

11. Using Topical Module Files

This chapter discusses procedures on using topical module public use files. Documentation for the topical modules are explained in this chapter as well as how to use the topical module files when performing common tasks. Those tasks include:

- Using the monthly interview status variables;
- Identifying people, households, and families;
- Using imputation flags; and
- Identifying states and metropolitan areas.

Before reading this chapter, users should read Chapter 9, “The SIPP Public Use Files”, for an introduction to Section II. Analysts using only one topical module file also should read about the use of sample weights (Chapter 8) and the computation of standard errors (Chapter 7). Those planning on merging data from a topical module to data from the core wave or full panel files should also read Chapter 10 for information about the core wave files, Chapter 12 for information about the full panel files for pre-96 Panels, and Chapter 13 for information about linking SIPP public use files.

This chapter focuses on the topical module files. It is written so that it can be used independently of the chapters describing the core wave and full panel files. Although there are many similarities across the three types of SIPP public use data files, important differences do exist. Because those differences are sometimes subtle, users familiar with the core wave and full panel files should read this chapter carefully, paying close attention to information about variable names and file structures. Tables, 9-2 and 9-3 in Chapter 9 summarize the differences between the core wave, topical module, and full panel longitudinal research files.

During the 1996 redesign most variable names changed from those used in previous panels. To aid users working with files from panels prior to 1996, this chapter presents names from both the Pre-1996 panels and 1996+ Panels. In the main body of the text, the Pre-1996 names are presented in parentheses following the 1996+ Panel variable names. For example, the sample unit ID variable name, which is SSUID in the 1996+ Panels, was SUID in pre-96 Panels, therefore, it is written in this chapter as SSUID (SUID). In tables, a variety of methods are used to present both the Pre-96 names and 1996+ Panel names.

Using the Technical Documentation of the Topical Module Files

Each data file purchased from the Census Bureau comes with a set of technical documentation and a data dictionary. The technical documentation includes:

- The item booklets (for the 1996, 2001, 2004, and 2008 Panels);
- The paper survey instrument (for panels prior to 1996);
- A glossary of selected terms;
- A cross-walk, mapping reference months into calendar months for each rotation group;
- A source and accuracy statement describing the sample weights and the computation of standard errors; and
- User Notes.

The survey instrument is vital to understanding what questions were asked, how they were asked, the order in which they were asked, to whom they were asked, and the way in which the answers were recorded. Some questions are conditioned on skip patterns (Chapter 3), so users should pay particular attention to which questions were skipped for which respondents. The skip patterns are best understood by consulting the data dictionary and referring to the universe statements and reviewing the survey instrument or by contacting the Income Surveys Branch, Demographic Surveys Division on 301-763-3819. The data dictionary can be found on the SIPP web site <http://www.census.gov/sipp/> under *Technical Information*. The questionnaire documentation can be found under *Survey Content*.

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

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

1. The definition,
2. The sample universe of the corresponding survey question,
3. The ranges for all legal values, and

4. The location (and size) in the file.

A machine-readable version of the data dictionary accompanies each data file. It can also be downloaded from the Internet

The data dictionary is formatted to facilitate processing by user-written computer programs. Figure 11-1 shows an excerpt from the data dictionary for the topical module from Wave 1 of the 2004 Panel. A “D” in the first column signifies that the next few lines define the variable: (1) the variable name; (2) the size (i.e., how many digits it contains); (3) the starting position; and (4) the definition. Lines beginning with a “T”, added with the 2004 Panel, contain short variable descriptions that can be used by many software packages as variable labels.

A “U” in the first column signifies that the next words describe the sample universe.¹ A “V” in the first column indicates that the next number and phrase describe one of the values of the variable. A blank in the first column denotes either a variable description or other comment. A period (.) before a word denotes the start of the value label.

Prior to the 1996 Panel, the dictionaries had a different format, shown in Figure 11-2. A “D” in the first column signifies that the next few lines define the variable: (1) the variable name; (2) the size (i.e., how many digits it contains); (3) the starting position; and (4) the definition. A “U” in the first column signifies that the next words describe the sample universe.² A “V” in the first column indicates that the next number and phrase describe one of the values of the variable. An asterisk in the first column denotes a comment. A period (.) before a word denotes the start of the value label.

¹ The universe definitions included in the data dictionaries prior to the 1996 Panel were not always accurate. Users of pre-1996 SIPP Panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

² See footnote 1.

