Survey of Income and Program Participation

2014 Panel Users' Guide

2019 Edition
Washington, D.C.

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

CONTENTS

1 INTRODUCTION TO THE 2014 PANEL 1

1.1 2014 PANEL 1
1.2 WHY REENGINEER SIPP 2
1.3 REENGINEERED SIPP FIELD TESTS 2
1.4 2014 PANEL AND BEYOND 3

2 SAMPLING DESIGN, ORGANIZING PRINCIPLES AND INTERVIEW PROCEDURES 4

2.1 SAMPLE DESIGN 4
2.1.1 Primary Sampling Units (PSUs) 4
2.1.2 Selection of Addresses in Sample PSUs 5
2.2 ORGANIZING PRINCIPLES 5
2.2.1 Panels 5
2.2.2 Waves 5
2.2.3 Reference Periods 5
2.3 INTERVIEW PROCEDURES 6
2.3.1 Nonresponse 7
2.3.2 Feedback/Dependent Interviewing 8

3 SURVEY CONTENT 10

3.1 ADULT WELL-BEING 10
3.2 ASSETS 11
3.3 CHILD CARE 16
3.4 CHILD WELL-BEING 18
3.5 DEMOGRAPHICS 19
3.5.1 Education 19
3.5.2 Family and Household 20
3.5.3 Language 24
3.5.4 Marital Status and Marital History 25
3.5.5 Nativity, Citizenship, and Parent Nativity 27
3.5.6 Parent Mortality 28
3.5.7 Residence 29
3.6 DEPENDENT CARE EXPENSES 31
3.7 DISABILITY 32
3.8 EDUCATION ENROLLMENT 34
3.9 EMPLOYMENT AND EARNINGS 34
3.9.1 Commuting and Work Schedules 38
3.10 FERTILITY 40
3.11 HEALTH CARE UTILIZATION AND MEDICAL EXPENDITURES 41
3.12 HEALTH INSURANCE 43
3.13 PROGRAMS AND INCOME TRANSFER 45
3.13.1 Disability Income Payment 45
3.13.2 Energy Assistance 48
3.13.3 Supplemental Nutrition Assistance Program (SNAP) 49
3.13.4 General Assistance 52

3.13.5 Life Insurance Retirement Income *55*
3.13.6 Lump Sum Severance Pay/Retirement Plan Income *56*
3.13.7 Miscellaneous Income *58*
3.13.8 Other Assistance Income *59*
3.13.9 Retirement Income *62*
3.13.10 School Meals *65*
3.13.11 Social Security Child Benefits *66*
3.13.12 Social Security Self Benefits *68*
3.13.13 SSI *69*
3.13.14 Survivor Income Benefits *72*
3.13.15 Support Received *74*
3.13.16 Support Paid *75*
3.13.17 TANF *76*
3.13.18 Tax Returns *79*
3.13.19 Unemployment Compensation Payment *81*
3.13.20 Veterans Benefits *83*
3.13.21 WIC *85*
3.13.22 Workers' Compensation *88*
3.14 INCOME AND POVERTY RECODES 91
3.14.1 Income Recodes *91*
3.14.2 Poverty Recodes *98*

4 LOCATING INFORMATION 101

4.1 SIPP WEBSITE 101
4.2 DATA 101
4.3 FTP SITE 101
4.4 ORLIN RESEARCH SIPP DATA TOOL 102
4.5 DATAFERRETT 102
4.6 EXTERNAL RESOURCES 103
4.6.1 Inter-university Consortium for Political and Social Research (ICPSR) *103*
4.6.2 National Bureau of Economic Research (NBER) *103*
4.7 RESEARCH PRODUCTS 103
4.7.1 P-70 Reports *103*
4.7.2 Table Packages *104*
4.7.3 Conference Papers and Presentations *104*
4.7.4 Working Papers *104*
4.7.5 Technical Documentation – Codebooks and Other Metadata *105*
4.7.6 Source and Accuracy Statements *105*
4.7.7 User Notes *105*

5 THE SIPP PUBLIC USE FILES 106

5.1 TYPES OF SIPP FILES 106
5.2 UNDERSTANDING THE ID VARIABLES IN SIPP 106
5.2.1 Sample Unit IDs (SSUID) *106*
5.2.2 Person Numbers (PNUM) *107*
5.3 IDENTIFYING PERSONS 107

5.3.1 Determining Monthly Household Composition *107*
5.3.2 Determining Monthly Family Composition *109*
5.3.3 Monthly Household Relationships *110*
5.3.4 Identifying Parents, Children, Spouses, and Partners at a Monthly Level *113*
5.3.5 Grouping Families *113*
5.3.6 Income Screener Questions *114*
5.3.7 Identifying Program Participation in Government Transfer Programs *115*
5.3.8 Collecting Information for Program Units *115*
5.3.9 Identifying Program Participation in Government Transfer Programs *116*
5.3.10 Coverage and Ownership Variables *116*
5.3.11 Detailed Means-Tested Program Participation Variables *119*
5.3.12 Means-Tested Program Participation in a Sample Household *123*

6 DATA EDITING AND IMPUTATION 127

6.1 DATA EDITING 127
6.1.1 Types of Missing Data *127*
6.1.2 Goals of Imputation *128*
6.1.3 Assessing the Influence of Imputed Data on Analysis *128*
6.1.4 Processing SIPP Data *129*
6.2 IMPUTATION 129
6.2.1 Model-Based Imputation *129*
6.2.2 Sequential Hot-Deck Imputation *132*
6.2.3 Specifying cold-deck or initial donor values *134*
6.3 STATUS FLAGS (IMPUTATION METHOD) 135
6.3.1 Confidentiality Procedures for the Public Use Files *135*

7 NONSAMPLING ERROR, SAMPLING ERROR AND WEIGHTING 137

7.1 NONSAMPLING ERROR 137
7.1.1 Effects of Nonsampling Error on Survey Estimates *138*
7.2 SAMPLING ERROR 139
7.2.1 Direct Variance Estimation *139*
7.2.2 Variance Units and Variance Strata, 1990–2014 Panels *140*
7.2.3 Replicate Weights for the 1996-2014 Panels *140*
7.2.4 Approximate Variance Estimates *141*
7.2.5 Variance Estimation with Imputed Data *142*
7.3 WEIGHTING 142
7.3.1 Choosing a Weight *142*
7.3.2 How Weights Are Constructed *143*
7.3.3 Reference Month Weights *144*
7.3.4 Panel and Calendar Year Weights *145*

APPENDIXES 147

A. ACRONYMS 148
B. GLOSSARY 150

REFERENCES 158

This page was intentionally left blank.

Chapter 1

1 Introduction to the 2014 Panel

The Survey of Income and Program Participation (SIPP) is a longitudinal household survey conducted by the U.S. Census Bureau. The survey's mission is to provide a nationally representative sample for evaluating: 1) annual and sub-annual income dynamics; 2) movements into and out of government transfer programs; 3) family and social context of individuals and households; and 4) interactions among these items.

This chapter introduces users to the reengineered 2014 SIPP Panel. Topics discussed include:

- The 2014 SIPP Panel and modifications compared to 2008 Panel
- Why reengineer SIPP
- Reengineered SIPP field tests
- The 2014 Panel and beyond (evaluating the reengineered 2014 Panel)

1.1 2014 Panel

SIPP is administered in panels and conducted in waves. Within a SIPP panel, the entire sample is interviewed over a 4-year period which includes a group of annual interviews conducted during a 4-month period. These groups of interviews are called waves. The first time an interviewer contacts a household is Wave 1; the second time is Wave 2, and so forth. The Census Bureau plans to conduct the 2014 panel in four waves, the first of which began in February 2014. For the 2014 panel, the Census Bureau reengineered the SIPP to accomplish several goals:

- Enhanced respondent recall via a new Event History Calendar (EHC)
- 12-month, calendar-year reference period instead of the previous 4-month reference period
- Reduced respondent burden
- Increased use of administrative records
- Updated operating system
- Reduced costs

The 2014 panel collects detailed information on cash and non-cash income (including participation in government transfer programs) once per year. A key feature of the reengineered SIPP is an annual data collection period.

To accomplish the shift to annual interviewing without a loss in data quality, the SIPP survey instrument collects several topics via an event history calendar (EHC). The EHC helps respondents recall information in a more natural "autobiographical" manner by using life events as triggers to recall other economic events. For example, a residence change may often occur contemporaneously with an employment change. The entire process of compiling the calendar focuses, by its nature, on consistency and sequential order of events, and attempts to correct for otherwise missing data. The EHC was developed through a series of annual field tests conducted between 2010 and 2013.

The 2014 panel design does not utilize topical modules as in prior production SIPP instruments; however, much of the traditional SIPP topical module content is integrated into the 2014 SIPP interview. Subsequently, all survey content is collected or validated during each wave of the panel. At the end of this chapter, Table 1-2 provides summary statistics for the 2014 panel, as well as summary statistics for all previous SIPP panels.

Table 1-1: Summary of 2014 Panel Changes

	2008 Panel	2014 Panel
Instrument	DOS-based	Blaise/C#
Interview Frequency	3x/year	Annual
Reference Period	Previous 4 months	Previous year
Sample Size	65,522 households	53,070 households
Questionnaire	Core + Topical Modules	Core
Collection Format	Traditional question/answer	Question/answer + EHC

1.2 Why Reengineer SIPP

In 2006, the U.S. Census Bureau began a complete redesign of the Survey of Income and Program Participation. The redesign followed a program review precipitated by a budgetary crisis. To stay viable, SIPP needed to modernize and more efficiently provide the critical information necessary to understand patterns and relationships in income and program participation. The reengineering set out to reduce respondent burden and costs, to improve data quality and timeliness, and to modernize the instrument and processing.

1.3 Reengineered SIPP Field Tests

To develop the instrument and provide information for use in evaluation, five SIPP- EHC field tests were conducted (in 2008, 2010, 2011, 2012, and 2013). The 2008 test was a paper-based reinterview of respondents from the 2008 SIPP panel. Following the success of the 2008 test, in 2010 the Census Bureau conducted a feasibility test with a sample of 7,982 addresses drawn from high poverty strata in 10 states. The 2010 test used a prototype instrument programmed in Blaise and C#.

For 2011, we interviewed a new test sample from all over the United States, using an improved instrument based on the 2010 design. We continued to follow this sample during the 2012 and 2013 field tests. The 2012 field test was a Wave 2 interview, and the 2013 field test was a Wave 3 interview of this same sample. For the 2012 sample, we added a Computer Audio Recorded Interview (CARI) component to the interviewing period to test recording of interviews. The purpose was to test the interaction between interviewer and respondent in order to enhance training and possible improve questions clarity.

An evaluation of the 2011-2013 field tests provides comparisons with data from the traditional three-interviews-per-year SIPP instrument and with administrative records. While estimates from the two survey instruments (SIPP-EHC and 2008 SIPP) differ statistically in some cases, these differences are

typically small. Rates of agreement with administrative data are better for SIPP-EHC (U.S. Census Bureau, 2013) than for the 2008 panel.

1.4 2014 Panel and Beyond

The Census Bureau will continue to evaluate the reengineered SIPP using collected respondent data, survey paradata, and administrative records data. Topics for future evaluation include:

- Comparing 2008 panel and 2014 panel data from overlapping reference periods
- Comparing SIPP, ACS, and CPS population estimates
- Analyzing the relationship between interviewer certification score and interviewer production
- Reviewing the average number of questions per interview, item-level don't know/refuse rates, and interviewer notes to identify areas for survey improvement
- Using CARI technology to conduct quality assurance by evaluating question delivery and comprehension as well as monitoring interviewer performance and conduct
- Comparing survey responses to administrative data, for areas where administrative data are available.

Table 1-2 Summary of the 1984-2014 SIPP Panels

Panel	Date of First Interview	Date of Last Interview	# of Wave 1 Eligible Households	# of Waves
1984	10/83	07/86	20,897	9
1985	02/85	08/87	14,306	8
1986	02/86	04/88	12,425	7
1987	02/87	05/89	12,527	7
1988	02/88	01/90	12,725	6
1989	02/89	01/90	12,867	3
1990	02/90	09/92	19,800	8
1991	02/91	09/93	15,626	8
1992	02/92	05/95	21,577	10
1993	02/93	01/96	21,823	9
1996	04/96	03/00	40,188	12
2001	02/01	01/04	50,500	9
2004	02/04	01/08	51,379	12
2008	09/08	12/13	52,031	16
2014	02/14	05/17	42,491	4

2 Sampling Design, Organizing Principles and Interview Procedures

The 2014 panel sampled approximately 53,070 designated Living Quarters (LQs); it yielded approximately 42,491 eligible LQs[1]. 29,825 households were interviewed between February and June 2014, resulting in 67,994 person interviews and a response rate of 70.19%. This chapter discusses the essential features of SIPP, including:

- Sample Design
- Panel organization
- Interview procedures
- Nonresponse
- Dependent data

2.1 Sample Design

The SIPP sample universe is the civilian, noninstitutionalized population of the United States. The sampling universe is based on addresses from multiple sources, chiefly the 2010 Decennial Census, and contains approximately 304.4 million individuals. SIPP sampled housing units from the current Master Address File (MAF), which is maintained by the U.S. Census Bureau and is the source of addresses for the American Community Survey, other demographic surveys, and the decennial census. The MAF is updated using the U.S. Postal Service's Delivery Sequence File and various automated, clerical, and field operations.

The 2014 SIPP sample is a multistage stratified sample of 53,070 housing units from 820 sample areas designed to represent the civilian, noninstitutionalized population of the U.S.

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 addresses within sample PSUs.

2.1.1 Primary Sampling Units (PSUs)

PSUs comprise one or more contiguous counties. Single counties are used as long as the county has a population of 7,500 or more. When the population threshold is not met, adjacent counties are combined. Larger populated PSUs are identified as self-representing (SR) PSUs, while the remaining PSUs are identified as non-self-representing (NSR). Generally, PSUs with 100,000 or more housing units are classified as SR. SR PSUs are in the SIPP sample with certainty, while the NSR PSUs are stratified and selected with a probability proportionate to their size. During the stratification process, NSR PSUs are grouped according to their similarity on specified poverty measures. Given that SIPP uses a state-based sample design, all strata are formed within state boundaries. During the PSU selection process, two NSR

[1] An LQ may be ineligible for interview for many reasons, most commonly the LQ is vacant or no longer a residential unit.

PSUs are selected from each stratum with their probability proportionate to their size in relation to the entire stratum in which they belong. The 2014 panel contains 344 SR PSUs and 476 NSR PSUs in the sample.

2.1.2 Selection of Addresses in Sample PSUs

The universe of addresses within each sample PSU is divided into two strata, one with a higher concentration of low-income households and the other with a lower concentration of low-income households. Addresses are sorted by geographic and demographic variables, and a systematic selection of units is taken from each stratum. A higher sampling rate is used in the stratum with the higher concentration of low-income households, thereby resulting in an oversample of low-income households. For the 2014 SIPP, the sampling rate for the low income stratum is 1.47 times the overall sampling rate in each PSU. This results in an 18 percent increase in the number of cases in and near poverty than without oversampling.

The 2014 SIPP includes a state expansion sample of 13,800 addresses selected from 16 states: Alabama, Arizona, Arkansas, Georgia, Illinois, Indiana, Kentucky, Louisiana, Michigan, Mississippi, New Mexico, North Carolina, Ohio, Pennsylvania, South Carolina, and Tennessee. The size of the expansion in each state was determined by the amount of sample needed in order to reach a target coefficient of variation[2] (CV) of 6 percent on the estimate of low income. The four largest states – California, Florida, New York, and Texas - met this requirement without any additional sample.

2.2 Organizing Principles

2.2.1 Panels

A Panel survey is a type of longitudinal study in which sampled persons are interviewed in intervals overtime. The SIPP 2014 panel will track individuals over a span of four years.

2.2.2 Waves

One full cycle of administering the questionnaire to the entire panel is a wave. The Census Bureau plans to conduct the 2014 panel in four waves.

2.2.3 Reference Periods

For the 2014 SIPP, the reference period is the preceding calendar year. However, key month level data that SIPP previously provided will still be captured in the reengineered SIPP. For example, SIPP asks each respondent if he or she had a health insurance plan at any time during the previous 12 months. If the answer is yes, the 2014 SIPP utilizes the event history calendar to develop spell and month level data for the respondent's coverage over each of the months in the reference period. For each wave, SIPP releases data covering the 12 individual months in the reference year. Over the course of a 4-wave panel, data are collected and released covering 48 consecutive months.

[2] The coefficient of variation equals the ratio of the standard deviation to the mean.

2.3 Interview Procedures

During Wave 1, the interviewer visits the sampled address, compiles a household roster, and attempts to interview all members of the household who are aged 15 years and older. If it is unclear, the instrument determines whether each person is a household member by asking a series of questions. Typically, a usual resident is one who sleeps in the household the majority of the time. While the Census Bureau prefers that all respondents who are present at the time of the interview answer for themselves, we will accept proxy interviews from another household respondent when necessary. Within each household, the instrument identifies a reference person, typically the owner or renter of the housing unit.

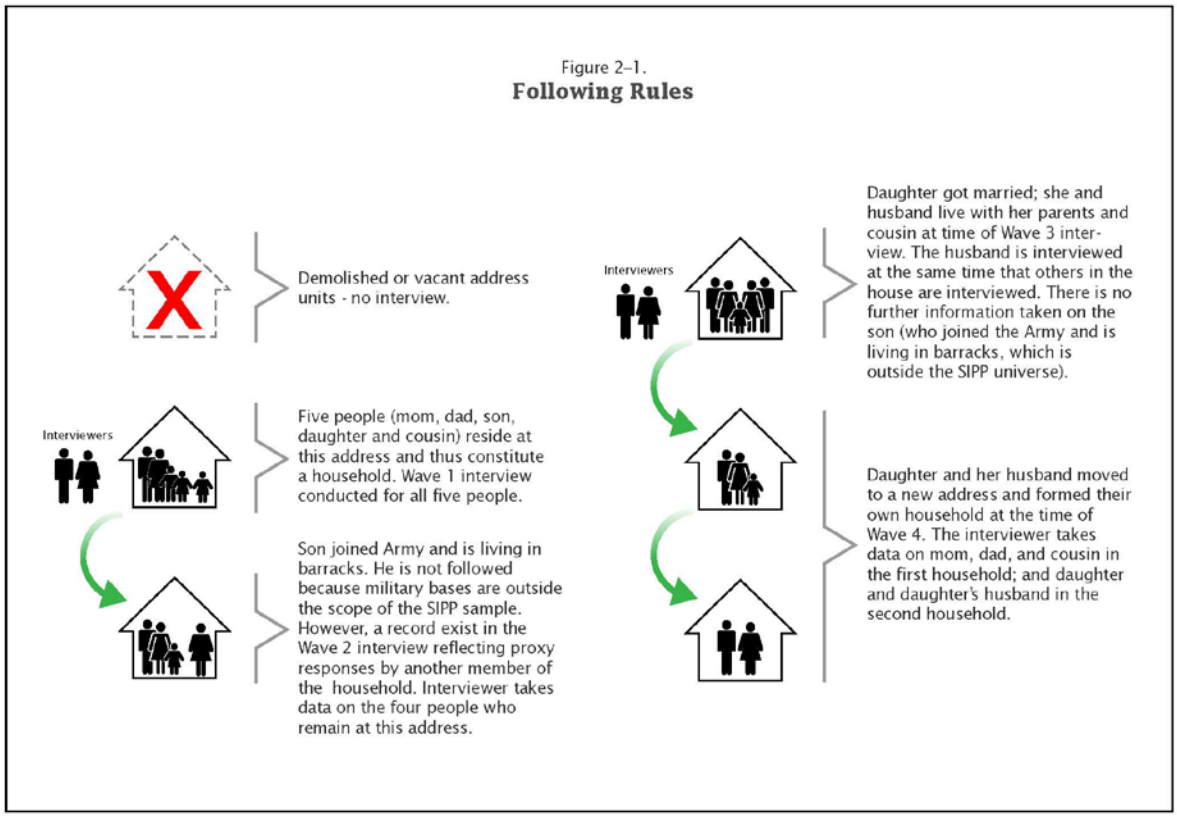
Beginning with Wave 2, SIPP switches from a household survey to a person-based survey. When visiting the original household, the interviewer updates the household roster for the housing unit, listing all people living or staying there, including anyone who may have joined the household, such as a new spouse or baby, and the dates they entered the household.

For those remaining at the same address, the interviewer verifies that certain previously collected information still applies, completes the questionnaire for each person aged 15 years and older, and collects certain information for children under age 15. Information is likewise collected for all new household members. Also noted are people who left the household and their dates of departure, along with their new address (if known).

Movers are interviewed at their new addresses, along with other household members who are living or staying there. When original sample persons (OSPs) 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 OSP. Similarly, when new individuals move in with original sample people after the first interview, they too become part of the SIPP sample for as long as they continue to live with an OSP. If no OSPs live at an address where a previous interview was conducted, SIPP does not collect information from the new occupants of that address.

If an entire household moves, the interviewer tries to find the original sample people and interview them at their new address(es). Should the entire household move more than 100 miles away from a SIPP PSU, Field Representatives (FRs) attempt to conduct the interview by telephone. If the household cannot be reached, we drop the sample members from the survey.

SIPP does not interview OSPs, 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, we will resume trying to interview them.



2.3.1 Nonresponse

The Census Bureau distinguishes between household and person nonresponse. Household nonresponse occurs either when the interviewer cannot locate the household or when the interviewer locates the household but cannot interview any adult household members. Person nonresponse occurs when the FR interviews at least one person in the household but cannot interview at least one other household member. This usually occurs because that person refuses to answer the questions or is unavailable and a proxy cannot provide the responses.

Since SIPP follows all OSPs from one wave to the next, those members that form new households are also included in the SIPP sample. This expansion of original households can be estimated within the interviewed sample, but is impossible to determine within the non-interviewed sample. Therefore, a growth factor based on the growth in the known sample is used to estimate the unknown expansion of the non-interviewed households. Growth factors account for the additional nonresponse stemming from the expansion of non-interviewed households. They are used to compute a more accurate estimate of the weighted number of non-interviewed household units (HUs) at each wave. The Census Bureau categorizes household nonresponse as Types A and D, and person-level nonresponse as Type Z.

2.3.1.1 Type A Household Nonresponse

Type A household nonresponse occurs when the interviewer finds an eligible household address, but obtains no interviews.

Examples of Type A nonresponse include:

- 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.
- Interviews cannot be conducted because of language barrier.

When this type of household nonresponse occurs in Wave 1, SIPP does not attempt to interview the household members at subsequent waves. For Type A nonresponse that occurs in subsequent waves, however, interviewers try to obtain interviews in the following waves.

2.3.1.2 Type D Household Nonresponse

Type D household nonresponse occurs when original sample members move to an unknown address or an address where the respondent is unable to be interviewed. This applies only to Wave 2 and beyond.

2.3.1.3 Type Z Person-level Nonresponse

Type Z person-level nonresponse occurs when a sample person is part of the household on the date of the interview but refuses to answer, or is not available for the interview and a proxy interview is not obtained. While household nonresponse is usually handled by weighting adjustments, Type Z cases are handled by imputation. (See discussion of imputation and weighting in Chapters 6 and 7.)

2.3.1.4 Item Nonresponse

Item nonresponse is an additional source of missing data; it occurs when respondents do not answer one or more survey questions, even though they complete most of the questionnaire. Respondents might refuse to answer a particular question or set of questions. Sometimes, item nonresponse occurs when respondents do not have the information requested. Although interviewers are trained to attempt to persuade respondents to answer all applicable questions, and will call back if a respondent can provide data later, those efforts are not always successful. Item nonresponse can also result from the post-interview 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 (see Chapter 6).

2.3.2 Feedback/Dependent Interviewing

Dependent interviewing is the process in which information from a previous interview(s) carries forward into the current survey instrument in order to streamline the interviewing process and maximize data quality.

In an effort to both improve data quality and reduce respondent burden, Waves 2+ of the 2014 SIPP instrument feature enhanced use of dependent data. The 2014 instrument utilizes previous wave information for over 500 items, allowing respondents to verify information for these items and reduce interview length.

Additionally, dependent data has been shown to mitigate the negative effects of seam bias on data quality. Seam bias is a common ailment of longitudinal surveys, where event changes are reported disproportionately at the "seam" between waves. The key to alleviating seam bias is to create overlapping periods where one wave's interview period includes a portion of the next year's reference period. Through dependent interviewing, the instrument already possesses data for the early part of the next wave's reference period. With this data from the previous wave, the Computer-Assisted Personal Interviewing (CAPI) instrument can tailor question wording to remind respondents of their situation during the previous wave. Therefore, the recall window shrinks and respondents are less apt to report changes at the transition between two reference periods. When SIPP introduced increased use of dependent interviewing in the 2004 panel, every area of the survey experienced a decrease in seam bias, with many topics cutting its effect in half (Moore, 2008).

Utilizing dependent data in a household survey, particularly one allowing for proxy respondents, creates the possibility of sharing information within a household that a respondent would rather keep private. To protect respondent confidentiality within a household, the Census Bureau instituted the Respondent Identification Policy (RIP) in 1998. RIP mandates that Census demographic surveys gain respondent consent to reference any previously collected data. Therefore, there is a small subsample of respondents for whom dependent data is not employed.

3 Survey Content

While 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, SIPP also collects extensive data on many additional facets of economic well-being. Employment status and income are not the sole determinants of households and individuals' financial situations. To create a more complete representation of national well-being, SIPP also collects extensive information concerning family dynamics, educational attainment, housing expenditures and conditions, asset ownership, health insurance, disability, childcare, and food security. These data put the income, program recipiency, and program eligibility into the child care, family, and social context. Thus, SIPP is the only Federal data source that can facilitate the examination of the ways in which these factors interact to influence financial well-being and movement into or out of government assistance programs. This chapter details the content areas in SIPP and specifically:

- The information collected within the content area
- Major content changes compared to the 2008 Panel
- Why the content is important
- The level of information provided
- Content variables and where they can be found
- Other concepts relevant to the content area

3.1 Adult Well-Being

What is collected in the Adult Well-Being content?

SIPP collects information on certain characteristics of the household (e.g., are there cracks in the ceiling or walls or holes in the floor), surrounding neighborhood (e.g., is the neighborhood safe from crime), and food security (e.g., did the household ever not have enough food). These questions are intended to be answered by the household reference person; however, any household member aged 15 years and older may act as a proxy.

Why is the Adult Well-Being information important? What assistance does it provide?

These measures offer additional insight into living conditions that more common measures (such as income or wealth) are unable to provide.

What level of information does the Adult Well-Being content provide?

All of these items are recorded at the household level, meaning the answers provided by the household respondent apply to the entire household. The information on adult well-being and food security can be used to examine differences in living conditions, financial hardships, and food security among the population.

What are the Adult Well-Being variables and where can I find them?

The Adult Well-Being content is asked of the reference person. A portion of the content is located after the EHC, with the remainder located at the end of the person-level interview.

What other concepts are relevant to the Adult Well-Being content area?

These questions refer to conditions experienced during the reference year. For respondents who lived in multiple locations during this period, they refer to the household where the reference person lived the majority of the time.

3.2 Assets

What is collected in the Assets content?

The Assets content provides detailed information on assets and liabilities for individuals and households. Three types of data are collected:

1. Asset ownership during the reference period, as well as type of ownership (joint-ownership and/or individual-ownership).
2. Value of assets, and any debts held against these assets as of the last day of the reference period.
3. Income received from each asset during the reference period.

Most questions are asked of all household members who were aged 15 years and older at the time of the interview (referred to as person-level variables), while a few questions are asked only of the household reference person (referred to as household-level variables).

The following table lists topics covered within the Assets section, as well as the microdata abbreviation used for naming the associated variables. Indicators for income questions and type of ownership questions (where applicable) are included as well.

Topics Covered in the 2014 SIPP Assets Content				
Topic	Microdata abbreviation	Level of data	Income questions	Type of ownership questions
ASSETS*				
Retirement Accounts				
IRA and KEOGH accounts	IRAKEO	Person		
401k, 403b, 503b, and Thrift Savings Plan accounts	THR401	Person		
Interest-Earning Assets				
Government securities	GOVS	Person	✓	✓
Interest-earning checking accounts	ICHK	Person	✓	✓

Topics Covered in the 2014 SIPP Assets Content				
Topic	Microdata abbreviation	Level of data	Income questions	Type of ownership questions
Savings accounts	SAV	Person	✓	✓
Money market accounts or funds	MM	Person	✓	✓
Certificates of deposit	CD	Person	✓	✓
Municipal and corporate bonds	MCBD	Person	✓	✓
Educational savings accounts	ESAV	Household		†
Other Income-Generating Assets				
Stocks	ST	Person	✓	✓
Mutual funds	MF	Person	✓	✓
Rental property	RP	Person	✓	✓
Annuities	ANN	Person	✓	
Trusts	TR	Person	✓	
Other Assets				
Regular (non-interest earning) checking accounts	CHK	Person		✓
Other real estate	RE	Person		✓
Businesses owned as a job	BUSJ(1-7)	Person		✓
Businesses owned as an investment only	BUSI(1-3)	Person		✓
Life insurance policies	LIFE	Person		
Primary residence (for non-mobile homes)	PR	Household		†
Primary residence (for mobile homes)	MH	Household		†
Cars, trucks, and vans	VEH	Household		†
Recreational vehicles	RECV	Household		✓
Other financial investments (such as coins, collectibles, jewelry, artwork, mortgages paid to him/her, other loans owed to him/her, and royalties)	OINV	Person	✓	
LIABILITIES				
Debts Secured by Assets				
Primary residence debt (for non-mobile homes)	PR	Household		†
Primary residence debt (for mobile homes)	MH	Household		†
Rental Property Debt	RP	Person		✓
Vehicle debt	VEH	Household		†
Recreational vehicle debt	RECV	Household		†
Debt on businesses owned as a job	BUSJ(1-7)	Person		†

Topics Covered in the 2014 SIPP Assets Content				
Topic	Microdata abbreviation	Level of data	Income questions	Type of ownership questions
Debt on businesses owned as an investment only	BUSI(1-3)	Person		†
Debts Not Secured by an Asset (Unsecured Debt)				
Credit card debt and store bills	CC	Person		✓
Student loans and educational-related expenses	EDUC	Person		✓
Other debt (includes medical bills not covered by insurance, loans obtained through a bank or credit union, money owed to private individuals, debt held against mutual funds or stocks)	OT	Person		✓
ADDITIONAL TOPICS				
Rent and mortgage payments	RENTMORT	Household		
Utility payments	UTILS	Household		

*All reported assets are those held outside of other pooled assets. For example, reported mutual funds are those held outside of retirement accounts.
†Joint ownership can be inferred from the reported list of owners

Why is the Assets information important? What assistance does it provide?

The Assets section produces a wide range of national estimates for wealth, debt, and distribution. Additionally, person-level and household-level wealth and income can be used to model eligibility for various government programs. The level of detail within SIPP also provides data users the flexibility to construct their own units of analysis (individuals, families, households, etc.), and study how changes in household structure and other conditions affect wealth over time.

Major changes between the 2008 and 2014 panels:

- The 2014 panel of SIPP now asks specific questions for the following topics:
    4. Annuities and trusts
    5. Educational savings accounts (e.g., 529 and Coverdell accounts)
    6. Businesses owned as an investment only
    7. Debt from student loans or education-related expenses

- Prior to the 2014 panel, values for savings accounts, money market deposit accounts, certificates of deposit, and interest-earning checking accounts were reported as one aggregate value (in TIAJTA and TIAITA). Starting in the 2014 panel, each of these assets has its own reported value. Similarly, the values for U.S. government securities and municipal and corporate bonds are now reported individually as well.

- Additional person-level recodes are now available for the Assets content. These recodes match previously provided household-level recodes. This facilitates the study of person-level wealth and within-household wealth dynamics. These recodes include, but are not limited to: TNETWORTH (net worth of the individual), TVAL_AST (person-level sum of all asset values), and TDEBT_AST (person-level sum of all debts).
- Ratios are now used during hot-deck imputation to retain relationships between value, debt, and income. For example, for a respondent who reported an amount for the value of an asset but did not report an amount for its debt, a value-to-debt ratio is imputed from the hot-deck, and the reported value of the asset is divided by this ratio to assign an amount for its debt.
- Detailed data on joint-ownership of assets are now available for those who do not have a spouse or civil union partner in the household. Data users can now identify whether these respondents owned any assets jointly, as well as their share of value/income/debt associated with these assets. Previously, data on joint accounts were only available for married respondents.
- The number of businesses as a job on which data are collected (percent owned, value, debt) has been expanded from 2 to 7.
- The survey now uses data from the National Automotive Dealers Association (NADA) to determine the value of cars, trucks, and vans for every wave of the survey. Previously, SIPP used this microdata to set vehicle value items in the first Real Estate, Dependent Care, and Vehicles Topical Module in each panel, and then updated vehicle values assuming a representative depreciation rate.
- For many questions, respondents now have the opportunity to select from a predefined range for amounts they are unsure of or refuse to provide a specific value for. Reported ranges are used to impute values that fall within the selected range.
- Variable names are now designed to make it easier to identify related variables using consistent stems for topics and characteristics. For example, **EOWN_ICHK** and **TJSICHKVAL** both refer to interest-earning checking accounts. **TJSICHKVAL** and **TJSSAVVAL** both refer to the value of an account held jointly with spouse or civil union partner as of the last day of the reference period (for an interest-earning checking account and a savings account, respectively). The following table details the convention used to name variables within the section:

Naming Conventions for Assets and Liabilities Variables							
Person-level assets and liabilities:							
Ownership of asset or liability:	**E**	+	**OWN_** *(for assets)* *or* **DEBT_** *(for liabilities)*	+	***Variable Abbreviation***		
Details for asset or liability:	**E** *or* **T**	+	***Joint/Own Prefix***	+	***Variable Abbreviation***	+	***Characteristic***

Naming Conventions for Assets and Liabilities Variables							
	(**T** *when variable is top-coded)*		(***JS***, ***JO***, *or* ***O***, *where applicable)*		*(such as* ***THR401***, ***SAV***, ***CHK****)*		(**INC** *for income*, **VAL** *for value*, **DEBTVAL** *for debt value*, *etc.)*
Household-level assets and liabilities:							
Ownership of asset:	**EOWN_**	+	***Variable Abbreviation***				
Details for asset: (includes details for liabilities)	**E** *or* **T** (**T** *when variable is top-coded)*	+	***Variable Abbreviation*** *(such as* ***VEH, PR, ESAV****)*	+	***Asset Number*** *(when multiple assets of one type are reported)*	+	***Characteristic*** (**VAL** *for value*, **DEBTVAL** *for debt value*, *etc.)*

What level of information does the assets content provide?

