ID variables. *See also specific variables*
additional household members, **9-**3, **10-**8, **10-**25
core wave files, **9-**3, **9-**12, **10-**6–**10-**14, **10-**20–**10-**28, **10-**29–**10-**30, **11-**11–**11-**12, **11-**13, **11-**27, **13-**9, **13-**14
description, **9-**2–**9-**4
family, **9-**12, **10-**11–**10-**14, **11-**17, **11-**18, **12-**18
family composition from, **9-**6–**9-**7, **9-**13, **10-**11, **10-**12, **10-**19, **11-**17, **11-**18
full panel files, **9-**3, **9-**12, **9-**14, **12-**6, **12-**23–**12-**28, **13-**9, **13-**15
household composition from, **9-**6, **10-**23–**10-**24
monthly characteristics from, **9-**8
mover identification, **9-**3, **9-**12, **10-**8, **10-**20, **10-**22–**10-**26, **11-**13, **11-**14, **11-**21–**11-**27, **12-**14, **12-**23–**12-**28
names by file type, **9-**2, **9-**3
person, **8-**17, **9-**4–**9-**8, **9-**11, **10-**7–**10-**9, **11-**11, **11-**13–**11-**15, **12-**13–**12-**15, **13-**23
purpose, **9-**2–**9-**4
topical module files, **9-**3, **9-**6, **11-**7, **11-**11–**11-**27, **13-**11, **13-**14, **13-**15
transfer program unit composition from, **9-**7

Immigration, **3-**12–**3-**13, **8-**5, **C-**8

Imputation. *See also* Sequential hot-deck imputation procedure
additional household members' records, **4-**6–**4-**7, **10-**36
age, race, and gender, **10-**37
carryover procedures, **4-**5, **4-**10, **4-**13, **4-**16, **10-**37, **E-**9
core wave files, **4-**2, **4-**4, **4-**6–**4-**7, **4-**13, **8-**16, **9-**15, **10-**6, **10-**25, **10-**36–**10-**37, **11-**9, **12-**10, **12-**37, **13-**1, **13-**6–**13-**7
cross-observation, **12-**37
cross-sectional, **4-**4, **4-**8–**4-**9
defined, **E-**5
dependent, **4-**13
disadvantages, **4-**3
effect on analyses, **4-**3, **4-**11, **4-**16, **7-**6, **8-**17, **13-**6–**13-**7, **13-**8, **13-**12
EPPFLAG, **4-**10, **4-**13, **4-**14, **10-**36–**10-**37
error, **12-**7, **13-**7, **13-**12, **13-**14
exiting sample members, **13-**17, **13-**19–**13-**20
flags, **4-**11, **4-**13–**4-**14, **4-**15, **10-**36–**10-**37, **11-**28, **12-**37, **13-**8, **13-**12, **13-**22, **E-**5
full panel files, **1-**8, **4-**3, **4-**5, **4-**14, **8-**17, **9-**15, **10-**37, **12-**7, **12-**10, **12-**17, **12-**37, **13-**1, **13-**8, **13-**11
goals of, **4-**2–**4-**3, **4-**11
income, **4-**4, **4-**7, **4-**9, **4-**10, **4-**15, **4-**16, **10-**37, **11-**28, **12-**37
item nonresponse, **2-**21, **4-**1, **4-**2, **4-**4, **4-**7, **4-**12, **4-**14, **6-**1, **6-**2, **7-**6
little Type Z, **4-**10, **4-**13, **10-**37
logical, s*ee* Data editing
longitudinal, **4-**8, **4-**16
and linking files, **4-**5, **13-**7, **13-**8, **13-**22
missing data, **1-**8, **2-**20, **2-**21, **4-**4, **7-**6, **9-**15, **11-**24, **13-**20
missing wave, **4-**5, **4-**16, **8-**7, **8-**17, **9-**5, **9-**15, **10-**36, **12-**7, **12-**10, **12-**17, **13-**11, **13-**16, **13-**22
nonmatches and, **13-**17, **13-**22
nonresponse adjustments, **2-**20, **4-**5, **8-**17, **10-**36, **C-**18
person nonresponse adjustments, **1-**8, **2-**20, **4-**1–**4-**2, **4-**6–**4-**7, **7-**6, **10-**36, **11-**11, **12-**7, **12-**13
personal demographic characteristics, **4-**4, **4-**6, **4-**12, **4-**16, **8-**6, **11-**11
program participation, **4-**7, **10-**28
redesign of 1996, **4-**1, **4-**5, **4-**6, **4-**7, **4-**13, **4-**15, **8-**17, **12-**37, **13-**1
sample unit characteristics, **4-**4, **4-**6, **8-**6
statistical, **4-**1, **4-**4, **4-**8, **4-**13
steps, **4-**4
topical modules, **4-**2, **4-**5, **4-**14, **9-**15, **11-**11, **13-**12
Type Z, **1-**8, **2-**20, **4-**2, **4-**6–**4-**7, **4-**13, **4-**14, **7-**6, **8-**5, **9-**5, **12-**7, **12-**10, **12-**13, **12-**17, **13-**8, **13-**12, **E-**13
variance estimation, **4-**3, **4-**11, **4-**12, **4-**16, **7-**6
weighting adjustments, **8-**4, **8-**5
whole record procedure, **13-**11
within-wave, **13-**11

Income. *See also* Program income
amounts, **1-**8, **3-**6, **12-**30
annual, **3-**8, **8-**18, **11-**21
asset, **3-**13, **4-**7, **10-**29, **12-**37
children's, **3-**6
core questions, **1-**8, **3-**3–**3-**4, **3-**6
core wave file structure, **13-**7

core wave file variables, **9**-12, **10**-19–**10**-20, **10**-21, **10**-27, **10**-37
CPS data, **1**-1, **1**-9, **1**-10
earned, **10**-32–**10**-35, **12**-37, **B**-1–**B**-4, **B**-7
errors in estimates, **6**-4
exiting sample members, **13**-19, **13**-20
family, **9**-12, **10**-19–**10**-20, **10**-21, **10**-35, **10**-36, **12**-23, **12**-36, **12**-37, **C**-18
full panel file variables, **9**-12, **12**-23, **12**-30–**12**-31, **12**-32–**12**-37
household, **7**-5, **9**-5, **10**-35, **10**-36, **10**-37, **11**-28, **12**-13, **12**-36, **12**-37, **C**-15
imputation, **4**-4, **4**-7, **4**-9, **4**-10, **4**-15, **4**-16, **10**-37, **11**-28, **12**-37, **13**-19
information resources, **5**-2, **5**-3, **5**-16
monthly, **12**-31, **12**-36
nonresponse, **6**-2
property, **3**-12, **6**-4
PSID data, **1**-10–**1**-11
subfamily, **12**-23
subpopulation variables, **11**-28
summary variables, **10**-29, **10**-35–**10**-36, **12**-36
taxes, **3**-8, **3**-14
topcoding, **4**-17, **9**-15, **10**-29, **10**-32–**10**-36, **11**-28, **12**-31, **12**-36–**12**-37, **B**-1–**B**-4, **B**-6–**B**-7
topical modules, **3**-8, **3**-12
types recorded in SIPP, **3**-3–**3**-4, **3**-5, **11**-21
unearned, **3**-3–**3**-4, **3**-5, **3**-6, **10**-29, **10**-32, **11**-28, **12**-30, **12**-32–**12**-36, **12**-37, **B**-6–**B**-7
unreported, **13**-19
variables, **9**-12, **12**-23, **12**-30–**12**-31, **12**-32–**12**-36
weighting adjustments, **13**-19

Income Survey Development Program (ISDP), **1**-1–**1**-2, **1**-13

Infants, **8**-17, **9**-5, **9**-8, **10**-25, **11**-24, **12**-26, **13**-16, **13**-17

Information resources. *See also* Microdata files; Technical documentation; Web sites
bibliography (online), **1**-13, **5**-15
directory of data and publications, **5**-15
P-70 series, **1**-13–**1**-14, **5**-1, **5**-2–**5**-3
Quality Profile, **1**-13, **5**-1, **5**-13
telephone numbers, **5**-16
User Notes, **5**-12, **5**-14, **10**-2, **11**-2, **12**-2
variable metadata, **5**-15
working papers, **1**-14, **5**-13, **5**-14, **5**-15

Institutionalized individuals **2**-6, **2**-9, **2**-15, **2**-16, **8**-7, **8**-18, **11**-11, **13**-16, **13**-17, **13**-20

Instrumental Activities of Daily Living (IADL) battery, **3**-10

Interest income, **10**-29

Internal data files, **1**-5, **5**-1

Inter-university Consortium for Political and Social Research (ICPSR), **1**-5–**1**-6, **5**-12