Figure 11 - 1. Excerpt from the Data Dictionary for a Topical Module File

Wave 1 of the 1996+ SIPP Panel		
D EENTAID	3	500
T PE: Address ID of hhld where person entered sample		
Address ID of the household that this person belonged to at the time this person first became part of the sample		
U All persons		
V	11:159	.Entry address ID
D EPPNUM	4	503
T PE: Person number		
Person number. This field differentiates persons within the sample unit. Person number is unique within the sample unit.		
U All persons		
V	101:1599	.Person number
D EPPINTVW	2	507
T PE: Person's interview status		
U All persons		
V	1	Interview (self)
V	2	Interview (proxy)
V	3	Noninterview - Type Z
V	4	Noninterview - pseudo Type Z.
V		Left sample during the reference period
V	5	Children under 15 during reference period
D EPOPSTAT	1	509
T PE: Population status based on age in 4th reference month		
Population status. This field identifies whether or not a person was eligible to be asked a full set of questions, based on his/her age in the fourth month of the reference period.		
U All persons		
V	1	Adult (15 years of age or older)
V	2	Child (Under 15 years of age)
D ENTRY 2	30	
Entry address ID		
Address of the household that person belonged to at the time person first became part of the sample		
U All persons, including children		
D PNUM	3	32
Person number		
U All persons, including children		
D FINALWGT	3	35
Person weight (interview month)		
There are four implied decimal places.		
U All persons, including children		

Figure 11 -2. Excerpt from the Data Dictionary a Topical Module File

Wave 3 of the 1993 SIPP Panel	
D	ENTRY 2 30
	Entry address ID
	Address of the household that person belonged to at the time person first became part of the sample
U	All persons, including children
D	PNUM 3 32
	Person number
U	All persons, including children
D	FINALWGT 3 35
	Person weight (interview month)
	There are four implied decimal places.
U	All persons, including children

Figure 11-3 shows sample SAS and FORTRAN syntax for reading the data described by the codebook fragments in Figure 11-1. Additional SAS program code could be used to associate value labels (a SAS “format”) with the INTVW variable.

Relationship of the Topical Module Data Files to the Survey Instrument

Each wave’s survey instrument includes one or more topical modules,³ as described in Chapter 3. The questions in those modules are often asked after the core survey questions and can be found toward the end of the survey instrument. The data from the topical modules in a wave are usually combined into one topical module data file for each SIPP wave.

The topical module data dictionary does not replicate the survey instrument. Thus, analysts should keep a few things in mind when using the data:

- The variables on the data files do not correspond one-to-one with the questionnaire items - the variables are listed in a different order, some are not included in the public use files, and some are created from a combination of other variables;
- The range of possible values of the variables on the data files does not always correspond one-to-one with the response categories shown on the survey instrument or in the data dictionary;

³ Prior to the 1992 Panel, there were no topical modules administered with the Wave 1 interview, although some topical content was included in the Wave 1 core questionnaire for the purpose of obtaining historical information. Since the 1992 Panel, Wave 1 has had topical modules.

Figure 11-3. Corresponding SAS and FORTRAN Syntax to Read Data from a Topical Module File

Wave 1 of the 2004 Panels	
SAS	
Input	<pre>@45 EENTAID 3. EPPNUM 4. EPPINTVW 2. EPOPSTAT 1. ;</pre>
LABEL	<pre>EENTAID = "Adrs ID where person entered sample" EPPNUM = "Person number" EPOPSTAT = "Population status based on age in fourth" EPPINTVW = "Person's interview status" ;</pre>
FORTRAN	
<pre>READ(INFILE,1000) EENTAID EPPNUM EPOPSTAT EPPINTVW 1000 FORMAT(T45, I3, I4, I1, I2)</pre>	
Wave 3 of the 1993 SIPP Panel	
SAS	
Input	<pre>@30 ENTRY 2. PNUM 3. @38 FINALWGT 9.4 ;</pre>
LABEL	<pre>ENTRY = "Entry address ID" PNUM = "Person number" FINALWGT = "Person weight (interview month)" ;</pre>
FORTRAN	
<pre>READ(infile,1000) ENTRY, PNUM, INTVW 1000 FORMAT(T457,I2,I3,I1)</pre>	

- The variable name in the data dictionary may not readily indicate the variable's content;
- Prior to the 1996 Panel, some variable names were used in different topical module files for different variables. For example, in the 1990 Panel, TM8400, was used in the Wave 2 topical module for a variable that indicates whether the respondent completed 12th grade. The same variable name was used in the Wave 6 topical module to indicate whether the respondent was a parent of children who were under 21 years of age living in the respondent's household.
- The complexity of the skip patterns may not be apparent just by looking at the data dictionary. Many questions were administered only to the household reference person, or to adults (age 15 years or older), or to adults 25 years or older, or to some other subset of

survey respondents.⁴ To avoid potential problems and confusion, analysts should become familiar with the survey instrument before using the data when working with the data, refer to both the survey instrument and the data dictionary.