Most asset questions are asked of all household members who were aged 15 years and older at the time of the interview. A few topics are collected at the household-level, including: educational savings accounts; primary residence (both mobile and non-mobile homes); cars, trucks, and vans; recreational vehicles; and rent, mortgage, and utilities payments.

What are the Asset variables and where can I find them?

All of the Asset data are collected after the event history calendar (EHC) in the post-EHC section.

What other concepts are relevant to the Assets content area?

Asset Value, Debt and Income **-** All asset and debt values are as of the last day of the reference period (December 31st of the preceding year). An asset value of zero means that either the value was zero or the asset was no longer owned as of the last day of the reference period. A debt value of zero means that the respondent did not hold the debt as of the last day of the reference period. Reported income is the total amount of income received between the first and last days of the reference period.

Educational Savings Accounts **-** The household reference person is asked whether anyone in the household was the owner or a beneficiary of an educational savings account. The survey then collects detailed information on up to three accounts with the largest balances, including: the owner and the beneficiary of each account, and the value of each account as of the last day of the reference period. Only those educational savings accounts where the owner resides in the household are included in the calculations of net worth.

Net worth and Wealth - Net worth is the sum of asset values minus the sum of liabilities for a person or household. Wealth is typically defined as only the sum of asset values for a person or household.

Primary Residence - Primary residence is the residence at which household members live most of the time. In the Assets content, primary residence variables are split into *primary residence that is not a mobile home* and *primary residence that is a mobile home*. This is to capture differences between the two home types. In general, the survey collects information on the value of the home or mobile home, the number of mortgages or loans (if any) on the property, and additional details pertaining to each mortgage or loan (such as the interest rate, whether it is fixed or variable, and if the debt is for the site if it is a mobile home).

Recreational Vehicles - SIPP asks the household reference person whether anyone in the household owned recreational vehicles and then asks to identify the type. It then collects information on the owners of each vehicle, vehicle value, and vehicle debt as of the last day of the reference period. If two of the same type of recreational vehicle are owned (such as owning two motorcycles), characteristics for the second recreational vehicle of that type are put into the 'other recreational vehicle' variables.

Respondent Burden Reduction - For person-level assets that can be owned jointly, the survey is designed to refrain from asking identical questions to spouses or partners. This is to reduce respondent burden. The survey tailors question text to ensure that proper values are collected. For example, if a married couple owns an asset jointly with a non-household member, the first spouse or partner is asked to report only "your and your spouse's share" of an asset's value, debt, and income. Then, the second spouse or partner is skipped for these questions. In the editing process, reported values are divided by two and copied to both spouses' or partners' records. In all cases, both spouses or partners are asked about individual ownership (ownership in own name).

3.3 Child Care

What is collected in the Child Care content?

SIPP collects information on child care arrangements for children under the age of 14. The 2014 SIPP instrument is designed to minimize respondent burden by streamlining the question format. Instead of running through a series of questions for each child, the questionnaire is organized by arrangement type. For example, first respondents are asked: 'did you use grandparent care?' If the respondent answers 'yes,' they are then asked which child used grandparent care. Questions regarding hours in a specific care arrangement, location, or payment by type of child care are not included in the new 2014 panel. Data are edited to store child care data on the parent and the child's record.

Major changes between the 2008 and 2014 panels:

- Reference period is December.
- Data collected annually.
- Data on hours in care and location are not collected.

- Payment for child care is based on all arrangements used for all kids. The 2008 panel collected payment information by the type of child care arrangement.
- Follow-up questions regarding Head Start enrollment were implemented to improve SIPP's estimate of Head Start usage.

Why is the Child Care information important? What assistance does it provide?

The child care data collected in the SIPP can be used to assess the number of children in parental, relative, and non-relative child care arrangements. SIPP child care data also allow us to assess the types of child care arrangements used by working and non-working parents. Child care is especially important for working families, and information collected in the child care questions and parts of the survey related to employment allow users to understand the types of child care working families use. Additionally, the information afforded by the child care questions shed light on how children spend their day when they are not in the care of their parent(s).

Lastly, the SIPP child care questions can be used to assess how much families pay for child care, if they receive government child care assistance and/or enrollment in federal Head Start programs, and if child care interfered with work and time lost from work due to child care interruptions.

What level of information does the Child Care content provide?

The data are edited so that data are available on the reference parent's record as well as the child's record. From the reference parent's record, measures can be extrapolated to a couple-level, family-level, household-level, and national-level. The reference period for all child care data is December of the reference year.

What are the Child Care variables and where can I find them?

The child care variables starting with an E are the edited variable stored on the parent's record. Variables starting with an R are recoded variables that are stored on the child's record and are based on the original variable stored on the parent's record.

A number of the variables added to the 2014 panel collect similar information that was collected in SIPP 2008. However, due to changes in the questionnaire construction, a number of variable names changed.

What other concepts are relevant to the Child Care content area?

<u>Head Start</u> – A federal subsidized child care program.

<u>Child care assistance</u> – Either federal or employer assistance.

3.4 Child Well-Being

What is collected in the Child Well-Being content?

The child well-being questions fall after the event history calendar, following the child care section. The universe of respondents consists of adults who are parents of children under 18 years old. In households where both parents are present the mother is the reference parent. The child well-being section covers a variety of topics such as meal times, reading, school engagement, outings and activities with nonresident parents. If the mother is not available for an interview, the father of the child can give proxy responses for her. In single-parent families, the resident parent is the reference parent. If neither parent is in the household, the householder is the reference parent.

Major changes between the 2008 and 2014 panels:

- The redesigned child well-being section of SIPP has fewer questions.
- The redesign is more efficient and less burdensome for the respondent.
- The 2014 SIPP uses a series of four questions used on an Urban Institute family survey to measure school engagement. Past research indicates, that the questions can be used to create a factor analysis score for school engagement.

Why is the Child Well-Being information important? What assistance does it provide?

The well-being of children is a growing area of interest to researchers and policy makers who focus on the social, cognitive, and economic security of children as they transition from preadolescents to young adults. These measures provide a look at parent's interactions with their children, as well as school engagement, which can provide a sense of how well the child is faring.

What level of information does the Child Well-Being content provide?

Child well-being data is stored on the reference parent's record and then a recode is created and placed on the child's record. Variables regarding meals, outings, reading, and religious service attendance are global questions and cover engagement with all children, not engagement child by child. Therefore, whatever value is on the parent's record is carried directly over to the child's record. For example, for the variable UDINLINGD/EDINRPAR, the reference parent with two children ages 5 and 10 tells us that she eats dinner with her children 5 nights a week. The value on EDINRPAR is then copied on to each child's record as a recoded variable (RDINRPR).

The data can be used at a child level, reference parent level (household level), or national level.

What are the Child Well-Being variables and where can I find them?

Variables starting with an E are the edited variable stored on the reference parent's record. Variables starting with an R are recoded variables that are stored on the child's record and are based on the original variable stored on the parent's record.

A number of the variables added to the 2014 panel collect similar information to what was collected in SIPP 2008. However, due to changes in the questionnaire construction, a number of variable names have changed.

What other concepts are relevant to the Child Well-Being content area?

Child Care – Child can be in parental, relative, and non-relative child care arrangements.

Child Care Assistance – Either federal or employer assistance.

Head Start – A federal subsidized child care program.

3.5 Demographics

3.5.1 Education

What is collected in the Demographic Educational Attainment content?

The demographic educational attainment content collects information for adults aged 15 and older about an individual's highest level of educational attainment. There is also information on whether high school graduates got their high school diploma by graduating from high school or by passing a GED exam. SIPP collects information about educational enrollment later in the instrument during the EHC portion.

In the 2014 panel, there is a new section on alternative credentials. This section includes information on whether respondents have a professional certification, state or industry license, or educational certificate. Respondents who reported a professional certification or license also report what organization or institution awarded the credential.

Why is the Demographic Education information important? What assistance does it provide?

This information is important because it provides details on the education, training, qualifications, and skill development of the adult population. Educational attainment is an important outcome in itself and can be used to show how education affects a variety of socioeconomic outcomes.

What level of information does the Demographic Education content provide?

All of these items are person-level items. The information can be used to examine differences in educational attainment, credentials, and skill development across subgroups of the population. The information on educational attainment and alternative credentials can be used with other SIPP topics such as labor force involvement to examine how attainment is related to a variety of socioeconomic outcomes.

What are the Demographic Education variables and where can I find them?

The educational attainment and GED questions for all household members are asked of the household respondent in the Demographics section. The questions related to professional certifications, licenses, and educational certificates are part of each individual interview.

What other concepts are relevant to the Demographic Education content area?

The new measures of credentials were developed by the federal Interagency Working Group on Expanded Measures of Enrollment and Attainment (GEMEnA), a group tasked with improving federal data collection on education, credentials, and job training. After an extensive development process, the GEMEnA group defined the concepts accordingly:

Professional certification – A credential awarded by a certification body based on an individual demonstrating through an examination process that he or she has acquired the designated knowledge, skills, and abilities to perform a specific job. The examination can be written, oral, or performance-based. Certification is a time-limited credential that is renewed through a recertification process.

License – A credential awarded by a licensing agency based on pre-determined criteria. The criteria may include some combination of degree attainment, certifications, certificates, assessment, apprenticeship programs, or work experience. Licenses are time-limited and must be renewed periodically.

Educational certificate – A credential awarded by an educational institution based on completion of all requirements for a program of study, including coursework and test or other performance evaluations. Certificates are typically awarded for life (like a degree). Certificates of attendance or participation in short-term training (e.g., 1 day) are not in the definitional scope for educational certificates.

3.5.2 Family and Household

What is collected in the Family and Household content?

Relationship to reference person, parent pointers, and parent type are collected in the 2014 panel and family and household recodes are created. The relationship of everyone in the household to everyone else is not directly collected. However, by using information collected in other questions, SIPP is able to identify household relationships. Although the relationships are not quite as detailed as those collected in the 2008 topical module, the 2014 panel has the advantage of having relationship variables for every month of the reference period. This will allow for a more nuanced study of changing household relationships over time. The 2014 data indicate cohabiting relationships that do not include the reference person throughout all months. Below is detailed information of changes to relationship to reference person, parent pointers, household and family recodes, and monthly relationships of people in the household.

Relationship to reference person

In 2008, there were 13 answer categories in the relationship to reference person question. In 2014, we have expanded that list to include 18 categories. This new list adds in-law relationships (mother-/daughter-/father-/son-in-law and brother-/sister-in-law) as well as a category for aunt/uncle,

niece/nephew to help improve the specificity of relationships. Additionally, unmarried partner and spouse categories have been split into separate answer categories for same-sex and opposite-sex couples. A simplified answer list was recoded in 2014 to match the answer categories of the 2008 panel in order to allow direct comparisons. This means relationship to reference person has two variables: an expanded answer list and a simplified answer list. However, in 2008 all same-sex couples are edited and shown in the unmarried partner category. In 2014, married same-sex couples are included in the spouse category of the simplified answer list. Relationship to reference person is a person-level variable collected in the front of the instrument with the basic demographics. This question is only asked relative to the reference person.

Monthly relationship of everyone in the household

For every month of the reference period, we create a monthly (person-month) recode of the relationship of each respondent in the household that month to every other person in the household that month. Each person with a record on the file will have a set of variables that show how the other household members are related to the record holder during the months they lived together. This recode uses a combination of residence ID, relationship to reference person, marital status, spouse pointers, partner pointers, parent pointers, parent type, Type 2 relationships, and biological child lists from the fertility section to code monthly relationships. For more information about residence ID, see the Residence content area. For more information about marital status and spouse pointers, see the Marital Status and Marital History content area. See the Fertility content area for further information about the biological child lists.

The relationships in this variable are reciprocal; for example, a grandparent and grandchild will both have an 8 as their relationship (grandparent/grandchild). Parent pointers and age variables can be used to determine which person is the grandparent and which is the grandchild. Parent pointers and age are both available at the person-month level.

There is an array of variables indicating which household member each monthly relationship variable is in reference to. These variables contain person numbers. It is important to use these variables in combination with the monthly relationship variables to map household relationships.

Parent pointers and parent type

In the 2008 panel's Demographics section, the reference person is asked if each person in the household has a mother present and a father present. In 2014, the reference person is asked if each person has a parent present. If he/she answers yes, he/she is asked if this person has a second parent present. Moving from asking about mothers and fathers to parents allows us to capture two-parent families in same-sex relationships rather than showing these families as single-parent families. This improves our ability to collect information about diverse family types without losing previously collected information.

Mothers and fathers can be identified by using parent pointers in conjunction with the sex variable. Similarly, type of mother and type of father (biological, step, adoptive) are now collected as type of first parent and type of second parent. Parent pointers are person-level variables on the child's record and contain the person number of their parent(s). In the 2014 data file, child pointers and child type are

recoded using the parent pointers and parent type. The child pointers contain the person number of all of the record holder's children who are present in the household during the interview month. They are located on the parent's record.

For 2014, in addition to the person-level parent pointers and parent type, monthly parent pointers and parent type variables will be recoded using monthly relationships, parent pointers, residence, and spouse pointers.

Household and family recodes

The 2014 panel has a recoded household-level variable for number of people in the household. There are additional household recodes, including the number of people in the household 65 and over and the number of people in the household under 18. Each of these variables has two versions: one counting only interviewed household members and one counting interviewed household members and Type 2 people (persons living in the household that month, but not at the time of interview).

In the 2008 panel, the family recode variable used the reference person and their relatives as the primary family and created subfamilies that did not include the household reference person. Because the 2014 SIPP does not collect a 'reference person' for each month of the reference period, family recodes do not have a reference person available to define a primary family. Instead, all people in the household that month are assigned a family. Because of this method of recoding families, several variables available in 2008 that identified subfamilies or number of own kids in families were not created in 2014. Like household recodes, family recodes each have two variables: one that includes only interviewed family members and one that includes interviewed family members and Type 2 people who are related to those family members.

Major changes between the 2008 and 2014 panels:

- Reference period is one year vs. four months
- We now collect information about people who lived in the household during the reference period, but no longer reside there (Type 2 people)
- Relationship of everyone to everyone else is shown monthly rather than as a topical module
- Answer lists for relationship questions have changed to distinguish opposite-sex and same-sex couples
- Parent pointers are gender-neutral, allowing for reporting of same-sex parents
- There is no householder in the monthly data, only in the interview-month record

Why is the Family and Household information important? What assistance does it provide?

SIPP is designed to observe individuals over time and explain changes in program participation, health insurance status, and other areas. Family and household information allows us to understand differences in poverty, program participation, work status, etc., by household composition.

What level of information does the Family and Household content provide?

The household content can be used to examine other content such as employment, education status, or poverty by demographic characteristics as well as to examine changes in household composition over time.

Relationship to reference person, parent pointers, and parent type are person-level variables collected at the interview month. The child pointer and type variables are recodes of the parent pointers and are person level variables. Monthly relationships of everyone in the household are person-month level. In addition, parent pointers and type are available for each month. Family and household composition variables are household-level and are available on a monthly basis.

What are the Family and Household variables and where can I find them?

Relationship to reference person, parent pointers, and parent type are collected in the demographics section at the front of the instrument. The child pointers are recodes of the parent pointers. Monthly relationships of everyone in the household to everyone else are recodes that use a combination of information from the following variables: residence ID to determine who is in the household each month, relationship to reference person, spouse pointers, partner pointers, marital status, parent pointers, parent type, fertility information, and Type 2 relationships. Family and household composition variables are recodes that use relationship variables, parent pointers, and spouse pointers in addition to information about who is in the household each month.

What other concepts are relevant to the Household and Family content area?

Relationship to reference person – This is the relationship between the reference person and the respondent at interview month.

Monthly relationships – These are the relationships between the respondent and everyone in the household that month, including Type 2 people. These relationships may change over the 12-month reference period. For example, cohabiting partners may get married. The list of relationships will also change as household composition changes and people leave or join the household.

Type 2 people – These are people that were not in the household at the time of interview but were in the household during the reference year and are currently not THHLDSTATUS = 3, 4, 5, or 6 respondents.

Parent pointers – These variables appear on the child's record and are the person number of the parent or parents present in the household at interview month. In order to determine if the parent is a mother or father, you will need to use parent pointers in conjunction with sex of the parent. Parent pointers are available for people of all ages.

Parent type – These variables appear on the child's record and reflect the type of relationship between parent and child, whether biological, step, or adoptive.

Child pointers – These are recodes of the parent pointers that place the person number of any child that is in the household at interview month on a parent's record.

Child type – These are recodes of the parent type variables that appear on the parent's record and reflect the type of relationship between parent and child, whether biological, step, or adoptive.

Monthly parent pointers and parent type – These are recodes of the parent pointers and parent type that identify and describe parents during the reference period.

Household variables – These are monthly recodes that show how many people are in the household and how many of those people are under 18 or 65 and over.

Family variables – These are monthly recodes that show family number, family type, family reference person, number of people in a family, and number of people under 18 in a family.

3.5.3 Language

What is collected in the Language content?

The content is important because language use and English ability are indicators of possible membership in a community with certain resources, while lack of English speaking ability can sometimes be a barrier. The content specifically includes information on whether people age 5 and older speak a language other than English at home. For people who speak a language other than English at home, there is additional information on what language they speak and how well they speak English. The data file also contains a recode variable that indicates households in which no one age 14 and older speaks English exclusively, or speaks a language other than English at home but still speaks English "very well."

Major changes between the 2008 and 2014 panels:

There were no new Language variables added or removed from the 2008 to 2014 panels.

Why is the Language information important? What assistance does it provide?

The content is important because it provides information on the language use and ability of the U.S. population. Additionally, language data allows researchers and policy makers to examine the association between English language proficiency and important socioeconomic outcomes.

What level of information does the Language content provide?

Most of the items are person-level items. Limited English speaking household is a variable indicating households where no one age 14 and older speaks English only or speaks a language other than English at home and speaks English "very well." The language content can be used to describe the languages spoken in the U.S. and to examine how language affects various outcomes. It can also be used to explore whether people who speak a language other than English at home differ from English-only speakers across a variety of socioeconomic outcomes.

What are the Language variables and where can I find them?

The language questions for all household members are asked of the household respondent in the main demographics section, before the EHC. If the interview is conducted in a language other than English, the household respondent is asked if others in the household speak a language other than English at

home. If the interview is conducted in English, the household respondent is asked whether anyone speaks a language other than English at home.

What other concepts are relevant to the Language content area?

Limited English Speaking Households – Households where no one age 14 and older speaks English or speaks a language other than English at home and speaks English "very well" (RLNGISOL).

3.5.4 Marital Status and Marital History

What is collected in the Marital Status and Marital History content?

All of the information that was collected in the 2008 panel is also collected in the 2014 panel. However, there are some minor differences in how the information is collected. Marital history information from the 2008 topical module is now collected in the SSA Supplement. There is also additional information collected in the 2014 panel that was not collected in the 2008 panel core.

Current marital status (never married, married, widowed, separated, or divorced) is part of the demographics section and asked of the reference person about every person aged 15 years and older living in the household during the interview month. For household members who are or have been married, the following information is collected:

- The spouse (for every person that is married, spouse present)
- Number of times married
- Whether they have ever been widowed or divorced
- The year of first marriage (if they are not currently in their first marriage)
- The year of current marriage (if they are currently married)

Marital status *changes* during the reference period are collected in the EHC and asked of each respondent. Up to three spells of marital status can be recorded. All respondents aged 15 and older have a monthly marital status. Additionally, SIPP collects data to identify respondents' spouses and partners, namely:

- Respondents married during any month, and living with their spouse, have a spouse pointer for that month indicating the person number of their spouse.
- Respondents who do not have a spouse present are asked whether they had a boyfriend, girlfriend or partner in the household.
- Respondents cohabiting in any month have a cohabitation pointer for that month and are asked if they are in a registered domestic partnership or civil union with their cohabiting partner.

Only one cohabiting partnership can be recorded during the reference period. If a person is identified as a cohabiting partner, he/she is edited as a partner in all months where both persons do not have a spouse present and are present in the household.

2014 SSA Supplement

The Social Security Administration Supplement on Retirement, Pensions, and Related Content, or SSA Supplement, was sponsored by the Social Security Administration and conducted via computer-assisted telephone interviewing (CATI) after the Wave 1 interview. It contains more detailed marital history information than the 2014 SIPP.

The SSA Supplement information is comparable to the 2008 Wave 2 SIPP Marital History topical module. All variables are person level.

For more information on the SSA Supplement content, please refer to Chapter 3 in the 2014 SSA Supplement Users' Guide.

Major changes between the 2008 and 2014 panels:

- Reference period is one year vs. four months
- Cohabitation pointers are available at the monthly level and at interview month
- Cohabiting couples can report being in a registered domestic partnership/civil union
- Same-sex and opposite-sex marriages are all included as married

Why is the Marital Status and Marital History information important? What assistance does it provide?

SIPP is designed to observe individuals over time and explain changes in program participation, health insurance status, and other areas. Detailed information about marital status, relationship status, and changes over time in relationship status allows us to make comparisons of different family and relationship types.

What level of information does the Marital Status and Marital History content provide?

The marital status and marital history information can be used to compare other content in the survey such as employment, education, or poverty by demographic characteristics as well as to examine changes in household composition over time. Marital status, spouse pointers, and cohabitation pointers are available at both the person and person-month levels. Registered domestic partnership/civil union is only a person-month variable. Times married, ever widowed, ever divorced, year of first marriage, and year of current marriage are all person-level variables.

What are the Marital Status and Marital History variables and where can I find them?

Marital status is asked of the reference person for every person aged 15 years and older, in the front end of the instrument. Spouses are identified for those with a marital status of married, spouse present.

Times married, ever widowed, and ever divorced are available for each person aged 15 years and older, year of current marriage is available for each person aged 15 years and older and currently married, and year of first marriage is available for each person aged 15 years and older that has ever been married.

Marital status changes are asked in the EHC section of each person aged 15 years and older. This spell level information is then recoded into a monthly marital status.

What other concepts are relevant to the Marital Status and Marital History content area?

Marital status – Both person-level (interview month) and person-month marital status variables are measured as married, spouse present; married, spouse absent; widowed; divorced; separated; or never married.

Spouse pointers – Both the person-month and the person file contain the person number of the record holder's spouse. The person-level variable indicates the spouse at interview month.

Cohabitation pointers – Both the person-month and the person file contain the person number of the record holder's cohabiting partner. The person-level variable indicates the partner at interview month.

Domestic partnership – This is only asked of those in a cohabiting relationship at some point during the reference period, and exists only as a monthly variable on the person-month file.

3.5.5 Nativity, Citizenship, and Parent Nativity

What is collected in the Nativity, Citizenship, and Parent Nativity content?

The nativity and citizenship questions are located in the Demographics section and asked of the reference person about everyone in the household. The parent nativity questions are located in the Parent section of the survey and are asked of all respondents whose biological parent(s) is/are not present in the household. Questions include:

- Whether a respondent was born in the United States
- State of birth
- Nativity status and country of origin for the respondent's biological parents (asked of each respondent).

For foreign-born respondents:

- Country of birth
- Citizenship status
- How the respondent became a citizen
- Immigration status on arrival
- Year of entry

Major changes between the 2008 and 2014 panels:

- Added questions on biological parents' place of birth.
- State and country of birth, immigration status on arrival, and year of entry asked of all applicable respondents.

Why is the Nativity, Citizenship, and Parent Nativity content important? What assistance does it provide?

Questions on citizenship are important for estimating the size of the foreign-born population, place of birth of native-born and foreign-born respondents, changes in citizenship status, and citizenship status on arrival to the U.S.

What level of information does the Nativity, Citizenship, and Parent Nativity content provide?

Questions on parents' nativity status can be used to calculate generational status of respondents. All nativity and citizenship questions are collected at the person level.

What are the Nativity, Citizenship, and Parent Nativity variables and where can I find them?

Nativity and citizenship are asked of the reference person for every person in the household in the demographics section.

Biological mother's and biological father's birth in the U.S. and country of birth are asked of each respondent whose biological parent(s) are not present in the household during the back end of the survey.

3.5.6 Parent Mortality

What is collected in the Parent Mortality content?

Respondents who do not have their biological mother/father in the household are asked the parent mortality questions, including:

- Whether a respondent's biological mother/father is still alive
- If a parent is deceased, the month and year of death
- The month and year that the respondent's biological mother/father was born

Major changes between the 2008 and 2014 panels:

The parent mortality content is new for the 2014 panel.

Why is the Parent Mortality content important? What assistance does it provide?

Questions on parents' mortality provide information on demographic trends in intergenerational relationships, including the length of parent-child relationships and the share of the population with aging parents. In addition, this information is useful for estimating eligibility for Social Security survivor benefits by minor or disabled children.

What level of information does the Parent Mortality content provide?

All parent mortality questions are collected for interview month at the person level.

What are the Parent Mortality variables and where can I find them?

The parent mortality variables contain information on whether one's biological mother and biological father are still alive, as well as the mother's and father's date of death and birth. These items are asked during the Parent section at the back end of the survey of each respondent whose biological parent(s) is/are not present in the household. In addition, for respondents whose biological parent(s) is/are present, 1) the parent(s) is/are marked as still alive, 2) the parent(s) birth date(s) is/are copied from their demographic information, and 3) the parent(s) death date(s) is/are marked as out of universe.

Thus, the parent mortality variables on the released file contain information on the biological parents of all respondents.

3.5.7 Residence

What is collected in the Residence section?

The Residence section collects information about an individual's residence history from the start of the reference year through the interview month, including:

- Where the respondent lived
- How long the respondent lived at the residence
- Whom the respondent lived with
- Why the respondent moved to the residence
- Tenure status
- Type of living quarters
- Receipt of housing assistance

2014 Residence Information

As with other data collected in the EHC, residence information is collected in reverse chronological order. That is, each respondent's residence history is collected beginning with the most current residence and working backward through the reference year. The instrument collects information on up to five residences at which the respondent lived during the reference year.

For the residence spell that includes January of the reference year, information on the year and month the respondent moved to the residence is collected along with the tenure status of the residence the respondent lived in prior to the residence that includes January of the reference period.

The instrument stores address information for each new residence entered. This information is then available for other respondents to select when reporting their own residence histories, or for reports of multiple residence spells at the same address. This improves the reliability and consistency of the reported address information. For each residence spell, respondents are always asked to report whom they lived with, the tenure status, type of living quarters, housing assistance receipt, and reason for move as this information may vary across spells.

In some instances where respondents in a household report living together for an entire reference period, residence history information for the reference period is copied from the first respondent to the other respondents who lived with the first respondent. The exception is when respondents report living together during January of the reference month. In these situations, each respondent is always asked the month and year of move and the reason for move to the residence, as these may vary across respondents. When respondents do not report living together for the entire reference period, each respondent reports residence history information separately.

The residences respondents lived at are assigned unique residence IDs, (ERESIDENCEID) which can be used to identify households across the reference year and across panels. Residence IDs are assigned in the order that they are entered into the instrument and may not correspond to the chronological ordering of residences for any given respondent. The address information and associated residence IDs are fed back in subsequent waves and are available for respondents to select.

In order to help users, variables in processing (recodes) are constructed from other variables available on the file. These include a monthly mover flag and geographic recodes such as region of residence, metropolitan status, and metropolitan principal city indicator. The monthly mover flag allows users to easily identify when a respondent changed residences and if the move was within the same county, to a new county in the same state, or to a new state.

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Respondents may report up to 5 residence spells
- Respondents may report residence spells in group quarters, medical institutions, emergency or transitional shelters, tent or trailer sites, cars or vans, and abroad
- Respondents report the reason for moving to a residence
- Metropolitan principal city indicator for large metropolitan areas (RPCI).

Why is the Residence information important? What assistance does it provide?

SIPP is designed to observe individuals over time and explain changes in income, program participation, health insurance status, and other areas. Where someone lives is associated with economic status, health, and access to various public assistance programs. The longitudinal nature of the survey makes it possible to study the timing of moves in relationship to other important events, such as birth, changes in marital status, changes in jobs, and changes in program participation or health status. The data on tenure and housing assistance are important for gauging changes in the housing situation of the nation and states. The residence data can be used to produce national estimates of migration rates, reason for move, tenure, and housing assistance.

What level of information does the Residence content provide?

Residence data are collected at the spell level in the EHC and are provided to data users in person-month format. The exceptions are tenure and housing subsidies for the sampled unit, which are collected during the front end of the survey and copied to the appropriate residence spell during the EHC portion of the survey.

What are the Residence variables and where can I find them?

All residence information is available at a monthly level. The BMONTH and EMONTH variables indicate the EHC beginning and end month of the residence spell.

Geographic information for each residence, including region, state of residence, metropolitan status, and metropolitan principal city indicator are available for each residence at a monthly level.

Type of living quarters, tenure, rent subsidy, and housing voucher receipt are available for each residence at a monthly level.

The mover flag identifies when a respondent changes residences while the reason for move is collected for each residence during the reference period.

For the January residence, year and month moved in are collected. Tenure status of the unit lived in prior to the January residence is also collected.

What other concepts are relevant to the Residence content area?

Residence ID – A variable that identifies unique households within a given original sample unit (ERESIDENCEID).

Mover flag – A monthly variable that identifies when a respondent changed residences and if the move was within the same county, to a new county in the same state, or to a new state (RMOVER).

3.6 Dependent Care Expenses

What is collected in the Dependent Care Expense content?

In multi-person households, the Dependent Care Expense section asks the household reference person if during December of the reference year anyone in the household paid for the care of a child or a disabled person so that a household member could attend work, attend training, or look for a job.

For respondents who report any dependent care expense, SIPP asks for the amount spent during the month of December. Should a respondent not be able to provide an exact amount, SIPP provides four option ranges to choose from:

1. less than $200
2. $200 to $399
3. $400 to $599
4. $600 or more

Major changes between the 2008 and 2014 panels:

SIPP now collects the amount paid.

2014 and 2008 Dependent Care Assistance Panel Year Crosswalk	
2014 Panel	**2008 Panel**
CAREPAY_ANYONE	RE34
DEPENDENT_EXP	RE35

Why is the Dependent Care Expense content important? What assistance does it provide?

The dependent care data collected in the SIPP can be used to measure the household financial burden of dependent care. For lower income households, this burden may constitute a large portion of the household monthly income.

What level of information does the Dependent Care Expense content provide?

All dependent care expense questions are collected at the household level and asked of the household reference person.

What are the Dependent Care Expense variables and where can I find them?

All of the dependent care data are collected after the event history calendar (EHC) in the post-EHC section.

3.7 Disability

What is collected in the Disability content?

The disability content covers whether the respondent has difficulty performing certain activities due to a physical, mental, or emotional condition. The section contains six functional limitation questions, along with additional questions specific to children or the working-age population. Many of the questions in the Disability section have been present in past SIPP topical modules, though not asked during each wave of the panel. This panel will be the first time a longitudinal trend of disability status and type will be available from SIPP.

The six functional limitation questions cover difficulties with hearing, seeing, cognitive activities, ambulatory activities, self-care activities, and independent living activities. This set of six questions is consistent with the standard disability questions implemented across multiple government surveys, including the American Community Survey (ACS) and the Current Population Survey (CPS).

The additional child questions ask about developmental delays (under 5 years old only), difficulty playing with other children (5 to 14 years old), and difficulty with school work (5 to 14 years old).

The work-related questions apply to individuals aged 15 years and older and cover difficulty finding or keeping a job, limitations in the kind or amount of work possible, and being prevented from working at all. The work disability items were present in the 2008 panel core questionnaire and are unchanged in the 2014 panel.

In order to help users, two disability recode variables are constructed that help determine the type and general measure of disability of the respondent:

- RDIS: Since 2008, a set of six standard disability questions has been used across government surveys, including ACS and CPS. These six questions are ESEEING, EHEARING, ECOGNIT, EAMBULAT, ESELFCARE, and EERRANDS. RDIS indicates the respondent has at least one of these limitations.

- RDIS_ALT: Because the SIPP also asks the child limitation questions EDDELAY, EPLAYDIF, and ESKOOLWK, and the work limitation questions EFINDJOB and EJOBCANT, another recode variable has been provided (RDIS_ALT) that indicates the respondent has at least one of these 5 child or work limitations, OR one of the six standard limitations.

2014 SSA Supplement

The Social Security Administration Supplement on Retirement, Pensions, and Related Content, or SSA Supplement, was sponsored by the Social Security Administration and conducted via computer-assisted telephone interviewing (CATI) after the Wave 1 interview. It contains more detailed marital history information than the 2014 SIPP.