Interview. *See also* Computer-assisted interviewing; Monthly interview status variable; Telephone interviews/ interviewing
additional household members, **2**-16, **2**-17
consistency checks, **2**-17, **3**-1
core questions, **3**-1, **3**-2–**3**-6, **6**-2
dates, by panel, **2**-2
face-to-face, **2**-17, **6**-2
household status code, **11**-12
identifying household members, **2**-6–**2**-7, **2**-16
intervals, **1**-4, **2**-1, **2**-9, **8**-8
mode, by wave, **6**-2
month, **E**-5
probes, **3**-3
procedures, **1**-4, **2**-16–**2**-17, **2**-21, **3**-1–**3**-2, **6**-2, **8**-19
skip patterns, **2**-17, **3**-2, **10**-2, **10**-6, **11**-2, **11**-6, **12**-2, **12**-3, **12**-6, **E**-11
telephone. *See* Telephone interviews/interviewing
topical questions, **3**-1, **3**-6–**3**-16

Interview month weights
calendar month estimation, **8**-14, **8**-15
core wave file, **8**-8–**8**-11, **8**-14, **8**-15
construction, **8**-4–**8**-5, **8**-6
format, **8**-8–**8**-9
household-level analyses, **8**-10–**8**-11
person-level analyses, **8**-9–**8**-10, **8**-16, **11**-28
population represented by, **8**-9, **8**-10, **8**-14
topical module file, **8**-16, **9**-8, **11**-28
by type of file, **8**-3
uses, **8**-8–**8**-11

Interviewer
discretion in identifying reference person, **10**-18, **11**-20
errors, **4**-2
experience, **8**-19

INTVW field, **4**-13–**4**-14

Item nonresponse
data editing, **4**-1
defined, **E**-5
errors, **6**-1, **6**-2
imputation, **2**-21, **4**-1, **4**-2, **4**-4, **4**-7, **4**-12, **4**-14, **6**-2, **7**-6
rates, **6**-2
sources, **2**-20–**2**-21, **4**-2

Iterative proportional fitting, **C**-5

Jackknife repeated replications, **7**-2

Labor force status. *See also* Employment; Unemployment; Work
core questions, **3**-3, **3**-4
errors in estimates, **6**-4
imputation, **4**-4, **4**-7, **4**-8–**4**-10, **4**-14, **10**-36–**10**-37
information resources, **5**-3, **5**-16
noninterview adjustments, **C**-18
spell estimation, **8**-18
and topcoding, **10**-32, **10**-33, **B**-3, **B**-4
weekly data, **2**-3
Liabilities
errors in estimates, **6**-4
topical questions, **3**-6, **3**-8
Linking files or data. *See also* Merging files or data
across waves, **13**-7, **13**-12, **13**-16
bias in analyses from, **13**-1–**13**-2
conceptual issues, **1**-9
core data from all waves, **4**-3
core wave file reformatting, **13**-3–**13**-4, **13**-5–**13**-6
core wave to full panel, **1**-9, **12**-28, **13**-8–**13**-11
editing/imputation effects, **4**-5, **13**-7, **13**-8
format changes for, **13**-3–**13**-4, **13**-5–**13**-6
households or families, **13**-1–**13**-2, **13**-11–**13**-12
husbands and wives, **10**-6, **12**-13
multiple core wave files, **4**-5, **5**-4, **13**-4, **13**-6–**13**-8
multiple topical module files, **13**-1, **13**-11–**13**-12
overview, **1**-9
parents and children, **10**-6, **12**-13
procedures, **13**-2–**13**-15
reasons for, **5**-4, **9**-9, **12**-13, **13**-1, **13**-4
topical module to core wave, **1**-9, **13**-12–**13**-14
topical module to full panel, **1**-9, **13**-14–**13**-15
unit composition changes and, **13**-1–**13**-2
within waves, **13**-7, **13**-16
Linking records across microdata files, **9**-4, **10**-7, **11**-13, **11**-16, **12**-13
Living conditions, topical modules, **3**-7
Longitudinal analyses
of core wave data, **13**-6–**13**-7, **13**-8
defined, **E**-6
editing, **4**-1
household or family charactistics, **13**-2
imputation effects, **7**-6, **8**-17, **13**-6–**13**-7
quarterly estimates, **8**-16
restrictions on, **9**-5, **12**-9–**12**-10, **12**-15, **12**-16, **13**-2, **13**-6–**13**-7
seam effect and, **6**-3
weights, **8**-3, **8**-4, **8**-16, **12**-7
Longitudinal research files. *See* Full panel files
Long record format, **13**-2
Long-term care, **3**-9, **3**-12
Loss of sample. *See also* Attrition
reasons for, **13**-16, **13**-17, **13**-18–**13**-19, **C**-15
rates, **2**-17–**2**-18, **2**-19
Marital history, **3**-12, **8**-18, **8**-19
Marital status, **11**-11, **11**-12, **11**-19
Marriages, **2**-11, **5**-16, **6**-4, **11**-24, **11**-27, **12**-26
Mean, defined, **7**-5
Measurement errors, **6**-2–**6**-3, **13**-12
Medicaid, **3**-4, **9**-7, **9**-14, **10**-27, **10**-29, **10**-30–**10**-31, **12**-29, **12**-30, **12**-31
Medical expenses, **3**-12
Medicare, **3**-4, **9**-7, **9**-14, **10**-27, **10**-28, **12**-29, **12**-30, **12**-31
Merging files or data. *See also* Linking files or data
aggregate records, **13**-13
attrition and, **13**-16, **13**-17, **13**-20–**13**-21
calendar month estimates, **8**-14–**8**-16, **8**-19
core wave with full panel, **10**-6, **12**-1, **12**-6, **12**-17, **12**-20, **12**-28, **12**-30, **13**-1, **13**-3, **13**-4
core wave with topical module, **1**-8, **3**-10, **9**-6, **9**-9, **10**-6, **11**-1, **11**-7, **11**-8, **11**-10, **11**-11, **11**-13, **11**-17, **11**-19, **12**-6, **12**-13, **13**-1, **13**-3, **13**-4, **13**-12, **13**-13, **13**-14, **13**-15
duplicated records, **13**-23
for family membership identification, **9**-6, **11**-13, **11**-17, **12**-17, **12**-20
format of output, **13**-2, **13**-3
households in pre-1996 panels, **9**-6, **12**-28, **13**-22–**13**-23
imputation and, **1**-8
multiple core wave files, **10**-1, **10**-6, **12**-13
multiple topical module files, **11**-13
nonmatches in, **1**-8, **13**-12, **13**-14, **13**-15–**13**-23
people exiting or entering the population and, **13**-17–**13**-20
person indentification and, **10**-6–**10**-7, **12**-13
procedures, **10**-1, **11**-1
program coverage, **12**-30
quarterly estimates, **8**-14–**8**-16
reasons for, **8**-14–**8**-16, **9**-9, **13**-1
redesign of SIPP and, **13**-22
topical module with full panel, **9**-6, **10**-6, **11**-1, **11**-7, **11**-13, **11**-19, **12**-1, **12**-6, **13**-12
types, **13**-2–**13**-3
variables from different files, **11**-11, **11**-19, **13**-4
weights, **5**-4, **13**-1, **13**-12
within core wave files, **1**-9, **12**-13, **13**-3–**13**-4, **13**-5–**13**-6, **13**-7
Methodology, information resources, **5**-16, **6**-3
Metropolitan area identification, **4**-17–**4**-18, **9**-15, **10**-38–**10**-39, **12**-38
Metropolitan Statistical Areas (MSAs), **10**-39