Structure of the Topical Module Files

The topical module files for the 1996+ Panels contain one record for each person who was in the sample with a completed (or imputed) interview in the fourth month of the wave's reference period (the month immediately prior to the interview). This arrangement is similar to the person month format of the core wave files, but only records for month four are included in the topical module files. Prior to the 1996 Panel, the topical module files contained one record for each person who was interviewed or for whom an interview was attempted in that wave (Table 11-1 shows one record for each such person; compared with Table 10-1, which shows up to four records per sample person in the core wave files).⁵

In general, each topical module file contains data for all topical module subject areas administered during a particular wave.⁶ Each topical module file also contains selected information from the SIPP core; thus, for some analyses, those files can be used independently from the core wave and full panel data files. When more detailed information from the SIPP core is needed, data from the topical modules must be merged with data from the core wave or full panel files. Chapter 13 provides a detailed discussion of merging SIPP files.

The topical module file structure differs from that of the core wave files in the following ways:

- For the 1996+ Panels, the topical module files contain one record for each person who was a SIPP sample member during month four of the wave; the core wave files contain one record per person for each month the person is in the sample.
- Prior to the 1996 Panel, the topical module files contain one record per person for each person present in a SIPP household *at the time of the interview*; the core wave files contain one record per person for each month the person was in the sample during the *previous 4 months*.

⁴ The universe definitions included in the data dictionaries prior to the 1996 Panel were not always accurate. Users of pre-1996 SIPP panels should check the skip patterns in the actual survey questionnaire to determine which subset of respondents was asked each question.

⁵ The variables shown - sample unit ID, current address ID, entry address ID, and person number - are discussed in detail later in this chapter.

⁶ Chapter 3 offers a detailed listing of the topical modules administered with each wave of each SIPP panel.

Table 11 -1. Example of the Topical Module File Structure

1996+ Panels			
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Entry Address ID (EENTAID)	Person Number (EPPPNUM)
123456789123	021	011	0101
123456789123	021	011	0102
123456789123	021	021	0201
123456789123	021	021	0202
Panels Prior to 1996			
Sample Unit ID (SUID)	Current Address ID (ADDID)	Entry Address ID (ENTRY)	Person Number (PNUM)
123451000	21	11	101
123451000	21	11	102
123451000	21	21	201
123451000	21	21	202

- Prior to the 1996 Panel, the topical module files included records for people whose entire household refused to be interviewed or left the sample;⁷ those people are excluded from the core wave files.
- Prior to the 1996 Panel, the structure of the topical module files was roughly similar to that of the full panel files, containing one record per person.

Reference Periods and Samples

Sample definitions and reference periods in the topical modules vary across panels, across topical modules within panels, and even within topical modules. Users should pay careful attention to those details in the topical module files they are using.

For the 1996+ Panels, most topical module questions were asked only of people who were in the SIPP sample during the fourth month of the wave's reference period. People who were members of SIPP households at the time of the interview (month five) but who were not members of SIPP households during the previous month were not asked the topical module questions in the 1996+ Panels. For the 1996+ Panels, many of the questions refer to just that month (month four). However, some topical module questions, and in some cases, entire topical modules refer to longer periods of time, such as the previous 4 months, the previous year, or, in the various history topical modules administered with Waves 1 and 2, the person's life before SIPP.

⁷ Panels that included topical modules in Wave 1, such as the 1993 and 1996 Panels, exclude those people from the Wave 1 topical module files.

Prior to the 1996 Panel, most topical module questions were asked of people who were in the SIPP sample at the time of the interview (month five). This included people who were household members

at the time of the interview but who were not members of SIPP households at any time during the previous 4 months, the reference period for SIPP core questions in that wave.⁸ Many questions asked about current (month five) conditions, although some asked about longer periods in the past.