For more information on the SSA Supplement content, please refer to Chapter 3 in the 2014 SSA Supplement Users' Guide.

Major changes between the 2008 and 2014 panels:

- The addition of 9 total (6 adult and 3 child) functional disability questions in addition to the three work disability questions previously present in the 2008 core questionnaire.
- A model-based imputation of general disability status for functional disability and work disability status for EFINDJOB, which is used for records that do not contain enough data to reliably determine disability status.
- The creation of two recode flags indicating the respondent has a disability: RDIS and RDIS_ALT.

Why is the Disability information important? What assistance does it provide?

As a panel survey, the SIPP is designed to observe individuals over time and explain changes in program participation, health insurance status, and other areas. Changes in disability status and/or work disability status as well as the types of disability present can impact eligibility for and receipt of benefits. Moreover, the longitudinal nature of the survey allows us to measure movements in to and out of disability, creating a fuller picture of the changing nature of disability and how it relates to other social and economic aspects that influence participation in available programs, making disability data critical to achieving a main goal of the SIPP.

What level of information does the Disability content provide?

All data related to disability status are available at the person level and represent status at the time of the interview.

What are the Disability variables and where can I find them?

One work disability question, "limitations in the kind or amount of work possible," is asked prior to the Disability section as a lead-in to questions about the receipt of disability income. The remaining questions are asked in a block later in the survey.

The Disability section begins by asking all individuals if they have any difficulty seeing or hearing. Children under 5 years old are asked an additional question related to having a developmental delay. All other questions are presented in a series based on the age category that the respondent belongs to.

What other concepts are relevant to the Disability content area?

The concept of work disability may be captured through the variables EFINDJOB, EJOBCANT, and EDISABL.

3.8 Education Enrollment

What is collected in the Education Enrollment content?

The Education Enrollment section of the EHC collects information from respondents age 3 or older on the months of enrollment, grade, type, full-time or part-time status, credential worked toward, and whether or not a grade was repeated.

Why is the Education Enrollment information important? What assistance does it provide?

This information is important because school enrollment is related to labor force participation for adults, and can be an outcome measure for children.

What level of information does the Education Enrollment content provide?

Most of these enrollment variables provide person-month information. That is, the information may vary for each month for each person. However, all information should be consistent within the same spell of enrollment. The variables EEDGREP and EEDENROLL do not vary across months.

What are the Education Enrollment variables and where can I find them?

The enrollment content is asked of the people in the household that are aged 3 years or older during the EHC section of the instrument.

What other concepts are relevant to the Education Enrollment content area?

These questions refer to school enrollment during the individual months of the reference year. "Enrollment spell" refers to a length of consecutive month(s) during which the respondent is continuously enrolled.

3.9 Employment and Earnings

What is collected in the Employment and Earnings content?

SIPP collects information about an individual's work history from the beginning of the reference year through the interview month, including:

- When a job was held
- Different types of employment earnings
- Employment and business characteristics
- Reasons the respondent did not hold a job
- Whether they looked for work

While the data collected still address the same major areas as prior panels, the mode of collection and dissemination has changed greatly. The data are collected at a spell level, instead of the person-month format in which they ultimately appear. Most of the information about an employment or non-employment spell does not vary over the time it was held. This means that much of the information about a job spell will be copied to every monthly record where the job was held - the exceptions are earnings from sources other than wage/salary or business, and the monthly recoded variables created in processing. While data were collected at a spell level in the 2008 panel, much of the information about a job spell was copied to every monthly record in the reference period, regardless of when that job was held.

For all of the variables regarding timing of an event about a job, the dates are collected in reverse chronological order to improve recall; however, on the output file these variables are reported chronologically. That is, each respondent's jobs are collected and recorded beginning with the most current job and working backward, but are reported starting with the first job held in the reference year and ending with the current job. The respondent may hold multiple jobs concurrently.

The Employment section begins by asking whether the respondent currently works for pay. If not, it asks whether the respondent worked at all since January of the reference year. Basic information about the job, such as beginning and ending dates, the type of arrangement (job for employer, self-employed business, or other), and the reason for the job ending (when applicable) are collected first. The next questions are characteristics of the job/business such as industry, occupation, union status, the number of employees, and incorporation status.

For industry and occupation (I&O), respondents are asked to provide kind of business or industry, kind of work, and usual activities at work. Industry data describe the kind of business conducted by a person's employing organization, or own business if self-employed. Occupation describes the kind of work the person does on the job. These written responses to the industry and occupation questions are converted to standard codes through both automated coding and the Census Bureau coding staff.

Next, we ask about the types of earnings the respondent received (wage/salary, commission, tips, overtime, and/or bonus) and the amounts earned. For wage/salary earnings, we ask about current or most recent pay rate, and allow the respondent to report up to two changes in pay rate over the job spell. This will account for any changes in pay rate due to an annual raise or movement in and out of part-time status, for example. We also ask for the number of hours worked per week, and similarly allow the respondent to report up to two changes in the number of hours worked. Earnings from commissions, tips, overtime, and bonuses are collected at the job spell level if they are received every month; if they are not received every month, we collect monthly amounts. Finally, we ask respondents to report any time they were away from the job without pay within the reported job spell.

The same questions are asked for each job, for up to seven jobs. This is different from prior panels, which only collected data for the first two jobs and sorted them by which one was held for more weeks or hours. No "main" job is determined in the 2014 panel, and jobs are sorted on the file by the earliest month they were held within the reference period. Ties are broken by which job was reported first by the respondent. Similarly, moonlighting is no longer distinguished separately from any other job – those

with a main job and an additional part-time job will simply have two jobs listed concurrently. In the same vein, the classification of "contingent workers" is no longer used. The questions about whether a job has an "other" work arrangement and whether that arrangement was definite are still present, but they are applied to a particular job and not to an individual. In this way we can identify occasional or indefinite work that happens in addition to any main job held.

If an individual held more than seven jobs during the reference period, some summary information is collected about the additional jobs. That information is included with the employment status recode and the earnings recodes (e.g., RMESR, TPEARN) but is not included on the edited file directly.

Jobs may be linked across waves using EJB(n)_JOBID. If a job was held at the time of the interview and the respondent allows their data to be fed back in later waves, then during the subsequent interview they will enter the job spell and answer questions about the job and earnings during the next wave. Some responses are carried forward such as industry and work arrangement, while others are asked again such as occupation and earnings.

Finally, for those periods in which the respondent was not employed, information is collected about the labor force status of the respondent during that period. This includes information about why they were not working, unpaid work in a family business or farm, time spent on layoff, and time spent looking for work. To facilitate consistency in reporting, if the respondent held a job during part of the year, the survey instrument will calculate gaps in employment automatically and collect information about each spell of non-work separately. Note that it is possible for separate spells of non-employment to be present in a single month if there is a short job spell within the month. Because the data are mapped to a person-month format, this can lead to multiple spells of non-employment being mapped to the same month. In these (rather rare) cases identified by the variable ROVERLAPMN, the output file will have some characteristics from both spells of non-employment, and the timing variables will relate to the earlier spell.

In order to help users, many variables in processing, including recodes, are constructed from other variables available on the file. These include weekly and monthly employment status variables, weekly wage and salary earnings, monthly total earnings from each job, monthly earnings from all jobs, and weekly and monthly hours worked per week at each job, among others. This gives users the flexibility to construct variables to suit their own purposes if necessary, while keeping the file usable.

Major changes between the 2008 and 2014 panels:

- Reference period of one year vs. four months.
- Allow respondents to report up to two changes in wage/salary pay rate and hours worked for each job over the reference period.
- Allow respondents to report up to three periods of time away without pay.
- Allow respondents to report detailed information for up to seven jobs.
- Collect and report the components of earnings as well as totals.
- The 2014 panel uses the 2012 Public Use Microdata Sample (PUMS) census industry code list. Since prior SIPP panels used a previous version of the industry code list, the industry data within these panels requires cross-walking to the 2012 PUMS industry code list for full comparability.

- The 2014 panel uses the 2012 PUMS census occupation code list. Since prior SIPP panels used a previous version of the occupation code list, the occupation data within these panels requires cross-walking to the 2012 PUMS occupation code list for full comparability.
- Information about a job or business appears only on monthly records when that job or business was held.
- Prior wave earnings are no longer fed back.

Why is the Employment and Earnings information important? What assistance does it provide?

Changes in labor force status, earnings, or usual hours of employment can all impact eligibility for and receipt of benefits. The longitudinal nature of the survey allows us to measure movements in and out of the labor force, movement between jobs, and changes in earnings, making the labor force data critical. The weekly nature of the employment data allows for observation of short duration spells of employment and non-employment.

What level of information does the Employment and Earnings content provide?

All data related to employment status are available at the weekly level. All information about spells of employment and non-employment are included on the monthly records, and so monthly statistics can be calculated for all items.

What are the Employment and Earnings variables and where can I find them?

All of the labor force data are collected in the EHC. Each job gets its own line in the calendar, and periods of time not working are recorded on a separate line.

What other concepts are relevant to the Employment and Earnings content area?

Away without pay – Note: the question specifies that the time away should be two weeks or more. However, if the respondent reports only a single week away without pay, we do include that information on the file.

Commission – A form of payment based on the amount of services the employee performed. For example, in sales positions, many times the sales employee is paid a percentage of the amount the customer paid or a percentage of the profit received by the firm. In the SIPP, we also ask respondents to include broader pay for unit of service when reporting commission amounts, which could include payments more commonly known as piece-rates. It could also include other types of pay schemes such as painters who are paid by the room or truck drivers who are paid by the mile. Commission can be in lieu of wage and salary payments ("straight commission") or in addition to these payments.

Earnings – Remuneration (pay, wages) of a worker for services performed during a specific period of time. The SIPP focuses on reporting monthly earnings, but for wage and salary earnings, data are available at the weekly level.

Employed Persons – Persons who report holding a job, business, or doing any other work for pay during the time period specified. This includes those who are temporarily absent, whether the absence was

paid or unpaid. Also included as employed are those who worked 15 hours or more unpaid in a related household member's family business or farm.

Full-Time Workers – Those who usually work 35 hours or more per week.

Industry – Describes the kind of business conducted by a person's employing organization, or own organization, if self-employed. Industry data are derived from questions about employer/business name, kind of industry, kind of work, and usual activities at work.

Labor Force – The sum of those classified as either unemployed or employed.

Occupation – Describes the kind of work that the individual does on the job. Occupation data are derived from questions about kind of work and usual activities at work.

On Layoff – Persons are considered to be "on layoff" if they are currently unemployed due to slack work, shortages, or other business-related concerns (rather than personal misconduct). Persons on layoff who are expecting to be recalled to their positions are considered unemployed whether they looked for work or not during the layoff.

Part-Time Workers **–** Those who usually work fewer than 35 hours per week.

Unemployed Persons - Persons who had no employment during the week and were either on layoff awaiting recall or looking for work.

Wages and Salaries – Payments per a unit of time, generally not tied to performance. Wages are payment per hour of work; salaries are usually given as an annual amount. Both wages and salaries can be paid in a number of ways; data for SIPP allow them to be reported as weekly, biweekly, monthly, or twice per month, as well as reporting the hourly wage, annual salary, or gross annual amount. Any other pay schedules are asked to report monthly averages.

Weekly Hours – The number of hours the respondent usually (more than half the time) worked per week at their job during the period specified. Respondents whose hours vary are instructed to attempt to give an average amount.

3.9.1 Commuting and Work Schedules

What is collected in the Commuting and Work Schedule content?

The commuting data include means of transportation to work, distance to work, minutes to work, parking and toll expenses, additional commuting expenses, and other job-related expenses. The work schedule data collected consist of days of the week worked, days of the week worked entirely at home, the start and end times of work, the type of schedule worked, and the reason for working said schedule.

The commuting and work schedule data are collected at the spell level for each job the respondent reported in the EHC, up to seven jobs for the entire calendar year. This means that the data associated with a given job will be copied to every monthly record for the specific spell of work at the job. As with other data in the EHC, the commuting and work schedule information is collected in reverse

chronological order. That is, each respondent's information is collected and recorded beginning with the most current job.

The Commuting section begins by asking the mode of transportation used to get to work. These modes include:

1. Drove own vehicle
2. Rider in someone else's vehicle/van pool
3. Bus
4. Rail
5. Other public transportation
6. Walked
7. Bicycled
8. Drove company car
9. Worked at home
10. Other

The specific mode of transportation determines the subsequent commuting questions that each respondent is asked. If the respondent drove his or her own vehicle or a company car, the respondent is asked about miles driven to work, miles reimbursed, and parking/tolls. If the respondent commutes using a mode other than drove alone/company car (and excluding worked from home), he or she is asked about other commuting expenses. With the exception of those who worked from home, workers are asked the one-way travel time to work in minutes. Finally, all workers are asked if they have any other job-related expenses not already reported, and, if so, the annual amount of these expenses.

The Work Schedule section immediately follows the commuting questions and begins by asking which days of the week the respondent worked for the given job, the start and end time of the job, and if there were any days worked entirely at home. If the respondent did work from home, he or she is asked to report the specific days worked from home. The section concludes with two questions on the type of work schedule and the reason for the work schedule that are asked of all respondents who report a job.

The same information is collected, for each respondent, for up to seven jobs, and for up to two spells per job for the entire calendar year. In prior panels, the commuting and work schedule data were collected in separate topical modules which were not asked during the same wave. The commuting data were collected for up to three jobs held during the reference month while the work schedule data were collected for up to two jobs held during the reference month.

Major changes between 2008 and 2014:

- Allow respondents to report up to seven jobs during the year
- Commuting and work schedule data are collected concurrently rather than in separate waves

Why is the Commuting and Work Schedule information important? What assistance does it provide?

As a panel survey, SIPP is designed to observe individuals over time and explain changes in income, program participation, health insurance status, and other areas. SIPP commuting and work schedule

information, combined with the wealth of additional SIPP information, is important for understanding the social demographic and economic underpinnings of changes in work and transportation patterns.

What level of information does the Commuting and Work Schedule content provide?

Commuting and work schedule data are collected at the spell level in the EHC and provided to data users in person-month format.

What are the Commuting and Work Schedule variables and where can I find them?

The Commuting and Work Schedule Variables are asked for each job spell held by a respondent and are edited in person-month format.

What other concepts are relevant to the Commuting and Work Schedule content area?

- Labor force content
- Industry and occupation content

3.10 Fertility

What is collected in the Fertility content?

The 2014 SIPP includes an expanded fertility section, including a more comprehensive fertility history, as well as new questions. The SIPP instrument asks for complete fertility histories from all respondents aged 15 years and older, including:

- The total children ever born/fathered
- The month and year of birth for each child (NOTE: month of birth is not released on the public use file)
- A direct question about multiple partner fertility, asked of all adults with more than one child.
- Whether their current union (married or cohabiting) is a childbearing one
- Whether they are a grandparent

Major changes between the 2008 and 2014 panels:

- **Multiple partner fertility** – A new item for the Census Bureau, providing the first national estimate of multiple partner fertility.
    - Men's fertility - The questions about men's children will provide a more comprehensive portrait of men's fertility than has previously been available.
    - Shared fertility - The information about other parents will allow parent partnerships between survey respondents to be identified, even with no children in the household. For example, for adults living without children in the household, we can determine whether the three children a wife reports having had are the same three children that her husband reports. Previous data would only allow speculation as to whether any or all of those children were shared.

- Child residence and mortality – All previously asked questions regarding where children live, and whether they are deceased, have been removed from the survey.

- **Maternity Leave –** All previously asked questions regarding women's work and leave at the time of their first birth have been removed from the survey.

Why is the Fertility information important? What assistance does it provide?

The fertility data collected can be used to assess the number of children ever born to adults, the onset of parenthood via age at first birth, and the number of childbearing partnerships. The fertility questions provide information about the complexity of families, which is relevant to family stability, financial security, and children's networks.

What level of information does the Fertility content provide?

Data are available at the person, family, and household levels.

What are the Fertility variables and where can I find them?

The Fertility questions are collected in the back end of the instrument, after Assets and Disability, but before any of the Child Care or Well-Being questions.

What other concepts are relevant to the Fertility content area?

Household relationships, demographics, marital history, and child well-being are other topics in the SIPP instrument with direct ties to the fertility content. As fertility and the timing of family formation are key predictors of economic well-being, fertility also has important implications for employment and social program use.

Additional other concepts related to Fertility include the number of biological children, childlessness, year of birth of first child, age at birth of first child, whether a respondent has multiple partner fertility, year entered multiple partner fertility, number of child-bearing unions, whether a current partnership is a childbearing union, and whether a respondent is a grandparent.

3.11 Health Care Utilization and Medical Expenditures

What is collected in the Health Care Utilization and Medical Expenditures content?

Health Care Utilization

The Health Care Utilization content collects information about:

- Respondent's health status (excellent, very good, good, fair, or poor)
- Number of days sick in bed and hospitalized
- Prescription medication use
- Number of visits to dentists and medical providers
- For uninsured respondents, visits to medical providers and dentists as well as uninsured respondents' income when they received medical services

Medical Expenditures

The Medical Expenditures content collects information about respondents' medical expenses, including:

- Health insurance premiums
- Over-the-counter medical items
- Non-over-the-counter medical products and services
- Whether the respondent had a Flexible Spending Account

Major changes between the 2008 and 2014 panels:

Over-the-counter and non-over-the-counter expenses are now collected under two separate variables.

Why is the Health Care Utilization and Medical Expenditures information important? What assistance does it provide?

In light of the implementation of the Patient Protection and Affordable Care Act (ACA) and in general, it is important to identify health care utilization and the amounts spent on premiums and other medical expenditures as well as characteristics of individuals with differing utilization of health care, amounts of expenditures, and prevalence of flexible spending accounts.

The longitudinal nature of the survey allows us to measure health care utilization and medical expenditures over time to estimate the impact of health reform and changes in private insurance markets, which is important for understanding topics related to public insurance and other government programs.

What level of information does the Health Care Utilization and Medical Expenditures content provide?

All data related to health care utilization and medical expenditures are available annually at the person level.

What are the Health Care Utilization and Medical Expenditures variables and where can I find them?

This content is collected after the EHC.

What other concepts are relevant to the Health Care Utilization and Medical Expenditures content area?

Additional concepts relevant to the content area include: Health Status, Health Care Utilization, Medical Expenditures, and Health Insurance Premiums.

Note on Infants

In the 2014 SIPP Panel, infants, defined as household members less than one year old at the time of the interview, do not receive proxy interviews. This change does not affect most topics in the SIPP. However, for the Health Care Utilization and Medical Expenditures section, infants are considered a separate entity with unique health care needs, since they have unusually high medical utilization and expenditures compared to other groups. Due to these large differences between infants and the rest of the population, we do not impute health care utilization and medical expenditure information for those who are less than one year old at the time of interview. Values for these variables will be set to not in universe except for the two exceptions described below.

1. The question for the amount paid out-of-pocket for household insurance premiums coverage indirectly collects information on infants. This question is asked of all individuals over age 15, and responses include premiums for policies that cover infants. This affects the household-level variable THIPAY, the amount paid for health insurance premiums.

2. For the variable EHLTSTAT, we impute values for infants. This variable collects self-reported (or proxy reported) health status. We impute values for this variable because we do not have any reason to believe that reported health status of infants is fundamentally different from that of other young children.

3.12 Health Insurance

What is collected in the Health Insurance content?

SIPP collects information on health insurance coverage, including:

- Private health insurance
- Additional private health insurance
- Medicare
- Medical assistance (Medicaid)
- Military coverage
- Any other coverage

SIPP asks about coverage using two screener questions:

1. Are you currently covered?
2. Were you covered at any time since January 1st (prior year)? (With the exception of Medicare, which only asks the first question)

The survey captures different spells of coverage within the year, each with complete plan details. Private insurance questions ask about up to two simultaneous sources of coverage in any given month, with a follow-up question about the more specific type/source (e.g., through employer, direct-purchase, etc.). Additional questions are asked about reasons the respondent did not have different types of coverage.

Major changes between the 2008 and 2014 panels:

- Questions on Marketplace coverage, including whether the respondent obtained health insurance coverage through a health insurance marketplace and whether the respondent's health insurance premium was subsidized, beginning in Wave 2.
- Military coverage is asked about separately (instead of being subsumed under private coverage).
- Both children and adults all receive the same set of health insurance questions. However, children are asked via adult proxy.
- In the 2014 panel, all health insurance data are collected at the month-level, with the exception of the reasons for not enrolling/purchasing insurance.

Questions on reasons the respondent did not take up/purchase different types of coverage are new to the 2014 panel. In prior panels, a single core question (EHIRSN) asked why respondents did not have private coverage. Additionally, one variable covering the reason the respondent did not obtain employer-sponsored insurance was collected as part of the Employer Provided Benefits topical module. The 2014 panel collects richer data than in the past by asking

1. Whether a respondent's employer offered health insurance coverage to its employees.
2. The reason employed respondents did not have employer-sponsored insurance coverage.
3. The reason respondents did not have private coverage.
4. The reason respondents did not have public coverage.
5. For every spell that ends within the reference period, respondents are asked why that coverage ended.

Why is the Health Insurance information important? What assistance does it provide?

The health insurance data collected in the SIPP can be used to measure and track the percentage of the population with private health insurance, medical assistance, Medicare, military coverage, and other coverage. Among other things, it provides information on who the plan holder is, whether the employer pays for premiums, and beginning in Wave 2+, whether the plan was purchased in a marketplace/exchange.

The panel nature of the data allows users to look at how health insurance changes over time, either on its own or in conjunction with other variables. While high quality health insurance data has always been important, it is particularly salient now as researchers are quantifying the impacts of the Patient Protection and Affordable Care Act.

What level of information does the Health Insurance content provide?

All health insurance information is available at the person level, and coverage information is available monthly. This information can be used to measure health insurance rates or health insurance change over time either on its own or in conjunction with other variables.

What are the Health Insurance variables and where can I find them?

The block of health insurance questions is located at the end of the EHC (the final six lines of the EHC), after the government program participation questions (e.g., TANF, General Assistance, and WIC).

After reporting health coverage and the beginning and ending month of the spell, the spell-specific questions follow; the questions include source of coverage (e.g., ESI, direct-purchase, etc.). More

specifically, the health insurance questions collected outside of the EHC include four questions on why respondents did not obtain health insurance: two questions on ESI take-up, one question on direct purchase, and one question on medical assistance.

What other concepts are relevant to the Health Insurance content area?

- Private health insurance
- Employer-sponsored health insurance
- Medicare (Part A, Part B, Supplement, Part C, Part D)
- Medical assistance (Medicaid, Children's Health Insurance Program (CHIP))
- Military coverage (TRICARE or CHAMPUS, CHAMPVA, VA)
- Other coverage
- Marketplace/Exchange (beginning in Wave 2+)
- Premium Subsidies

Note on Infants

Beginning with the redesigned 2014 SIPP Panel, individuals who are less than one year old at the time of interview do not receive proxy health insurance interview questions. However, some health insurance information is collected on these infants. Each individual age one and older (or a proxy respondent) is asked about their health insurance coverage. For each type of coverage, if the person is covered, the interviewer asks who else in the household had that coverage the same months as the individual. Thus, if the infant has the same coverage as another household member for the same months as that person, the infant will have his or her health insurance reported. However, coverage will not be captured if the infant has a different type of coverage or is covered different months from the household member who reported the coverage. Infants are particularly unlikely to be covered the same months as other household members, since they are born during the reference year. We impute the missing health insurance information for infants using the same model based imputation methods as with other missing health insurance data. However, users should be aware that imputation rates will be unusually high for infants.

3.13 Programs and Income Transfer

3.13.1 Disability Income Payment

What is collected in the Disability Income Payment content?

SIPP collects person-level data about whether an individual between the ages of 15 and 69 who reported having a health condition received Disability Income at any time during the reference year.

The data collected in the 2014 panel include the same major fields as in the 2008 and prior panels, specifically the type of disability payment and the amount received during each month of the reference year.

New in 2014, the Disability Income content includes the monthly amount received by Disability Income benefit type.

The Disability Benefit section begins by asking if the respondent received any income due to a health condition during the reference year.

Respondents are then asked which type of disability benefit was received:

1. Payments from a sickness, accident, or disability insurance policy
2. Employer disability payments
3. Pension from a company or union including income from a profit-sharing plan
4. Federal Civil Service or other Federal civilian employee pension
5. State government pension
6. Local government pension
7. U.S. Military retirement pay
8. U.S. Government Railroad Retirement
9. Black Lung benefits
10. Other disability income

For each benefit type reported, the instrument asks for the months of receipt and amount received.

Amounts are collected by moving backwards chronologically from the interview month to the start of the reference period. For example, "How much do you receive in local government pension now?", "When did that start?"

If the start month for the amount reported is after the month in which benefit receipt started, the instrument then asks how much was received prior to the most current amount.

For each benefit type, respondents may report up to four amount changes during the reference period. After the fourth amount, the instrument asks for the amount received during the first month of receipt.

Major changes between the 2008 and 2014 panels:

- Collect the amount received by benefit type
- Collect up to four changes in the amount

Why is the Disability Income Payment information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources such as Disability Income sources. Since income received from a job source or a disability benefit affects an individual's economic situation, the data collected for Disability Income is important to understanding an individual's economic situation.

What level of information does the Disability Income Payment content provide?

Disability Income data collected in SIPP can be used to create estimates on the national, household, and person-levels.

Disability Income is collected on the person-level for all household members who were between the ages of 15 and 69 during the reference year. Teenagers who were age 15 and older during the reference year are in universe for the Disability Income section. Children and teenagers under the age of 15 at the

end of the reference year and adults over the age of 69 at the beginning of the reference year are not in universe for the Disability Income section.

What are the Disability Income Payment variables and where can I find them?

Disability Income data are collected after the EHC.

What other concepts are relevant to the Disability Income Payment Content area?

Black Lung Benefits – The United States Government Program enacted in 1973 to provide monthly payments and medical benefits to coal miners who have disability from pneumoconiosis as a result of employment in and around the United States' coal mines. The act also provides monthly benefits to a miner's dependent survivors if black lung disease caused the miner's death.

Disability – Disability refers to a person's physical, mental, developmental, or emotional impairment.

Disability Insurance – An insurance policy that provides the policyholder benefits in the event that he or she becomes sick or injured.

Federal Civilian Employee Pension – The Federal Employees Retirement System (FERS) refers to a civilian retirement pension for civilian federal employees who began their employment after 1987.

Federal Civil Service Pension – This refers to a retirement pension for civilian federal employees who began their civilian federal employment between 1920 and 1987.

Local Government Pension – Also known as a public pension plan, this is the Local Government employee retirement plan for vested local government employees.

Military Retirement Pay – The retirement pension plan provided for active duty military personnel with benefits that start when the employee retires.

National Guard Retirement – An Army National Guard retirement plan provided for employees who have 20 years service in the National Guard. By age 60, this retirement pension may combine with any additional income or retirement.

Other Disability Income – Provides supplementary income in the event of illness or accident that prevents the insured from working at their employment.

Pension – A pension is an employee benefits plan where employees and employers pay into a group trust fund. The payment benefits are based on a formula factoring an employee's final pay and employment tenure with an employer.

Reserve Forces Retirement – A retirement pension plan for Reserve Service Members who completed 20 years of service and who are eligible to begin receiving Reserve Forces Retirement at age 60 or depending on their active service are eligible to begin receiving their Reserve Forces Retirement before age 60.

Sickness and Accident Insurance – A type of disability insurance that provides coverage when the policyholder becomes sick, injured, or dies from an accident.

State Government Pension – Also known as a public pension plan, this is the State Government employee retirement plan for vested state government employees.

U.S. Government Railroad Retirement – A federally administered retirement program for railroad employees and employers. Railroad employees with 10 or more years of service or having at least 5 years of service since 1995 are eligible for this retirement program. Tier 1 benefits take the place of Social Security and are first payable at age 62 or after 30 years of service for early retirement. The first full retirement benefits are payable for retired employees between the age of 65 and 67. The Tier 2 benefits are similar to a private pension.

3.13.2 Energy Assistance

What is collected in the Energy Assistance content?

The Energy Assistance section collects information from the household respondent about the payment of utilities and receipt of energy assistance at any time during the reference period, including:

- Whether the household payed separately for utilities (for households receiving a rental subsidy or housing voucher)
- Whether the household received energy assistance
- Whether assistance was in the form of checks sent to the household; coupons or vouchers sent to the household; and/or payments sent directly to the utility company, fuel dealer, or landlord (for household that reported receipt of energy assistance)

Major changes between the 2008 and 2014 panels:

- Energy assistance amount is no longer collected.

Why is the Energy Assistance information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Households often receive financial assistance from programs such as Energy Assistance. Since assistance received affects a household's economic situation, the data collected for Energy Assistance are important to understanding a household's overall economic situation.

What level of information does Energy Assistance content provide?

Energy assistance data are collected at the household level. The reporting period is the entire reference period.

What are the Energy assistance variables and where can I find them?

While all data are available in a monthly format, these variables hold the same value over the entire reference period. The following variables are available:

- Indicator of whether the household payed separately for utilities (EUTILITIES)
- Indicator of whether the household received energy assistance (EENERGY_ASST)
- Energy assistance received in the form of checks sent directly to the household (EENERGY_PMT1)
- Energy assistance received in the form of coupons or vouchers sent to the household (EENERGY_PMT2)
- Energy assistance received in the form of payments sent directly to the utility company, fuel dealer, or landlord (EENERGY_PMT3)

What other concepts are relevant to the Energy Assistance content area?

Energy Assistance – Government energy assistance programs that help pay heating and cooling costs. This assistance is generally paid directly to the electric company, fuel dealer, or landlord. Eligibility varies slightly from state to state, but all recipients have limited incomes.

Household Respondent – There are several sections of the SIPP interview. Some sections are asked of only the household respondent, while others are asked of every eligible household member. The household respondent is the first eligible adult household member interviewed.

3.13.3 Supplemental Nutrition Assistance Program (SNAP)

What is collected in the SNAP content?

The SNAP section collects information about SNAP benefit receipt from the start of the reference period through the interview month. Variables on the public use file provide the following information:

- Monthly and reference period coverage indicators
- Months received
- Year receipt began if receiving in January of the reference period (wave 1)
- Reason(s) receipt began
- Reason(s) receipt stopped
- Payment amount
- Person in whose name the benefit is received (benefit owner)
- Continuation flag for spells that include the last month of the reference period

To reduce respondent burden, SIPP does not collect information about the receipt of SNAP separately for children under age 15. Respondents age 15 or older with a household income below 200 percent of the poverty line are asked to report whether they or anyone in their family (i.e., children) received SNAP benefits. In many cases, the screener clump selects one adult in the family to report SNAP receipt for all people in the nuclear family. Adults not in the screener clump are each asked whether they or anyone in their family received SNAP benefits.

As with other data collected in the EHC, SNAP information is collected in reverse chronological order. That is, each respondent is first asked if they or someone in their family are currently receiving SNAP. If a respondent reports current receipt, the month of the reference period or interview year receipt began is collected. If no one in the respondent's family is currently receiving SNAP, the respondent is then asked if they received SNAP at any time since the first month of the reference period. If receipt is reported

during that time, the months that receipt started and stopped are collected. If a SNAP spell includes January of the reference period the first time a respondent is interviewed, information on the year the respondent first began receiving benefits is collected.

For each SNAP spell, respondents are asked to report why receipt began, why receipt stopped (if applicable), payment amount, who was covered by the benefit, and the person in whose name the benefit is received (benefit owner), as this information may vary across spells.

Respondents may report up to four different payment amounts in a spell to allow for changes in payments resulting from factors such as cost of living adjustments and earnings variation. The amounts are recoded into a single monthly variable.

A continuation flag for respondents who reported receiving SNAP in the last month of the reference period provides information on whether the spell: (1) ended in the last month of the reference period; (2) ended after the last month of the reference period but before the interview month; or (3) was ongoing as of the interview month.

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Reference period indicator for coverage
- Respondents may report up to three SNAP spells
- A continuation flag for respondents who reported receiving SNAP in the last month of the reference period

Why is the SNAP content important? What assistance does it provide?

SIPP is designed to observe individuals over time and show changes in income, program participation, and other areas. SNAP is an important source of income for some low-income families with children. The longitudinal nature of the survey makes it possible to study the timing of SNAP receipt in relationship to other important events, such as changes in marital status, employment status, and/or participation in other assistance programs.

The SNAP data can be used to produce national estimates of coverage and payments. Although the United States Department of Agriculture regularly publishes data on SNAP participation and benefits, detailed SNAP data are published less frequently. The SIPP collects more detailed information on SNAP receipt than both the ACS and CPS and provides it at a more detailed level – the SIPP provides monthly and annual data while the other surveys only provide annual data. When combined with data on earnings and other income, data on SNAP benefits help provide a more comprehensive picture of a family's or household's economic well-being over the length of the panel.

What level of information does SNAP information provide?

SNAP data are collected at the spell level in the EHC and are provided to users in a person-month format.

What are the SNAP variables and where can I find them?

While all data are in a monthly format, the value of some variables are the same for entire reference period (person-level variables). The value of other variables may be the same throughout an entire spell (spell-level variables) or may vary from month to month (monthly-level variables).