Microdata files. *See also* Core wave files; Full panel files; Topical module files
confidentiality procedures, **1**-5, **4**-4, **4**-5, **4**-17–**4**-18, **7**-2, **10**-6, **10**-8, **11**-13, **12**-14
construction of variables, **9**-8
contents, **5**-3–**5**-4, **5**-6–**5**-11
creation, **4**-4, **4**-5
defined, **E**-6
differences among types, **9**-10, **9**-11–**9**-15, **11**-8, **11**-11–**11**-12
extracts from, **5**-13
formats, **5**-3–**5**-5, **5**-11, **5**-12
ID variables, **9**-2–**9**-4
monthly family composition, **9**-6–**9**-7
monthly household composition, **9**-6
monthly interview status variable, **9**-4–**9**-5
monthly transfer program unit composition, **9**-7
multiple file usage, **9**-9
person identification, **9**-4–**9**-8
sources for obtaining, **5**-1, **5**-3, **5**-4, **5**-12–**5**-13
technical documentation, **1**-14, **5**-12, **5**-14
types, **1**-8, **5**-3, **9**-1–**9**-2, **9**-11
User Notes, **5**-12, **5**-14, **12**-2
variable metadata, **5**-15
website, **1**-6
weight selection, **9**-8
Migration history, **3**-12–**3**-13, **5**-16
Military barracks
original sample members in, **2**-9, **2**-10, **2**-11, **2**-15, **10**-25, **11**-24, **12**-25–**12**-26, **13**-16, **13**-17
Missing data
adjustments for, s*ee* Data editing; Sequential hot-deck procedures
code for linking files, **13**-3, **13**-4
defined, **E**-6
flagging, **11**-9, **12**-10
imputation, **1**-8, **2**-20, **2**-21, **4**-4, **7**-6, **9**-15, **11**-24, **13**-20
model-based approaches, **13**-22
panel weights, **8**-17, **13**-22
problems caused by, **4**-2
selection of replacement values, **4**-8, **4**-13, **4**-15
statistical packages, **13**-21
substituting the mean for, **13**-20–**13**-21
topical modules, **4**-5, **5**-4
types of, **4**-1–**4**-2
weighting adjustments, **13**-21, **13**-22
Missing waves
defined, **E**-6
full panel files, **12**-10, **13**-22
imputation, **4**-5, **4**-16, **8**-7, **8**-17, **9**-5, **9**-15, **10**-36, **12**-7, **12**-10, **12**-17, **13**-11, **13**-16, **13**-17, **13**-22
weighting adjustments, **8**-7, **13**-22
Monthly
cross-sectional weights, **5**-4
employment income, **10**-32–**10**-35
family composition, **9**-6–**9**-7, **9**-8, **12**-17–**12**-18, **12**-20
household composition, **9**-6, **9**-8
program income variables, **12**-30, **12**-36, **12**-37
transfer program unit composition, **9**-7, **9**-8
variables, **9**-3–**9**-4, **9**-8
Monthly interview status variable
core wave files, **9**-4, **9**-5, **9**-11, **11**-9, **11**-11, **11**-12
defined, **E**-6
full panel files, **1**-8, **9**-4, **9**-5, **9**-11, **11**-11, **12**-6, **12**-7, **12**-8, **12**-9–**12**-10, **12**-11–**12**-12, **12**-13, **12**-15, **12**-16, **12**-18, **12**-20, **12**-23, **12**-29
name, by file type, **9**-4, **11**-11, **12**-15
noninterview code, **9**-5
number of occurrences, **12**-6, **12**-9
person-level, **11**-9–**11**-11, **11**-12, **12**-16
program participation, **12**-29
purpose, **9**-4, **9**-11, **11**-9, **12**-9
realigned by calendar month, **12**-11–**12**-12
restrictions on use, **9**-5, **12**-9–**12**-10
topical module files, **9**-4–**9**-5, **9**-11, **11**-9–**11**-11, **11**-12
values, **9**-5, **11**-9, **11**-10, **12**-9–**12**-10
Mothers, **10**-15
Moves/movers. *See also* Following rules
abroad, **2**-9, **2**-15, **10**-25, **11**-24, **12**-26, **13**-16, **13**-17, **13**-20
additional household members, **4**-6–**4**-7, **8**-6, **10**-8, **10**-20, **11**-24, **12**-24–**12**-25
defined, **E**-6
distance considerations, **2**-15, **2**-20, **C**-15
identification, **9**-3, **9**-12, **10**-8, **10**-20, **10**-22–**10**-26, **11**-13, **11**-14, **11**-21–**11**-27, **12**-14, **12**-23–**12**-28
interview procedures, **1**-4, **2**-17
nonmatches in merged files, **13**-16, **13**-17, **13**-20
nonresponse, **2**-17, **2**-20
patterns of, **5**-3
person identification and, **9**-11, **9**-12, **10**-6, **11**-14, **12**-14, **13**-23
temporarily absent members distinguished from, **2**-15–**2**-16
tracing, **2**-9, **2**-15, **2**-16
weighting adjustments, **8**-4, **8**-5, **8**-6, **13**-20, **C**-13–**C**-15, **C**-16, **C**-19
MSA-Place Status, **8**-5
Multiple files
reasons for working with, **9**-9
Multivariate statistics, **13**-20–**13**-21

National Center for Health Statistics (NCHS), **6**-4
National Longitudinal Survey (NLS), **E**-7
National Research Council, Committee on National Statistics, **1**-2
New-construction frame, **2**-6
New construction noninterview adjustment factor, **C**-1, **C**-12
Noninterviews. *See also* Household noninterviews; Person nonresponse
adjustment factors, **C**-1, **C**-2–**C**-3, **C**-12, **C**-13, **C**-18–**C**-19
departure, **E**-2
monthly interview status variable code, **9**-5
person-level, **1**-8, **4**-6–**4**-7, **9**-5, **11**-11
Type D, **2**-15
Type Z, **4**-1–**4**-2, **4**-14, **11**-9, **12**-13, **13**-8, **13**-11, **13**-12
Nonresponse. *See also* Household nonresponse; Item nonresponse; Person nonresponse
bias, **2**-17, **4**-2, **6**-1
movers, **2**-17, **2**-20
imputation adjustments, **2**-20, **4**-5, **8**-17, **10**-36
nonsampling error, **6**-1–**6**-2
and quality of data, **2**-18
rates, **2**-17–**2**-18, **2**-20, **4**-3, **6**-2
refusals, **2**-17, **2**-18, **2**-20, **4**-2, **4**-7, **10**-36, **12**-13
subpopulations, **6**-4
unit, **4**-1, **4**-3, **4**-4
wave, **4**-5, **7**-6
weighting adjustments, **2**-17, **2**-18, **4**-1, **6**-2, **6**-4, **8**-4, **8**-5, **8**-6, **8**-8, **C**-3
Nonsampling errors
effects on survey estimates, **6**-3–**6**-4, **8**-19
information resources, **5**-13, **5**-16
measurement errors, **6**-2–**6**-3
nonresponse, **6**-1–**6**-2
and pooling data, **8**-19
recall period and, **8**-18
sources, **1**-6–**1**-7, **6**-1
undercoverage of subpopulations, **1**-6, **6**-1
Nursing homes, **2**-16, **3**-14, **8**-18, **13**-20

Old-Age, Survivors, and Disability Insurance (OASDI), **7**-4
Original sample members
age, **2**-7
births to, **2**-14
defined, **E**-7
following rules, **1**-4, **2**-7, **2**-9–**2**-15, **10**-25, **11**-24, **13**-15
marriage, **2**-11
merged households, **10**-25
in military barracks, **2**-9, **2**-10, **2**-11, **2**-15, **10**-25
moves, **9**-3, **10**-22, **C**-13
noninterview rates, **6**-2
number, by panel, **2**-2
person numbers, **10**-8, **10**-9, **10**-20, **11**-14, **12**-14
reentering sample universe, **13**-16, **13**-17
separation/divorce, **2**-14
temporarily absent, **2**-15–**2**-16
weights for, **8**-6, **8**-7
Oversampling
defined, **2**-8, **E**-7
1990 panel, **2**-8, **8**-2
1996 panel, **1**-3, **2**-8–**2**-9
rate, **2**-9

P-70 series reports, **1**-13–**1**-14, **5**-1, **5**-2–**5**-3, **E**-7
Panel files. *See* Full-panel files; Partial-panel files
Panel Study of Income Dynamics (PSID), **1**-10–**1**-11, **E**-8
Panel weights, **8**-16–**8**-17, **8**-18–**8**-19
Panels
attrition by, **2**-19
composition, **2**-8–**2**-9
core content differences, **3**-3–**3**-6
date of interview by, **2**-2
defined, **2**-1, **E**-7
followup to 1992 and 1993, **1**-11, **2**-2
household number by, **1**-2, **2**-2, **2**-8, **8**-20, **12**-7
length of, **2**-1–**2**-2, **8**-16, **8**-19
nonresponse by, **2**-19, **E**-8
number of waves by, **2**-2, **12**-6, **12**-7
organizing principles, **2**-1–**2**-3
original sample members in Wave 1 by, **2**-2
overlapping, **1**-3, **2**-1, **8**-19, **8**-20, **9**-9
oversampling, **1**-3, **2**-8–**2**-9
pooling data from, **8**-19–**8**-21
structure, **1**-2, **1**-3, **2**-1, **12**-6, **12**-7
topical modules by, **3**-7, **3**-8–**3**-15, **5**-4, **5**-6–**5**-11, **11**-6
variance units and strata by, **7**-2–**7**-3
weights, **8**-16–**8**-17, **8**-18–**8**-19, **C**-17–**C**-25
Parents, **10**-7, **10**-15, **10**-17, **10**-18, **10**-19, **11**-12, **11**-13, **11**-16, **11**-19, **11**-20, **11**-21, **11**-22, **12**-13, **12**-21, **12**-22
Partial panel files, **5**-12, **9**-3, **E**-8
Person. *See also* Reference person
associated sample, **C**-13, **C**-14
monthly interview status variable, **11**-9–**11**-11, **11**-12, **12**-16
noninterview records, **1**-8, **4**-6–**4**-7, **9**-5, **11**-11
out of scope, **12**-13