Using a Person's Monthly Interview Status Variables

A person's monthly interview status variable is used to determine whether the data for that person in a given month should be used. Some analysts refer to it as the in sample variable to distinguish it from the household interview status variable, EOUTCOME (ITEM36B), and another variable that indicates the type of interview or noninterview for the person, EPPINTVW (INTVW). The interview status variable has three possible values: 0, 1, and 2. A value of 1 indicates that the person was both in-scope for the survey (a member of the population that the SIPP sample is intended to represent) and, aside from some item nonresponse, provided complete answers to the SIPP core questions for the reference month in question.⁹

Monthly Interview Status in the Topical Module Files for the 1996+ Panels

There is only one interview status variable in the topical module files from the 1996+ Panels. That variable, EPPMIS4, identifies a person's status in the fourth reference month of the wave. Because the topical module files from the 1996+ Panels contains records only for people for whom this variable is equal to 1 (and so equals 1 on all records in the file), EPPMIS4 can be safely ignored when working with topical module files from the 1996-2008 Panels. When using FERRET (Federated Electronic Research, Review, Extraction and Tabulation Tool), users should select the variable, SREFMON, from the *Sample Unit Variable* file for topical module analyses. The variable, SREFMON, must be set to 4, the fourth reference month and unselected for all other options.

⁸This has important implications for procedures used to merge the topical modules to data from the core. Core data that correspond to the same reference month as a topical module must often be merged from the subsequent wave rather than from the same wave as the topical module, as discussed in Chapter 13.

⁹The only exception is for Type Z noninterviews. For Type Z noninterviews prior to the 1996 Panel, complete records for the SIPP core were imputed and the monthly interview status variable was set to 1, indicating that, for most analytic purposes, the responses should be treated as though they were provided by the respondent. This exception is handled similarly in the 1996+ Panels when there is no prior wave information. When prior wave information exists, items are imputed using the same hot-deck methods applied to instances of item nonresponse.

Monthly Interview Status in the Topical Module Files from Panels Prior to 1996

The topical module files for panels prior to 1996 are different. On those files, a person's interview status variable is labeled PP-MIS1, PP-MIS2, PP-MIS3, PP-MIS4, and PP-MIS5. These variables refer to the four reference months of the wave (PP-MIS1 to PP-MIS4) and the interview month itself (PP-MIS5).

The monthly interview status is the only reliable guide to whether the data for a given person should be used in a given month. Analysts should use data for only those months in which a person's interview status (PP-MIS) is equal to 1.¹⁰

Any data present for months when a person's interview status is coded either 0 or 2 should be ignored. A code of 0 indicates that the person was not in the sample that month, and a code of 2 indicates a noninterview for that month.

On the topical module files for panels prior to 1996, the topical module questions were asked only of sample members with PP-MIS5 equal to 1:¹¹ that is, the topical module questions were asked only of those who were in the SIPP sample at the time of the interview. Because the reference periods of the topical module questions vary, some topical module questions contain information about people who had been secondary sample members during previous months, even though they were no longer part of the SIPP sample at the time of the interview. The variables PP-MIS1 to PP-MIS4 are useful when working with topical module questions that refer to previous months. The four variables are also useful when merging topical module data with data from the core, a topic discussed in Chapter 13.

Four sample members are shown in Table 11-2. Two were present in the interview month (PPMIS5 = 1), and two were not present (PP-MIS5 = 2). Analysts interested in just the interview month should use data only for people with PP-MIS5 = 1. In this example, only persons 101 and 201 would be included.

If the research focuses on January, analysts should use data only for people with $PP-MIS_x = 1$, where x corresponds to the reference month that contains information about January (which varies by wave and rotation group). Assuming an analyst is interested in January 1994, the example represents Wave 4 and rotation group 1 of the 1993 Panel (see Table 11-3 for the reference months); the analyst would use only the people with $PP-MIS_1 = 1$. Thus, only persons 101 and 102 would be

¹⁰ As a safeguard against inadvertently using data for months where PP-MIS is not equal to 1, all monthly variables in the user's data extract should be set to a missing value for months when PP_MIS is not equal to 1. Most statistical packages allow certain values to be flagged as missing. Once flagged, those values are excluded from computations.

¹¹ In some cases, questions are asked of all household members over 14 years old. In other cases, they may be asked only of the household reference person. There are also topical modules in which other subsets of household members are interviewed

included.

Table 11-2. Monthly Interview Status Variables in the 1984-1993 SIPP Panels

Sample Unit ID (ID)	Current Address ID (ADDID)	Entry Address ID (ENTRY)	Person Number (PNUM)	Rotation Group (ROTATION)	PP-MIS				
					1	2	3	4	5
123451000	11	11	101	1	1	1	1	1	1
123451000	11	11	102	1	1	1	2	2	2
123451000	11	11	201	1	2	2	2	2	1
123451000	11	11	202	1	0	0	2	2	2

Table 11 -3. Interview Month and Reference Months for Each Rotation Group in Wave 4 of the 1993 Panel