Person-level variables

- A reference period coverage indicator identifies whether an individual was covered by SNAP in at least one month of the reference period (RFSCOV)
- Year receipt began, for respondents with a SNAP spell that begins in January of the reference year in respondent's first interview (EFSLCY)
- The continuation flag identifies the status of a respondent's spell if it was on-going as of the last month of the reference period (RFS_CONTFLG)

Month-level variables

- Payment amount (TFS_AMT)

Spell-level variables

- Begin and end month of the SNAP spell (EFS_BMONTH and EFS_EMONTH)
- Owner of the SNAP benefit (EFSOWN)
- Reason(s) for beginning SNAP receipt (EFSBRSN1 and EFSBRSN2)
- Reason(s) for ending SNAP receipt, if applicable (EFSERSN1 and EFSERSN2)
- Monthly coverage indicator (RFSYN)

For each SNAP benefit unit, one adult in the household is considered the owner of the SNAP benefits (the person in whose name the benefits are received), but it is not necessarily the person who reported the SNAP benefit. The SNAP owner may or may not be covered by the SNAP benefit. All details of the SNAP spell are available on the record of the benefit owner. The beginning and ending month of the SNAP spell, the continuation flag, and the owner of the SNAP benefit are found on the records of all people covered by the SNAP benefit. The monthly and reference year indicators of SNAP receipt are found on all people's records.

What other concepts are relevant to the SNAP content area?

Benefit Owner – A monthly variable that identifies the person number of the individual in whose name the benefit is received.

Continuation Flag – A person-level variable that indicates whether the year-end spell (1) ended at the end of the reference year, (2) continued at the end of the reference year but ended before the interview month, or (3) was ongoing during the interview month.

Income Screeners – The income screener questions, which are asked in the pre-EHC portion of the instrument, limit the universe for some means-tested programs. When certain income-related parameters are met, the instrument sets the EHC sections for SNAP, TANF, GA, and WIC to not in universe.

Screener Clump – The screener clump is an internally derived instrument variable that limits the number of people who are in universe for the income screener questions and the family- or household-based EHC screener questions (SNAP, TANF, and GA). The intention is to ask only one respondent the family- or household-based program questions when resources were shared during the entire reference period.

Spell – The duration of a benefit or event reported by a respondent in the Event History Calendar section of the survey that is defined by a start month, end month, and a continuous span between the start and end month.

Supplemental Nutrition Assistance Program (SNAP) **–** This federal aid program administered by each state and funded by the U.S. Department of Agriculture provides food purchasing assistance for low income residents. SNAP may also be known by its former name, the Food Stamp Program. The program was enacted by Congress in the Food Stamp Act of 1964 and later renamed to SNAP (Supplemental Nutrition Assistance Program) by the Farm Bill of 2008, when the distribution system of coupon and stamp was replaced and updated with the Electronic Benefit Transfer (EBT) System. The amount of benefits received depends on the size, income, and expenses of the recipient's household. The benefits distribute monthly in direct deposit to the EBT. Recipients may use the EBT to pay for food at participating food retail markets, which may include supermarkets, grocery stores, farmers' markets, and farmers' market roadside stands.

3.13.4 General Assistance

What is collected in the General Assistance content?

The GA section collects information about GA benefit receipt from the start of the reference period through the interview month. Variables on the public use file provide the following information:

- Monthly and reference period coverage indicators
- Months received
- Year receipt began if receiving in January of the reference period (wave 1)
- Reason(s) receipt began
- Reason(s) receipt stopped
- Payment amount
- Person in whose name the benefit is received (benefit owner)
- Continuation flag for spells that include the last month of the reference period

To reduce respondent burden, SIPP does not collect information about the receipt of GA separately for children under age 15. Respondents age 15 or older with a household income below 200 percent of the poverty line are asked to report whether they or anyone in their family (i.e., children) received GA benefits. In many cases, the screener clump selects one adult in the family to report GA receipt for all people in the nuclear family. Adults not in the screener clump are each asked whether they or anyone in their family received GA benefits.

As with other data collected in the EHC, GA information is collected in reverse chronological order. That is, each respondent is first asked if they or someone in their family are currently receiving GA. If a respondent reports current receipt, the month of the reference period or interview year receipt began is

collected. If no one in the respondent's family is currently receiving GA, the respondent is then asked if they received GA at any time since the first month of the reference period. If receipt is reported during that time, the months that receipt started and stopped are collected. If a GA spell includes January of the reference period the first time a respondent is interviewed, information on the year the respondent first began receiving benefits is collected.

For each GA spell, respondents are asked to report why receipt began, why receipt stopped (if applicable), payment amount, who was covered by the benefit, and the person in whose name the benefit is received (benefit owner), as this information may vary across spells.

Respondents may report up to four different payment amounts in a spell to allow for changes in payments resulting from factors such as cost of living adjustments and earnings variation. The amounts are recoded into a single monthly variable.

A continuation flag for respondents who reported receiving GA in the last month of the reference period provides information on whether the spell: (1) ended in the last month of the reference period; (2) ended after the last month of the reference period but before the interview month; or (3) was ongoing as of the interview month.

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Reference period indicator for coverage
- Respondents may report up to three GA spells
- A continuation flag for respondents who reported receiving GA in the last month of the reference period

Why is the General Assistance content important? What assistance does it provide?

SIPP is designed to observe individuals over time and show changes in income, program participation, and other areas. GA is an important source of income for some low-income families with children. The longitudinal nature of the survey makes it possible to study the timing of GA receipt in relationship to other important events, such as changes in marital status, employment status, and/or participation in other assistance programs.

The SIPP collects more detailed information on GA receipt than both the ACS and CPS and provides it at a more detailed level – the SIPP provides monthly and annual data while the other surveys only provide annual data. When combined with data on earnings and other income, data on GA benefits help provide a more comprehensive picture of an individual's economic well-being over the length of the panel.

What level of information does the GA content provide?

GA data are collected at the spell level in the EHC and are provided to users in a person-month format.

What are the GA variables and where can I find them?

While all data are in a monthly format, the value of some variables are the same for entire reference period (person-level variables). The value of other variables may be the same throughout an entire spell (spell-level variables) or may vary from month to month (monthly-level variables).

Person-level variables

- A reference period coverage indicator identifies whether an individual was covered by GA in at least one month of the reference period (RGACOV)
- Year receipt began, for respondents with a GA spell that begins in January of the reference year in respondent's first interview (EGALCY)
- The continuation flag identifies the status of a respondent's spell if it was on-going as of the last month of the reference period (RGA_CONTFLG)

Month-level variables

- Payment amount (TGA_AMT)

Spell-level variables

- Begin and end month of the GA spell (EGA_BMONTH and EGA_EMONTH)
- Owner of the GA benefit (EGAOWN)
- Reason(s) for beginning GA receipt (EGABRSN1 and EGABRSN2)
- Reason(s) for ending GA receipt, if applicable (EGAERSN1 and EGAERSN2)
- Monthly coverage indicator (RGAYN)

For each GA benefit unit, one adult in the household is considered the owner of the GA benefits (the person in whose name the benefits are received), but it is not necessarily the person who reported the GA benefit. The GA owner may or may not be covered by the GA benefit. All details of the GA spell are available on the record of the benefit owner. The beginning and ending month of the GA spell, the continuation flag, and the owner of the GA benefit are found on the records of all people covered by the GA benefit. The monthly and reference year indicators of GA receipt are found on all people's records.

What other concepts are relevant to the General Assistance content area?

Benefit Owner – A monthly variable that identifies the person number of the individual in whose name the benefit is received.

Continuation Flag – A person-level variable that indicates whether the year-end spell (1) ended at the end of the reference year, (2) continued at the end of the reference year but ended before the interview month, or (3) was ongoing during the interview month.

General Relief or General Assistance Programs – General Assistance programs are funded and administered at the state and local government levels. Some states do not have General Assistance programs and other states that have General Assistance programs allow localities to choose whether or not to implement the programs. The programs serve people who do not qualify for federal assistance. The populations these programs serve vary by locality. Examples include disabled, elderly, unemployed, children, families with children, and employable individuals without children. The needs of each recipient may also vary. Assistance may be a one-time emergency or a regular monthly benefit and/or

payment. Payments received may be in cash or in the form of a voucher. Also, the benefit and payment received may be intended for a specific need, such as medical expenses, burial expenses, or to meet general needs.

Income Screeners – The income screener questions, which are asked in the pre-EHC portion of the instrument, limit the universe for some means-tested programs. When certain income-related parameters are met, the instrument sets the EHC sections for SNAP, TANF, GA, and WIC to not in universe.

Screener Clump – The screener clump is an internally derived instrument variable that limits the number of people who are in universe for the income screener questions and the family- or household-based EHC screener questions (SNAP, TANF, and GA). The intention is to ask only one respondent the family- or household-based program questions when resources were shared during the entire reference period.

Spell – The duration of a benefit or event reported by a respondent in the Event History Calendar section of the survey that is defined by a start month, end month, and a continuous span between the start and end month.

3.13.5 Life Insurance Retirement Income

What is collected in the Life Insurance Retirement Income content?

SIPP collects person-level data for respondents age 30 or older who reported being retired, ever being retired, or receiving retirement income from a paid-up life insurance policy or annuity during the reference year.

Specifically, the Life Insurance Retirement Income content collects:

- Was the payment was received as a lump-sum
- Each month of receipt
- The total amount received during the reference year

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Collect the total amount received
- Collect the month when the income benefit received

Why is the Life Insurance Retirement Income information important? What assistance does it provide?

Since income received from a job source or a life insurance benefit affects an individual's economic situation, the data collected for Life Insurance Income is important to understanding an individual's economic situation.

What level of information does the Life Insurance Retirement Income content provide?

Life Insurance Retirement Income is collected on the person level for all household members who were age 30 or older and who reported being retired or ever being retired during the reference year.

Individuals under age 30 or who reported not being retired or ever being retired are not in universe for the Life Insurance Income Benefit section.

What are the Life Insurance Retirement Income variables and where can I find them?

All of the Life Insurance Income Benefit data are collected in the instrument after the EHC. The Life Insurance Income Benefit section collects up to two types of life insurance income benefits, total amount received, and the month(s) when the payments were received.

What other concepts are relevant to the Life Insurance content area?

Annuity – An annuity is a pension plan purchased by a policyholder at their retirement or during their working years to guarantee a regular income after retirement. Some annuity contracts provide guaranteed distribution for a set number of payments and some provide payment until death. SIPP includes all such annuities purchased by the person, except those already reported as part of the employee's pension.

Paid-up Life Insurance Policy – A whole life insurance policy, universal life insurance policy, or permanent policy that provides a benefit upon death of the policyholder and accumulates cash value over time, enabling benefits to be paid in the event that the policyholder voluntarily terminates the insurance policy before death or the insured event occurred. The paid-up value is the sum of money that the insurance company will pay to the policyholder in the event that the policyholder voluntarily terminates the life insurance policy.

Retirement Income – Payments received based on length of service at retirement, age of retirement, and other employer retirement requirements. Examples of retirement income include U.S. government railroad retirement; pension from a company, union, Federal civilian employee pension; U.S. military retirement pay; retainer pay; National Guard retirement, Reserve Forces retirement; state government pension, local government pension; and any pension type other than Social Security.

Retirement Income Life Insurance – Individuals with a sufficient number of years of service or who have a disability may retire before reaching age 65 with a pension plan that includes life insurance. Individuals employed in hazardous occupations such as law enforcement may be eligible to retire with a pension that includes a paid-up life insurance policy or annuity as early as age 40 if they have completed 20 or more years of service.

3.13.6 Lump Sum Severance Pay/Retirement Plan Income

What is collected in the Lump Sum Severance Pay/Retirement Plan Income content?

SIPP collects person-level data from respondents aged 15 years and older about receipt of lump sum severance pay and retirement plan income at any time during the reference year.

For respondents who report receiving lump sum severance pay/retirement plan income, SIPP collects the type of payment:

- Pension/retirement pay

- Severance pay
- Deferred payment/final paycheck
- Something else (respondents are then asked to specify the type)

Respondents are then asked to report the total amount received during the reference year.

Respondents are next asked if any of money was rolled over to an IRA or other kind of retirement plan. For respondents who rolled over money (or who plan to), the amount rolled over is then collected.

Major changes between the 2008 and 2014 panels:

- Collect the amount received
- Collect the amount rolled over into another retirement plan

Why is the Lump Sum Severance Pay/Retirement Plan Income information important? What assistance does it provide?

Since income received from a job source or a lump sum benefit affects an individual's economic situation, the data collected for lump sum income is important to understanding an individual's economic situation.

What level of information does the Lump Sum Severance Pay/Retirement Plan Income content provide?

Each Lump Sum Income Benefit section is collected on the person level for all household members who were aged 15 years and older during the reference year. Children under the age 15 by the end of the reference year are not in universe for the Lump Sum Income Benefits section.

What are the Lump Sum Severance Pay/Retirement Plan Income variables and where can I find them?

All of the Lump Sum Income Benefits data are collected in the instrument after the EHC. The Lump Sum Income Benefits section collects up to four types of lump sums, total dollar amount received from the lump sum income benefits, two roll over lump sum income benefit types, and the total dollar amount rolled over.

What other concepts are relevant to the Lump Sum Severance Pay/Retirement Plan Income content area?

<u>Deferred payment</u> – An employer or employee arrangement in which the compensation is paid after the 1- or 2-week span that the income is earned. Pensions and retirement plans are examples of deferred payment plans.

<u>Individual Retirement Account (IRA)</u> – An Individual Retirement Account is a personal retirement plan that allows employees, self-employed individuals, and other individuals to set aside money each year for retirement in a tax-deferred account. Earnings on all amounts contributed to any IRA accumulate are on a tax-deferred basis.

Pension – A pension is an employee benefits plan where employees and employers pay into a group trust fund. The payment benefits are based on a formula factoring an employee's final pay and employment tenure with an employer.

Retirement plan – A financial arrangement offered by employers, insurance companies, and unions that replaces employment income during retirement.

Severance pay – Employer offered compensation to newly unemployed employees. The payment may be offered to retired, resigned, laid-off, or dismissed with cause employees and may wave the employees right to pursue a legal claim against the former employer or work for a competitor of the employer.

3.13.7 Miscellaneous Income

What is collected in the Miscellaneous Income content?

SIPP collects person-level data from respondents age 15 or older regarding miscellaneous income receipt at any time during the reference year.

The section begins by asking if the respondent received money or income from any of the following sources during the reference year:

- Community or religious charity
- Family or friends
- Roomers or boarders
- Estates
- Incidental or casual earnings
- Miscellaneous cash income, such as lottery winnings
- National Guard or Reserve Pay

If a respondent reports receiving any of the seven miscellaneous income sources, the respondent is then asked how much income was received during the reference year from all the miscellaneous sources.

Major changes between the 2008 and 2014 panels:

- The Miscellaneous Income content in 2008 and prior panels only collected the type of income received during the reference year.
- New for 2014, SIPP also collects the total amount received during the reference year.

Why is the Miscellaneous Income information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources. Since income received from a miscellaneous source affects an individual's economic situation the data collected for Miscellaneous Income is important to understanding an individual's economic situation.

What level of information does the Miscellaneous Income content provide?

Miscellaneous Income is collected at the person level on an annual basis for all household members who were age 15 or older during the reference year. Children under the age of 15 by the end of the reference year are not in universe for the Miscellaneous Income section.

What are the Miscellaneous Income variables and where can I find them?

The Miscellaneous Income section is collected after the EHC.

What other concepts are relevant to the Miscellaneous Income content area?

Community Charity – An organization formed for the purpose of collecting and distributing benefits to support individuals, families, and children. Community charities include churches, social organizations, and government assistance programs.

Estates – This is a description for an individual's property, entitlements, and obligations.

Incidental or Casual Earnings – Income or profit generated from informal, occasional, or irregular periods of receipt or employment.

Miscellaneous Cash Income – This includes any cash income that does not fall into a specific category.

National Guard or Reserve Pay – Any pay received while on duty (active or temporary) as a member of the National Guard or the Guard Reserve of any branch of the Armed Forces.

Religious Charity – A profit or non-profit organization formed for purposes where both membership and leadership is in the observance of faith or vow of faith. A religious charity may manage a house of worship, provide payment to religious leadership, or provide payment to a religious organization.

Roomers or Boarders – Housing unit residents who rent a housing unit or section for one night or more. The housing rental agreement may include meals and dining facilities.

Short-Term Cash Assistance – Any assistance received from a government agency, charitable organization such as a church, community organization, or family and friends. This does not include TANF, General Assistance, or long-term assistance from any source.

Trusts – A contract administered by a trustee for the distribution of money or property by a beneficiary.

3.13.8 Other Assistance Income

What is collected in the Other Income Assistance content?

Although the programs in this section are not necessarily means-tested, SIPP only collects these data from respondents aged 15 years and older whose household incomes fall below 200% of the poverty line. The questions focus on informal types of assistance (food, transportation, other (clothing and housing), and training) that fall outside the government's main social welfare programs. For each kind of assistance or topic, variables identify the type of assistance received during the reference period, the source(s) of assistance received during the reference period, and the months assistance was received during the reference period.

Food Assistance

SIPP asks respondents whether they received various types of food assistance (not including SNAP, which is collected separately) during the reference period:

- Money, vouchers, or certificates to buy groceries
- Bags of groceries or packaged foods
- Meals from a shelter, soup kitchen, Meals-on-Wheels, or other charity
- Something else

Respondents who report receiving food assistance are asked about the source(s) of assistance during the reference period:

- Government social service agency
- Community or religious charitable organization
- Family or friends
- Someplace else

Respondents are then asked to identify the months they received food assistance during the reference period.

Transportation Assistance

SIPP asks respondents whether they received various types of transportation assistance during the reference period:

- Gas vouchers
- Bus or subway tokens or passes
- Help repairing, registering, or insuring a car
- Rides to a doctor's office or medical appointment
- Some other kind

Respondents who report receiving gas vouchers or bus or subway tokens or passes are asked about the source(s) of assistance during the reference period:

- Government social service agency
- Someplace else

Respondents are then asked to identify the months they received transportation assistance during the reference period.

Clothing and Housing Assistance

SIPP asks respondents whether they received various types of clothing assistance during the reference period:

- Free or reduced price clothes
- Money or vouchers
- Both free or reduced price clothes and money or vouchers

Respondents who report receiving clothing assistance are asked about the source(s) of this assistance during the reference period:

- Government social service agency
- Community or religious charitable organization
- Family or friends
- Employer
- Someplace else

SIPP asks respondents whether they received housing assistance to help pay for housing during the reference year (not including residing in public housing, receiving a housing voucher, and receiving energy assistance, which are collected separately).

Respondents who report clothing and/or housing assistance are then asked to identify the months they received clothing and/or housing assistance during the reference period.

Training Assistance

This section asks respondents ages 18-64 with a household income under 200% of the poverty line whether they participated in any of the following types of training programs to improve or enhance their job readiness skills during the reference period:

- Classes or training to improve basic reading or math skills
- Job readiness training to learn about resume writing, job interviewing, or building self esteem building
- Job search programs or job clubs, or job resource centers to find out about jobs, to schedule interviews, or to fill out applications
- Training to learn a specific job skill, such as computers, car repair, nursing, day care work, or some other job skill

Respondents who report receiving TANF during the reference period are asked:

- Whether the training or use of job search resources was a requirement of the state or county welfare office, their choice, or both.

Respondents are then asked to identify the months they participated in training assistance programs during the reference period.

Major changes between the 2008 and 2014 panels:

- Information on receipt of other welfare assistance is no longer collected in this section
- Amounts for food assistance, clothing assistance, gas vouchers, and bus or subway tokens or passes are no longer collected

Why is the Other Income Assistance information important? What assistance does it provide?

SIPP collects information on participation in a number of well-known government assistance programs such as GA, SSI, TANF and WIC. However, some individuals receive less formal assistance that falls outside these established programs. The Other Income Assistance questions identify the types of

assistance these individuals receive as well as the source(s) of the assistance and the months of receipt. This information provides data users with a more complete picture of the programs and services that contribute to an individual's income and overall well-being. Additionally, collecting information on months of receipt allows users to study movements on and off these programs and the demographic and economic changes that coincide with these movements.

What level of information does the Other Income Assistance content provide?

This information is collected in the post-EHC section and is provided to users in a person-month format.

What are the Other Income Assistance variables and where can I find them?

For each topic, SIPP has an initial mark all that apply question that asks whether the respondent received different types of during the reference period. Each type is edited as a separate "yes" or "no" variable. To identify whether a respondent received assistance in at least one month of the reference period each "yes" or "no" variable for that topic must be checked.

To determine the months of receipt during the reference period, monthly "yes" or "no" variables are available for each topic. If respondents report more than one type or source of assistance for a single assistance program (e.g. reporting transportation assistance for repairing a car and rides to medical appointments), users are not able to identify which months of receipt were associated with each specific type or source of assistance.

While all data are in a monthly format, the value of some variable is the same for entire reference period (person-level variables). The value of other variables may vary from month to month (monthly-level variables).

Person-level variables

- Type(s) of assistance
- Source(s) of assistance

Monthly-level variables

- Month(s) of assistance

What other concepts are relevant to the Other Income Assistance content area?

Income Screener – Households with an annual and monthly income above 200% of the poverty level are screened out of the Other Income Assistance Content.

3.13.9 Retirement Income

What is collected in the Retirement Income content?

SIPP collects person-level data regarding whether an individual age 30 or older received retirement income during the reference year, including the type of income and amount received during each month of the reference year.

For respondents who report receiving Retirement Income, SIPP collects which type(s), specifically:

- Federal Civil Service or other Federal Civilian employee pension
- Local government pension
- Military retirement pay
- National Guard or Reserve Forces retirement
- Other retirement income
- Pension from a company or union including income from a profit-sharing plan
- State government pension
- U.S. Government Railroad Retirement

SIPP then collects separate monthly amounts for each type of retirement income. Respondents can report up to four amounts received and start months for each of the retirement benefits. If the respondent received more than four different monthly amounts throughout the reference year for any one retirement income benefit type, the instrument asks the amount of money that the respondent received during the first start month.

2014 SSA Supplement

The Social Security Administration Supplement on Retirement, Pensions, and Related Content, or SSA Supplement, was sponsored by the Social Security Administration and conducted via computer-assisted telephone interviewing (CATI) after the Wave 1 interview. It contains more detailed marital history information than the 2014 SIPP.

For more information on the SSA Supplement content, please refer to Chapter 3 in the 2014 SSA Supplement Users' Guide.

Major changes between the 2008 and 2014 panels:

New in 2014, SIPP collects the monthly amount received, as well as any changes in amount, by retirement income benefit type.

Why is the Retirement Income information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources such as retirement income benefits. Since payments received affect an individual's economic situation, the data collected for retirement income benefits are important to understanding an individual's economic situation.

What level of information does the Retirement Income content provide?

Retirement income information is collected at the person level for all respondents who report ever being retired.

What are the Retirement Income variables and where can I find them?

All of the retirement income data are collected after the event history calendar (EHC) in the post-EHC section.

What other concepts are relevant to the Retirement Income content area?

Federal Civil Service Pension – The Civil Service Retirement Act of 1920 provided a retirement system for federal employees beginning civil service employment between 1920 and 1987. CSRS is a contributory retirement system. Employees share in the expense of annuities.

Federal Civilian Employee Pension – The Federal Employers Retirement System (FERS) provides retirement system for federal employees beginning civil service employment after 1987. FERS provides retirement income benefit from three sources: a basic benefit plan, Social Security, and Thrift Savings Plan. Two of the three parts of FERS can go with employees to their next job if they leave the federal agency. FERS withholds the cost of the basic benefit and Social Security from employee pay as payroll deductions.

Local Government Pension – Also known as a public pension plan, the local government employee retirement plan is a pension plan for vested local government employees.

Military Retirement Pay – Retirement pension plan provided for active duty military personnel with benefits that start when the employee retires.

National Guard Retirement – Army National Guard employee retirement plan provided for employees who have 20 years service while serving the Guard by age 60 and may combine with any additional income or retirement pension plan.

Other Retirement Income – Income from a pension plan provided for employee retirement., such as an Individual Retirement Account (IRA).

Pension – A pension is an employee benefits plan in which employees and employers pay into a group trust fund. The payment benefits are based on a formula factoring an employee's final pay and employment tenure with an employer.

Reserve Forces Retirement – A retirement pension plan is for service members who complete 20 years of service and who become eligible for retirement from the Reserve Forces at age 60. Reserve service members may begin retirement benefits before age 60 if they deploy for war or national emergency. Effective January 2008, for every consecutive 90 days that service members are mobilized, their annuity start date is reduced three months.

State Government Pension – Also known as a public pension plan, the state government employee retirement plan is a pension plan for vested state government employees.

U.S. Government Railroad Retirement – A federally administered retirement program established for railroad employees and employers. The premiums are employee and employer paid. The first annuity and benefits were distributed in 1936. The program covers railroad employees for retirement, unemployment, sickness, disability, spousal, and survivor benefits. Employees with 10 or more years of service or individuals having at least 5 years of service since 1995 are eligible for the retirement

program. Tier 1 benefits take the place of Social Security and are first payable at age 62 or after 30 years of service for early retirement. The first full retirement benefits are payable for retired employee between the ages of 65 and 67. The Tier 2 benefits are similar to a private pension retirement plan.

3.13.10 School Meals

What is collected in the School Meals content?

The school meals section collects information about the receipt of school lunches and breakfasts at any point during the reference year, including:

- Whether any children in the family usually got the school lunch/breakfast that their school provided
- Whether the lunches/breakfasts received were free or reduced-price because the family qualified for the School Lunch/Breakfast Program, or full-price because the family did not qualify for the School Lunch/Breakfast Program
- Indicator of receipt of school lunches/breakfasts is available on the record of eligible children

Major changes between the 2008 and 2014 panels:

- EFREELUN and EFRERDLN were combined into one variable indicating whether the school lunches received were free or reduced-price because the family qualified for the School Lunch Program or full-price because the family did not qualify for the School Lunch Program.
- EFREEBRK and EFRERDBK were combined into one variable indicating whether the school breakfasts received were free or reduced-price because the family qualified for the School Breakfast Program or full-price because the family did not qualify for the School Breakfast Program.
- ELUNCH_YN is on the child's record and indicates whether the child received free school lunches, reduced-price school lunches, full-price school lunches, or did not receive the lunch that their school provides.
- EBREAK_YN is on the child's record and indicates whether the child received free school breakfasts, reduced-price school breakfasts, full-price school breakfasts, or did not receive the breakfast that their school provides.

Why is the School Meals information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Households often receive financial assistance from programs such as free and reduced-price school meals. Since meals received affect a household's economic situation, the data collected for school meals are important to understanding a household's overall economic situation.

What level of information does the School Meals content provide?

Information about the receipt of school meals are asked of designated parents of children between the ages of 5 and 18 who have not yet graduated from high school. Information about receipt of school meals is stored on the parent's record. Additionally, all children between the ages of 5 and 18 who have not yet graduated from high school have indicators of receipt of school breakfasts and lunches on their record. The reporting period is the entire reference period.

What are the School Meals variables and where can I find them?

While all data are available in a monthly format, these variables hold the same value over the entire reference period.

The following variables are available on the record of designated parents of eligible children:

- Indicator that one or more children in the family usually got the lunch their school provided (ESCHOOLLUNCH)
- Indicator that one or more children in the family usually got the breakfast their school provided (ESCHOOLBREAK)
- Whether the lunches received were free or reduced-price because the family qualified for the School Lunch Program, or full-price because the family did not qualify for the School Lunch Program (EFREE_LUNCH)
- Whether the breakfasts received were free or reduced-price because the family qualified for the School Breakfast Program, or full-price because the family did not qualify for the School Breakfast Program (EFREE_BREAK)

The following variables are available on the record of children between the ages of 5 and 18 who have not yet graduated from high school:

- Whether respondent got free lunches, reduced-price lunches, full-price lunches, or did not receive school lunches (ELUNCH_YN)
- Whether respondent got free breakfasts, reduced-price breakfasts, full-price breakfasts, or did not receive school breakfasts (EBREAK_YN)

What other concepts are relevant to the School Meals content area?

Federal School Breakfast/Lunch Program – The federal school breakfast and lunch program is a program sponsored by the Food and Nutrition Service of the U.S. Department of Agriculture. This program subsidizes the cost of all school meals in the country through the National School Lunch Program. Recipients are children in school who live in households with limited incomes. The benefit generally comes in the form of a discounted or free lunch and/or breakfast every school day during the school year, but in some localities the program is extended through the summer months.

3.13.11 Social Security Child Benefits

What is collected in the Social Security Child Benefits content?

The Social Security Child Benefits section collects person-level data from respondents who lived with at least one child under the age of 18 during the reference period and who received Social Security benefits on behalf of a child at any time during the reference year.

The section begins by asking about any Social Security benefit receipt on behalf of a child during the reference year. If the respondent reports any receipt and there is more than one child in the household, SIPP asks the respondent to identify the child(ren) in whose name(s) benefits were received. SIPP next asks why the respondent began receiving Social Security benefits.

Monthly amounts are collected by asking about the amount received in the most recent month of receipt, which can be the interview month. For example, "How much do you receive in Social Security benefits on behalf of these children now?" Then the instrument asks, "Did you get that amount during all months of receipt?"

If the respondent replies that the amount received was the same in all months, then the instrument moves to the next survey topic. If the respondent replies that the amount received varied, the instrument then asks the amount received prior to the most recent amount. The instrument continues to work backwards through the reference period until the amount for each month of receipt is collected. After a respondent has reported four different receipt amounts, the instrument completes the monthly amount section by asking the amount received during the first month of receipt during the reference period.

Major changes between the 2008 and 2014 panels:

- Collect the monthly amount received
- Collect monthly changes in the amount

Why is the Social Security Child Benefits content important? What assistance does it provide?

Data from job earnings alone are not representative of a household's economic situation. Households often receive income from non-job sources such as Social Security child benefits. Since income received from a job source or Social Security child benefits affect a household's economic situation, the data collected for the Social Security Child benefits section is important to understanding a household's economic situation.

What level of information does the Social Security Child Benefits content provide?

Monthly Social Security child benefit amounts are collected on the person level for all household members age 18 or older with a child who resided in the house at any time during the reference year. Teenagers under the age of 18 at the end of the reference year are not in universe for the Social Security Child Benefits section.

What are the Social Security Child Benefits variables and where can I find them?

All of the Social Security Child Benefits data are collected after the Event History Calendar (EHC) in the post-EHC Other Assistance section. The Social Security Child Benefits section collects amount(s) received and when the amount was received. It collects up to four amount changes during the reference year.

What other concepts are relevant to the Social Security Child Benefits content area?

<u>Child</u> –Refers to family relationship and/or the age of the person. In family relationships, a child references a parent's biological, step, adopted, or guardianship daughter or son. Child also references anyone under age 18.

Disabled – An impairment that may be physical, cognitive, mental, hearing, vision, sensory, emotional, developmental, chronic disease, or some combination thereof.

3.13.12 Social Security Self Benefits

What is collected in the Social Security Self Benefits content?

The Social Security Self Benefits collects person level data for each household member age 18 or older who received Social Security benefits at any time during the reference year.

The section begins by asking about any Social Security benefit receipt on behalf of oneself during the reference year. If the respondent reports any Social Security benefit receipt, SIPP then asks why the respondent began receiving Social Security benefits.

If the respondent reports a disability as the reason for receipt, SIPP then asks is at what age did the respondent begin receiving Social Security due to their disability.

For married respondents, SIPP asks if the Social Security benefits were received jointly with the spouse.

Amounts are collected by asking about the amount received in the most recent month of receipt, which can be the interview month. For example, "How much do you receive in Social Security benefits?" Then the instrument asks, "Did you get that amount during all months of receipt?"

If the respondent replies that the amount received was the same in all months, then the instrument moves to the next survey topic. If the respondent replies that the amount received varied, the instrument then asks the amount received prior to the most recent amount. The instrument continues to work backwards through the reference period until the amount for each month of receipt is collected. After a respondent has reported four different receipt amounts, the instrument completes the monthly amount section by asking the amount received during the first month of receipt during the reference period.

Major changes between the 2008 and 2014 panels:

- Collect the monthly amount received
- Collect monthly changes in the amount

Why is the Social Security Self Benefits content important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources such as Social Security benefits. Since income received from a job source or Social Security benefits affect an individual's economic situation, the data collected for Social Security benefits is important to understanding an individual's economic situation.

What level of information does the Social Security Self Benefits content provide?

Monthly Social Security benefit amounts are collected on the person level for all household members age 18 or older during the reference year. Teenagers under the age of 18 at the end of the reference year are not in universe for the Social Security Self Benefits section.

What are the Social Security Self Benefits variables and where can I find them?

All of the Social Security Self data are collected after the Event History Calendar (EHC) in the post-EHC section. The Social Security Self section collects amount(s) received and when the amount was received. It collects up to four amount changes during the reference year.

What other concepts are relevant to the Social Security Self Benefits content area?

Disabled – An impairment that may be physical, cognitive, mental, hearing, vision, sensory, emotional, developmental, chronic disease, or some combination.

Retired – Employment status or point in time when an employee terminates their employment or reduces their employment hours due to their age or tenure. Retirement is usually between the ages of 50 and 70. However, early retirement can be at any age, before the age or before the tenure needed for eligibility for support and funds from the employer and government.

Social Security on Behalf of Self – Social Security is managed by the Social Security Administration and is designed to provide retirement, healthcare, disability, and survivors insurance for U.S. workers. As insured employees work, they pay Social Security taxes and earn credits that count toward eligibility for Social Security benefits. Most workers need 10 years of work to qualify for benefits. Social Security benefits may include monthly income benefits, health insurance (Medicare, Medicaid, or both), lump sum payments, and cost of living adjustments. Workers at least age 62 can retire with reduced benefits and can receive higher benefits at age 65, 66, or 67, depending on their birthdate.