Person identification. *See also* Person Number
core wave files, **9-**11, **9-**15, **10-**6–**10-**9, **11-**11, **13-**9, **13-**23
examples, **11-**14, **11-**15
full panel file, **8-**17, **9-**11, **9-**15, **12-**13–**12-**15, **13-**23
and merging files or data, **10-**6–**10-**7, **12-**13, **13-**23
moves and, **9-**11, **9-**12, **10-**6, **11-**14, **12-**14, **13-**23
reasons for, **10-**6–**10-**7, **12-**13
topical module files, **9-**11, **9-**15, **11-**11, **11-**13–**11-**15, **13-**23
variables, **8-**17, **9-**4–**9-**8, **9-**11, **10-**7–**10-**9, **11-**11, **11-**13–**11-**15, **12-**13–**12-**15, **13-**23
Person-month
format, **1-**8, **5-**4, **5-**5, **8-**8, **9-**1, **9-**3, **9-**5, **9-**6, **9-**11, **10-**6, **10-**7, **10-**25, **11-**7, **13-**2, **13-**3–**13-**4, **13-**5–**13-**6, **13-**7, **13-**9, **13-**13, **13-**15, **E-**8
record, **8-**8, **8-**15
Person nonresponse (Type Z)
core questions, **4-**2, **13-**22
defined, **E-**8, **E-**12
errors, **6-**1, **6-**2
forms of, **2-**20
imputation adjustments, **1-**8, **2-**20, **4-**1–**4-**2, **4-**6–**4-**7, **7-**6, **10-**36, **11-**11, **12-**7, **12-**13, **13-**22
rates, **6-**2
sources of, **2-**15, **2-**18, **2-**20, **4-**1–**4-**2, **12-**13
Person Number
additional household members, **10-**25, **11-**14, **11-**24
changes in, **10-**26, **11-**27, **12-**14, **12-**26, **13-**22
core wave files, **1-**8, **9-**3, **10-**6, **10-**7, **10-**8, **10-**9, **10-**10, **10-**13–**10-**14, **10-**15, **10-**21, **10-**22, **10-**28, **11-**11, **11-**12, **11-**23, **13-**3, **13-**7
components, **9-**4, **10-**6, **11-**14, **12-**14
family identification, **10-**13–**10-**14, **10-**21, **11-**18, **12-**20, **12-**23
family-level income, **12-**23
full panel files, **1-**8, **12-**7, **12-**8, **12-**11–**12-**12, **12-**14, **12-**15, **12-**16, **12-**20, **12-**23–**12-**27, **12-**37
household composition, **10-**10, **10-**15, **10-**16, **10-**19, **10-**23–**10-**24, **11-**16, **11-**19, **11-**21, **11-**22, **12-**16
income topcodes, **10-**36, **12-**37
merged households, **10-**25, **13-**22
movers, **10-**20, **10-**22, **10-**23–**10-**24, **10-**25, **11-**14, **11-**22, **11-**23, **11-**25–**11-**26, **12-**23–**12-**27
multigeneration household members, **11-**21, **11-**22
newborns, **11-**24, **12-**26
original sample members, **10-**8, **10-**9, **10-**20, **10-**25, **11-**14, **12-**14
purpose, **9-**4, **11-**14
recombined households, **10-**26
reference person, **10-**16
sorting files for linking, **13-**3, **13-**4, **13-**9, **13-**14, **13-**15
spouses, parents, and guardians, **12-**21, **12-**22
topical module files, **11-**7, **11-**10, **11-**11, **11-**12, **11-**13, **11-**14, **11-**15, **11-**16, **11-**18, **11-**19, **11-**21, **11-**22, **11-**24, **11-**25–**11-**26, **11-**27
transfer program recipient, **10-**28
variable names, **9-**3
by wave, **10-**8–**10-**9, **12-**14
Person-record
duplicates, **13-**23
format, **9-**4, **9-**5, **9-**7, **9-**11, **10-**6, **10-**7, **13-**2, **13-**3–**13-**4, **13-**5–**13-**6, **13-**7, **13-**9, **13-**13
Person weights
adjustments, **C-**5
base, **C-**2
construction, **8-**4–**8-**5
cross-sectional, **8-**16, **11-**28
final, **8-**2, **8-**3, **8-**4
full panel file, **8-**3, **8-**17
household, family, subfamily weights from, **8-**6, **8-**10, **8-**11, **8-**12
husbands and wives, **8-**10
initial, **8-**5
interview month, **8-**8, **8-**9–**8-**10, **8-**16, **11-**28
population represented by, **8-**16
reference month, **8-**8–**8-**12, **8-**16
topical module files, **11-**11, **11-**12, **11-**28
by type of file, **8-**3, **9-**15, **11-**11, **11-**12
variable name, **11-**12
zero, **9-**5, **9-**8
Personal demographic characteristics, **3-**2
editing, **13-**8
imputation, **4-**4, **4-**6, **4-**12, **4-**16, **8-**6, **11-**11
Personal history topical module, **3-**6, **3-**7, **3-**15
Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), **1-**3, **9-**7, **10-**27
Perturbation factors, **7-**3
Pooling data
family-level income, **10-**20
from multiple panels, **8-**19–**8-**21
from multiple waves, **8-**15
nonsampling errors and, **8-**19
reasons for, **9-**9
Population control adjustments, **1-**6, **6-**1, **C-**3–**C-**4
Population mean, **7-**5
Population variance, **7-**5
Post Enumeration Surveys, **2-**6
Poststratification adjustment, **8-**4

Poverty status
CPS estimates, **1**-9, **6**-4
determining, **2**-8–**2**-9
errors in estimates, **6**-4
information resources, **5**-2, **5**-3, **5**-16
SPD estimates, **1**-11
weights, **8**-5, **8**-6, **C**-2, **C**-18
Primary individuals, **8**-11, **8**-12, **9**-4, **9**-6, **10**-11, **11**-17, **11**-18, **12**-17, **12**-19, **12**-20, **E**-8
Primary recipient ID, **9**-8, **9**-14
Primary sampling units (PSUs)
address selection, **2**-6
defined, **E**-8
imputation role, **4**-11
moves 100+ miles from, **2**-15
non-self-representing, **2**-5, **C**-12, **E**-7
person identification, **10**-8, **11**-13, **12**-14
selection of, **2**-**6**, **7**-2
self-representing, **2**-5, **E**-11
variance estimation role, **7**-1, **7**-2
with-replacement assumption, **7**-2
Program income
authorized recipient, **10**-7, **10**-27, **10**-28, **12**-29
core questions, **3**-3, **3**-5
errors in, **6**-4
monthly, **12**-30, **12**-36, **12**-37
person-level amount, **9**-14
recipient for family, **10**-7, **10**-27, **10**-28, **12**-13
topcodes, **10**-36
variables, **9**-14, **10**-27, **12**-30, **12**-31, **12**-32–**12**-36, **12**-37
weighting adjustments, **C**-18
Program participation
administrative records compared to responses, **6**-3
core questions, **1**-8, **3**-3, **3**-4, **3**-5, **3**-6
CPS data, **1**-9
disability and, **3**-10
economics of, **5**-3; *see also* Program income
eligibility, **3**-9, **3**-15, **10**-38, **11**-29, **12**-38
imputation, **4**-7, **10**-28
primary recipient ID, **9**-8, **9**-14
P-70 publications, **5**-2, **5**-3
recipiency history, **3**-13, **3**-15, **8**-18, **10**-26, **10**-27
recipient characteristics, **5**-2
SPD data, **1**-11
spell estimation, **8**-18, **12**-7
variables describing, **9**-14, **10**-27, **12**-29, **12**-31–**12**-36
weights, **9**-5, **12**-13
Program units
composition, **9**-7, **9**-8
constructing characteristics of, **9**-8
core wave files, **9**-14, **10**-26–**10**-29, **10**-30–**10**-31
coverage, **4**-16, **9**-14, **10**-26–**10**-28, **10**-29, **10**-30–**10**-31, **12**-28, **12**-30–**12**-31
defined, **E**-9
examples, **10**-30–**10**-31
full panel files, **9**-14, **12**-28–**12**-30
identification, **9**-14, **12**-28–**12**-30
longitudinal household problem, **13**-2
Property. *See also* Real estate ownership; Vehicle ownership
income, **3**-13, **6**-4
taxes, **3**-12, **3**-13
topcoding, **11**-28, **B**-6
Proxy respondents, **2**-10, **2**-16, **3**-1, **6**-2, **10**-6, **10**-25, **11**-24, **E**-9
Pseudo-families, **9**-6, **10**-11, **10**-15, **11**-17, **12**-17
Public use files, **E**-9. *See also* Microdata files

Quality Profile, **1**-6, **1**-13, **2**-5, **2**-8, **2**-18, **5**-1, **5**-13, **6**-3
Quality of data
accuracy of definitions in data definitions, **11**-6
CAI and, **1**-3, **3**-1, **6**-2, **8**-16
interview consistency checks, **2**-17, **3**-1
matched records containing imputed data, **1**-9
nonresponse and, **2**-18
Quarterly estimates, **8**-14–**8**-16
Questionnaires. *See also* Computer-assisted interviewing
core items, **2**-3, **3**-1, **3**-2–**3**-6
correspondence of variables to items on, **10**-4–**10**-6, **11**-6, **12**-5–**12**-6
data dictionary correspondence to, **10**-4–**10**-6, **11**-6, **12**-5–**12**-6
design, **5**-16, **8**-19
documentation, **5**-14, **11**-2
edits, **2**-17, **4**-6
paper instrument, **2**-17, **3**-1, **3**-2, **4**-6, **4**-15, **8**-6, **10**-2, **10**-6, **11**-2, **12**-2
rostering, **2**-7, **3**-2
screens, **5**-14

Race/ethnic origin
imputation, **10**-37
income topcoding, **10**-32, **10**-33, **B**-2–**B**-3, **B**-4
reference person, **8**-5, **C**-2
variable name, **11**-12
weighting, **8**-5, **8**-6, **C**-3–**C**-4
Railroad Retirement, **3**-5, **6**-4, **9**-7, **9**-14, **10**-27, **10**-28, **12**-29
Raking procedure, **8**-5, **C**-4, **C**-5, **C**-10, **C**-11, **C**-12, C-24
Real estate ownership, **3**-3, **3**-8, **3**-12, **11**-28