Rotation Group	Reference Months for Core Questions	Interview Month
2	Oct., Nov., Dec. 1993; Jan. 1994	Feb. 1994
3	Nov. Dec. 1993; Jan. , Feb. 1994	Mar. 1994
4	Dec. 1993; Jan. , Feb., Mar. 1994	Apr. 1994
1	Jan. , Feb., Mar., Apr. 1994	May 1994

As demonstrated by this example, the topical module files for panels conducted before 1996 contain a record for each person for whom no interview data were collected, either because the person refused to be interviewed (and no proxy interview was obtained) or because the person left the survey sample (e.g., died or entered the Armed Forces or an institution). Those individuals have PP-MIS5 = 2 and PP-MISj = 1 for j = 1, 2, 3, or 4 or INTVW = 3 or 4. Their demographic information was gathered from the previous time that they were successfully interviewed; if they have topical module information, it was completely imputed by the Census Bureau.

Comparison of Variables in the Topical Module and Core Wave Files

The topical module files contain a number of variables that are also present in the core wave files. These include variables needed to identify the household and the person. Also included are selected demographic characteristics. In the 1996+ Panels, the values for the demographic characteristics correspond to the month-four values in the core wave file for the same wave for the 1996+ Panels. Variables common to the core wave and topical module files are generally given the same names in both files. For example, SSUID is used for the sample unit identifier, SHHADID is the current

address ID, and EPPNUM is the person number on both files.¹² Among the demographic variables, TAGE is used on both files for the respondent's age, and EMS is used for the respondent's marital status. Table 11-4 shows the 27 variables that are common to the core wave file and topical module file from Wave 1 of the 1996+ Panels.

Prior to the 1996 Panel, the demographic data on the topical module files corresponded to the interview month (month five), not to any of the 4 reference months for the core interview. For that reason, the information in variables such as AGE, RRP, and MS (the respondent's age, relationship to the household reference person, and marital status) could differ from the core wave file variables of the same names for the wave in which the topical module was administered. This would indicate that a change occurred between the last month of the reference period (month four) and the interview month (month five). Some variables included on both the core wave and topical module files have different names. As shown in Table 11-5, sample unit ID, rotation group, state, interview status in month five, and the person-level weight are contained in both files but have different variable names.

Identifying People

There are many occasions when it is necessary to identify which records belong to each individual in the SIPP data files. This need arises, for example, when

- Merging data from topical module files to data from the core wave or full panel files,
- Merging data from two or more topical module data files,
- Linking husbands and wives, and
- Linking parents and children.

In the 1996+ Panels, two variables are needed to uniquely identify a person: the sample unit ID and the person number.¹³ For files from panels prior to 1996, three variables are needed to uniquely identify a person: the sample unit ID, entry address ID, and person number. Table 11-6 shows the variable names used in the topical module files for the 1996+ Panels and for the pre-1996 Panels.

¹² Use of common names facilitates merging of the core wave and topical module files from the 1996+ Panels. Merging files is discussed extensively in Chapter 13.

¹³ Users should note that in the 1996+ Panels, the entry address ID is no longer needed for unique identification. Its continued use will not create any problems; it is simply redundant information. That is a change from earlier panels, in which the entry address ID was key to uniquely identifying a person.

Table 11 - 4. Variables Common to the Core Wave and Topical Module Files from Wave 1 of the 1996+ Panels

Variable Name	Description
EEDUCATE	Highest degree received or grade completed
EENTAID	Address ID of household where person entered
EMS	Marital status
EORIGIN	Ethnic origin of this person
EOUTCOME	Interview status code for this household
EPNDAD	Person number of father
EPNGUARD	Person number of guardian
EPNMOM	Person number of mother
EPNSPOUS	Person number of spouse
EPOPSTAT	Population status based on age
EPPINTVW	Person's interview status
EPPNUM	Person number
ERACE	Race of this person
ERRP	Relationship to reference person
ESEX	Gender of this person
RDESGPNT	Designated parent or guardian flag
RFID	Family ID number
RFID2	Family ID excluding related subfamily
SHHADID	Household address ID-differentiates households
SPANEL	Sample code-indicates panel year
SROTATON	Rotation of data collection
SSUID	Sample unit identifier
SSUSEQ	Sequence number of sample unit - primary
SWAVE	Wave of data collection
TAGE	Age as of last birthday (topcoded)
TFIPSST	FIPS state code (topcoded)
WPFINWGT	Person weight

Table 11 -5. Examples of Same Variables with Different Names in the Core Wave and Topical Module Files Prior to the 1996 Panel