3.13.13 SSI

What is collected in the Supplemental Security Income (SSI) content?

The SSI section collects information about an individual's benefit receipt from the start of the reference period through the interview month. Variables on the public use file provide the following information:

- Reference period indicator
- Months received
- Year receipt began if receiving in January of the reference period (wave 1)
- Continuation flag for spells that include the last month of the reference period
- Reason(s) receipt began
- Reason(s) receipt stopped
- Payment source
- Payment amount
- Person in whose name the benefit is received (for respondents under age 18)

As with other data collected in the EHC, SSI information is collected in reverse chronological order. That is, each respondent (or proxy respondent if the person is under age 15) is first asked if they are currently receiving SSI. If a respondent reports current receipt, the month of the reference period or interview year receipt began is collected. If a respondent is not currently receiving SSI, they are asked if they received it any time since the first month of the reference period. If receipt is reported during that time, the months of the reference period or interview year receipt started and stopped is collected. If an SSI

spell includes January of the reference period the first time a respondent is interviewed, information on the year the respondent first began receiving benefits is collected. If an SSI spell includes the last month of the reference period a continuation flag provides information on whether the spell: (1) ended in the last month of the reference period; (2) ended after the last month of the reference period but before the interview month; or (3) was on-going as of the interview month.

For each SSI spell, respondents are asked to report why receipt began, why receipt stopped (if applicable), payment source, and payment amount as this information may vary across spells. Additionally, the person in whose name the benefit is received is asked if the respondent is under age 18.

Respondents may report whether they receive their monthly SSI benefit in one or two payments. For each payment, the monthly payment amounts are collected first and the payment source is collected second. Respondents may report up to five different payment amounts for each source they report in a spell to allow for changes in payments resulting from factors such as cost of living adjustments and earnings variation. The source of each payment is available for analysis, but the amounts are recoded into a single monthly variable. The primary reason for this is respondents who received a combined Federal and state payment are not asked to report which portion of the benefit was a Federal payment and which portion of the benefit was a supplemental state payment. In 2013 nearly one-quarter (24 percent) of SSI beneficiaries received a combined payment.

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Reference period indicator for coverage
- Respondents may report up to 3 SSI spells
- Proxy respondent reports about child SSI receipt on the child's record
- A single amount variable versus separate variables for Federal payments and supplement state payments
- A continuation flag for respondents who reported receiving SSI in the last month of the reference period

Why is the SSI information important? What assistance does it provide?

SIPP is designed to observe individuals over time and explain changes in income, program participation, and other areas. SSI is an important source of income for low-income individuals who are aged, blind, or disabled. The longitudinal nature of the survey makes it possible to study the timing of SSI receipt in relationship to other important events, such as changes in disability status and/or resources.

The SSI data can be used to produce national estimates of coverage and payment amounts. Although the Social Security Administration regularly publishes data on SSI participation and benefits, detailed SSI data (particularly in regards to supplemental state payments) are published less frequently. The SIPP collects more detailed information on SSI receipt than both the American Community Survey (ACS) and Current Population Survey (CPS) and provides it at a more detailed level – the SIPP provides monthly and annual data while the other surveys only provide annual data. When combined with data on earnings

and other income, SSI benefits help provide a more comprehensive picture of an individual's – and the people they reside with – economic well-being over the length of the panel.

What level of information does the SSI content provide?

SSI data are collected at the spell level in the EHC and are provided to users in a person-month format.

What are the SSI variables and where can I find them?

While all data are in a monthly format, the value of some variables is the same for entire reference period (person-level variables). The value of other variables may be the same throughout an entire spell (spell-level variables) or may vary from month to month (monthly-level variables).

Person-level variables

- A reference period coverage indicator identifies whether an individual received SSI in at least one month of the reference period (RSSICOV)
- Year receipt began for respondents with a SSI spell that begins in January of the reference year in respondent's first interview (TSSI_LC_YR)
- The continuation flag identifies the status of a respondent's spell if it was on-going as of the last month of the reference period (RSSI_CONTFLG)

Spell-level variables

- Begin and end month of the SSI spell (ESSI_BMONTH and ESSI_EMONTH)
- Reason(s) for beginning SSI receipt (ESSI_BRSN1 and ESSI_BRSN2)
- Reason(s) for ending SSI receipt (ESSI_ERSN1 and ESSI_ERSN2)
- Payment source(s) (ESSI_SRC1 and ESSI_SRC2)
- For SSI recipients under the age of 18, the person number of the benefit owner (ESSI_OWNER)

Monthly-level variables

- Payment amount (TSSI_AMT)

What other concepts are relevant to the SSI content area?

Benefit Owner – A monthly variable that identifies the person number of the individual receiving benefits on behalf of a child. This information is only collected for SSI recipients under the age of 18.

Continuation Flag – A person-level variable that indicates whether a spell was right censored. This variable, adopted by multiple EHC programs, identifies whether a spell ended in the last month of the reference period (not right censored), continued after the reference period ended but ended before the interview month (right censored), or was on-going at the time of the interview (right censored).

SSI Users' Note

In wave 1 of the 2014 panel, matching SIPP data to SSA administrative records revealed that nearly one-half of respondents who reported SSI receipt did not have a corresponding administrative record indicating receipt of SSI payments.

A users' note has been posted to the SIPP website detailing possible respondent confusion regarding SSI and subsequent edited data.

3.13.14 Survivor Income Benefits

What is collected in the Survivor Income Benefits content?

SIPP collects person-level data about whether an individual aged 15 years and older received survivor income benefits at any time during the reference year. Respondents report the type received:

- Black Lung benefits
- Federal Civil Service or other Federal civilian employee pension
- Income from a paid-up life insurance policy or annuity
- Local government pension
- Military retirement pay
- National Guard or Reserve Forces retirement
- Other (asked to specify)
- Payments from an estate or trust
- Pension from a company or union including income from a profit-sharing plan
- State government pension
- U.S. Government Railroad Retirement
- Veterans' compensation/ pension
- Workers' Compensation

For each benefit type reported, the instrument asks for the months of receipt and amount received. Amounts are collected by moving backwards chronologically from the interview month to the start of the reference period. For example, "How much do you receive in local government pension now?" and "When did that start?"

If the respondent began receiving the amount reported after the month in which benefit receipt started, the instrument then asks how much was received prior to the most current amount. For each benefit type, respondents may report up to four amount changes during the reference period. After the fourth amount, the instrument asks for the amount received during the first month of receipt.

Major changes between the 2008 and 2014 panels:

The data collected in the 2014 panel address the same major fields as the 2008 and prior panels, the monthly receipt and type of benefit. New for 2014, the Survivor Benefit content includes the monthly amount received by each survivor income benefit type.

Why is the Survivor Income Benefits information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources such as survivor income Benefits. Since income received from a job source or a survivor benefit affects an individual's economic situation, the data collected for survivor income benefits is important to understanding an individual's economic situation.

What level of information does the Survivor Income Benefits content provide?

Survivor Income Benefit information is collected at the person level for each adult household member who is a widow/widower household. Children under the age of 15 by the end of the reference year are not in universe for the Survivor Income Benefits section.

What are the Survivor Income Benefits variables and where can I find them?

All of the survivor benefit data are collected in the instrument after the EHC.

What other concepts are relevant to the Survivor Income Benefits content area?

Annuity – An annuity is a pension plan purchased by a policyholder at their retirement or during their working years to guarantee a regular income after retirement. Some annuity contracts provide guaranteed distribution for a set number of payments and some provide payment until death. SIPP includes all such annuities purchased by the person, except those already reported as part of the employee pension.

Black Lung Benefits – Black Lung Disability refers to the federal program beginning in 1973, that provides pension to coal miners who have disability from pneumoconiosis or to their dependent survivors.

Estates – This is a common law description for an individual's property, entitlements, and obligations.

Federal Civil Service Pension – The Civil Service Retirement Act of 1920 provided a retirement system for federal employees who began their civil service employment between 1920 and 1987. CSRS is a contributory retirement system and employees share in the expense of annuities.

Federal Civilian Employee Pension – The Federal Employees Retirement System (FERS) is the retirement system for federal employees who began their civil service employment after 1987. FERS provides retirement income benefits from three sources: a basic benefit plan, Social Security, and Thrift Savings Plan. Two of the three parts of FERS can go with employees to their next job if they leave the Federal Government. The federal agency withholds the cost of the basic benefit and Social Security from their pay as payroll deductions.

Local Government Pension – Also known as a public pension plan, this is the local government employee retirement plan for vested local government employees.

3.13.15 Support Received

What is collected in the Support Received content?

SIPP collects person-level data about monetary payments that were received for foster child care, child support, and alimony (spousal support) at any time during the reference year.

- Foster Child Care – Asked of guardians aged 15 and older, about whether they received foster child support payments for their foster child(ren) under 18 years of age, from the state in which they lived.
- Child Support – Asked of parents or legal guardians aged 15 and older, who did not previously report receiving pass-through child support payments about whether they received child support payments for their child(ren) under 21 years of age from the child(ren)'s parent(s) who lived outside their household. Respondents are asked if child support payments were ever court ordered or informally agreed upon.
- Alimony – Asked of respondents aged 15 and older, who are currently divorced or separated or who have ever been divorced, about whether they received alimony or spousal support payments from a former spouse.

For each type of support, respondents are initially asked if they received any payments since the beginning of the reference period. If the response is in the affirmative, the respondent is asked they month in which they last received it, the month they started receiving it, and which months in between they received it. The respondent is then asked how much support was received in the reported last month, and then if they received that same amount every month. If the same amount was not received every month, the individual amount of support received every month is collected.

Why is Support Received information important? What assistance does it provide?

The collection of Support Received information is important as we attempt to measure the amounts of inter-household transfer payments that occur, as well as for whom those payments are supporting. Collecting amounts of support received allows us to compare and contrast with amounts reported by respondents who pay support for their child(ren) and other persons living outside their household. The collection of monthly Support Received income amounts allows users to identify regular and irregular receipt of these income sources.

What level of information does the Support Received content provide?

Support Received content is collected at the person level. Amounts of support received are available at the person-month level.

What are the Support Received variables and where can I find them?

Support Received data are collected in the instrument after the EHC.

Major changes between 2008 and 2014 panels:

- Collect information about individual months and amounts for each month received.

- No topical module for child support received, meaning no detailed information about child support agreements, health care, custody, paternity, and other child-support agreement specific information.

What are the Support Received variables and where can I find them?

All of the Support Received data are collected in the instrument after the EHC.

What other concepts are relevant to the Support Received content area?

Child Support – Financial help received from the other parent(s) may include payments made directly or indirectly to provide support for child(ren) in the form of rental or mortgage payments, medical costs or health insurance paid, costs associated with child care, school, or camp, or other tangible items to provide for the needs of the child(ren).

3.13.16 Support Paid

What is collected in the Support Paid content?

SIPP collects the following information from parents of children under 21 years of age living outside their household:

- The number of children under 21 years of age living outside their household
- If the parents made payments to support these children
- How much support was paid during the previous calendar year
- How often the respondent spent time with the children during the previous calendar year

All respondents aged 15 years and older are asked if they provided financial support to people living outside their household, including:

- Their parent or parents
- Children 21 years and older
- Other related person(s)
- Current or ex-spouse(s)
- Other unrelated people.

For each category, the number of persons supported is collected, as well as the amount paid during the previous calendar year.

Why is the Support Paid information important? What assistance does it provide?

The collection of Support Paid information is important as we attempt to measure the amounts of inter-household transfer payments that occur, as well as who those payments support. Collecting amounts of child support paid allows us to compare and contrast amounts received as reported by custodial parents.

What level of information does the Support Paid Content provide?

The support paid data are collected at the person level.

What are the Support Paid variables and where can I find them?

Support Paid data are collected in the middle of the instrument, after the Event History Calendar (EHC).

What other concepts are relevant to the Support Paid content area?

Support paid to children under 21 living elsewhere includes payments made directly to the other parent/guardian, through a court or agency, or withheld from a person's paycheck.

3.13.17 TANF

What is collected in the TANF content?

The TANF section collects information about TANF benefit receipt from the start of the reference period through the interview month. Variables on the public use file provide the following information:

- Monthly and reference period coverage indicators
- Months received
- Year receipt began if receiving in January of the reference period (wave 1)
- Reason(s) receipt began
- Reason(s) receipt stopped
- Payment amount
- Person in whose name the benefit is received (benefit owner)
- Whether benefits covered children only, both children and adults, or only a pregnant woman
- Continuation flag for spells that include the last month of the reference period

Additionally, when a spell of TANF is reported, information about receipt of pass-through child support is collected, including:

- Monthly indicators of whether pass-through child support was received
- Which children under age 22 were covered by pass-through child support payments
- Whether the TANF amounts reported include pass-through child support payments
- Amount of pass-through child support payments received
- Amount of child support collected by the agency on respondent's behalf

To reduce respondent burden, SIPP does not collect information about the receipt of TANF separately for children under age 15. Respondents age 15 or older with a household income below 200 percent of the poverty line are asked to report whether they or anyone in their family (i.e., children) received TANF benefits. In many cases, the screener clump selects one adult in the family to report TANF receipt for all people in the nuclear family. Adults not in the screener clump are each asked whether they or anyone in their family received TANF benefits.

As with other data collected in the EHC, TANF information is collected in reverse chronological order. That is, each respondent is first asked if they or someone in their family are currently receiving TANF. If a respondent reports current receipt, the month of the reference period or interview year receipt began is collected. If no one in the respondent’s family is currently receiving TANF, the respondent is then asked if they received TANF at any time since the first month of the reference period. If receipt is reported

during that time, the months that receipt started and stopped are collected. If a TANF spell includes January of the reference period the first time a respondent is interviewed, information on the year the respondent first began receiving benefits is collected.

For each TANF spell, respondents are asked to report why receipt began, why receipt stopped (if applicable), payment amount, who was covered by the benefit, and the person in whose name the benefit is received (benefit owner), as this information may vary across spells. Additionally, respondents indicate whether the benefit covered only children, both children and adults, or only a pregnant woman.

Respondents may report up to four different payment amounts in a spell to allow for changes in payments resulting from factors such as cost of living adjustments and earnings variation. The amounts are recoded into a single monthly variable.

A continuation flag for respondents who reported receiving TANF in the last month of the reference period provides information on whether the spell: (1) ended in the last month of the reference period; (2) ended after the last month of the reference period but before the interview month; or (3) was ongoing as of the interview month.

Within each reported TANF spell, respondents are asked whether pass-through child support from a state or county welfare program was received. If pass-through child support was received, respondents are asked which children under the age of 22 covered by pass-through child support payments and whether the TANF amounts reported include pass-through child support payments. Finally, respondents report the amount of pass-through child support payments received and the amount of child support collected by the agency on respondent's behalf. As with the TANF amounts, up to four different payment amounts may be reported. The amounts are recoded into monthly variables.

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Reference period indicator for coverage
- Respondents may report up to three TANF spells
- A continuation flag for respondents who reported receiving TANF in the last month of the reference period

Why is the TANF content important? What assistance does it provide?

SIPP is designed to observe individuals over time and show changes in income, program participation, and other areas. TANF is an important source of income for some low-income families with children. The longitudinal nature of the survey makes it possible to study the timing of TANF receipt in relationship to other important events, such as changes in marital status, employment status, and/or participation in other assistance programs.

The TANF data can be used to produce national estimates of coverage and payments. Although the Administration for Families and Children regularly publishes data on TANF participation and benefits, detailed TANF data are published less frequently. The SIPP collects more detailed information on TANF receipt than both the ACS and CPS and provides it at a more detailed level – the SIPP provides monthly

and annual data while the other surveys only provide annual data. When combined with data on earnings and other income, data on TANF benefits help provide a more comprehensive picture of a family's or household's economic well-being over the length of the panel.

What level of information does the TANF provide?

TANF data are collected at the spell level in the EHC and are provided to users in a person-month format.

What are the TANF variables and where can I find them?

While all data are in a monthly format, the value of some variables are the same for entire reference period (person-level variables). The value of other variables may be the same throughout an entire spell (spell-level variables) or may vary from month to month (monthly-level variables).

Person-level variables

- A reference period coverage indicator identifies whether an individual was covered by TANF in at least one month of the reference period (RTANFCOV)
- Year receipt began, for respondents with a TANF spell that begins in January of the reference year in respondent's first interview (ETANFLCY)
- The continuation flag identifies the status of a respondent's spell if it was on-going as of the last month of the reference period (RTANF_CONTFLG)

Month-level variables

- TANF benefit payment amount (TTANF_AMT)
- Pass-through child support payments received (TPT_AMT)
- Pass-through child support collected by the agency on respondent's behalf (TPTAG_AMT)

Spell-level variables

- Begin and end month of the TANF spell (ETANF_BMONTH and ETANF_EMONTH)
- Owner of the TANF benefit (ETANFOWN)
- Indicator of whether benefits covered children only, both adults and children, or only a pregnant woman (ETANFADLTKID)
- Reason(s) for beginning TANF receipt (ETANFBRSN1 and ETANFBRSN2)
- Reason(s) for ending TANF receipt, if applicable (ETANFERSN1 and ETANFERSN2)
- Monthly coverage indicator (RTANFYN)
- Indicator of whether pass-through from a state or county welfare program was received (EPTCS)

For each TANF benefit unit, one adult in the household is considered the owner of the TANF benefits (the person in whose name the benefits are received), but it is not necessarily the person who reported the TANF benefit. The TANF owner may or may not be covered by the TANF benefit. All details of the TANF spell are available on the record of the benefit owner. The beginning and ending month of the TANF spell, the continuation flag, and the owner of the TANF benefit are found on the records of all people covered by the TANF benefit. The monthly and reference year indicators of TANF receipt are found on all people's records.

What other concepts are relevant to the TANF content area?

Benefit Owner – A monthly variable that identifies the person number of the individual in whose name the benefit is received.

Continuation Flag – A person-level variable that indicates whether the year-end spell (1) ended at the end of the reference year, (2) continued at the end of the reference year but ended before the interview month, or (3) was ongoing during the interview month.

Income Screeners – The income screener questions, which are asked in the pre-EHC portion of the instrument, limit the universe for some means-tested programs. When certain income-related parameters are met, the instrument sets the EHC sections for SNAP, TANF, GA, and WIC to not in universe.

Screener Clump – The screener clump is an internally derived instrument variable that limits the number of people who are in universe for the income screener questions and the family- or household-based EHC screener questions (SNAP, TANF, and GA). The intention is to ask only one respondent the family- or household-based program questions when resources were shared during the entire reference period.

Spell – The duration of a benefit or event reported by a respondent in the Event History Calendar section of the survey that is defined by a start month, end month, and a continuous span between the start and end month.

Temporary Assistance for Needy Families – TANF is an income supplement program serving low-income single mothers and their children. Some states serve married couples. TANF is funded at the federal level and administered at the state level. While women receive TANF, court-ordered child support goes directly to the state, which decides whether and how much of the payment goes to the mother. This is called pass-through child support, bonus child support, or disregard payments.

3.13.18 Tax Returns

What is collected in the Tax Returns content?

SIPP collects person-level data from respondents age 15 or older regarding federal tax returns during the reference year, specifically:

- Did the individual file a tax return
- Did the individual plan to file a return (if no return was filed)
- Filing status
    - Single
    - Married filing jointly
    - Married filing separately
    - Head of household
- Did anyone claim the respondent as a dependent (respondents between age 15 and 25)
- Did the individual receive an Earned Income Tax Credit (EITC)

Major changes between 2008 and 2014 panels:

- New to 2014, collect the dependent status for individuals between ages 15 and 25.
- The 2008 panel and prior asked each individual to reference a copy of their previous year tax forms. We no longer ask respondents to reference tax returns.
- No longer collect:
    - Number of exemptions
    - Gross income
    - Net tax liability
    - Tax deductions
    - EITC amount
    - Capital gain and loss
    - Paid property taxes

Why is the Tax Returns information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources. Since tax return status does affect an individual's economic situation, the data collected for tax returns are important to understanding an individual's economic situation.

What level of information does the Tax Returns content provide?

Each Tax Return section is collected at the person level for all household members age 15 or older during the reference year. Children under the age of 15 are not in universe.

What are the Tax Returns variables and where can I find them?

Household income and program receipt, which determine eligibility for EITC, are collected prior to the EHC. The specific Tax Return information is collected after the EHC.

What other concepts are relevant to the Tax Returns content area?

Annual Poverty Threshold – An annual set of income thresholds that vary by family size and composition to determine who is in poverty. If a family total income is less than the threshold, the family and each member in the family group is in poverty. The annual poverty threshold is based on before-tax income. The annual poverty threshold does not include non-cash benefits (such as public housing, Medicaid, and food stamps).

Deductions – There are two types of tax return deductions. (1) A standard deduction is a dollar amount that reduces taxable income. (2) An itemized deduction is a combination of the standard deduction and the tax year dollar amount of tax expenses such as medical expenses, charitable contributions, and property taxes.

Earned Income Tax Credit – A tax credit earned by an individual with a tax return filing status of single, head of household, married filing separately, or married filing jointly who has earned income and adjusted gross income within certain limits.

Head of Household – An individual that is unmarried or considered unmarried who lived with one or more family members and who has paid more than half the cost of keeping the home for the tax year.

Married Filing Jointly – An individual that is married or considered married. To file jointly both the individual and their spouse must agree to file a joint return. On a joint return, you combine income and deduct your combined allowable expenses. A couple can file a joint return even if one had no income or deductions. A joint return tax may be lower than your combined tax for other filing statuses.

Married Filing Separately –A married individual who is responsible only for their own taxes or because it results in less tax than filing a joint return.

Monthly Poverty Threshold – A monthly set of income thresholds that vary by family size and composition to determine who is in poverty. If a family's total income is less than the threshold, then the family and each member in the family group is in poverty. The monthly poverty threshold is based on before-tax income. The monthly poverty threshold does not include non-cash benefits (such as public housing, Medicaid, and Food Stamps).

Net Tax Liability – The total dollar amount of annual income taxes owed or paid.

Single – An individual who is unmarried, considered unmarried, or a widow/widower who does not qualify as head of household, married filing separately, or married filing jointly.

3.13.19 Unemployment Compensation Payment

What is collected in the Unemployment Compensation Payment content?

SIPP collects person-level data from respondents aged 15 or older regarding Unemployment Compensation (UC) receipt at any time during the reference year. Additional questions categorize UC receipt into three different types (regular, supplemental, and other).

The data collected in the 2014 panel cover the same major topics as in the 2008 and prior panels, specifically the type of UC payment and the amount received during each month of the reference year.

The UC section begins by asking if the respondent received any income from unemployment compensation during the reference year. If a respondent replies that they have received UC income during the reference year, then the instrument asks about which types of UC were received. The instrument collects data for three types of compensation. The first type is regular, which is the most commonly known type, received from state unemployment agencies and the federal government. The second is supplemental, which includes unemployment insurance paid for privately. The third type is denoted as other in the questionnaire, and corresponds to union benefits that respondents may be eligible for if they are members of a union.

For each benefit type reported, the instrument asks for the months of receipt and amount received. For respondents who received multiple types of UC, once the data about the first type are collected, the instrument moves on to ask about the next type.

For each type, the months of receipt are determined first. Amounts are collected by asking about the amount received in the most recent month of receipt, which can be the interview month. For example,

"How much did you receive in regular Unemployment Compensation in October?" Then the instrument asks, "Did you get that amount during all months of receipt?"

If the respondent replies that the amount received was the same in all months, then the instrument moves to the next type received, or if only one type was received, to the next topic in the survey. If the respondent replies that the amount received varied, then the questionnaire asks the amount received for each month of receipt, month by month, starting with the first month of receipt.

Major changes between the 2008 and 2014 panels:

Collect amount data month by month for respondents who have different amounts in some months.

Why is Unemployment Compensation Payment information important? What assistance does it provide?

Data from job earnings alone are not representative of an individual's economic situation. Individuals often receive income from non-job sources such as unemployment compensation. Especially for unemployed individuals, unemployment compensation can constitute a large share of available income, and have a great impact on their economic situation.

What level of information does the Unemployment Compensation Payment content provide?

Unemployment Compensation data collected in SIPP can be used to create estimates on the national, household and person-levels. Unemployment Compensation data are collected on the person level for all household members who were 15 and older during the reference year. Teenagers who were aged 15 and older during the reference year are in universe for the Unemployment Compensation Income section.

What are the Unemployment Compensation Payment variables, and where can I find them?

Unemployment Compensation data are collected after the EHC in the SIPP questionnaire.

What other concepts are relevant to the Unemployment Compensation Payment content area?

<u>Covered Employment</u> – "Covered" here refers to being covered by state unemployment insurance programs. Most work for wages falls under the category of covered employment. Exceptions vary by state, but can include independent contractors, sole proprietors or members of a partnership, railroad workers subject to Federal RR Unemployment Insurance Act, some real estate brokers, some insurance brokers, some agricultural workers, elected officeholders, some church employees, and other exceptions on a state-by-state basis.

<u>Other Unemployment Insurance</u> – This category of unemployment insurance is a catchall for unemployment compensation not from regular or supplemental unemployment insurance programs. The most common "other" programs correspond to union benefits that respondents may be eligible for if they are members of a union.

Supplemental Unemployment Insurance – This refers to all unemployment insurance plans paid for privately. Some of these are paid for by employees themselves, or by the self-employed. They can be used as an additional wage replacement to regular unemployment compensation, because state UI payments are generally capped at a weekly or monthly amount that may be lower than workers' full wage rate. These programs are also provided by employers in sectors where employment is not covered. Employer-sponsored plans in non-covered employment make up the bulk of supplemental unemployment insurance plans.

Unemployment Insurance (UI) – This is the name of the programs administered by every state and the District of Columbia that provide assistance to jobless people who are looking for work. In these programs, employers are required to pay a certain amount to the state unemployment agency monthly. The amount paid is tied to each worker's wage rate and hours worked. These payments fund the unemployment insurance agency. Not all types of employment are covered by UI.

For workers who lose their job, file for UI, and are determined to be eligible for it, these programs provide a payment equal to a percentage of previous earnings. The amount of the payment varies by state. The number of weeks that a worker can receive the benefit also varies by state. In SIPP, payments from this type of program are coded as regular Unemployment Compensation payments.

3.13.20 Veterans Benefits

What is collected in the Veterans Benefits content?

The Veterans Benefits section collects person-level data from household members age 17 or older during the reference year who reported previously serving in the U.S. Armed Forces but are not currently enlisted.

SIPP collects the type of veterans benefits and amount received during each month in the reference year. The section begins by asking if the respondent received veterans benefits at any point during the reference year.

For respondents who report receiving veterans benefits, SIPP asks what type of benefits were received during the reference year. Respondents may report receipt of five separate benefit types:

1. Service-connected disability compensation plan
2. Veterans pension
3. Other VA payments
4. G.I. Bill benefits
5. Insurance proceeds

SIPP asks each VA benefit recipient if he or she was required to complete an Eligibility Verification Report, or EVR, during the reference year in order to receive VA benefits. For respondents receiving service connected disability compensation benefits, SIPP collects their disability rating:

1. 0 percent
2. 10-20 percent
3. 30-40 percent

4. 50-60 percent
5. 70+ percent

The instrument collects monthly receipt amounts separately for each benefit type.

Monthly amounts are collected by asking about the amount received in the most recent month of receipt, which can be the interview month. For example, "How much do you receive in veterans pension benefits now?" Then the instrument then asks, "Did you get that amount during all months of receipt?"

If the respondent replies that the amount received was the same in all months, then the instrument moves to the next benefit type or survey topic. If the respondent replies that the amount received was not the same in each month of receipt, the instrument then asks the amount received prior to the most recent amount. The instrument continues to work backwards through the reference period until the amount for each month of receipt is collected. After a respondent has reported four different receipt amounts, the instrument completes the monthly amount section by asking the amount received during the first month of receipt during the reference period.

Major changes between the 2008 and 2014 panels:

- Collect the monthly amount received
- Collect monthly changes in the amount

2014 and 2008 Veterans Benefit Panel Year Crosswalk	
2014 Panel	**2008 Panel**
VA_ANY	VAYN, VAMILRET
VA_TYPE	VAYNTYPE
VAQUES	VAQUES
VA_NOW	
VA_START_REC	
VA_LAST_REC	
VA_1STSTART	
VA_1STAMT	
VA_GUESS	
VA_2NDSTART	
VA_2NDAMT	
VA_3RDSTART	
VA_3RDAMT	
VA_4THSTART	
VA_4THAMT	
VA_STARTAMT	

Why is the Veterans Benefits content important? What assistance does it provide?

Data from job earnings alone are not representative of a household's economic situation. Individuals often receive income from non-job sources such as veterans benefits. Since income received from a job

source or veterans benefits affect an individual's economic situation, the data collected for veterans benefits is important to understanding an individual's economic situation.

What level of information does the Veterans Benefits content provide?

Monthly benefit amounts for each type of benefit reported are collected at the person-level for all household members age 17 or older during the reference year that reported previously serving in the U.S. Armed Forces but are not currently enlisted.

What are the Veterans Benefits variables and where can I find them?

All of the Veterans Benefits data are collected after the Event History Calendar (EHC) in the post-EHC section. The Veterans Benefits section collects the type of benefit(s), amount(s) received, and when the amount was received for up to four amount changes during the reference year.

What other concepts are relevant to the Veterans Benefits content area?

Military Retirement Pay – A retirement pension plan provided for active duty military personnel with benefits that start when the employee retires.

National Guard Retirement – An Army National Guard employee retirement plan provided for employees who have 20 years service in the National Guard. By age 60, this retirement plan may combine with additional income or retirement pension.

Pension – A pension is a benefit plan in which employees and employers pay into a group trust fund. The pension benefits are based on the factoring of an employee's final pay and the employment tenure with the employer.

Veterans Benefits – This refers to a monthly pension administered by the United States Department of Veteran Affairs to active duty military branch members, their dependents, and their survivors.

Veterans Compensation – This refers to a monthly pension administered by the United States Department of Veteran Affairs to veterans, their dependents, and their survivors.

Veterans Pension – This refers to a monthly pension administered by the United States Department of Veteran Affairs for veterans who served in the military but did not qualify for military retirement pay.

Reserve Forces Retirement –A retirement pension plan for service members who complete 20 years of service and/or who become eligible for retirement from the Reserve Forces at age 60 or in some cases before age 60 depending on active service.

Service-Connected Disabilities – A disability benefit paid to veterans disabled by injury or illness during active duty. Monthly benefits vary with the type of disability and the number of eligible dependents.

3.13.21 WIC

What is collected in the WIC content?

The WIC section collects information about WIC benefit receipt from the start of the reference period through the interview month. Variables on the public use file provide the following information:

- Monthly and reference period coverage indicators
- Months received
- Year receipt began if receiving in January of the reference period (wave 1)
- Reason(s) receipt began
- Reason(s) receipt stopped
- Whether benefits covered only children, only a pregnant or nursing woman, or both
- Payment amount[3]
- Person in whose name the benefit is received (benefit owner)[4]
- Continuation flag for spells that include the last month of the reference period

To reduce respondent burden, SIPP does not collect information about the receipt of WIC separately for children under age 15. Female respondents age 15 or older, or male respondents with a child under age 5 in the household, with a household income below 200 percent of the poverty line are asked to report whether they or anyone in their family (i.e., children) received WIC benefits.

As with other data collected in the EHC, WIC information is collected in reverse chronological order. That is, each respondent is first asked if they or someone in their family are currently receiving WIC. If a respondent reports current receipt, the month of the reference period or interview year receipt began is collected. If no one in the respondent's family is currently receiving WIC, the respondent is then asked if they received WIC at any time since the first month of the reference period. If receipt is reported during that time, the months that receipt started and stopped are collected. If a WIC spell includes January of the reference period the first time a respondent is interviewed, information on the year the respondent first began receiving benefits is collected.

For each WIC spell, respondents are asked to report why receipt began, why receipt stopped (if applicable), and which children and adults were covered by the benefit. Additionally, respondents indicate whether the benefit covered only children, only a pregnant or nursing woman, or both.

A continuation flag for respondents who reported receiving WIC in the last month of the reference period provides information on whether the spell: (1) ended in the last month of the reference period; (2) ended after the last month of the reference period but before the interview month; or (3) was ongoing as of the interview month.

Major changes between the 2008 and 2014 panels:

- Longer reference period (one year vs. four months)
- Reference period indicator for coverage

---

[3] The SIPP questionnaire does not ask respondents to report the amount of their WIC benefit. WIC participants are not given a specific amount of money, rather they are allowed to purchase a certain package of food products. Often recipients do not know the monetary value of this food package. WIC participants are assigned a WIC benefit amount based on average WIC expenditures.

[4] The SIPP questionnaire does not ask respondents to report the owner of the WIC benefit. For data editing purposes, the respondent who reports the WIC spell is assigned as the benefit owner.

- Respondents may report up to three WIC spells
- A continuation flag for respondents who reported receiving WIC in the last month of the reference period

Why is the WIC content important? What assistance does it provide?

SIPP is designed to observe individuals over time and show changes in income, program participation, and other areas. WIC is an important source of income for some low-income families with children. The longitudinal nature of the survey makes it possible to study the timing of WIC receipt in relationship to other important events, such as changes in marital status, employment status, and/or participation in other assistance programs.