Recall, **1**-6, **1**-9, **2**-3, **6**-2, **8**-18
Record Check Studies, **6**-3–**6**-4
Redesign (1996) of SIPP
address clusters, **2**-6
confidentiality procedures, **4**-17–**4**-18, **10**-6, **10**-38
core content, **3**-3–**3**-4
data dictionaries, **12**-3
defined, **E**-9
editing and imputation procedures, **4**-1, **4**-5, **4**-6, **4**-7, **4**-13, **4**-15, **8**-17, **12**-37, **13**-1
entry address ID, **9**-4, **10**-7, **10**-8, **10**-9, **11**-13, **12**-13, **13**-3
full panel files, **4**-16, **9**-3, **9**-11–**9**-15, **13**-1
household characteristics, **8**-6, **10**-10, **11**-14, **11**-16
interview procedures, **2**-17, **3**-1, **8**-6, **8**-16
and merging files, **13**-22
monthly interview status code, **9**-5
overview, **1**-2–**1**-3
panel structure, **1**-2, **2**-1, **2**-2, **8**-16
program unit IDs, **10**-28
questionnaires, **10**-5
rotation groups, **2**-4–**2**-5
state identification, **11**-29
topcoding, **10**-29, **10**-32–**10**-35, **12**-31, **B**-1–**B**-2
topical module files, **3**-10, **5**-4, **9**-5, **11**-6, **11**-7, **11**-8, **11**-9, **11**-11, **11**-17, **11**-29
variable names, **8**-1, **9**-1, **9**-3, **10**-1, **10**-5, **10**-6, **11**-1, **13**-1, **13**-2, **A**-10–**A**-17
weights, **7**-3, **8**-1, **8**-3, **8**-5, **8**-6, **8**-9, **8**-16, **12**-37, **C**-1, **C**-2–**C**-3
Reference month weights
calendar month estimation, **8**-14, **8**-15
construction, **8**-4–**8**-6
core wave files, **8**-3, **8**-4–**8**-5, **8**-6, **8**-8–**8**-13, **8**-14, **8**-15, **10**-37
family-level analyses, **8**-11–**8**-12, **8**-13
format, **8**-8–**8**-9
household-level analyses, **8**-10–**8**-11
number per person, **8**-8
person-level analyses, **8**-8, **8**-9–**8**-10
population represented by, **8**-10
second-stage calibration adjustment, **8**-6, **C**-16–**C**-17
subfamily-level analyses, **8**-11–**8**-12, **8**-13
variable, **8**-8–8-9
Reference period
aligned to calendar months, **12**-7, **12**-9, **12**-10, **12**-11–**12**-12
core wave files, **9**-2, **10**-7, **11**-8, **13**-4, **13**-7
CPS, **1**-9
cross-walk, **10**-2, **11**-2, **12**-2
defined, **2**-1, **2**-3, **E**-9
for household composition, **11**-14
interview month used in estimates with, **8**-9
length of, **1**-2, **2**-3, **2**-4–**2**-5
organizing principles, **2**-3–**2**-4
by panel, **12**-7
and recall errors, **2**-3
by rotation group, **2**-4–**2**-5, **10**-2, **11**-2, **11**-10, **12**-9, **12**-10, **12**-11–**12**-12
topical modules, **3**-7, **11**-8, **11**-10, **11**-11, **11**-19, **11**-21, **13**-13
weighting adjustments for pooled data by, **8**-21
Reference person
changes in, **8**-10, **10**-18, **12**-21
defined, **3**-11, **10**-16, **11**-20, **E**-9
family, **3**-11, **8**-11–**8**-12, **9**-6, **10**-11, **10**-12, **10**-15, **10**-16
group quarters, **8**-12
household, **8**-10–**8**-11, **8**-12, **10**-11, **10**-12, **10**-15, **10**-16–**10**-19, **11**-6, **11**-12, **11**-16, **11**-17, **11**-19–**11**-21, **12**-17, **12**-21
identification of, **2**-16, **10**-16
interviewer discretion in identifying, **10**-18, **11**-20
nonfamily household, **8**-12
primary individual, **10**-11, **11**-17
proxy interviews with, **2**-16, **3**-1
race, **8**-5, **C**-2, **C**-15
relationships of household members to, **8**-10–**8**-11, **10**-11, **10**-15, **10**-16–**10**-19, **11**-12, **11**-19–**11**-21, **12**-17, **12**-21, **12**-22
topical questions, **3**-7, **3**-8
two people designated as, **11**-21
unmarried partner of, **10**-17, **11**-20
variable name, **10**-16
weights, **8**-6, **8**-10, **8**-11, **C**-2, **C**-15, **C**-16
Replicability of published estimates, **5**-1
Reservation wage, **3**-13
Respondents. *See also* Reference person
absent for consecutive waves, **4**-5, **4**-16, **7**-6
age, **1**-2, **2**-7, **2**-16, **3**-1, **3**-6, **3**-7, **3**-9, **3**-10, **11**-6, **11**-10
burden on, **2**-3
"donors," **1**-5, **2**-20, **4**-1, **4**-3, **4**-7, **4**-9, **4**-10, **4**-13, **10**-37
misinterpretation of questions, **6**-3
proxy, **2**-10, **2**-16, **3**-1, **6**-2, **10**-6, **10**-25, **11**-24
referral to records, **3**-3, **3**-14, **6**-3
in scope, **8**-5, **8**-7, **8**-16, **9**-8, **11**-9, **E**-5
topical modules, **3**-7, **11**-6, **11**-10
Responses
administrative records compared to, **6**-3–**6**-4
error sources, **1**-6–**1**-7, **6**-3
Retirement expectations, **3**-13
Retirement/pension accounts, **3**-3, **3**-5, **3**-7, **3**-8, **3**-13–**3**-14, **5**-2, **5**-16, **11**-21
Roomers/boarders, **10**-17, **11**-20
Rostering, **2**-7, **2**-16, **3**-2

Rotation group, **1**-2
calendar month estimation by, **8**-12, **8**-14, **8**-15, **9**-9
defined, **2**-1, **2**-3, **E**-9
format, **2**-3, **8**-8, **10**-7
and nonsampling errors, **6**-2, **6**-3
quarterly estimates by, **8**-15
reference period by, **2**-4–**2**-5, **10**-2, **11**-2, **11**-10, **12**-9, **12**-10, **12**-11–**12**-12
skipped, **2**-3
variable, **11**-10, **11**-11, **11**-12
weights, **8**-5, **8**-8, **8**-12, **8**-14, **8**-16, **C**-16
Rural addresses, **2**-6

Sample design
comparison of surveys, **1**-10
oversampling, **2**-8–**2**-9
selection of sampling units, **2**-5–**2**-7
and variance estimates, **7**-1
Sample population
comparison with other surveys, **1**-9, **1**-10
entries and exits, **13**-17–**13**-20. *See also* Attrition
size considerations, **1**-2, **1**-3, **2**-2, **6**-2, **8**-5, **9**-9, **12**-7, **C**-19
universe, **13**-17
Sample Unit IDs
additional household members, **9**-3, **10**-8, **10**-9, **11**-13, **12**-14
changes in, **10**-26, **11**-13, **11**-27, **12**-14, **12**-26
components, **9**-2, **11**-13
core wave files, **9**-3, **10**-7, **10**-8, **10**-9, **10**-10, **10**-11, **10**-13–**10**-14, **10**-21, **10**-22, **10**-23–**10**-24, **11**-11, **11**-12, **11**-13, **11**-23, **13**-3, **13**-7, **13**-9
family identification, **10**-11, **10**-13–**10**-14, **10**-21, **11**-17, **11**-18, **12**-18, **12**-20, **12**-23
family-level income, **12**-23
full panel files, **9**-3, **12**-7, **12**-8, **12**-11–**12**-12, **12**-14, **12**-15, **12**-16, **12**-18, **12**-20, **12**-23–**12**-28, **12**-29, **13**-9
household composition, **9**-6, **10**-10, **10**-23–**10**-24, **11**-14, **11**-16, **11**-25–**11**-26, **12**-15, **12**-16, **12**-25, **12**-26
merged households, **12**-28
movers, **9**-3, **10**-8, **10**-20, **10**-22, **10**-23–**10**-24, **11**-13, **11**-22, **11**-23, **12**-14, **12**-23–**12**-28
newborns, **10**-25
parents and spouses, **12**-22
program participation, **12**-29
purpose, **9**-2–**9**-3, **9**-4, **10**-8, **11**-13, **11**-14, **12**-14
secondary sample persons, **9**-3
sorting files for linking, **13**-3, **13**-4, **13**-7, **13**-9, **13**-14, **13**-15
topical module files, **9**-3, **11**-7, **11**-10, **11**-11, **11**-12, **11**-13, **11**-14, **11**-15, **11**-17, **11**-18, **11**-25–**11**-26, **11**-27
transfer program unit composition, **9**-8, **10**-28
variable names, **8**-1, **9**-1, **9**-3, **10**-1, **10**-10, **11**-1, **11**-11, **12**-15, **13**-2
by wave, **10**-9
Sample units. *See also* Primary sampling units
imputation of characteristics, **4**-4, **4**-6, **8**-6
merged, **10**-25, **10**-26, **11**-27, **12**-26
selection of, **2**-5–**2**-7
Sampling errors
bias in estimates of, **1**-7, **2**-5
direct variance estimation, **7**-1–**7**-3
GVFs, **7**-4–**7**-6
imputation and, **7**-6
information resources, **5**-13, **5**-16
magnitude of, **7**-4
nonresponse and, **6**-2
survey design considerations, **7**-1
SAS reformatting code, **13**-3–**13**-4, **13**-5–**13**-6, **13**-9, **13**-10
SAS syntax, **10**-4, **10**-5, **11**-4, **11**-5, **12**-3, **12**-5
School. *See also* Education and training
enrollment, **3**-4, **3**-14
lunch program participation, **3**-4, **3**-6
Seam effect, **1**-6–**1**-7, **4**-16, **6**-3, **6**-4, **8**-16, **8**-19, **E**-9
Secondary individuals, **8**-11–**8**-12, **9**-6, **10**-11, **11**-17, **11**-18, **12**-17, **E**-9
Secondary sample members, **9**-3, **9**-4, **11**-10, **13**-15–**13**-16, **13**-17, **E**-9
Security, of telephone interviews, **2**-17
Self-employment, **3**-3, **3**-4, **3**-6, **4**-7, **10**-32, **C**-18
Sequential hot-deck imputation procedure
allocation flags, **4**-11, **4**-13–**4**-14
classes/adjustment cells, **4**-8, **4**-9–**4**-10, **4**-12
cold-deck values, **4**-8, **4**-11–**4**-12
core wave data, **4**-4, **11**-9
cross-sectional, **4**-8, **4**-9
data editing compared, **4**-8
donors, **4**-1, **4**-8, **4**-9, **4**-10
geographic sort variables, **4**-8, **4**-11
identifying records with no item nonresponse, **4**-8
longitudinal, **4**-8, **4**-9, **4**-10
overview, **1**-5, **4**-8–**4**-11
preprocessing sample file, **4**-11–**4**-12
redesign, **4**-5, **4**-7
selecting replacement values, **4**-8, **4**-13
steps, **4**-8, **4**-11–**4**-14
topical module data, **4**-5, **4**-14
types, **4**-8–**4**-9