Description	Variable Name in the Core Wave File	Variable Name in the Topical Module File
Sample unit ID	SUID	ID
Rotation group	ROT	ROTATION
State of residence	HSTATE	STATE
Monthly interview status in the interview month	MIS5	PP-MIS5
Person-level weight in the interview month	P5WGT	FINALWGT

Table 11- 6. Variables Used to Uniquely Identify a Person in the Topical Module Files

1996+ Panels	
Variable Name	Description
SSUID	Sample unit ID
EPPNUM	Person number
Pre-1996 Panels	
Variable Name	Description
ID	Sample unit ID
ENTRY	Entry address ID
PNUM	Person number

The variables can be described as follows:

- SSUID (ID) uniquely identifies each initially sampled dwelling unit.¹⁴ Every person in a core wave file was either a member of one of those units (an original sample member) or lives with someone who was a member of an initially sampled dwelling unit. A person's connection to that unit is an attribute of that person and does not change over time.¹⁵ This means that as people move from address to address, their SSUID (ID) stays the same. As new people join the homes of original sample members, they receive the SSUID (ID) of the original sample members.

¹⁴ The SSUID (ID) is a random recode of three other variables in the Census Bureau's internal (not public use) files: the respondent's sampling area (primary sampling unit), the cluster of housing units within that area (called the "segment"), and a sequentially assigned serial number. Those three variables are omitted from the public use files to protect the confidentiality of the respondents.

¹⁵ There is one rare exception to this rule for panels prior to 1996, which is described in the section entitled "Identifying Movers" later in this chapter.

- EENTAID (ENTRY) identifies the address where the person lived at the time he or she was first interviewed. It does not change even if the person moves.¹⁶ Prior to the 1996 Panel, it was used in conjunction with the person number and the sample unit ID to uniquely identify people within the sampling unit. It is not needed to uniquely identify people in the 1996+ Panels. Values for this variable are unique only within sample units. The entry address ID has two components. The first part of the ID number (two digits in the 1992 and 1996 Panels, and one digit in all others) identifies the wave in which SIPP interviews were first conducted at the address. The second part of the number (one digit in all panels) sequentially numbers addresses within a sample unit [SSUID (ID)] that enter the sample in the same wave. See Chapter 10 for a more complete discussion.
- Prior to the 1996 Panel, PNUM uniquely identified a person within the sample unit and entry address ID. For the 1996+ Panels, EPPPNUM uniquely identifies a person within the sample unit. EPPPNUM (PNUM) does not change even if the person moves.¹⁷ The first part of EPPPNUM (PNUM) (two digits in the 1992 and 1996+ Panels, and one digit in all others) indicates the wave in which the person was first interviewed.¹⁸ The remaining two digits are sequentially assigned within the household. Thus, original sample members are assigned person numbers ranging from 100 to 199. Individuals who enter the SIPP sample in Wave 2 are assigned a person number ranging from 200 to 299. Those who enter in Wave 10 are assigned person numbers ranging from 1001 to 1099.

Table 11-7 illustrates how the combination of SSUID (ID), EENTAID (ENTRY), and EPPPNUM (PNUM) uniquely identifies people and provides information about when they first entered the SIPP sample. In this example, there are eight individuals: five are original sample members, one person joined the SIPP sample in Wave 4, one person joined in Wave 7, and one person joined in Wave 10.

To uniquely identify a household or group quarters in the topical module files, analysts should use the two variables shown in Table 11-8.

People with the same SSUID (ID) (sample unit ID) and SHHADID (ADDID) (current address ID) values live in the same household (or group-quarters location) in the relevant month. For the 1996+ Panels, household membership refers to month four of the wave's reference period. For the pre-1996 Panels, household membership refers to the interview month. The eight individuals shown in Table

¹⁶ See footnote 7.

¹⁷ For cases in the 1996 Panel for whom prior wave information did not exist for a person-level noninterview (such as in Wave 1 or in Waves 2-12 when the person was new to the sample), the whole record may have been imputed. To identify such cases, users need to check both person number (to distinguish wave of entry into the sample) and EPPINTVW, which will be 3 or 4 for these cases.

¹⁸ Chapter 4 contains a discussion of how analysts can determine whether these special imputation procedures were used.

11-9 make up four households. The first household contains the first four individuals. The second household contains one person. The third household contains one person. The fourth household contains two people. (Figure 2-1 illustrates concepts of household and changes in household.)