The WIC data can be used to produce national estimates of coverage and payments. Although the United States Department of Agriculture regularly publishes data on WIC participation and benefits, detailed WIC data are published less frequently. The SIPP collects more detailed information on WIC receipt than the CPS and provides it at a more detailed level – the SIPP provides monthly and annual data while the CPS only provide annual data. When combined with data on earnings and other income, data on WIC benefits help provide a more comprehensive picture of a family's or household's economic well-being over the length of the panel.

What level of information does WIC information provide?

WIC data are collected at the spell level in the EHC and are provided to users in a person-month format.

What are the WIC variables and where can I find them?

While all data are in a monthly format, the value of some variables are the same for entire reference period (person-level variables). The value of other variables may be the same throughout an entire spell (spell-level variables) or may vary from month to month (monthly-level variables).

Person-level variables

- A reference period coverage indicator identifies whether an individual was covered by WIC in at least one month of the reference period (RWICCOV)
- Year receipt began, for respondents with a WIC spell that begins in January of the reference year in respondent's first interview (EWICLCY)
- The continuation flag identifies the status of a respondent's spell if it was on-going as of the last month of the reference period (RWIC_CONTFLG)

Month-level variables

- Payment amount (TWIC_AMT)

Spell-level variables

- Begin and end month of the WIC spell (EWIC_BMONTH and EWIC_EMONTH)
- Owner of the WIC benefit (EWICOWN)
- Indicator of whether benefits covered children only, only a pregnant or nursing woman, or both (EWICCOVTYP)

- Reason(s) for beginning WIC receipt (EWICBRSN1 and EWICBRSN2)
- Reason(s) for ending WIC receipt, if applicable (EWICERSN1 and EWICERSN2)
- Monthly coverage indicator (RWICYN)

For each WIC benefit unit, one adult in the household is considered the owner of the WIC benefits (the person in whose reported the WIC spell). The WIC owner may or may not be covered by the WIC benefit. All details of the WIC spell are available on the record of the benefit owner. The beginning and ending month of the WIC spell, the continuation flag, and the owner of the WIC benefit are found on the records of all people covered by the WIC benefit. The monthly and reference year indicators of WIC receipt are found on all people's records.

What other concepts are relevant to the WIC content area?

Benefit Owner – A monthly variable that identifies the person number of the individual in whose name the benefit is received.

Continuation Flag – A person-level variable that indicates whether the year-end spell (1) ended at the end of the reference year, (2) continued at the end of the reference year but ended before the interview month, or (3) was ongoing during the interview month.

Income Screeners – The income screener questions, which are asked in the pre-EHC portion of the instrument, limit the universe for some means-tested programs. When certain income-related parameters are met, the instrument sets the EHC sections for SNAP, TANF, GA, and WIC to not in universe.

Spell – The duration of a benefit or event reported by a respondent in the Event History Calendar section of the survey that is defined by a start month, end month, and a continuous span between the start and end month.

WIC **–** Women, Infants, and Children (WIC) is a federal grant program administered by the state and at the federal level by the Food and Nutrition, U.S. Department of Agriculture in 1974. WIC was formerly known as the Special Supplemental Nutrition Program. The WIC program provides supplemental food, healthcare referrals and nutrition education for low income program women (up to 6 months after birth or after pregnancy ends), infants, and children up to age 5 who are at nutrition risk. Recipients of WIC receive monthly benefits which may include checks, vouchers, or electronic cards to purchase food. WIC benefits also include supplemental foods, nutrition education, nutrition counseling, as well as screening and referrals for health, welfare, and social services.

Farmers' Market Nutrition Program (FMNP) – FMNP is a federal grant program that began in 1992. The FMNP program provides locally grown food to WIC participants with the goal of expanding awareness and use of farmers' markets for low income women (up to 6 months after birth or after pregnancy ends), infants, and children up to age 5 who are at nutrition risk.

3.13.22 Workers' Compensation

What is collected in the Workers' Compensation content?

The Workers' Compensation section collects person-level data for household members age 15 or older regarding workers' compensation benefits received at any time during the reference year.

The section begins by asking if the respondent received workers' compensation benefits at any point during the reference year. If a respondent replies that they have received workers' compensation during the reference year, the instrument then asks about which types of workers' compensation were received. The instrument collects data for three types of compensation. The first type is regular government-provided, which is the most common type. The second is supplemental employer-provided. The third type is other, which includes union benefits that respondents may be eligible for if they are members of a union.

For each benefit type reported, the instrument asks for the months of receipt and amount received. For respondents who received multiple types of workers' compensation, once the data about the first type are collected, the instrument moves on to ask about the next type.

For each type, the months of receipt are determined first. Amounts are collected by asking about the amount received in the most recent month of receipt, which can be the interview month. For example, "How much do you receive in regular Workers' compensation now?" Then the instrument asks, "Did you get that amount during all months of receipt?"

If the respondent replies that the amount received was the same in all months, then the instrument moves to the next benefit type or survey topic. If the respondent replies that the amount received was not the same in each month of receipt, then the instrument then asks the amount received prior to the most recent amount. The instrument continues to work backwards through the reference period until the amount for each month of receipt is collected. After a respondent has reported four different receipt amounts, the instrument completes the monthly amount section by asking the amount received during the first month of receipt during the reference period.

Major changes between the 2008 and 2014 panels:

- Collect the monthly amount received
- Collect monthly changes in the amount

2014 and 2008 Workers' Compensation Benefit Panel Year Crosswalk	
2014 Panel	**2008 Panel**
WC_ANY	WCYN
WC_NOW	
WC_START_REC	
WC_LAST_REC	
WC_1STSTART	
WC_1STAMT	
WC_GUESS	
WC_2NDSTART	
WC_2NDAMT	
WC_3RDSTART	
WC_3RDAMT	

WC_4THSTART	
WC_4THAMT	
WC_STARTAMT	

Why is the Workers' compensation content important? What assistance does it provide?

Data from job earnings alone are not representative of a household's economic situation. Individuals often receive income from non-job sources such as workers' compensation. Since income received from a job source or workers' compensation affect an individual's economic situation, the data collected for Workers' compensation is important to understanding an individual's economic situation.

What level of information does the Workers' compensation content provide?

Monthly benefit amounts for each type of benefit reported are collected at the person-level for all household members age 15 or older during the reference year. Children under the age of 15 by the end of the reference year are not in universe for the Workers' Compensation section.

What are the Workers' compensation variables and where can I find them?

All of the Workers' compensation data are collected after the Event History Calendar (EHC) in the post-EHC section. The Workers' compensation section collects the type of benefit(s), amount(s) received, and when the amount was received for up to four amount changes during the reference year.

What other concepts are relevant to the Workers' compensation content area?

Workers' compensation – Employer's insurance providing wage and medical benefits to an employee who is injured during the course of employment. Workers' compensation insurance is required by almost every state. Rules and requirements may vary by state. An employer pays for workers' compensation and the employer is covered for work-related accidents. Workers' compensation benefits can help pay for medical and hospital bills. They also can be provided if the employee cannot return to work. If an employee dies due to a work-related accident, workers' compensation insurance can also pay an insurance benefit to the family.

Workers' compensation Payments – Plans and coverage may vary. Plans can be made for weekly payments in place of wages (similar to disability insurance), compensation for loss in wages, and reimbursement or payment of medical and health care expenses (similar to health insurance). Workers' compensation benefits can be payable to families of workers killed during employment (similar to life insurance). Most states have workers' compensation coverage requirements; however, some states do not require workers' compensation coverage for specific industries such as agriculture and some small private-sector employers.

Job-Related Injury – This refers to an injury that occurred while working or because of a work related role.

Job-Related Illness – This refers to an occupational related disease or disability that occurred as a result of a work related role or during employment.

3.14 Income and Poverty Recodes

Unlike the above Survey Content areas, data on total income and poverty status are not collected directly from the SIPP Instrument, but are aggregated or assigned based on existing survey content.

3.14.1 Income Recodes

What is included as an Income Recode?

The SIPP instrument creates a number of income recodes at the person, family, and household level. The Income Recodes aggregate reported income from the Assets, Employment and Earnings, and Programs and Income Transfer sections of the SIPP.

Major changes between the 2008 and 2014 panels:

In the 2014 panel, monthly income from non-means tested programs (TPSCININC) was separated from means-tested transfers (TPTRNINC). In the 2008 panel, income from both sources were reported under TPTRNINC.

In the 2014 panel, TPEARN no longer includes income from moonlighting since this is not measured separately in the 2014 panel. Similarly, TPEARN in the 2014 panel no longer includes income from severance payments since this is not measured separately from lump sum payments from a retirement plan in the 2014 panel. In the 2014 panel, TPEARN values reflect the reported pay schedule of respondents while accounting for monthly fluctuations in earnings given that there are fewer aggregate hours and weeks typically worked in some months relative to others. In the 2008 panel, month durations were not factored into calculating monthly earnings.

Why is the Income Recode information important? What assistance does it provide?

Total monthly personal income, TPTOTINC, is one of the most frequently used variables by SIPP users as it provides a metric to evaluate an individual's economic well-being. Income is the measure most commonly used because it affects both the quantity and quality of the goods and services an individual can purchase.

Additional variables allow data users to measure total income resources shared at the family or household level.

What level of information does the Income Recode content provide?

The Income Recodes are calculated monthly by summing the following income sources. For income sources reported annually, reported income is allocated evenly across the months of the reference year

by dividing by 12. To avoid double counting, transfer payment amounts that are family based are only placed on the record of the family member identified as the owner of the benefit.

For detailed information on the individual components of the Income Recodes, see their respective sections of the User Guide.

Total Personal Income (TPTOTINC) Derived from:

TPEARN	Total personal earnings
TPPRPINC	Total personal investment/property income
TPTRNINC	Total personal means-based transfer income
TPSCININC	Total personal social insurance payments
TPOTHINC	Total personal other income

Total Personal Earnings (TPEARN) Derived from:

Income earned from all jobs worked during the month, including wage and salary income, bonus payments, commissions, overtime payments, tips, other income from self-employed businesses, self-employed business profits, and accounting for time spent away from a job without pay.

Total Personal Investment & Property Income (TPPRPINC) Derived from:

TINC_BANK	Income earned from interest-earning assets held at financial institutions	
	TOICHKINC	Total interest income earned from individually-owned checking accounts
	TJSICHKINC	Share of total interest income earned jointly-owned checking accounts [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOICHKINC	Share of total interest income earned from jointly-owned interest-earning checking accounts [for respondents without a spouse or civil union partner in the household]
	TOSAVINC	Total interest income earned from individually-owned savings accounts
	TJSSAVINC	Share of total interest income earned from jointly-owned savings accounts [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOSAVINC	Share of total interest income earned from jointly-owned savings accounts [for respondents without a spouse or civil union partner in the household]
	TJSCDINC	Share of total interest income earned from jointly-owned certificates of deposit[for respondents with a spouse or civil union partner in the household; this joint amount question is

		asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOCDINC	Share of total interest income earned from jointly-owned certificates of deposit [for respondents without a spouse or civil union partner in the household]
	TOCDINC	Total interest income earned from individually-owned certificates of deposit
	TJSMMINC	Share of total interest income earned from jointly-owned money market accounts and money market mutual funds [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse or partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOMMINC	Share of total interest income earned from jointly-owned money market accounts and money market mutual funds [for respondents without a spouse or civil union partner in the household]
	TOMMINC	Share of total interest income earned from individually-owned money market accounts and money market mutual funds
TINC_BOND	Income earned from other interest-earning assets	
	TJSGOVSINC	Share of total interest income earned from jointly-owned government securities [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOGOVSINC	Share of total interest income earned from jointly-owned government securities [for respondents without a spouse or civil union partner in the household]
	TOGOVSINC	Total interest income earned from individually-owned government securities
	TJSMCBDINC	Share of total interest income earned from jointly-owned municipal and corporate bonds [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOMCBDINC	Share of total interest income earned from jointly-owned municipal and corporate bonds [for respondents without a spouse or civil union partner in the household]
	TOMCBDINC	Total interest income earned from individually-owned municipal and corporate bonds
TINC_STMF	Income earned from stocks and mutual funds	

	TJSMFINC	Share of dividend income earned from jointly-owned mutual funds [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOMFINC	Share of dividend income earned from jointly-owned mutual funds [for respondents without a spouse or civil union partner in the household]
	TOMFINC	Total dividend income earned from individually-owned mutual funds
	TJSSTINC	Share of dividend income earned from jointly-owned stocks [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse/partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJOSTINC	Share of dividend income earned from jointly-owned stocks [for respondents without a spouse or civil union partner in the household]
	TOSTINC	Total dividend income earned from individually-owned stocks
TINC_RENT	Net income from rental properties	
	TJSRPNETINC	Share of net income earned from jointly-owned rental property [for respondents with a spouse or civil union partner in the household; this joint amount question is asked of only one spouse or partner, and this response is divided by 2 and copied to both spouses'/partners' records]
	TJORPNETINC	Share of net income earned from jointly-owned rental property [for respondents without a spouse or civil union partner in the household]
	TORPNETINC	Net income earned from individually-owned rental property
TINC_OTH	Net income from other assets	
	TANNINC	Amount of income received from annuities
	TTRINC	Amount of income received from trusts
	TOINVINC	Amount of income received from other financial investments

Total Means-Tested Transfer Payments (TPTRNINC) Derived from:

TTANF_AMT	Monthly TANF benefit amount
TPT_AMT	Monthly pass-through child support amount
TSSI_AMT	Monthly SSI benefit amount
TGA_AMT	Monthly GA benefit amount
TVA2AMT	Amount received from Veterans' pension

Total Social Insurance Payments (TPSCININC) Derived from:

TVA1AMT	Amount received from VA benefit payments for a service-connected disability
TVA3AMT	Amount received from other VA payments
TVA4AMT	Amount received for G.I. Bill benefits
TVA5AMT	Amount received from VA Insurance proceeds
TWCAMT	Amount received in workers' compensation
TUC1AMT	Amount of regular unemployment compensation received
TUC2AMT	Amount of supplemental unemployment compensation received
TUC3AMT	Amount of other unemployment compensation received
TSSSAMT	Amount received in Social Security benefit payments
TSSCAMT	Amount received in Social Security benefit payments on behalf of a child(ren)

Total Other Income (TPOTHINC) Derived from:

TSUR1AMT	Amount received in survivor benefits from a pension from a company or union including income from a profit-sharing plan
TSUR2AMT	Amount received in survivor benefits from Veterans' compensation or pension
TSUR3AMT	Amount received in survivor benefits from a Federal civil service or other Federal civilian employee pension
TSUR4AMT	Amount received in survivor benefits from U.S. Government Railroad Retirement in month
TSUR5AMT	Amount received in survivor benefits from a state government pension
TSUR6AMT	Amount received in survivor benefits from a local government pension
TSUR7AMT	Amount received in survivor benefits from income from a paid-up life insurance
TSUR8AMT	Amount received in survivor benefits from military retirement pay
TSUR9AMT	Amount received in survivor benefits from Black Lung Benefits
TSUR10AMT	Amount received in survivor benefits from workers' compensation
TSUR11AMT	Amount received in survivor benefits from a payments from an estate or trust
TSUR12AMT	Amount received in survivor benefits from National Guard or Reserve Forces retirement
TSUR13AMT	Amount received in survivor benefits from other survivor income
TRET1AMT	Amount received in retirement income from a pension from a company or union including income from a profit-sharing plan
TRET2AMT	Amount received in retirement income Federal civil service or other Federal civilian employee pension
TRET3AMT	Amount received in retirement income from State government pension
TRET4AMT	Amount received in retirement income from local government pension
TRET5AMT	Amount received in retirement income military retirement pay
TRET6AMT	Amount received in retirement income from U.S. Government Railroad Retirement
TRET7AMT	Amount received in retirement income from National Guard or reserve forces retirement
TRET8AMT	Amount received in retirement income from other retirement income
TDIS1AMT	Amount received in payments from a sickness, accident, or disability insurance policy

TDIS2AMT	Amount received in payments from an employer disability payments
TDIS3AMT	Amount received in pension from a company or union including income from a profit sharing plan
TDIS4AMT	Amount received payments from Federal civil service or other federal civilian employee pension
TDIS5AMT	Amount received from a state government pension
TDIS6AMT	Amount received from a local government pension
TDIS7AMT	Amount received from U.S. military retirement pay
TDIS8AMT	Amount received from U.S. government railroad retirement
TDIS9AMT	Amount received in payments from Black Lung Benefits
TDIS10AMT	Amount received from other disability income
TFCCAMT	Amount of foster child care payments received
TCSAMT	Amount of child support payments received
TALIAMT	Amount of alimony payments received
TLMPAMT	Amount of lump sum payment received
TDEFERAMT	Amount received as a deferred payment or final paycheck from a job or business that ended during the reference period
TLIFEAMT	Amount received from a paid up life insurance policy or annuity
TMINC_AMT	Amount of miscellaneous income received

What are the Income Recode variables and where can I find them?

The SIPP instrument creates aggregate monthly income recodes at the person, family, and household level. Additional recodes are created at the person level based on income source.

Aggregate			Description
Person	**Family**	**Household**	
TPTOTINC	TFTOTINC	THTOTINC	Total monthly income excluding Type 2 people
	TFTOTINCT2	THTOTINCT2	Total monthly income including Type 2 people
Components			**Description**
Person	**Family**	**Household**	
TPEARN			Monthly earnings: The sum of the amount of gross earnings, wages, and salary, and/or the amount of monthly income (positive or negative) from self-employment for each job and/or business recorded for the reference month.
TPPRPINC			Monthly investment and property income: The sum of dividend, interest, and property/rental income
TPTRNINC			Monthly means-tested transfer payments: The sum of all means-tested cash transfers (e.g., SSI, GA, TANF)
TPSCININC			Monthly social insurance payments: Monthly income from non-means-tested social insurance programs (e.g., Social Security, unemployment compensation)
TPOTHINC			Monthly income from other sources: Includes support payments (e.g., alimony, child support), retirement income (e.g., pensions), and any other income not captured elsewhere

What other concepts are relevant to the Income Recode content area?

Family – Family-level income variables include the income of all household members related by birth, marriage, or adoption in a given month. 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.

Household – Household-level income variables include the income of all household members in a given month.

Type 2 Person – These are people that were not in the household at the time of interview but were in the household during the reference year and are currently not THHLDSTATUS = 3, 4, 5, or 6 respondents.

Distribution of Annual Income – For income sources reported annually, reported income is allocated evenly across the months of the reference year by dividing by 12.

Reporting of Joint Income – To avoid double counting, transfer payment amounts that are family based are only placed on the record of the family member identified as the owner of the benefit. For respondents with a spouse or civil union partner in the household; income from jointly owned accounts are asked of only one spouse/partner and this response is divided by 2 and copied to both spouses'/partners' records.

3.14.2 Poverty Recodes

What is included as a Poverty Recode?

Poverty Recodes are assigned based on data collected on household and family composition as well as the aggregate Income Recodes.

Major changes between the 2008 and 2014 panels:

Unlike in the prior 2008 panel when only poverty thresholds were provided, the 2014 panel also includes income-to-poverty ratios, removing confusion about the universe and assignment of poverty status.

The 2014 panel also includes annual measures of income-to-poverty ratios. For individuals in the poverty universe all year, monthly thresholds and income are summed across all 12 months of the calendar year to create an annual measure of poverty.

Subfamily poverty thresholds are no longer produced in the 2014 panel as the concept of subfamilies is no longer applicable. When two or more distinct families (individuals related by birth, marriage, or adoption) reside in the same household in the 2014 panel each family receives the appropriate threshold based on composition.

Why is the Poverty Recode information important? What assistance does it provide?

Monthly family poverty status (TFINCPOV) is one of the most frequently used variables by SIPP users as this provides a monthly measure of poverty, allowing data users to evaluate the movement of families into and out of poverty over the course of the reference period.

Additional variables at the household level and including Type 2 individuals allow data users to measure the ability of a family or household to meet their basic needs.

What level of information does the Poverty Recode content provide?

The monthly poverty thresholds are assigned based on the composition of a family or household in a given month, depending on the variable being used. The monthly poverty thresholds are also adjusted to reflect changes in the Consumer Price Index for all Urban Consumers (CPI-U) across months.

The "base" annual poverty thresholds from 1982 are available on the Census Bureau's website, https://census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html. These annual thresholds are converted to monthly thresholds by dividing by 12 and adjusting for inflation based on the difference in the reference month CPI-U and the 1982 annual CPI-U value.

Once the appropriate monthly threshold is assigned, income-to-poverty ratios are calculated by dividing topcoded total family (or household) income by the appropriate threshold. If the resulting ratio is less than 1, then that family (or household) and all individuals in it, are in poverty for the given month.

For individuals with a valid poverty threshold in each month of the reference period, annual income-to-poverty ratios are calculated. Values are based on the annual sum of the individual's monthly total family (or household) income divided by the annual sum of the individual's monthly poverty thresholds (reflecting changing family or household composition over the reference period). The resulting annual values are then divided to calculate the individual's annual income-to-poverty ratio.

What are the Poverty Recode variables and where can I find them?

Poverty Recodes are available at both the monthly and person level, providing thresholds as well as poverty status.

Monthly

Thresholds	
RFPOV	Monthly family poverty threshold excluding Type 2 persons
RFPOVT2	Monthly family poverty threshold including Type 2 persons
RHPOV	Monthly household poverty threshold excluding Type 2 persons
RHPOVT2	Monthly household poverty threshold including Type 2 persons
Poverty Status	
TFINCPOV	Monthly family income-to-poverty ratio excluding Type 2 persons
TFINCPOVT2	Monthly family income-to-poverty ratio including Type 2 persons
THINCPOV	Monthly household income-to-poverty ratio excluding Type 2 persons
THINCPOVT2	Monthly household income-to-poverty ratio including Type 2 persons

Annual

Poverty Status	
TFCYINCPOV	Annual family income-to-poverty ratio excluding Type 2 persons
TFCYINCPOVT2	Annual family income-to-poverty ratio including Type 2 persons
THCYINCPOV	Annual household income-to-poverty ratio excluding Type 2 persons
THCYINCPOVT2	Annual household income-to-poverty ratio including Type 2 persons

What other concepts are relevant to the Poverty Recode content area?

Family – A family is defined as all household members related by birth, marriage, or adoption in a given month. In other words, primary family members, including related subfamily members, are treated as one family by the Census Bureau when creating family assignments. When poverty status is determined at the family level, all family members have the same poverty status.

Household – Household-level poverty status is based on the income of all household members in a given month. When poverty status is determined at the household level, all household members have the same poverty status.

Type 2 Person – These are people that were not in the household at the time of interview but were in the household during the reference year and are currently not THHLDSTATUS = 3, 4, 5, or 6 respondents.

Income-to-Poverty Ratios – Income-to-poverty ratios are calculated by dividing total family or household income by the appropriate threshold. Resulting values below 1 indicate a family/household is "in poverty" with values closer to 1 indicating a family/household is closer to meeting their threshold. Values equal to or above 1 indicate a family is not in poverty.

4 Locating Information

This section details where data users may find the data and its corresponding documentation, including links for all publicly available Survey of Income and Program Participation (SIPP) data.

4.1 SIPP Website

The SIPP website presents a number of resources related to SIPP, from its inception in 1984 to the present, including:

- SIPP data files for all panels and waves
- Research products based on SIPP data
- Technical documentation
- The SSA Supplement

All SIPP related information is located at: www.census.gov/sipp

4.2 Data

Users may directly access SIPP data in multiple ways:

- FTP
- Orlin Research SIPP Data Tool
- DataFerrett
- External Sources

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.

The SIPP data page is located at: http://www.census.gov/programs-surveys/sipp/data.html

4.3 FTP Site

The SIPP FTP site contains datasets for all panels and waves of SIPP. For each wave of a specific panel, the FTP site contains the core data file, topical module data files, replicate weight file, data dictionary, and available SAS input statements.

Additionally, the FTP site contains the longitudinal weight data file, longitudinal weight file data dictionary, longitudinal weight SAS input statements, longitudinal replicate weight file data dictionary, longitudinal replicate weight SAS input statements, longitudinal replicate weight record layout, longitudinal replicate weight files for each calendar year of the panel, and longitudinal replicate weight files through each year of the panel.

Data files can be downloaded as compressed files. The three data compression types available on the FTP site are: DOS/Windows, GNU gzip, and Standard Unix. The FTP site also contains brief user notes that can also be found within the main data page section for the corresponding panel and wave.

The SIPP FTP site is http://thedataweb.rm.census.gov/ftp/sipp_ftp.html

4.4 Orlin Research SIPP Data Tool

The SIPP Data Tool is a software package designed by Orlin Research, Inc. under contract to the U.S. Census Bureau. The tool is designed with an eye towards simplicity and user-friendliness. It allows analysts and data users to easily join multiple waves of SIPP content, including both core and topical module files. Additionally, you can pool multiple SIPP panels into a single dataset for analysis. Data users can create recodes and summarize data across file types and records. The data analysis happens directly in the Data Tool. It incorporates the R statistical package, and so it allows you to run a variety of analyses, from basic descriptive statistics to sophisticated time-series and spell analysis models. In addition to the data, the tool integrates SIPP documentation like codebooks, questionnaires, and other metadata with dynamic links into the data itself. As the new tool is nearing release, more information will be posted to the SIPP website.

4.5 DataFerrett

DataFerrett is a web-based data analysis and extraction tool to customize national, state, and local data. It was developed by the U.S. Census Bureau to equip users with the ability to analyze large amounts of data and create customized reports and tables. Through DataFerrett, users can create customized spreadsheets and turn those spreadsheets into graphs and maps without any additional software. DataFerrett draws upon the DataWeb, a network of public and private databases providing a vast amount of statistical information that is constantly updated and expanded.

There are two basic types of data that users can access through DataFerrett:

- **Microdata** – A data record that represents a survey response or an administrative record.
- **Aggregated Data** –A variable that contains a count of an estimate of a characteristic (e.g., the number of factories in a county or the number of people in the labor force).

DataFerrett allows users to:

- Select the dataset and the variables within the dataset
- Review and recode selected variables
- Either download the data for use in a different software program or make a table using DataFerrett's built-in spreadsheet tool
- Tabulate data from two or more datasets (provided they have a common dimension or variable)
- Analyze standard geographies (e.g. state, county, tract)
- Create thematic maps on the fly for any table that contains a mappable geography in one of the dimensions
- Extract data for use in mapping software, containing the fully qualified geographic codes

DataFerrett has several tutorials and a Users' Guide in order to help users navigate the website.

The DataFerrett website is http://dataferrett.census.gov/

4.6 External Resources

Besides the Census Bureau-maintained data resources described above, other institutions have also created resources that may be of interest to researchers using SIPP data. (Note that the U.S. Census Bureau has no control over and takes no responsibility for the content of these resources.)

4.6.1 Inter-university Consortium for Political and Social Research (ICPSR)

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.

ICPSR SIPP data are located at https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/135

4.6.2 National Bureau of Economic Research (NBER)

The National Bureau of Economic Research provides downloadable access to all SIPP panels and waves going back to 1984. Data users have the option of downloading data in SAS, SPSS, or STATA preformatted versions.

NBER SIPP data are located at http://www.nber.org/data/survey-of-income-and-program-participation-sipp-data.html

4.7 Research Products

The SIPP website contains a number of research products analyzing SIPP data. These include P-70 reports, conference papers and presentations, table packages, working papers, and a SIPP bibliography. The website is updated on a regular basis to include the most recent and up-to-date research related to SIPP data.

The SIPP research products are located at: https://www.census.gov/programs-surveys/sipp/library/publications.html

4.7.1 P-70 Reports

The P-70 report series was designed to inform the public about the economic status and well-being of American households using data from the SIPP. The P-70 reports cover a wide range of topics, including living conditions, child care arrangements, health insurance coverage, participation in government programs, poverty, income distribution, and employment.

The Census Bureau's P-70 series is the primary source for published estimates from SIPP. These may help data analysts in a number of ways:

- Published estimates may contain estimates needed for the research project at hand, thus saving users the need to generate those estimates themselves.
- 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, which are often impossible for external data users to replicate because the internal files have not been subjected to top-coding and other confidentiality protection techniques.

P-70 series are located at: http://www.census.gov/programs-surveys/sipp/library/publications/sipp-pubs-auto.html

4.7.2 Table Packages

SIPP table packages cover a variety of topics, including the dynamics of poverty, social insurance programs, and disability rates. Most table packages include summary reports of the statistics with tables and graphs to explain the main highlights from the data, other related research reports and presentations, and the complete data tables in Excel format.

The full list of table packages are located at: http://www.census.gov/programs-surveys/sipp/library/publications/tables.html

4.7.3 Conference Papers and Presentations

The SIPP website contains a number of presentations and conference papers related to SIPP. Census Bureau analysts produce a substantial amount of research drawn from SIPP data and are active participants at both national and regional conferences, including the annual meetings of the Population Association of America, the American Economic Association, and the American Sociological Association.

Conference Papers and Presentations are located at: https://www.census.gov/programs-surveys/sipp/about/re-engineered-sipp/sipp-conferences.html

4.7.4 Working Papers

Many researchers and scholars use SIPP data for research about a range of issues related to the well-being and economic status of the nation. The SIPP website makes this research available to other scholars through a series of working papers.

SIPP's series of working papers are located at: http://www.census.gov/programs-surveys/sipp/library/working-papers.html

4.7.5 Technical Documentation – Codebooks and Other Metadata

The technical documentation section of the SIPP website contains a wealth of information to assist analysts. This includes the complete technical documentation for the core and topical modules for each wave within a panel (2008 and before), source and accuracy statements, data dictionaries, questionnaires, and the complete listing of all topical modules by wave within a panel. The technical documentation for the 2014 panel does not distinguish between core and topical modules because SIPP has integrated some topical module content into the 2014 interview going forward. Technical documentation information for the SSA Supplement will be listed separately.

SIPP's technical documentation is located at: https://www.census.gov/programs-surveys/sipp/tech-documentation.html

4.7.6 Source and Accuracy Statements

Source and Accuracy Statements are available for each wave within a panel. These documents summarize the source of data, sampling, estimation procedures, population controls, how to use the weights, possible types of error in SIPP data, and how to calculate different types of standard errors.

The Source and Accuracy Statements are located at: http://www.census.gov/programs-surveys/sipp/tech-documentation/source-accuracy-statements.html

4.7.7 User Notes

User notes for SIPP contain information on possible changes in the data or in the questionnaire. They provide guidance on how to account for any changes that may have occurred. This includes providing documentation and code to correct any possible inconsistencies in the data and any other necessary clarifications.

The SIPP user notes are located on the Wave page within the Data section for each specific dataset.

5 The SIPP Public Use Files

The 2014 SIPP incorporates a completely new data structure compared to past SIPP panels. Previously, SIPP utilized an oscillating survey structure, with a core set of questions that were present in all waves, and a supplemental topical module of questions that varied between waves. The 2014 panel does not feature topical modules; each wave contains a full set of survey questions. Users of data from previous panels should refer to Chapter 9 of the previous SIPP Users' Guide. This chapter provides a general introduction to the SIPP public use files, including:

- Types of data files
- Indicator variables, and how to identify persons and families
- Key income variables
- Identifying program participation
- Selecting weights
- State analysis

5.1 Types of SIPP Files

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

- Wave files are in person-month format with one record per person for each month of the 12-month reference period that the person is in the sample.
- 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.

5.2 Understanding the ID Variables in SIPP

To fully utilize SIPP's longitudinal survey framework, it is necessary to identify sample members across data files.

Although in the past SIPP has employed different ID variables for the different types of public use files, the 2014 panel contains uniform ID variables for the wave and longitudinal files.

5.2.1 Sample Unit IDs (SSUID)

During the initial Wave 1 interview, each sampled dwelling unit is assigned a SSUID identifier, a unique (random) sample unit ID. The same SSUID is assigned to each person within the household and never changes throughout the panel. If an original sample person (OSP) moves to a different address, the person keeps the same SSUID. If new people join an OSP at either the original or a new address, they are assigned the same SSUID as the OSP and become secondary sample members by virtue of their association with the OSP. At the conclusion of the panel, all people who have ever resided at the

originally sampled unit or with a member of that sample unit share the same SSUID. That SSUID is their common link to the original sample unit.

5.2.2 Person Numbers (PNUM)

All sample members are assigned a person number when they first enter the SIPP panel. The person number differentiates persons within the sample unit and does not change throughout the panel. The first digit indicates the wave in which a person joined the SIPP sample, and the second two digits correspond to the person's place on the household roster.

5.3 Identifying Persons

Each person in SIPP can be uniquely identified by the combination of a sample unit ID and a person number. These ID variables are useful when linking the records for a single person across multiple SIPP data files.

Table 5-1: How to Uniquely Identify Persons

Sample Unit ID (SSUID)	Person Number (PNUM)	Notes
123456789123	101	Original sample member
123456789123	102	Original sample member
123456789123	301	Enters SIPP sample in Wave 3
123456789123	401	Enters SIPP sample in Wave 4
123456789123	402	Enters SIPP sample in Wave 4
321456789123	101	Original sample member
321456789123	102	Original sample member
321456789123	201	Enters SIPP sample in Wave 2

5.3.1 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. For many purposes, a household can be thought of as people living at a common address. A person's current address ID for a given month, together with his or her sample unit ID, identifies the household in which that person is a member for that month.

A key factor in the strength of the SIPP longitudinal survey framework is collecting residence history information for all respondents during the reference year and following sample members when they move between waves. The variables necessary for identifying sample members who have moved are shown below.