updating hot-deck values, **4-**13
Severence pay, **3-**3, **3-**5
Shelter. *See* Housing
Simple random sample (SRS), **1-**7, **2-**5, **7-**1
Single parents, **8-**19, **C-**22–**C-**25
Social Security, **3-**3, **6-**4, **9-**7, **9-**14, **10-**27, **10-**28, **10-**29, **10-**30–**10-**31, **10-**36, **12-**29, **B-**5
Sorting operations, **4-**11
Source and accuracy statement, **5-**14, **7-**4, **7-**5, **10-**2, **10-**37, **11-**2, **11-**29, **12-**2, **12-**38, **13-**21, **E-**11
Special places. *See* Group quarters frame
Spell durations, **6-**4
Spell estimations, **6-**4, **8-**18–**8-**19, **12-**7, **13-**20
Spouses, **8-**10, **10-**15, **10-**17, **10-**19, **11-**12, **11-**13, **11-**16, **11-**19, **11-**20, **11-**21, **11-**22, **12-**13, **12-**21, **12-**22, **C-**3, **C-**6, **C-**10, **C-**11, **C-**12, **C-**20, **C-**22–**C-**25
Standard errors
bias in estimates of, **2-**5, **13-**21
computation of, **5-**14, **10-**1, **10-**2, **11-**1, **11-**2, **12-**2, **13-**21
of estimated numbers, **7-**4–**7-**5
of mean, **7-**5–**7-**6
overlapping panel structure and, **2-**2
tables of, **7-**4
Standard of living, **3-**8, **3-**10
State identification, **4-**17–**4-**18, **9-**15, **10-**38, **11-**11, **11-**12, **11-**29, **12-**38
State-level estimates, **10-**38, **11-**29, **12-**38
State variable, **9-**15, **10-**38, **11-**11, **11-**29, **12-**38
Subfamily(ies)
analyzing people in, **10-**12
defined, **8-**11, **10-**11, **12-**17
as distinct family unit, **10-**12, **12-**19
edited relationships, **10-**15
excluding for analysis purposes, **10-**12, **10-**13–**10-**14, **10-**15, **11-**17, **12-**19, **12-**20
ID variables, **10-**11–**10-**14, **10-**21, **11-**17, **12-**18, **12-**20, **12-**23
including with primary family, **10-**13–**10-**14, **10-**21, **12-**19, **12-**20
income variables, **10-**19–**10-**20, **10-**21, **12-**23
number in household, **10-**15, **10-**21, **11-**17
related, **3-**11, **8-**4–**8-**5, **8-**11–**8-**12, **8-**13, **9-**7, **9-**12, **10-**11, **10-**13–**10-**14, **10-**15, **10-**19–**10-**20, **10-**21, **11-**16, **11-**17, **12-**17, **12-**20, **12-**23, **E-**9
type, **10-**13–**10-**14
unrelated, **3-**11, **8-**11, **9-**6, **9-**7, **10-**11, **10-**12, **11-**16, **12-**17, **12-**19, **12-**20, **E-**13
weights, **8-**4–**8-**5, **8-**6, **8-**8, **8-**11–**8-**12, **8-**13
Subpopulations. *See also* Race/ethnicity
income topcoding, **11-**28
nonresponse, **6-**4
oversampling, **8-**2
poverty status, **2-**8–**2-**9
PSID coverage, **1-**11
undercoverage, **1-**6, **6-**1, **6-**4, **C-**17
weighting, **8-**2, **C-**1, **C-**8–**C-**9
Subsampling, address, **2-**6, **C-**2
Supplemental Security Income (SSI) program, **6-**4, **9-**14
definition of qualifiying disabling conditions, **10-**28, **12-**30
federal/state administration, **10-**28
history, **3-**15
income variables, **12-**30, **12-**34–**12-**36
program units, coverage, and recipiency, **10-**29, **10-**30–**10-**31, **12-**29, **12-**30, **12-**31
user-created monthly variables, **12-**30, **12-**34–**12-**36
variables describing participation, **10-**27, **10-**28, **12-**29
variance functions, **7-**4
Supplemental unemployment benefits, **3-**5
Support. *See also* Child support
nonhousehold members, **3-**14
Survey of Program Dynamics (SPD), **1-**10, **1-**11, **2-**2, **E-**11
Surveys-on-Call, **1-**6, **5-**12–**5-**13, **E-**11
Survival analysis, **8-**18
Survivors' income, **3-**3
Systematic bias, **6-**3

Tax returns, **1-**10, **3-**14
Taxes
income, **3-**8, **3-**13, **3-**14
property, **3-**13
Taylor-series approximation, **7-**2
Technical documentation
core wave files, **10-**2–**10-**4
defined, **E-**11
description of, **1-**14, **5-**12, **5-**14
full panel files, **12-**2–**12-**5, **12-**9
instrument screens and program code, **10-**2, **11-**2
source, **3-**1
topical module files, **3-**7, **11-**2–**11-**5
Telephone interviews/interviewing
callbacks, **2-**17, **2-**21
movers, **2-**15, **C-**15
procedures, **2-**17
quality of data, **6-**2
security/confidentiality of, **2-**17
Telephone numbers, **5-**16