Table 11 -7. How to Uniquely Identify a Person in the Topical Module Files

1996+ Panels				
Sample Unit ID (SSUID)	Entry Address ID (EENTAID)	Person Number (EPPPNUM)	Current Address ID (SHHADID)	Notes
123456789123	011	010	07	Original sample member
123456789123	011	0102	071	Original sample member
123456789123	011	0401	071	Enters SIPP sample in Wave 4
123456789123	071	0701	071	Enters SIPP sample in Wave 7
321456789123	011	0101	031	Original sample member
321456789123	011	0102	032	Original sample member
321456789123	011	0103	101	Original sample member
321456789123	101	1001	101	Enters SIPP sample in Wave 10
Pre-1996 Panels				
Sample Unit ID (ID)	Entry Address ID (ENTRY)	Person Number (PNUM)	Current Address ID (ADDID)	Notes
123456789	11	101	71	Original sample member
123456789	11	102	71	Original sample member
123456789	11	401	71	Enters SIPP sample in Wave 4
123456789	71	701	71	Enters SIPP sample in Wave 7
321456789	11	101	31	Original sample member
321456789	11	102	32	Original sample member
321456789	11	103	101	Original sample member
321456789	101	1001	101	Enters SIPP sample in Wave 10 (1992 Panel)

^a Not needed to uniquely identify a person in the 1996 - 2008 Panels.

Table 11-8. Variables Used to Uniquely Identify a Household or Group Quarters in the Topical Module Files

1996+ Panels	
Variable Name	Description
SSUID	Sample unit ID
SHHADID	Current address ID in month 4 (in month 5)
Pre-1996 Panels	
Variable Name	Description
SUID	Sample unit ID
ADDID	Current address ID in month 4 (in month 5)

Table 11-9. How to Uniquely Identify a Household in the Topical Module Files

1996+ Panels			
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Person Number (EPPPNUM)	Notes
123456789123	071	0101	Four people in this household
123456789123	071	0102	
123456789123	071	0401	
123456789123	071	0701	
321456789123	031	0101	One person in this household
321456789123	032	0102	One person in this household
321456789123	101	0103	Two people in this household
321456789123	101	1001	
Pre - 1996 Panels			
Sample Unit ID (ID)	Current Address ID (ADDID)	Person Number (PNUM)	Notes
123456789	71	101	Four people in this household
123456789	71	102	
123456789	71	401	
123456789	71	701	
321456789	31	101	One person in this household
321456789	32	102	One person in this household
321456789	101	103	Two people in this household
321456789	101	1001	

Identifying Families

The term *family*, as used in Census Bureau publications, refers to a group of two or more people related by birth, marriage, or adoption who reside together; all such individuals are considered members of one family.

The Census Bureau distinguishes among several types of families:

- A *primary family* is a family containing the household reference person and all of his or her relatives. This means that a household composed of a husband and wife, their son, and their son's wife (i.e., the daughter-in-law) is classified as a primary family containing four people.
- A *related subfamily* is a nuclear family that is related to but does not include the household reference person. For example, the son and his wife (i.e., the daughter-in-law) in the preceding example are a related subfamily.
- An *unrelated subfamily* (sometimes called a secondary family) is a nuclear family that is not related to the household reference person. Thus, a husband and wife who live in a friend's house are classified as an unrelated subfamily. A mother and daughter who live in the mother's boyfriend's apartment are classified as an unrelated subfamily.
- A *primary individual* is a household reference person who lives alone or lives with only nonrelatives. Primary individuals are sometimes treated by the Census Bureau as families of only one person and are referred to as pseudo-families.
- A *secondary individual* is not a household reference person and is not related to any other people in the household. Secondary individuals are sometimes treated by the Census Bureau as families of only one person and are referred to as pseudo-families.

In the topical module files for the 1996+ Panels, the variables shown in Table 11-10 can be used to uniquely identify families.

Table 11-10. Variables Used to Uniquely Identify a Family in the Topical Module Files for the 1996+ Panels

Variable Name	Description
SSUID	Sample unit ID
SHHADID	Current address ID
<i>and one of the following:</i>	
RFID	Family ID in month four of the wave
RFID2	Family ID in month four (excluding related subfamily members; RFID2=0 for related subfamily members)

The Census Bureau has two principal methods for distinguishing families that are based on the variables and numbering schemes shown in Table 11-10. Analysts must remember to choose which type of family classification they want and then use the appropriate method.