Table 5-2: Variables for Identifying Movers

Name	Description	Values[5]
SSUID	Sample Unit Identifier: Uniquely identifies each initially sampled dwelling unit. Every person sampled is either an original member of one of those units (an original sample member, or OSP) 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. This means that as people move from address to address, their SSUID stays the same. As new people join the homes of original sample members, they receive the SSUID of the original sample members.	000000000000: 999999999999
PNUM	Person Number: Differentiates persons within the sample unit. Does not change throughout panel. The first digit indicates the wave in which a person joined the SIPP sample; the second two correspond to the person's place on the household roster.	101:499
ERESIDENCEID	Current Address ID: Identifies the specific residence where the person lived during each month. When members report living at or moving to previously un-sampled households, it changes to signify new (spawned) sample households and the wave spawned.	100001:400999

The residence ID (ERESIDENCEID) is a unique six-digit identifier for addresses within original sample units (SSUIDs). While the sample unit identifier and person number (PNUM) values remain the same for the entire panel, movers both within and across waves are identified by a change in the ERESIDENCEID field. The first digit of ERESIDENCEID identifies the first wave in which the address appears in the data.

Addresses entered in Wave 1 start with 1, those entered in Wave 2 start with 2, and so on. The second and third digits identify households that have spawned between waves (Waves 2+ only; this will be 00 for Wave 1). The second digit indicates whether a case is a parent or a child. 0 is the parent; letters A, B, C are the first child, second child, third child, etc. The third digit indicates the same thing but for grandchild cases (spawn of a spawn). The last three digits of ERESIDENCEID represent a sequential numbering of addresses associated with an original sample unit (SSUID) that enter the sample in the same wave.

Note

There are fewer ERESIDENCEID values of 100002 than 100003. Due to an artifact of an earlier version when that slot had to be saved, the instrument rarely assigns 100002.

[5] Values based on 2014 Panel 4 wave schedule

Table 5-3 shows an example household.

Wave 1: Persons 101 (father), 102 (mother), 103 (son), 104 (daughter), and 105 (grandma) are all original sample members.

- Grandma lived at the interview address during the interview month but not during the reference year.
- Grandma lived at two different residences during the reference year but did not live with any of the respondents from the interview address.
- Son moved into the interview address with parents during the month of July.

Table 5-3: Identifying Households and Movers

Person	SSUID	PNUM	Current ERESIDENCEID	1	2	3	4	5	6
Wave 1 Original Home, Parents' Household									
Father	101111103123	101	100001	100001	100001	100001	100001	100001	100001
Mother	101111103123	102	100001	100001	100001	100001	100001	100001	100001
Son	101111103123	103	100001	100005	100005	100005	100005	100005	100005
Daughter	101111103123	104	100001	100001	100001	100001	100001	100001	100001
Grandma	101111103123	105	100001	100004	100004	100004	100004	100004	100004

Person	SSUID	PNUM	Current ERESIDENCEID	7	8	9	10	11	12
Wave 1 Original Home, Parents' Household									
Father	101111103123	101	100001	100001	100001	100001	100001	100001	100001
Mother	101111103123	102	100001	100001	100001	100001	100001	100001	100001
Son	101111103123	103	100001	100001	100001	100001	100001	100001	100001
Daughter	101111103123	104	100001	100001	100001	100001	100001	100001	100001
Grandma	101111103123	105	100001	100004	100004	100003	100003	100003	100003

5.3.2 Determining Monthly Family Composition

The term *family*, as used at the Census Bureau, 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, a grandparent living with a grandchild would be considered a family, as would a pair of cousins, a parent with an adopted child, an aunt with a nephew, or a blended family of married partners with a mix of biological and step children.

SIPP groups related persons into groups via the variable RFAMNUM. This variable will have the same numeric value for everyone in the household that month who are related to the family reference person by birth, marriage, or adoption. If there is a second group of people in the household who are related to each other, but not to the first family, then they will have a value of '2' on RFAMNUM. People who are unrelated to anyone else in the household will have a unique value of RFAMNUM.

5.3.2.1 Rules for Assigning RFAMNUM and RFAMKIND

When grouping sample members into RFAMNUM groups and assigning RFAMKIND, the instrument institutes the following process through the household roster:

1. The edit checks the roster for the first married couple. The edit then looks for all persons related to the first married couple and assigns them all RFAMNUM=1. The edit then moves through the household roster, looking for a second married couple not yet assigned to a family and initiates the same grouping process, grouping members in RFAMNUM=2. The edit continues through the roster, scanning for married couples who are not part of a previously-assigned RFAMNUM group. Each person in the group with a married couple will have the family type variable, RFAMKIND, set to a value of 1 for "married couple".
2. Once all married couples are in families, the edit looks for the first mother-child pair not yet assigned to a family and groups anyone related to that pair (and not already assigned to a family) into a RFAMNUM group. Each person in that group will have the RFAMKIND value of 2, for "female, no spouse present".
3. Once all mother-child pairs are in families, the edit looks for the first father-child pair not yet assigned to a family and groups anyone related to that pair (and not already assigned to a family) into a RFAMNUM group. Each person in that group will have the RFAMKIND value of 3, for "male, no spouse present".
4. Once all father-child pairs are in families, the edit looks for remaining household members not yet assigned to a family. If a person has any relatives not already assigned to a family, they are grouped in the same RFAMNUM group. Those with no relatives are grouped in their own RFAMNUM group. Each person in a family of more than one will have RFAMKIND set to a value of 2 for "female, no spouse present" or 3 for "male, no spouse present" according to the sex of the family reference person.

Table 5-4 provides an example of how household members are assigned to families.

Table 5-4: Identifying a Family

Person	Relationship	SSUID	PNUM	ERESIDENCEID	RFAMNUM	RFAMKIND
John Doe	Self	123456789123	101	100001	1	1
Jane Doe	Spouse	123456789123	102	100001	1	1
Jimmy	Child	123456789123	103	100001	1	1
James	Unrelated	123456789123	104	100001	2	3
Joan Moe	James' Girlfriend	123456789123	105	100001	3	2

5.3.3 Monthly Household Relationships

The 2014 SIPP edit creates an array of recoded variable which denotes the monthly relationship of each household member to every other household member. Previous SIPP panels only showed the relationship of each person to the household reference person. For help analyzing relationships in prior SIPP panels, see the previous SIPP Users' Guide.

SIPP allows for a possible 30 persons on the roster (up to 20 interviewed household members plus up to 10 Type 2 people). Consequently, for each month there are 30 RREL variables (RREL1-RREL30) on each person's file. The instrument creates a matrix denoting the following household relationships.

Table 5-5: Monthly Household Relationship Values

99 = self	10 = half siblings
1 = opposite-sex spouse	11 = step siblings
2 = opposite-sex unmarried partner	12 = adoptive siblings
3 = same-sex spouse	13 = other siblings
4 = same-sex unmarried partner	14 = parent-/child-in-law
5 = biological parent/child	15 = brother-/sister-in-law
6 = stepparent/stepchild	16 = aunt/uncle/niece/nephew
7 = adoptive parent/child	17 = other relative
8 = grandparent/grandchild	18 = foster parent/ child
9 = biological siblings	19 = other nonrelative

The following table shows the matrix values for a household containing a husband, wife, their two adult children, their grandchild, the husband's brother, and the girlfriend of the husband's brother.

Table 5-6: Monthly Household Relationship Matrix

	Husband	Wife	Adult Child #1	Adult Child #2	Grandchild (Adult Child #1's Child)	Husband's Brother	Girlfriend (of Husband's Brother)
Husband	99	1	5	5	8	9	19
Wife	1	99	5	5	8	15	19
Adult Child #1	5	5	99	9	5	16	19
Adult Child #2	5	5	9	99	16	16	19
Grandchild (Adult Child #1's Child)	8	8	5	16	99	17	19
	Husband	Wife	Adult Child #1	Adult Child #2	Grandchild (Adult Child #1's Child)	Husband's Brother	Girlfriend (of Husband's Brother)
Husband's Brother	9	15	16	16	17	99	3
Girlfriend (of Husband's Brother)	19	19	19	19	19	2	99

Note

Due to possible changes in monthly household composition, it is necessary to also reference RREL_PNUM when constructing relationship matrices.

5.3.4 Identifying Parents, Children, Spouses, and Partners at a Monthly Level

RREL and the relationship matrix allow for the identification of various relationships among household members, but a simpler way exists for identifying a person's parent(s), child(ren), spouse, or partner. Four variables (EPNSPOUS_EHC, EPNCOHAB_EHC, EPNPAR1_EHC, and EPNPAR2_EHC) identify the person number of the spouse, person number of the partner, person number of the first parent, and person number of the second parent. In each case, the person is identified only if she or he is living at the same address as the focal person in that month.

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 5-7 displays a household containing a husband, wife, their two adult children, and their grandchild.

Table 5-7: Identifying, Parents, Children, and Spouses

Person	Relationship	PNUM	EPNSPOUS_EHC	EPNPAR1_EHC	EPNPAR2_EHC
John Doe	Self	101	102	9999	9999
Jane Doe	Spouse	102	101	9999	9999
John Doe Jr.	Child	103	9999	101	102
Joan Doe	Child	104	9999	101	102
Jimmy Doe	Joan's Child	105	9999	104	9999

5.3.5 Grouping Families

In an effort to reduce respondent burden, SIPP groups certain related person into "clumps" whereby the household respondent's answer applies to all members of the clump. Thus, in the subsequent interviews of fellow "clump" members, certain questions will not be asked.

The clump is an internally-derived instrument variable. When certain parameters are met, the SIPP instrument limits the number of people who are in universe for the income screener questions and the family- or household-based EHC screener questions (SNAP, TANF, and GA). The intention is to ask only one respondent the family- or household-based program questions when resources were shared during the entire reference period.

Who is included in the screener clump?

- Household respondent
- Spouse of the household respondent
- Children of the household respondent under age 22 who did not have dependents living with them

The clump always includes the household respondent, who is the first person interviewed. If the household respondent resided with a spouse during the entire reference period, then the spouse is included in the clump. Similarly, if the household respondent resided with their children under age 22 who did not have dependents of their own living with them during the entire reference period, then the children are included in the clump.

Who is not included in the screener clump?

- Non-nuclear families
- Parents of household respondents between the ages of 15 and 21

The clump intentionally excludes multigenerational households, unmarried partner households, and other types of non-nuclear family forms because of their complexity. This conservative approach allows for the income dynamics across different family forms to be captured. For example, in a multigenerational household, a daughter and granddaughter might receive TANF, while a father and mother did not. Additionally, the screener clump intentionally excludes the parents of a household reference person between the ages of 15 and 21. This conservative approach ensures the collection of the most accurate information from all adults in the household.

5.3.6 Income Screener Questions

The income screener questions, which are asked in the pre-EHC portion of the instrument, limit the universe for some means-tested programs. When certain parameters are met, the instrument sets the EHC sections for SNAP, TANF, GA, and WIC to not in universe.

The income screener questions ask about:

- Annual income below a specified amount
- Monthly income below a specified amount
- A "just to be sure we didn't miss anything" catchall for programs

The income amounts are based on 200% of the poverty threshold for the number of people – including Type 2 – residing in the sample unit. Annual amounts are rounded to the nearest $1,000, while monthly amounts are rounded to the nearest $500. The same threshold is used for the household respondent of the clump and individuals who are not part of a clump. This conservative approach ensures that resource sharing between household members is captured. For example, in an unmarried partner household with one child, the father may earn more than 200% of the poverty threshold for an individual but not more than 200% of the poverty threshold for three people. In other words, if the household as a whole or one or members could have been eligible for means-tested programs based on income, the intention is to ask about receipt of means-tested programs.

If respondents report having annual income above the specified threshold, they move on to the monthly question. If they report having monthly income above the specified threshold, they move on to the catchall program question ensuring they (or in the case of clump respondents anyone in the clump) did not receive some type of means-tested assistance. If respondents report having income below a specified threshold or having some type of means-tested assistance, they remain in universe for the SNAP, TANF, GA, and WIC questions. Otherwise, if respondents satisfy all three program screener questions, they are set to not in universe for those programs.

5.3.7 Identifying Program Participation in Government Transfer Programs

The SIPP collects data on a number of government transfer programs:

EHC Programs

- General Assistance (GA)
- Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)
- Supplemental Nutrition Assistance Program (SNAP), formerly called food stamps
- Supplemental Security Income (SSI)
- Temporary Assistance for Needy Families (TANF)

Non-EHC Programs

- Disability benefits
- Energy assistance
- Food assistance
- Foster/child/spousal support
- Retirement income
- School breakfast/lunch
- Social Security (adult)
- Social Security (on behalf of a child)
- Survivor benefits
- Transportation assistance
- Unemployment/workers' compensation
- VA benefits

5.3.8 Collecting Information for Program Units

Transfer programs vary in a number of ways that make the collection of benefits a challenge. Some programs are means-tested, requiring applicants' resources to fall below a defined threshold to qualify for benefits. Non-means-tested programs waive income and asset restrictions but may have other eligibility restrictions such as employment or job loss. Some programs are designed to provide benefits to an individual, while others determine eligibility at the family or household level. To reduce respondent burden when asking about means-tested programs that may cover more than just one person, SIPP uses a screener clump and income screener questions to reduce respondent burden. Additionally, when a program may cover more than one person, the use of rosters allows respondents to select any person residing in the household as being covered by the benefits they reported.

5.3.8.1 Rosters

The universes for means-tested programs that may cover more than one person exclude:

- Respondents under age 15
- Respondents in a family clump who are not the household respondents

Respondents who are in universe may select any household member as being covered by the benefit they reported. Household respondents – who by definition are in the clump – may select household members who are a part of the clump, as well as those who are not. Similarly, respondents who are not a part of a clump may select household members who are a part of the clump, as well as those who are not. This ensures any program unit structure can be easily reported.

5.3.9 Identifying Program Participation in Government Transfer Programs

The SIPP collects data on a number of government transfer programs. Participation in the major means-tested programs is asked about in the Event History Calendar (EHC), which collects data at the spell level – or for each set of continuous months a benefit was received. The EHC asks about benefit receipt from:

- Supplemental Security Income (SSI)
- Supplemental Nutrition Assistance Program (SNAP)
- Temporary Assistance for Needy Families (TANF)[6]
- General Assistance (GA)
- Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)

5.3.10 Coverage and Ownership Variables

Two key concepts for understanding how to measure program participation in the SIPP are coverage and ownership. The coverage variable on the public use file that has the same value for the entire reference period is:

- Recode indicating coverage in at least one month of the reference period

The coverage and ownership variables on the public use file that have the same value for each month of a spell include:

- Benefit coverage this month
- Benefit ownership this month

Table 5-8 shows how to identify coverage and ownership for each program. Benefits from SSI cover an individual, while benefits from the other programs (SNAP, TANF, GA, and WIC) cover a "program unit" including all household members who meet the eligibility criteria. Program units may include just one person, more than one person but not all household members, or all household members. It varies based on program rules and household composition. As a result, there are slight variations in the way variables are assigned values between SSI and the other programs.

For SSI, information about spells is available on each monthly record included in a spell for the respondent covered by the benefit. There are two types of coverage indicators. A reference period coverage indicator (which has the same value in each month of the reference period) shows whether a respondent was covered by SSI benefits in one or more months of the reference year (RSSICOV). The variable ESSI_BMONTH, which indicates the first month of a spell, identifies whether a respondent was

[6] The TANF section collects information about the receipt of pass-through child support payments. For more information about this topic, see Chapter 3.13.17 in the Users' Guide.

covered by SSI in a particular month. The value of this variable changes if there are multiple spells and/or if there are months in the reference period not included in a spell. If the respondent is under age 18, ESSI_OWNER points to the owner of the benefit. Unlike other EHC programs, all of the information about SSI spells is available on the record of the child covered by the benefit (see Tables 9 and 10).

Table 5-8: Measuring Coverage and Ownership in Means-Tested Programs in the 2014 SIPP

	Reference Period Variable	Monthly Variables	
	Coverage	Coverage	Ownership
Supplemental Security Income (SSI)			
Benefit owner			
Covered.....	RSSICOV=1	ESSI_BMONTH=1-12[7]	ESSI_OWNER=.N[8]
Not covered.....	RSSICOV=2	ESSI_BMONTH=.N	ESSI_OWNER=.N
Not Benefit Owner			
Covered.....	RSSICOV=1	ESSI_BMONTH=1-12	ESSI_OWNER=PNUM
Not covered.....	RSSICOV=2	ESSI_BMONTH=.N	ESSI_OWNER=.N
Supplemental Nutrition Assistance Program (SNAP)			
Benefit owner			
Covered.....	RFSCOV=1	RFSYN=1	EFSOWN=PNUM
Not covered.....	RFSCOV=2	RFSYN=2	EFSOWN=PNUM
Not Benefit Owner			
Covered.....	RFSCOV=1	RFSYN=1	EFSOWN=PNUM[9]
Not covered.....	RFSCOV=2	RFSYN=2	EFSOWN=.N
Temporary Assistance for Needy Families (TANF)			
Benefit owner			
Covered.....	RTANFCOV=1	RTANFYN=1	ETANFOWN=PNUM
Not covered.....	RTANFCOV=1	RTANFYN=2	ETANFOWN=PNUM
Not Benefit Owner			
Covered.....	RTANFCOV=1	RTANFYN=1	ETANFOWN=PNUM[3]
Not covered.....	RTANFCOV=2	RTANFYN=2	ETANFOWN=.N
	Reference Period Variable	**Monthly Variables**	
	Coverage	**Coverage**	**Ownership**
General Assistance (GA)			
Benefit owner			
Covered.....	RGACOV=1	RGAYN=1	EGAOWN=PNUM
Not covered.....	RGACOV=2	RGAYN=2	EGAOWN=PNUM
Not Benefit Owner			
Covered.....	RGACOV=1	RGAYN=1	EGAOWN=PNUM
Not covered.....	RGACOV=2	RGAYN=2	EGAOWN=.N
Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)			
Benefit owner			
Covered.....	RWICCOV=1	RWICYN=1	EWICOWN=PNUM
Not covered.....	RWICCOV=2	RWICYN=2	EWICOWN=PNUM
Not Benefit Owner			
Covered.....	RWICCOV=1	RWICYN=1	EWICOWN=PNUM
Not covered.....	RWICCOV=2	RWICYN=2	EWICOWN=.N

[7] ESSI_BMONTH refers to the month during the reference year that an SSI spell began. For SSI, it is the indicator that the respondent was covered by SSI benefits in a given month.

For the other EHC programs, a respondent age 15 or older in each program unit is identified as the benefit owner for each spell. The ownership variable is available on the records of the benefit owner and all household respondents covered by the benefit in the months covered by a spell (e.g., EFSOWN in Table 5-8). Detailed information about the spell is located on the record of the benefit owner in the months covered by a spell, regardless of whether this person was covered by the benefit (see Tables 2 and 3). Limited information about the spell is on the records of respondents in the months they were a part of a program unit but not the owner.

Table 5-8 shows the variable names for the two types of coverage indicators. A reference period coverage indicator identifies whether a respondent was covered by the benefit in one or more months of the reference period (e.g., RFSCOV in Table 5-8). The monthly coverage indicators identify which household members were covered by a benefit in a given month (e.g., RFSYN). The value of this variable may change across the reference period.

5.3.11 Detailed Means-Tested Program Participation Variables

Several variables on the public use file provide information about when a program spell began and ended. These variables, which have the same value for each month of a spell, include:

- Begin month of benefit receipt
- End month of benefit receipt
- Year benefit receipt began if receiving in January of the reference period in a wave 1 interview
- Continuation flag for spells that include the last month of the reference period

As Table 5-9 shows, for SSI the months receipt began and ended, the year benefit receipt began, and the continuation flag are available on the monthly records included in a spell for the person receiving the SSI benefit. For SNAP, TANF, GA, and WIC, the months receipt began and ended and the continuation flag are available on the monthly records included in a spell for any household member who was covered by a benefit and/or was the owner of a benefit. The year benefit receipt began is only available on the monthly records included in a spell for the benefit owner.

[8] This information is collected for SSI recipients under the age 18. Once SSI recipients turn age 18 they may receive benefits in their own name.

[9] For people covered by benefits who are not the owner, the ownership variables (e.g. EFSOWN) hold the PNUM value of the benefit owner.

Table 5-9: Identifying Spell Duration in Means-Tested Programs in the 2014 SIPP

	Spell begin month	Spell end month	Continuation flag[10]	Left-censored year[11]
Supplemental Security Income (SSI)				
Benefit owner				
Covered.....	ESSI_BMONTH=1-12	ESSI_EMONTH=1-12	RSSI_CONTFLG=1-3	ESSILCY≥0
Not covered.....	ESSI_BMONTH=.N	ESSI_EMONTH=.N	RSSI_CONTFLG=.N	ESSILCY=.N
Not Benefit Owner				
Covered.....	ESSI_BMONTH=1-12	ESSI_EMONTH=1-12	RSSI_CONTFLG=1-3	ESSILCY≥0
Not covered.....	ESSI_BMONTH=.N	ESSI_EMONTH=.N	RSSI_CONTFLG=.N	ESSILCY=.N
Supplemental Nutrition Assistance Program (SNAP)				
Benefit owner				
Covered.....	EFS_BMONTH=1-12	EFS_EMONTH=1-12	RFS_CONTFLG=1-3	EFSLCY≥0
Not covered.....	EFS_BMONTH=1-12	EFS_EMONTH=1-12	RFS_CONTFLG=1-3	EFSLCY≥0
Not Benefit Owner				
Covered.....	EFS_BMONTH=1-12	EFS_EMONTH=1-12	RFS_CONTFLG=1-3	EFSLCY=.N
Not covered.....	EFS_BMONTH=.N	EFS_EMONTH=.N	RFS_CONTFLG=.N	EFSLCY=.N
Temporary Assistance for Needy Families (TANF)				
Benefit owner				
Covered.....	ETANF_BMONTH=1-12	ETANF_EMONTH=1-12	RTANF_CONTFLG=1-3	ETANFLCY≥0
Not covered.....	ETANF_BMONTH=1-12	ETANF_EMONTH=1-12	RTANF_CONTFLG=1-3	ETANFLCY≥0
Not Benefit Owner				
Covered.....	ETANF_BMONTH=1-12	ETANF_EMONTH=1-12	RTANF_CONTFLG=1-3	ETANFLCY=.N
Not covered.....	ETANF_BMONTH=.N	ETANF_EMONTH=.N	RTANF_CONTFLG=.N	ETANFLCY=.N
General Assistance (GA)				
Benefit owner				
Covered.....	EGA_BMONTH=1-12	EGA_EMONTH=1-12	RGA_CONTFLG=1-3	EGALCY≥0
Not covered.....	EGA_BMONTH=1-12	EGA_EMONTH=1-12	RGA_CONTFLG=1-3	EGALCY≥0
Not Benefit Owner				
Covered.....	EGA_BMONTH=1-12	EGA_EMONTH=1-12	RGA_CONTFLG=1-3	EGALCY=.N
Not covered.....	EGA_BMONTH=.N	EGA_EMONTH=.N	RGA_CONTFLG=.N	EGALCY=.N
Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)				
Benefit owner				
Covered.....	EWIC_BMONTH=1-12	EWIC_EMONTH=1-12	RWIC_CONTFLG=1-3	EWICLCY≥0
Not covered.....	EWIC_BMONTH=1-12	EWIC_EMONTH=1-12	RWIC_CONTFLG=1-3	EWICLCY≥0
Not Benefit Owner				
Covered.....	EWIC_BMONTH=1-12	EWIC_EMONTH=1-12	RWIC_CONTFLG=1-3	EWICLCY=.N
Not covered.....	EWIC_BMONTH=.N	EWIC_EMONTH=.N	RWIC_CONTFLG=.N	EWICLCY=.N

[10] This variable is only available for spells that end in the last month of the reference period (e.g., ESSI_EMONTH=12). Like other detailed spell variables, it is present for each month of the spell.

[11] This variable is only available for spells that include January of the reference period (e.g., ETANF_BMONTH=1) the first time a respondent is interviewed. Like other detailed spell variables, it is present for each month of the spell

Additional detailed program participation variables include:

- Reasons benefit receipt started
- Reasons benefit receipt stopped if receipt ended during the reference year
- Whether the benefit covered only children, only adults, or both (available for TANF and WIC)
- Monthly benefit amount

The reasons benefit receipt started, reasons benefit receipt stopped, and type of benefit coverage (for TANF and WIC) have the same value for each month of a spell, but may change values between spells. For SSI, the reasons benefit receipt started and stopped are located on the monthly records included in a spell for the person covered by the SSI benefit. For other EHC programs, the reasons benefit receipt started and stopped and the type of benefit coverage (for TANF and WIC) are available on the benefit owner's monthly records included in a spell, regardless of whether this respondent was covered by the benefit.

Benefit amount may vary from month to month within a spell. Amount variables contain the monthly payment received by the individual or program unit. For SSI, the amount variable is on the monthly records included in a spell for the person covered by the SSI benefit. When an amount is collected for a program unit (SNAP, TANF, GA, or WIC), the value is placed on the benefit owner's monthly records included in a spell to avoid double counting income when creating family and household-level income recodes (see Table 5-10).

Table 5-10: Identifying Detailed Spell Information for Means-Tested Programs in the 2014 SIPP

	Why began[12]	Why ended[13]	Coverage type	Amount[14]
Supplemental Security Income (SSI)				
Benefit owner				
Covered.....	ESSI_BRSN1=1-4 ESSI_BRSN2=1-4 or N	ESSI_ERSN1=1-7 or N ESSI_ERSN2=1-7 or N	n.a.	TSSI_AMT≥0
Not covered.....	ESSI_BRSN(1,2)=.N	ESSI_ERSN(1,2)=.N	n.a.	TSSI_AMT=.N
Not Benefit Owner				
Covered.....	ESSI_BRSN1=1-4 ESSI_BRSN2=1-4 or N	ESSI_ERSN1=1-7 or N ESSI_ERSN2=1-7 or N	n.a.	TSSI_AMT≥0
Not covered.....	ESSI_BRSN(1,2)=.N	ESSI_ERSN(1,2)=.N	n.a.	TSSI_AMT=.N
	Why began	**Why ended**	**Coverage type**	**Amount**
Supplemental Nutrition Assistance Program (SNAP)				
Benefit owner				
Covered.....	EFSBRSN1=1-9 EFSBRSN2=1-9 or N	EFSERSN1=1-7 or N EFSERSN2=1-7 or N	n.a.	TFS_AMT≥0
Not covered.....	EFSBRSN1=1-9 EFSBRSN2=1-9 or N	EFSERSN1=1-7 or N EFSERSN2=1-7 or N	n.a.	TFS_AMT≥0
Not Benefit Owner				
Covered.....	EFSBRSN(1,2)=.N	EFSERSN(1,2)=.N	n.a.	TFS_AMT=.N
Not covered.....	EFSBRSN(1,2)=.N	EFSERSN(1,2)=.N	n.a.	TFS_AMT=.N
Temporary Assistance for Needy Families (TANF)				
Benefit owner				
Covered.....	ETANFBRSN1=1-9 ETANFBRSN2=1-9 or N	ETANFERSN1=1-9 or N ETANFERSN2=1-9 or N	ETANFADLTKID=1-3	TTANF_AMT≥0
Not covered.....	ETANFBRSN1=1-9 ETANFBRSN2=1-9 or N	ETANFERSN1=1-9 or N ETANFERSN2=1-9 or N	ETANFADLTKID=1-3	TTANF_AMT≥0
Not Benefit Owner				
Covered.....	ETANFBRSN(1,2)=.N	ETANFERSN(1,2)=.N	ETANFADLTKID=.N	TTANF_AMT=.N
Not covered.....	ETANFBRSN(1,2)=.N	ETANFERSN(1,2)=.N	ETANFADLTKID=.N	TTANF_AMT=.N
General Assistance (GA)				
Benefit owner				
Covered.....	EGABRSN1=1-9 EGABRSN2=1-9 or N	EGAERSN1=1-9 or N EGAERSN2=1-9 or N	n.a.	TGA_AMT≥0
Not covered.....	EGABRSN1=1-9 EGABRSN2=1-9 or N	EGAERSN1=1-9 or N EGAERSN2=1-9 or N	n.a.	TGA_AMT≥0
Not Benefit Owner				
Covered.....	EGABRSN(1,2)=.N	EGAERSN(1,2)=.N	n.a.	TGA_AMT=.N
Not covered.....	EGABRSN(1,2)=.N	EGAERSN(1,2)=.N	n.a.	TGA_AMT=.N
Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)				
Benefit owner				
Covered.....	EGABRSN1=1-9 EGABRSN2=1-9 or N	EWICERSN1=1-7 or N EWICERSN2=1-7 or N	EWICCOVTYP=1-3	TWIC_AMT≥0
Not covered.....	EGABRSN1=1-9 EGABRSN2=1-9 or N	EWICERSN1=1-7 or N EWICERSN2=1-7 or N	EWICCOVTYP=1-3	TWIC_AMT≥0
Not Benefit Owner				
Covered.....	EWICBRSN(1,2)=.N	EWICERSN(1,2)=.N	EWICCOVTYP=.N	TWIC_AMT=.N

Not covered.....	EWICBRSN(1,2)=.N	EWICERSN(1,2)=.N	EWICCOVTYP=.N	TWIC_AMT=.N

5.3.12 Means-Tested Program Participation in a Sample Household

Table 5-11 shows a sample household consisting of four related household members. A husband (person 101) and wife (person 102) own the house. Additionally, their adult daughter (person 103) and grandson (person 104) lived with them for the duration of the reference year. Three examples are presented to demonstrate:

- SNAP receipt that covers everyone in the household with a change in payment amount (Table 5-12)
- SSI receipt for an individual (Table 5-13)
- Multiple TANF spells across the reference period with a change in program unit (Table 5-14)

Table 5-11: Sample Household

	Husband	Wife	Adult daughter	Grandson
PNUM	101	102	103	104
TAGE	60	53	20	2

<u>SNAP Example</u>

The entire household was covered by SNAP benefits from January through December of the reference period (Table 5-12). From January through June of the reference period (months 1 to 6), the benefit amount was $340. Due to an increase in household income the benefit amount decreased to $295 for the rest of the reference period (months 7 to 12). The benefit owner was the wife (person 102). As Table 5-12 shows, the benefit amount is only present on the wife's (person 102) monthly records. The other household respondents point to person 102 as the owner (EFS_OWN=102), indicating that the full detailed spell information can be found on that person's record. This example includes a subset of SNAP variables (see Tables 5-8 to 5-10 for a full listing). Note that the amount variable is structured similarly to other detailed spell variables in terms of whose records are in universe and have values (see Table 5-14 for an example using the full set of spell variables).

[12] For each reported spell, respondents are asked why benefit receipt began. Respondents select reasons from a mark all that apply list. Since most respondents select a single reason, the public use file only has two variables that correspond to the first two reasons reported (when only one reason is reported the respondent has a value of .N for the second reason).

[13] For each reported spell that ends during the reference period, respondents are asked why benefit receipt stopped. Respondents select reasons from a mark all that apply list. Since most respondents select a single reason, the public use file only has two variables that correspond to the first two reasons reported (when only one reason is reported the respondent has a value of .N for the second reason).

[14] When program benefits cover multiple household members, the benefit amount is only recorded on the owner's record to avoid double counting amounts when calculating family and household-level income.

Table 5-12: Identifying Monthly Program Participation: SNAP

Month	Variables					
	RFSCOV[15]	EFS_ BMONTH	EFS_ EMONTH	RFSYN	EFS_OWN	TFS_AMT
Husband (101)						
1	1	1	12	1	102	.N
2	1	1	12	1	102	.N
3	1	1	12	1	102	.N
4	1	1	12	1	102	.N
5	1	1	12	1	102	.N
6	1	1	12	1	102	.N
7	1	1	12	1	102	.N
8	1	1	12	1	102	.N
9	1	1	12	1	102	.N
10	1	1	12	1	102	.N
11	1	1	12	1	102	.N
12	1	1	12	1	102	.N
Wife (102)						
1	1	1	12	1	102	340
2	1	1	12	1	102	340
3	1	1	12	1	102	340
4	1	1	12	1	102	340
5	1	1	12	1	102	340
6	1	1	12	1	102	340
7	1	1	12	1	102	295
8	1	1	12	1	102	295
9	1	1	12	1	102	295
10	1	1	12	1	102	295
11	1	1	12	1	102	295
12	1	1	12	1	102	295
Adult Daughter (103)						
1	1	1	12	1	102	.N
2	1	1	12	1	102	.N
3	1	1	12	1	102	.N
4	1	1	12	1	102	.N
5	1	1	12	1	102	.N
6	1	1	12	1	102	.N
7	1	1	12	1	102	.N
8	1	1	12	1	102	.N
9	1	1	12	1	102	.N
10	1	1	12	1	102	.N
11	1	1	12	1	102	.N
12	1	1	12	1	102	.N
Grandson (104)						
1	1	1	12	1	102	.N
2	1	1	12	1	102	.N
3	1	1	12	1	102	.N
4	1	1	12	1	102	.N

[15] RFSCOV has the same value in each month because it is a reference period indicator.

Month	Variables					
	RFSCOV[15]	EFS_ BMONTH	EFS_ EMONTH	RFSYN	EFS_OWN	TFS_AMT
5	1	1	12	1	102	.N
6	1	1	12	1	102	.N
7	1	1	12	1	102	.N
8	1	1	12	1	102	.N
9	1	1	12	1	102	.N
10	1	1	12	1	102	.N
11	1	1	12	1	102	.N
12	1	1	12	1	102	.N

SSI Example

The wife (person 102) reported receiving $700 in SSI benefits during each month of the reference period (Table 5-13). Since SSI provides benefits to individuals, the example only includes the records for person 102. This example includes a subset of SSI variables (see Tables 5-8 to 5-10 for a full listing).