Temporary Assistance for Needy Families (TANF), **1**-3, **3**-5, **3**-15, **9**-7, **9**-14, **10**-27, **10**-30
Time-in-sample bias, **1**-7, **2**-2, **6**-3, **8**-19, **E**-12
Topcoding
adjustments for inflation and real growth, **10**-32, **10**-34, **B**-1
age, **4**-17, **B**-4–**B**-5
algorithms, **10**-33–**10**-34
computations, **B**-1, **B**-2–**B**-3
core wave files, **9**-15, **10**-6, **10**-29, **10**-32–**10**-36, **11**-28
creating means for, **B**-3–**B**-4
defined, **E**-12
earned income, **10**-32–**10**-35, **B**-1–**B**-4, **B**-7
examples, **10**-34–**10**-35, **B**-2
full panel files, **9**-15, **12**-31, **12**-36–**12**-37
gender and, **10**-32, **10**-33, **B**-2, **B**-4
income, **4**-17, **9**-15, **10**-29, **10**-32–**10**-36, **11**-28, **12**-31, **12**-36–**12**-37, **B**-1–**B**-4, **B**-6–**B**-7
internal files, **5**-2
labor force status and, **10**-32, **10**-33, **B**-3, **B**-4
matrix, **B**-1, **B**-2–**B**-3
1996 Panel, **10**-29, **10**-32–**10**-35, **12**-31, **B**-1–**B**-2
pre-1996, **10**-35–**10**-36, **12**-31
purpose, **10**-29, **11**-27–**11**-28, **12**-31
property-related, **11**-28, **B**-6
race and, **10**-32, **10**-33, **B**-2–**B**-3, **B**-4
specifications, **B**-1–**B**-7
topical module files, **9**-15, **11**-27–**11**-28
unearned income, **10**-29, **10**-32, **11**-28, **B**-6–**B**-7
universe of cases, **11**-28
variables required, **B**-1, **B**-6–**B**-7
worker characteristics and, **10**-32
Topical content, **3**-1, **3**-6–**3**-7, **E**-12
Topical data, for skipped rotation groups, **2**-3
Topical items, **3**-1
Topical module files
allocation flags, **11**-28
content, **1**-4–**1**-5, **1**-8, **5**-4–**5**-11, **11**-7, **11**-10
core wave files compared, **9**-11–**9**-15, **11**-7, **11**-8, **11**-11–**11**-12, **13**-13
creation, **4**-5
data dictionary, **9**-11, **11**-2–**11**-5, **11**-6, **12**-3
defined, **E**-12
family composition variables, **9**-6, **9**-12, **9**-13, **9**-15, **11**-16–**11**-18, **11**-19–**11**-21, **11**-22
full panel files compared, **9**-11–**9**-15, **11**-8
full panel files linked with, **1**-9, **9**-6, **11**-1, **11**-7, **11**-8, **11**-13, **12**-1, **12**-6, **13**-14–**13**-15
household composition variables, **9**-12, **9**-13, **11**-16, **11**-19–**11**-21, **11**-22
household identification, **9**-11, **9**-15, **11**-11, **11**-14, **11**-15–**11**-16
ID variables, **9**-3, **9**-6, **11**-7, **11**-11–**11**-27, **13**-11, **13**-14, **13**-15, **13**-23
imputed data, **4**-14, **9**-15, **11**-11
linking family members, **11**-13
linking two or more, **13**-1, **13**-11–**13**-12
linking with core wave files, **1**-9, **13**-12–**13**-14
linking with full panel files, **1**-9, **13**-14–**13**-15
merging two or more, **11**-13
merging with core wave files, **1**-8, **3**-10, **9**-6, **9**-9, **10**-6, **11**-1, **11**-7, **11**-8, **11**-10, **11**-11, **11**-13, **11**-17, **11**-19, **12**-6, **12**-13, **13**-1, **13**-3, **13**-4, **13**-12, **13**-13, **13**-14, **13**-15
merging with full panel files, **9**-6, **10**-6, **11**-1, **11**-7, **11**-13, **11**-19, **12**-1, **12**-6
metropolitan area identification, **11**-29
monthly interview status variable, **9**-4, **9**-5, **9**-11, **11**-9–**11**-11
mover identification, **11**-13, **11**-14, **11**-21–**11**-27, **13**-23
overview, **1**-8
person identification, **9**-11, **9**-15, **11**-11, **11**-13–**11**-15, **13**-23
pre-1996, **11**-9–**11**-11
public use version, **9**-2, **9**-3, **11**-1–**11**-29
questionnaire correspondence to, **11**-6
redesign of 1996, **3**-9–**3**-10, **5**-4, **9**-5, **11**-6, **11**-8, **11**-9, **11**-11, **11**-17, **11**-29
state identification, **9**-15, **11**-11, **11**-29
structure, **5**-4, **5**-11, **9**-2, **9**-11, **11**-7–**11**-8, **13**-11, **13**-13
technical documentation, **11**-2–**11**-5
topcoding, **9**-15, **11**-27–**11**-28
variable names, **9**-3, **9**-15, **11**-1, **11**-6, **11**-11–**11**-12, **11**-13, **13**-11
weights, **8**-3, **8**-16, **9**-8, **9**-15, **11**-1, **11**-2, **11**-28–**11**-29, **13**-12, **13**-22
Topical modules, **1**-4
categories, **3**-7
core data merged with, **1**-8, **3**-10, **9**-9, **11**-8, **11**-10
data editing, **4**-4, **13**-12
defined, **3**-1, **3**-6
frequency and timing, **3**-6
“history” modules, **3**-9, **3**-15, **11**-8
household member relationships, **9**-6, **11**-11, **11**-19
imputation procedures, **4**-2, **4**-5, **4**-14, **9**-15, **11**-11, **13**-12, **E**-12
missing data, **4**-5, **5**-4
by panel and wave, **3**-7, **3**-8–**3**-16, **5**-4, **5**-6–**5**-11, **11**-6
purpose of, **3**-6
reference period for, **3**-7, **11**-8, **11**-10, **11**-11, **11**-19, **11**-21, **13**-13
respondents, **3**-7, **11**-6, **11**-10
sample definitions, **11**-8
title-content relationship, **3**-7

topics, **3-**6, **3-**7, **3-**8–**3-**16, **5-**6–**5-**11
Transfer programs, **9-**7. *See also* Program participation; Program units; *individual programs*

Undercoverage, **1-**6, **6-**1, **6-**4, **C-**17, **E-**13
Unemployment
compensation, **3-**3, **3-**5, **6-**4
CPS computations, **1-**9
length of, **3-**15
insurance, **3-**3
P-70 publications, **5-**2
reasons for, **3-**8, **3-**13, **3-**15
spell duration, **8-**18, **13-**20
Unit frame, **2-**6
University of Michigan, **1-**10
U.S. Government Printing Office, **5-**1
User Notes, **5-**12, **5-**14, **10-**2, **11-**2, **E-**13
Uses of SIPP, **1-**3–**1-**4
Usual place of residence, **E-**14

Variable metadata, **5-**15, **E-**14
Variables. *See also* ID variables
auxiliary, **4-**11, **4-**12
construction of, **9-**8
content, **5-**15
core wave files, **9-**1, **9-**13, **10-**1, **10-**4, **10-**8, **10-**11, **11-**11–**11-**12, **13-**9, **A-**1–**A-**34
covariances among, **4-**11, **4-**13
crosswalk of 1993 and 1996 names, **A-**1–**A-**34
dash characters in names, **13-**9
description of, **10-**2, **11-**2; *see also* Data dictionary
differences by file type, **9-**10, **9-**11–**9-**15
duplicate names for different variables, **13-**11
family composition, **9-**13, **10-**15–**10-**20, **11-**16–**11-**18, **11-**19–**11-**21, **11-**22, **12-**21–**12-**22
family identification, **8-**11, **10-**11–**10-**14, **12-**17–**12-**18
family-level income, **10-**19–**10-**20, **10-**21, **12-**23
file position, 1993 and 1996, **A-**18–**A-**34
full panel files, **1-**8, **8-**16–**8-**17, **9-**13, **12-**5, **13-**9
geographic sort, **4-**11
household composition, **4-**16, **8-**10, **9-**11, **9-**12, **9-**13, **9-**15, **10-**8, **10-**10, **10-**15–**10-**20, **10-**23–**10-**24, **11-**19–**11-**21, **11-**22, **12-**21–**12-**22
household identification, **10-**10
imputed, **4-**7, **4-**11, **4-**16, **12-**37
in-sample, **11-**9, **12-**9, **E-**5
interview month weights, **8-**9, **8-**10
length of names, **13-**4
merging from other files, **11-**11, **11-**19, **13-**4
monthly, **9-**3–**9-**4, **9-**8; *see also* Monthly interview status variable
name changes, **8-**1, **9-**1, **9-**3, **9-**15, **10-**1, **10-**6, **11-**1, **11-**11, **13-**1, **13-**2, **13-**11, **A-**1–**A-**34. *See also* ID variables
name–content correspondence, **10-**6, **11-**6, **12-**5
number of occurrences, **12-**3, **12-**6
previous wave, **11-**27, **13-**23
program income, **9-**14, **10-**27, **12-**30, **12-**31, **12-**32–**12-**36, **12-**37
program participation, **9-**14, **10-**27, **12-**29, **12-**31–**12-**36
questionnaire item correspondence, **10-**4–**10-**5, **11-**6, **12-**5–**12-**6
reference month weights, **8-**8–**8-**9
reference person, **10-**16
rotation group, **11-**10, **11-**11, **11-**12
subfamily, **8-**11
summary, **5-**15, **10-**29, **10-**35–**10-**36
for topcoding, **B-**1, **B-**6–**B-**7
topical module files, **8-**16, **9-**13, **11-**4, **11-**6, **11-**11–**11-**12, **11-**13–**11-**15
unearned income, **12-**30, **12-**32–**12-**36
values, **10-**5, **10-**12, **11-**4, **11-**9
variance estimation, **7-**3
weight, **9-**15
Variance estimation. *See also* Generalized variance functions (GVFs)
approximation methods, **7-**4–**7-**6
core wave files, **7-**3
degrees of freedom, **7-**2
direct methods, **7-**1–**7-**3
Fay's formula, **7-**3
imputation and, **4-**3, **4-**11, **4-**12, **4-**16, **7-**6
1990–1993 panels, **7-**2–**7-**3
1996 panel, **7-**3
OASDI, **7-**4
replication methods, **7-**2, **7-**3
sample design and, **1-**7, **7-**1
software, **7-**2, **7-**3, **7-**5
SRS formulas, **7-**1
SSI, **7-**4
strata, **7-**1, **7-**2–**7-**3
units, **7-**2–**7-**3
variables, **7-**3
Vehicle ownership, **3-**8, **3-**12
Veteran's benefits, **10-**27, **12-**29
Veterans Compensation and Pensions, **6-**4, **9-**7, **9-**14
VPLX software, **7-**3