Table 5-13: Identifying Monthly Program Participation-SSI

Month	Variables				
	RSSICOV[16]	ESSI_ BMONTH	ESSI_ EMONTH	ESSI_OWNER[17]	TSSI_AMT
Wife (102)					
1	1	1	12	.N	700
2	1	1	12	.N	700
3	1	1	12	.N	700
4	1	1	12	.N	700
5	1	1	12	.N	700
6	1	1	12	.N	700
7	1	1	12	.N	700
8	1	1	12	.N	700
9	1	1	12	.N	700
10	1	1	12	.N	700
11	1	1	12	.N	700
12	1	1	12	.N	700

TANF Example

The adult daughter (person 103) reported receiving TANF benefits in January through April of the reference period that covered her and her son (months 1 to 4), but stopped receiving benefits for a few months (months 5 to 8) (Table 5-14). In September, she reported receiving TANF benefits that only covered her son (person 104) through the remainder of the reference period (months 9 to 12). Detailed information about each spell, including the benefit amount, is available on the benefit owner's record (person 103). Since the adult daughter (person 103) and grandson (person 104) are the only people covered by the TANF benefit, this example only includes the records for these two household members.

[16] RSSICOV has the same value in each month because it is a reference period indicator.

[17] ESSI_OWNER only has values for respondents under age 18.

Table 5-14: Identifying Monthly Program Participation-TANF

Month	Variables										
	RTANF COV[18]	ETANF_ BMONTH	ETANF_ EMONTH	RTANFYN	ETANF_ OWN	TTANF_ AMT	ETANF LCY	RTANF_ CONTFLG	ETANF ADLTKID	ETANF BRSN1[19]	ETANF ERSN1[20]
Adult Daughter (103)											
1	1	1	4	1	103	190	2012	.N	2	4	6
2	1	1	4	1	103	190	2012	.N	2	4	6
3	1	1	4	1	103	190	2012	.N	2	4	6
4	1	1	4	1	103	190	2012	.N	2	4	6
5	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
6	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
7	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
8	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
9	1	9	12	2	103	90	.N	3	1	9	.N
10	1	9	12	2	103	90	.N	3	1	9	.N
11	1	9	12	2	103	90	.N	3	1	9	.N
12	1	9	12	2	103	90	.N	3	1	9	.N
Grandson (104)											
1	1	1	4	1	103	.N	.N	.N	.N	.N	.N
2	1	1	4	1	103	.N	.N	.N	.N	.N	.N
3	1	1	4	1	103	.N	.N	.N	.N	.N	.N
4	1	1	4	1	103	.N	.N	.N	.N	.N	.N
5	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
6	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
7	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
8	1	.N	.N	2	.N	.N	.N	.N	.N	.N	.N
9	1	9	12	1	103	.N	.N	3	.N	.N	.N
10	1	9	12	1	103	.N	.N	3	.N	.N	.N
11	1	9	12	1	103	.N	.N	3	.N	.N	.N
12	1	9	12	1	103	.N	.N	3	.N	.N	.N

[18] RTANFCOV has the same value in each month because it is a reference period indicator.

[19] Up to two reasons for receipt beginning are available on the public use file (see Table 5-10).

[20] Up to two reasons for receipt ending are available on the public use file (see Table 5-10).

6 Data Editing and Imputation

This section 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
- Statistical (or stochastic) imputation is used for some types of unit nonresponse and some types of item nonresponse.

This 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.

6.1 Data Editing

All respondent reported data is edited for logical conformity. For example, the respondent reported an age of 67 years old. However, the respondent is enrolled in 10th grade and is the biological child of a 47 years old household member. This is a simple example of when data editing would correct the reported age. Variables starting with an E are the edited variable.

During processing, there are also variables created with values based on the value of one or more other variables. These are recoded variables and start with an R.

6.1.1 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 (individual household member) occurs in SIPP when one or more of the people residing at a sample address are not interviewed and no proxy interview is obtained. Most types of unit nonresponse are dealt with through weighting adjustments (see Chapter 7).

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

Item nonresponse data are generally imputed for all topics.

6.1.2 Goals of Imputation

There are two key problems caused by missing data:

- 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
- Nonresponse is unlikely to be completely random, so estimates of population parameters are biased due to a potential non-representative sample

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. 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.

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.

6.1.3 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 comprehensive discussion of the treatment of imputed data in sample surveys is in Little and Rubin's 2014 book *Statistical analysis with missing data.*

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. Also, as the percentage of eligible sample members reinterviewed decreases, the pool from which donors 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.

6.1.4 Processing SIPP Data

At the conclusion of each wave of interviewing, the data collected during that wave are processed.

SUMMARY OF PROCESSING

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

1. The data collected during each wave are edited and imputed by topic.
2. A public use version of the wave data is then created from the resulting internal wave data. The public use file is the same as the Census Bureau's internal file except that it has certain information suppressed or top-coded to protect the confidentiality of survey respondents (see sections on Top-coding and Suppression of Geographic Information later in this chapter).
3. These steps are repeated at the conclusion of each wave of interviews.

6.2 Imputation

SIPP uses three imputation strategies:

1. Model-Based Imputation
2. Hot-deck Imputation
3. Cold-deck Imputation

Model-based imputation creates topic flags that determine whether a respondent should answer questions about a specific content area (e.g. Social Security or TANF). All other variables are imputed using hot-deck and cold-deck imputations.

6.2.1 Model-Based Imputation

For situations where a respondent has not provided information for an entire topic, SIPP uses a statistical model-based imputation process to predict whether the respondent should have reported data for that topic. The output of this prediction is a Y/N topic flag, where Y indicates that there should be data related to the topic.

The imputation process developed to handle missing topic flags (explained below) involves estimating sequential models that predict values for a given topic flag, conditional on demographic data, all other topic flags, and IRS and SSA administrative data on earnings and benefits. These data are non-monotonically missing, meaning the topic flags cannot be ordered from least missing data to most missing data. Rather, the data look like “Swiss cheese,” with pockets of missing data scattered throughout. Because of this feature, we iterate our process multiple times, estimating the sequential models repeatedly and always conditioning on the most up-to-date imputations for any explanatory variable in a model. This process is commonly called Sequential Regression Multiple Imputation (SRMI) and was initially developed by Raghunathan et al. (2001).

SRMI is based on Bayesian statistical theory and is a method of estimating an approximation of the posterior predictive distribution (PPD), a conditional probability distribution that describes the data generation process and captures the relationships between the variables. To impute missing values, we

take a draw from this PPD; in other words, we predict a value for the missing topic flag conditional on everything else we observe or have previously imputed for that record.

This modeling method has several advantages over the previously used method of hot-deck. We can include many more explanatory variables in the models than can be included as stratifiers in a hot-deck. This means we can condition the imputation for a given topic flag on the imputed values for every other topic flag, hopefully approximating a joint distribution of values instead of a series of independent imputations. Parent and spouse variables can also be used as regressors or conditioning variables in the models, which allows us to better preserve the relationships among the topic flags of household members. Finally, non-SIPP data can be used to mitigate the problem of respondents with missing values being different in unobservable ways from respondents with non-missing values.

6.2.1.1 Creating Topic Flags

The first step in creating a topic flag is to determine which respondents are in universe to be asked the screener questions for that topic. For some topics, the universe is determined strictly by age (for example, job topics, education enrollment, SSI, and unemployment compensation). For other topics, gender and presence of children are also universe determinants (for example WIC, school meals, Social Security for kids). Another group of topics is only asked of household respondents and the answers covered a group of individuals linked through family relationships to that household respondent (for example TANF, general assistance, and SNAP). Private health insurance is the most complicated topic with respondents allowed to choose to report coverage for themselves, for a family member, or both. The creation of the private health insurance topic flag requires a complete accounting and reconciliation of reports from all respondents within a sampling unit to create an accurate person-level indicator of coverage.

The key element in this process is determining which respondents are missing data needed to create the topic flag and which respondents were not or should not have been asked the questions about a particular topic. If the universe for a topic flag is not correctly determined, we would potentially have too much data that appears to be missing and we would impute values for people that should not have them. Hence, we make great efforts to specify the universe exactly as it was implemented in the survey instrument.

6.2.1.2 Building a Topic Flag Model

The first step in the model for each topic flag is the identification of important related variables for each topic flag. These variables fall into two categories, stratifying and regressor variables. A stratifying variable is an indicator variable that divides respondents into homogeneous groups where each group might be expected to have different models. A regressor variable is any variable that is potentially related to the topic flag and could go on the right hand side of a regression. For each group defined by the stratifying variables, the topic flag is regressed on the regressor variables for the cases where the topic flag is not missing. The resulting regression coefficients are then used to predict the topic flag for the missing cases. Once a particular topic flag has been imputed, it can be used as a predictor in the model for another topic flag, entering into the regression for the non-missing cases, and used to impute

values for the missing cases. This process is how dependency between topic flags is built into the models. We rely on multiple iterations of our process to prevent the order of the topic flag modeling from affecting the imputations. The final result is a set of topic flags with no missing values for every in-universe SIPP respondent because all the originally missing values have been replaced by imputed values.

6.2.1.3 Choosing a list of stratifying variables

A stratifying variable should be a categorical characteristic(s) that best predicts the presence or absence of a topic. We create an optimal set of stratifying variables and then alternative lists in case the cell sizes created by the optimal list are too small for regressions. Any sub-sample created by the stratifying variables that does not contain at least 100 observations will not be used for regressions. Instead, it will be combined with other sub-samples that are too small and divided again using the next smallest set of stratifying variables.

6.2.1.4 Choosing a list of regressor variables

Regressor variables are the additional variables used to predict the presence of a topic. Regressors may be binary, categorical, or continuous. Continuous variables are particularly useful on the regressor list since too many dummy variables in the regression can cause estimation problems in logistic regressions. For example, age categories are excellent as stratifying variables, but as regressors, it is preferable to use age, age squared, age cubed rather than a series of dummy variables defining the age category. When variables are dropped from the optimal stratifying list, we generally include them on the regressor list so that they can still provide predictive value to the modeling process even though they are no longer a stratifying variable.

6.2.1.5 Administrative Records in Model-Based Imputation

The model-based imputation makes use of six different sources of data shared with the Census Bureau by the Social Security Administration (SSA) for stratifying and regressor variables in the models. Once the topic flags are imputed, all administrative records are stripped from the data and no administrative records are on public-use files. We use two types of earnings records derived from W-2 forms filed with SSA by employers. The Detailed Earnings Record (DER) extract reports uncapped income-taxable earnings for each employer that filed a W-2 record from 1978-2012. It also contains a report of earnings that were not income taxable and were deferred into accounts like 401(k) plans. We utilize the DER to create a measure of total earnings in a given year and to count the number of jobs an individual held. From the DER we also create measures of self-employed earnings and an indicator of any deferred earnings. We also utilize the Summary Earnings Record (SER) extract which contains total earnings capped at the FICA taxable maximum from 1951-2012 to create a count of how many years an individual has worked over his or her lifetime.

We use the Master Beneficiary Record (MBR) and Payment History Update System (PHUS) extracts to create indicators for whether an individual was eligible for and received OASDI payments due to retirement, disability, spouse retirement or death, parent retirement or death, or some combination of

reasons. These extracts contain both present benefit receipt and historical information so we are able to tell what year an individual started receiving benefits and whether they ever stopped. The Supplemental Security Record (SSR) provides the same information about SSI benefits. These three files combined together give us a very accurate picture of who was receiving OASDI benefits, SSI benefits, or both. This information in turn is very helpful in predicting reports of OASDI and SSI receipt.

We make use of the Numident, a register of all Social Security Numbers (SSNs) ever issued in the United States, along with the MBR and SSR, as an administrative source of birth date information. If a person is receiving benefits we utilize the birth date from the benefits files in order to create an age for the individual during the survey reference period. If the person is not receiving benefits, we use the birth date from the Numident. While this does not replace the survey reported age on the final public use data, we do use this age derived from administrative data as an explanatory variable in our models.

Some SIPP respondents do not match to administrative data. This happened because they refused to consent to having their data linked or they did not provide enough information for the Census PVS system to find a Protected Identity Key (PIK) for them. The PVS system relies on matching name, date of birth, gender, and address to administrative data files to find an SSN which is then replaced with a PIK. When administrative data is missing, it is imputed during same model-based imputation process that imputes the topic flags.

6.2.2 Sequential Hot-Deck Imputation

The statistical imputation method used to impute missing items is known as a sequential hot-deck procedure. While multiple imputation determines if a respondent with missing data should have data for a topic (e.g. unemployment insurance), hot-deck imputation determines the value (e.g. monthly unemployment insurance benefit) for a specific amount associated with the topic. In some cases a ratio is imputed. This ratio is used to derive the value instead of imputing the value itself. This is done to preserve relationships between certain variables (e.g., asset value and income).

In a general sense, the sequential hot-deck 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 non-missing data from the same case.

The hot-deck is cross-sectional; only values from current wave responses are used in the definition of the hot-deck cells. SIPP hot-deck procedures are designed to preserve the univariate distribution of each variable subjected to imputation. However, they do not generally preserve the covariances among variables. 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.

The hot-deck procedure used in SIPP for questions is sequential because the selection of replacement values is implemented one record at a time from an ordered file.

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

1. Specifying cold-deck values
2. Sorting the sample cases
3. Identifying records with no item nonresponse or topic flag set and creating 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 non-missing for all cases (this step is omitted in the initial processing of the key demographic items: race, gender, etc.)
5. Selecting replacement values from donor cases to impute item-missing data

6.2.2.1 Sorting the Sample Cases

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

6.2.2.2 Identifying records with no item nonresponse or topic flag set and creating hot-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 yet imputed, the first record in the sorted file with consistent and non-missing 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 non-missing information update the previous set of consistent and non-missing 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.

6.2.2.3 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.

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 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 defines 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).

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

6.2.2.4 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 non-missing from a record in that imputation class. If the value of an item in the current record is non-missing, 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 non-missing case.

The updating is done item by item. Missing items in one record receive the current set of replacement values. Then the non-missing 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.

6.2.3 Specifying cold-deck or initial donor values

Cold-deck values are the values for which each cell in the hot-deck matrix is initialized. We never intend for an imputed item to receive the cold-deck value, but it is there as a last resort, if we cannot assign a

value via logical imputation or the hot-deck imputation process. The cold-deck value is usually the most commonly reported value for that particular item.

6.3 Status Flags (Imputation Method)

Prior to the 2014 panel, these flags were referred to as allocation (imputation) flags. These flags will continue to start with an "A" just like prior panels. For example, the status flag for the ERACE variable is ARACE. A status flag is associated with each item subject to imputation. When an item has been imputed, a status flag for that item is set.

For the 2014 panel, the status flags contain 10 possible values. A status flag with a value of 0 now indicates that the item is "Not in universe", which means that item did not meet the criteria to get data.

Table 6-1: Imputation Status Flag Values

Status Flag Value	Meaning	Description
0	Not in universe	The item is not imputed and that no data are reported by the respondents.
1	In universe, as reported	The item is not imputed and that data are reported by the respondents.
2	Statistical imputation (hot-deck)	The item is imputed a value from hot-deck imputation.
3	Logical imputation	The item is imputed logically.
4	Model-based imputation	The item has gone through a model-based imputation procedure.
5	Cold-deck value imputation	The item is imputed to a cold-deck value. This cold-deck value is what each cell in the hot-deck matrix gets initialized to. It is the most common reported value a respondent would be likely to report for that particular cell. So if an item could not get a hot-deck value, this initialized value is used to set the value for that item.
6	Imputed from a range	The item is imputed based on a range.
7	Combination of 1 and 2/3/5/6	The item is imputed using reported value and any combination of imputation method values 2, 3, 5, and 6 (described above).
8	Combination of 2/3/5/6	The item is imputed using any combination of imputation method value 2, 3, 5, and 6 (described above).
9	Can be determined from the allocation flags for the components of this recode	Used for recodes that are entirely made up of variables on the public use file.

6.3.1 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: top-

coding of selected variables (income, assets, and age) and suppression of geographic information. Because of these procedures, estimates based on data from the public use files will differ slightly from the Census Bureau's published estimates.

Top-coding

One piece of information that might reveal a respondent's identity is very high income or assets. For that reason, the Census Bureau top-codes these variables before making that information publicly available. This is done in different ways for different variables. For incomes and amounts with a very skewed distribution, such as annual salary or home values, the amount is replaced by an average of top-coded values[21]. For other income variables, such as program income where a threshold is determined by institutional rules, amounts over a certain maximum value are recoded to that maximum. In other words, income on the public use data files has a ceiling value for these variables. As in the past, summary income variables for persons, families, and households are the sum of the component variables after they have been topcoded. Although income is the primary variable that is top-coded, other variables that may disclose a respondent's identity, such as age, are also top-coded. Others, such as occupation, may be aggregated into broader categories. A few variables, such as starting dates for employment, may be bottom-coded if they pose a disclosure risk.

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, individual metropolitan areas and specific non-metropolitan areas (such as counties outside of metropolitan areas) are never identified. SIPP does identify metropolitan and nonmetropolitan status for respondents living in states where both the metropolitan and non-metropolitan populations are over 250,000 or states where a state's metropolitan or non-metropolitan population is 0. Respondents living in states that do not fit these criteria are coded as non-identified on metropolitan status. For some states, SIPP also identifies metropolitan principal city, metropolitan remainder, micropolitan, and nonmetropolitan status within metropolitan/nonmetropolitan areas, provided each of these areas meets the 250,000 threshold. If one of these areas does not meet the 250,000 population threshold, respondents will be coded as not identified, but the respondent will still have a value for metropolitan/nonmetropolitan status, provided that the population is over 250,000.

[21] For hourly wages, the median above the threshold is output on the data file instead of the mean. For most variables using this methodology, the mean output is the within-cell average based on person-characteristics. More information about these characteristics and the distributions above the top-coding threshold, such as mean, median and variance, is available. For specific variables and waves, see the SIPP webpage user notes, or the labor force or assets web pages.

7 Nonsampling Error, Sampling Error and Weighting

Statistics from surveys are subject to both nonsampling and sampling errors. This chapter addresses:

- Sources of nonsampling error and the effect of nonsampling error on estimates
- Sampling error and incorporating SIPP's complex sampling design when computing variance estimates
- SIPP weighting, available weights, and using weights to produce reliable estimates

For further information on the source of the data and accuracy of the estimates, including standard errors and confidence intervals, see the SIPP Source and Accuracy Statements.

7.1 Nonsampling Error

All surveys are subject to non-sampling errors. SIPP contains non-sampling errors common to most surveys, as well as errors unique to SIPP's longitudinal design. Under-coverage in household surveys is due primarily to within-household omissions; the omission of entire households is less frequent. SIPP experiences some differential under-coverage of demographic subgroups; for example, the coverage ratio[22] of black males ages 15 to 64 years is lower than that for white males in the same age group. To compensate for this differential under-coverage, the Census Bureau adjusts SIPP sample weights to population control totals. However, it is not certain to what extent those adjustments eliminate biases.

Sample attrition (when respondents leave the sample) is another source of error. Attrition reduces the available sample size and to the extent that those leaving the sample are systematically different from those who remain, survey estimates could be biased.

Response errors in SIPP take on a number of forms. Many recall errors are thought to be the source of seam bias. This effect results from the respondent's tendency to project current circumstances back onto each of the months that constitute the SIPP reference period (4 months for 2008 and earlier panels, and the prior calendar year for the 2014 panel). This causes any changes in respondent circumstances that occurred during the reference period appear to have happened in the first month of the reference period. The effect is a disproportionate number of changes that appear to occur between the last 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

[22] The coverage ratio is the estimated population before the post-stratification ratio (second stage) adjustment divided by the independent population control.

alter their responses to the survey questions in an effort to conceal sensitive information or to minimize the length of the interview.

Additionally, measurement error can occur when respondents misinterpret the questions.

7.1.1 Effects of Nonsampling Error on Survey Estimates

A considerable amount of research has been conducted to investigate the various sources of non-sampling error in SIPP. The results of the research are summarized in the SIPP Quality Profile, 3rd edition. Additional findings about SIPP data quality, especially for more recent panels, are summarized by John Czajka in Appendix A of the National Research Council's 2009 report Reengineering the Survey of Income and Program Participation. Despite the volume of methodological research, it remains difficult to quantify the combined effects of non-sampling errors on SIPP estimates. This problem is made more complex because the effects of different types of non-sampling error on survey estimates vary, depending on the estimate under consideration. However, there are some findings about non-sampling 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 under-coverage 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. However, those procedures may not fully correct for all potential biases (SIPP Quality Profile, 3rd Ed., Chapter 8).
- 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).
- SIPP estimates of the number of births compare favorably with CPS estimates. Both surveys, however, provide estimates that are low relative to records from the National Center for Health Statistics (NCHS). SIPP estimates of the number of marriages are fairly comparable with NCHS counts, but SIPP estimates of the number of divorces are consistently lower than NCHS estimates (SIPP Quality Profile, 3rd Ed., Chapter 10).
- Across all age groups, particularly children and the elderly, SIPP continues to identify more sources of family income than CPS. SIPP's greater effectiveness than CPS in capturing income from multiple sources among retired workers demonstrates an important way in which SIPP appears to provide a better tool for policy analysis (Czajka et al., 2008).
- In 2005, SIPP captured a higher share of aggregate annual benefits than CPS for Food Stamps, AFDC/TANF, OASI, and SSI, but was only marginally better for SSDI. In 1987, SIPP was on par with CPS for AFDC/TANF and SSDI. Whether because of poor recall or because respondents sometimes answer on the basis of their current situation, CPS estimates of persons who ever participated in a program sometimes line up with SIPP estimates of average monthly participants (Czajka, 2009).
- When compared to the Survey of Consumer Finances (SCF) by the Federal Reserve Board for late 1998 and early 1999, SIPP is much more effective in capturing liabilities than assets. SIPP's estimate of aggregate assets was 55 percent of the SCF estimate of $34.1 trillion, but its estimate of aggregate liabilities was 90 percent of the SCF estimate of $5.0 trillion (Czajka, 2009).

- Average monthly estimates of health insurance coverage from SIPP compare closely to estimates of health insurance coverage obtained in the National Health Interview Survey (NHIS) and CPS (Czajka, 2009).
- When examining the use of housing unit controls versus population controls, a team at the Census Bureau concluded that the weighting adjustment for within-household under-coverage when using population controls by age, sex, and race tended to be higher than the weighting adjustment for housing unit coverage when using housing unit controls, which focus on coverage of housing units (including whole households). Thus, when population-control-based weights are applied to characteristics such as household relationship, the estimate of householders (family plus nonfamily) will almost always be higher than the corresponding housing-unit-control-based weights that are applied to obtain the estimate of occupied housing units (Cresce et al., 2013).
- There seems to be evidence of potential bias due to nonresponse for some key statistics in SIPP. The Census Bureau has done nonresponse bias studies to investigate the effect of decreasing response rates, but more work needs to be done to truly quantify that bias (McMillan & Culver, 2013).

7.2 Sampling Error

This section discusses methods for obtaining the sampling error estimates derived from the SIPP panels. The sample selected for each SIPP panel is a stratified multistage probability sample. This complex sample design must be taken into account when calculating 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. Finally, sets of replicate weights are also provided in the SIPP data files and can be used to estimate more accurate standard errors and variances for SIPP estimates. While replicate weighting methods require more computing resources, many statistical software packages have procedures that simplify the use of replicate weights for users.

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 analysis assume simple random sampling for variance estimation. When applied to SIPP estimates, SRS formulas for variances typically underestimate the true variances. This section describes how appropriate variance estimates, which take into account the complex sample design, can be obtained for SIPP estimates.

7.2.1 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, which are sampled with probability proportional to size within strata. Some PSUs, called self-representing (SR) PSUs, are so large that they are included in the sample with certainty. Because no sampling is involved, the SR PSUs are, in

fact, not PSUs but strata. Smaller PSUs, called non-self-representing (NSR) PSUs, are stratified, and two NSR PSUs per stratum are sampled without replacement, so that no PSU is selected more than once for the sample.

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.

7.2.2 Variance Units and Variance Strata, 1990–2014 Panels

For the 1990–2014 panels, sample member records contain data on the PSU and stratum from which a person was sampled. Software packages use this information to estimate variance. However, to avoid potential identification of sampled persons in small areas, the original PSU and stratum codes are not included in the SIPP public data files. 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. However, the variance estimates are somewhat less precise than those obtained from using PSUs and strata that have not been combined.

7.2.3 Replicate Weights for the 1996-2014 Panels

Analysts should use Fay's modified balanced repeated replication (BRR) method for estimating variances for the SIPP panels. 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

(Figure 7-1)

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

where

$G =.N$umber of replicates;

$1 - k =$ perturbation factor;

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

$\theta_i = i^{th}$estimate of the parameter θ based on the observations included in the ith replicate;

$\theta_0 =$ survey estimate of the parameter θ based on the full sample.

The 1996 and 2001 SIPP panels use 108 replicate weights. The SIPP 2004 and 2008 panels use 120 and the SIPP 2014 panel uses 240 replicate weights. All replicate weights are calculated based on a perturbation factor of 0.5 (k = 0.5). For example, inserting the 2014 panel values into Equation (7-1) results in the variance formula of

$$Var(\theta_0) = \{1/(240 * 0.5^2)\} \sum_{i=1}^{240} (\theta_i - \theta_0)^2.$$

The Census Bureau uses VPLX and SAS software to compute the replicate weights that are available through DataFerrett, the SIPP FTP Site, and the SIPP website.

Any analysis must account for the complex survey design and sampling structure. Below is an example of SAS code that can produce unbiased state specific estimates of personal monthly income (note use of the DOMAIN statement as opposed to a WHERE statement; this more accurately accounts for the population subsampling and thus produces more appropriate variance estimates):

```
proc surveymeans data=l08puw1 varmethod = BRR (Fay=0.5) mean cv sum cvsum var T;
var RPTOTINC;
weight WPFINWGT;
repweights REPWGT1-REPWGT240;
Domain TFIPSST;  /*  For all states */
run;
```

7.2.4 Approximate Variance Estimates

The Census Bureau provides three forms for approximate variance estimation: GVFs, tables of standard errors (the square root of the variance) for different estimates, and design effects (DEFF) for person and household level estimates in multiple domains. 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

(Figure 7-2)

$$s = (ax^2 + bx)^{1/2},$$

where s represents the standard error and x represents 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 SIPP Source and Accuracy Statements that accompany the SIPP data files for each 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 design effects and summary tables of general standard errors. These estimates are also available in the Source and Accuracy Statements (more details and examples for how to use these methods are given in the SIPP Source and Accuracy Statements).

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 estimated totals with less than 10 million people, 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.

7.2.5 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 (in earlier panels, intermittent responses bounded by two responding waves). Imputation fills in gaps in the data set, making data analysis easier, and allowing more people to be retained as panel members for longitudinal analysis. 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. It is important to recognize this fact when sizable proportions of values are imputed.

7.3 Weighting

Person weights estimate the number of people in the target population that a person represents. In general, since population units may be sampled with different selection probabilities and since response rates and coverage rates may 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.

SIPP weights vary due to differential sampling rates as a result of oversampling and because response and coverage rates vary across subpopulations. For example, in Wave 1 of the 2014 panel, the final person lower quartile weight is 2,600 and the upper quartile weight is 5,800 (the maximum weight is 16,100). A respondent with a final person weight of 2,600 represents 2,600 people in the U.S. population for the reference month, whereas a respondent with a weight of 5,800 represents 5,800 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.

7.3.1 Choosing a Weight

The decision of which weight to use for a given analysis depends on the population and duration of interest for that analysis. Useful guidance for choosing the correct set of weights is to consider 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);
- Year (e.g., the calendar year weights in the longitudinal files); and
- Panel (e.g., the panel weight(s) in the longitudinal files).

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;
- A longitudinal sample that provides monthly data over the life of a panel

Analysts can use longitudinal samples to follow the same people over time and hence study issues such 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 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 post-stratification adjustments to make the weighted sample totals conform to population totals for key variables.

Table 7-1: Choosing Appropriate Weights

Time Duration	Person Level	Example
Monthly	WPFINWGT	Income in July, 2013
Calendar Year	WPFINWGT in December of the reference year	Total TANF receipt in 2013
Panel	FINPNL# where # = 2, 3, 4 (For Wave 1, the Panel and Calendar Year weight is the same as Dec. 2013)	Median duration spell of unemployment from June 2013 to June 2016

7.3.2 How 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 post-stratification (second-stage calibration) adjustment to correct for departures from known population totals.

7.3.3 Reference Month Weights

Reference month final weights are provided on the SIPP wave files for persons. The weights for persons are constructed first. The household, family, and 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 month. Positive weights are computed for all interviewed persons in the SIPP universe for the specified month.

A number of factors lead to fluctuations in sample size (and therefore weights) from month to month. They include births, deaths, and 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. In Wave 1, the weight for each sample person per month is a product of four components:

- Wave 1 base weight - This weight is the inverse of the probability of a sample person's address being selected.

- Weighting-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

- Wave 1 nonresponse adjustment -This adjustment compensates for different rates of household noninterview within adjustment classes. 512 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); and household size (1, 2, 3, 4+ people). In addition, the within-primary-sampling-unit poverty stratum (poverty, non-poverty) was added starting with the 1996 panel.

- 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, state, and household type. A husband/wife equalization is also included. A raking procedure is used to ensure that the weights agree with all the control totals included for calibration.

In subsequent waves, each person receives an initial weight that is carried over from Wave 1. This weight is adjusted to compensate for changes in the sample between waves resulting from movers and nonresponse. Then it is realigned to match the population totals for the reference or interview month:

- Wave 2+ initial weight -This weight is the product of the Wave 1 base weight, the weighting-control factor, and the Wave 1 nonresponse adjustment factor. It is for each original sample person who is a reference person or is living 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 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. Some adjustment classes are combined to ensure that the adjustment for each class does not exceed a factor of 4 in the 2014 SIPP 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.

7.3.4 Panel and Calendar Year Weights

Panel and calendar year weights are provided on the longitudinal data files for eligible sample members. For earlier panels, 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. Starting with the 2001 panel, sets of both panel weights and calendar weights were produced at the end of each calendar year covered by the panel. In these panels, each panel weight covers the longitudinal reference period from Wave 1 through the last wave of the corresponding calendar year. 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 (or longitudinal reference period) for which they were in scope for the survey. Other people are assigned weights of zero. Most people with nonzero final panel weights have provided data for all months of the panel (or all months in the waves of the longitudinal reference period).

In the 2008 SIPP Panel, final calendar year weights are assigned to all persons with EPPMIS=1 (person month in month, EPPMIS=1) in all calendar months of the same year. On the other hand, 2014 SIPP final calendar year weights are computed only for people who had an interview covering December of the reference year 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, third, and/or fourth calendar year, if data are obtained for those periods. Starting in the 2014 panel, calendar year weights are the same as the December month weights of the same year, and the first panel weight covering the period January through December 2013 is also the same as the December 2013 month weight.

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 month. This group consists of (a) people who enter the sample universe after the first wave and (b) people who were in the sample universe in the first wave but did not have an interview covering the control month. For example, infants born after December 2013 and people leaving institutions after December 2013 both enter the sample universe after Wave 1 and are assigned full panel and calendar year 1 weights of zero. Note that these same people will receive positive calendar year 2 (CY2) weights if they are in the sample universe in the second wave and have an interview covering the control month for CY2 (i.e. December 2014). However, people who provided data up to the point that they left the survey (through death or because they moved to an ineligible address) are assigned nonzero final panel weights.

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

- Initial weight -This weight is the household noninterview adjusted cross-sectional month weight for the control month of the appropriate year.

- 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.

- 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 month of the relevant year. Those for the panel weights are the population estimates for a designated control month in the first year of the panel.

Appendixes

A. 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
EHC	=	Event History Calendar
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
PSUs	=	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

B. 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 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.

FERRETT - 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 Guard Retirement - An Army National Guard retirement plan provided for employees who have 20 years service in the National Guard. By age 60, this retirement pension may combine with any additional income or retirement.

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 O*riginal 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 in the interview month.

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. Related subfamilies can only be identified in the interview month, since a reference person is not identified in the monthly data.

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.

Top-coding - 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. Unrelated subfamilies can only be identified in the interview month, since a reference person is not identified in the monthly data.

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

Bye, B., & Gallicchio, S. (1989). Two Notes on Sampling Variance. SIPP Working Paper 79.

Cresce, A. R., Cheng, Y., & Grieves, C. (2013). Household Estimates Conundrum: Effort to Develop More Consistent Household Estimates Across Surveys.

Czajka, J. L., Mabli, J., & Cody, S. (2008). Sample Loss and Survey Bias in Estimates of Social Security Beneficiaries: A Tale of Two Surveys. Mathematica Policy Research.

Little, R. J., Rubin, D. B. (2014). Statistical analysis with missing data. John Wiley & Sons. John Wiley and Sons.

McMillan, S., & III, R. C. (2013). Nonresponse Bias Analysis for Wave 1 2008 Survey of Income and Program Participation (ALYS-12). U.S. Census Bureau Internal Memorandum from Killion to Fields.

Moore, J. (2008). Seam Bias in the 2004 SIPP Panel: Much Improved, but Much Bias Still Remains. Washington, D.C.: U.S. Census Bureau.

Participation, P. O. (2009). Reengineering the Survey of Income and Program Participation. The National Academies.

Raghunathan, T., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology , 85-96.

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

U.S. Census Bureau. (2013). SIPP-EHC 2011 and 2012 Field Test Evaluation. Washington D.C..