Wages and salaries. *See also* Earnings
gross pay, **4-**9–**4-**10
imputation, **4-**7, **4-**9
reservation wage, **3-**13
topcoded, **10-**32–**10-**36, **12-**37

Waves. *See also* Missing waves
attrition rates by, **2**-19
bounded, **8**-7
combining, **8**-14–**8**-16
comparability of responses among, **8**-19
defined, **1**-2, **2**-1, **2**-3, **E**-14
interviewing mode by, **6**-2
nonresponse by, **2**-17–**2**-18, **2**-19, **7**-6
number of, **1**-3, **2**-2, **2**-3, **12**-6, **12**-7
organizing principles, **2**-3
overlapping, **8**-19, **8**-21, **9**-9
person identification by, **10**-8–**10**-9, **11**-14, **12**-14
short, **2**-2, **E**-11
size of sample, **1**-2, **2**-2
topical modules by, **3**-7, **3**-8–**3**-16, **5**-6–**5**-11
variable name, **11**-12
Web sites
Census Bureau, **1**-6, **5**-12
SIPP, **1**-6, **1**-13, **4**-1, **5**-1, **5**-12, **5**-13, **5**-14, **5**-15, **10**-2, **11**-2, **12**-2
variance estimation software, **7**-2
Weighting procedures
attrition adjustments, **8**-4, **8**-19, **13**-22
calendar month estimation, **8**-12, **8**-14–**8**-15, **8**-19, **9**-8, **12**-7, **13**-1, **13**-8
calendar year estimates, **8**-3, **8**-7–**8**-8, **8**-16–**8**-17, **8**-18, **9**-5, **9**-8, **12**-37–**12**-38, **13**-21, **C**-17–**C**-25
cell collapsing, **C**-2–**C**-3, **C**-4, **C**-5–**C**-6, **C**-8, **C**-16, **C**-19, **C**-23
children, **8**-17, **C**-4, **C**-7, **C**-10, **C**-19, **C**-24–**C**-25
control-total computation, **C**-4, **C**-8–**C**-9, **C**-16–**C**-17, **C**-20, **C**-23, **C**-25
core wave files, **5**-4, **8**-8–**8**-16, **10**-37
duplication control factor, **8**-4, **8**-5, **13**-23, **C**-1, **C**-2
first-stage ratio estimate factor, **C**-1, **C**-12, **C**-13
full panel files, **8**-16–**8**-19, **12**-1, **12**-37–**12**-38, **13**-22
household noninterview adjustment factor, **C**-1, **C**-2–**C**-3, **C**-15
imputation adjustments, **8**-4, **8**-5
information resources, **5**-16
later wave noninterview adjustments, **C**-12–**C**-13, **C**-15–**C**-16, **C**-17
missing waves, **8**-7, **13**-22
mover adjustment, **8**-4, **8**-5, **8**-6, **13**-20, **C**-13–**C**-15, **C**-16, **C**-19
new construction noninterview adjustment factor, **C**-1, **C**-12, **C**-13
noninterview adjustment factors, **C**-1, **C**-2–**C**-3, **C**-12, **C**-13, **C**-18–**C**-19
nonresponse adjustment factors, **2**-17, **2**-18, **4**-1, **6**-2, **6**-4, **8**-4, **8**-5, **8**-6, **8**-8, **C**-3
overview, **1**-7
panel, **C**-17–**C**-25
population control adjustments, **1**-6, **6**-1, **6**-4, **8**-6, **C**-3–**C**-4
pooled data from multiple panels, **8**-19–**8**-21
pre-1996 factors, **C**-1, **C**-12
quarterly estimates, **8**-15–**8**-16
raking, **8**-5, **C**-4, **C**-5, **C**-8, **C**-9, **C**-10, **C**-12, **C**-23, **C**-24, **C**-25
ratio adjustments, **C**-4, **C**-5, **C**-8, **C**-9, **C**-10, **C**-11, **C**-12, **C**-23, **C**-24, **C**-25
rotation group inflation, **8**-14
sample cut factor, **C**-13
second-stage calibration adjustments (post-stratification), **8**-4, **8**-5, **8**-6, **8**-8, **13**-21, **C**-1, **C**-3–**C**-12, **C**-13, **C**-16–**C**-17, **C**-20–**C**-25
spell estimations, **8**-18–**8**-19
subsampling of housing unit clusters, **8**-4, **8**-5
topical module files, **8**-16, **11**-28–**11**-29
Wave 1, **8**-5, **8**-9, **8**-10, **8**-14, **C**-1–**C**-12, **C**-13, **C**-14
Wave 2+, **8**-5–**8**-6, **8**-8, **C**-12–**C**-17
Weights. *See also* Reference month weights; Interview month weights; Person weights
additional household members, **8**-5, **8**-7, **8**-17, **9**-5, **9**-8
age-related, **8**-5, **C**-3–**C**-4
base, **8**-4, **8**-5, **C**-1–**C**-2, **C**-12, **C**-14
choosing, **8**-3–**8**-4, **9**-8, **10**-37, **13**-12
components, **8**-4
construction of, **8**-4–**8**-8
core wave files, **5**-4, **8**-3, **8**-4–**8**-5, **8**-7, **8**-8–**8**-13, **9**-8, **9**-15, **10**-1, **10**-2, **13**-8, **13**-22, **C**-1–**C**-25
cross-sectional, **5**-4, **8**-4, **8**-7, **11**-28, **C**-12–**C**-13, **C**-17
defined, **8**-1–**8**-2, **E**-14
effects on estimates, **1**-6, **8**-2
exiting sample members, **13**-17, **13**-19–**13**-20
family, **8**-4–**8**-5, **8**-6, **8**-8, **8**-11–**8**-12, **8**-13, **9**-15
final, **C**-1
full panel files, **8**-3, **8**-7–**8**-8, **8**-16–**8**-19, **9**-15, **12**-1, **12**-2, **12**-13, **12**-37–**12**-38, **13**-14, **13**-22, **C**-1–**C**-25
household, **8**-2, **8**-4–**8**-5, **8**-6, **8**-8, **8**-10–**8**-12, **8**-13, **8**-18, **9**-5, **9**-8, **9**-15, **C**-2–**C**-3
initial, **8**-6, **8**-7, **C**-12, **C**-13, **C**-15, **C**-17, **C**-18
longitudinal, **8**-3, **8**-4
merging, **5**-4, **13**-1, **13**-12
monthly cross-sectional, **5**-4, **8**-4
number per person record, **8**-8
panel, **8**-16–**8**-17, **8**-18–**8**-19
positive, **12**-13
program participation, **9**-5, **12**-13
purpose, **8**-1–**8**-2
redesign of SIPP and, **7**-3, **8**-1, **8**-3, **8**-5, **8**-6, **8**-9, **8**-16, **12**-37, **C**-1, **C**-2–**C**-3
reference person, **8**-6, **8**-10, **8**-11

replication, **7**-3
rotation group, **8**-5, **8**-8, **8**-12, **8**-14, **8**-16
source and accuracy statements, **5**-14, **10**-2, **11**-2, **11**-28, **12**-2, **12**-38
subfamily, **8**-4, **8**-6, **8**-8, **8**-11–**8**-12, **8**-13, **9**-15, **10**-37
topical module files, **8**-3, **8**-16, **9**-15, **11**-2, **13**-12, **13**-22
uses, **8**-8–**8**-21, **9**-8
variable names by file type, **9**-15
zero, **9**-5, **9**-8, **12**-13, **C**-19

Welfare. *See also* Program participation
history, **3**-15
reform, **1**-3, **2**-2–**2**-3, **3**-3, **3**-7, **3**-15, **5**-11, **9**-7, **10**-27

Well-being
adult, **3**-8, **5**-16, **11**-21
children, **3**-7, **3**-9, **5**-16, **11**-21
extended measures of, **3**-8, **3**-10, **5**-2, **5**-3
information resources, **5**-2, **5**-3, **5**-16
topical modules, **3**-7, **3**-8, **11**-21

What's Available from the Survey of Income and Program Participation, **5**-15

WIC program, **4**-16, **9**-7
authorized recipient, **10**-28
ID variables, **9**-14, **10**-27, **10**-28, **12**-29, **12**-30, **12**-31
imputed coverage, **10**-28, **12**-28
infant population, **8**-17
program units, coverage, and recipiency, **10**-29, **10**-30–**10**-31, **12**-28, **12**-29, **12**-30, **12**-31
unit totals, **10**-29

Wide-record format, **13**-2, **13**-6, **13**-7, **13**-9

Women, **5**-16

Work. *See also* Employment; Labor force status
disability, **3**-11, **3**-12, **3**-15
expenses related to, **3**-15
history, **3**-9, **3**-15, **5**-2
at home, **3**-6, **3**-16
moonlighting, **3**-3
part-time, **4**-8
schedule, **3**-4, **3**-7, **3**-16
time spent looking for, **3**-3

Working papers, **1**-13, **5**-13, **5**-14, **5**-15