- The first method defines a family as all persons who are related and living together. The family ID variable RFID is used with this definition. RFID groups the household reference person with all related household members by assigning them the same ID number. This family group corresponds to the Census Bureau's definition of primary family. RFID groups members of each unrelated subfamily (and primary and secondary individuals) separately.
- The second method is similar to the first in defining a family, but the family excludes related subfamilies. The family ID variable RFID2 is used with this definition. RFID2 equals zero for related subfamilies. RFID2 groups members of each unrelated subfamily (and primary and secondary individuals) in the same way as RFID - each group has a unique number.¹⁹

Table 11-11 illustrates the difference between the RFID and RFID2 variables. Those variables refer to month four of the wave's reference period. For example, a mother, a father, and a child would be family 1 (RFID = 1). The first household in the table contains a primary family of five people. The primary family contains members of related subfamilies. However, the topical module files for the 1996- 2004 Panels do not contain the variables needed to determine whether all subfamily members are members of the same subfamily. To determine that, an analyst would need to merge the RSID variable from the month four records in the core wave file.

The second "household" is actually three households, each containing a primary family that originally formed one household. The third household contains a primary family and two unrelated subfamilies. The fourth household contains a primary family and one unrelated subfamily. The fifth household contains a primary individual and an unrelated subfamily. The fifth household contains only a primary individual. The sixth household is a group quarters containing two people.

Other Variables Describing Household and Family Composition

The topical module files contain several additional variables from the SIPP core that describe household and family composition.²⁰ The household composition variables included in the topical

¹⁹ The variables included on the topical module files do not allow analysts to distinguish among different related subfamilies living in the same household. If needed, the RSID variable (which groups each related and unrelated subfamily separately) can be merged from the core wave files. Chapter 10 discusses the core wave files, and Chapter 13 discusses the merging of multiple SIPP files.

²⁰ Detailed information about the relationships between members is collected in the Household Relationships topical module. For the 1996+ Panels, those data provide extensive information about household composition during month four of the wave's reference period. For earlier panels, the topical module provides information about household composition at the time of the interview.

module files from the 1996+ Panels and from the Pre-1996 Panels are shown in Table 11-12. Additional variables from the core wave files and the full panel files can be merged with data from the topical module files when added detail is needed (Chapters 10, 12, and 13).

Table 11-11. Uniquely Identifying Families in the Topical Module Files in the 1996+ Panels

Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Family ID, Including Related Subfamily (RFID)	Family ID, Excluding Related Subfamily (RFID2)	Person Number (EPPNUM)	Notes
110011111123	11	1	1	0101	This household contains a primary family of five people. The primary family contains one or more related subfamilies.
110011111123	11	1	0	0102	
110011111123	11	1	0	0103	
110011111123	11	1	0	0104	
110011111123	11	1	0	0105	
110077777723	11	1	1	0101	Three households formed by people who were originally members of the same originally sampled household (SSUID of 110077777723). Two subfamilies split off from the original household to become two new primary families at addresses 21 and 22.
110077777723	21	1	1	0102	
110077777723	21	1	1	0103	
110077777723	22	1	1	0104	
110077777723	22	1	1	0105	
122210000123	11	1	1	0104	This household contains a primary family and two unrelated subfamilies.
122210000123	11	2	2	0305	
122210000123	11	2	2	0306	
122210000123	11	3	3	0307	
122210000123	11	3	3	0308	
55555555123	21	1	1	0101	This household contains a primary individual and an unrelated subfamily.
55555555123	21	2	2	0201	
55555555123	21	2	2	0202	
55555555123	21	2	2	0203	
61000000123	32	1	1	0101	Primary individual. Group quarters with two secondary individuals
897454644123	11	1	1	0101	Group quarters with two secondary individuals.
897454644123	11	2	2	0102	

Table 11 - 12. Household and Family Composition Variables in the Topical Module Files

1996+ Panels	
Variable Name	Description
ERRP	Relationship to household reference person in month four
EMS	Marital status in month four
EPNMOM	Person number of mother in month four
EPNDAD	Person number of father in month four
EPNGUARD	Person number of guardian in month four
EPNSPOUS	Person number of spouse in month four
RDESGPNT	Designated parent or guardian in month four
Pre-1996 Panels	
Variable Name	Description
RRP	Revised relationship to the household reference person (living with relatives, child of household reference person, etc.)
PNSP	Person number of spouse
PNPT	Person number of parent

Using the Relationship to Reference Person [ERRP (RRP)] Variable

As Table 11-13 shows, ERRP (RRP) provides a summary description of how each individual is related to the household *reference person*.²¹

The ERRP (RRP) variable contains summary information about each person's relationship to the household reference person. Analysts should take into consideration that the household description depends upon the identity of the household reference person. For example, the household in Table 11-14 contains a mother, her daughter, and her daughter's son. If the mother is the household reference person [ERRP = 1 (RRP = 1)], her daughter is listed as a child of the household reference person [ERRP = 4 (RRP = 4)] and the daughter's son is listed as a grandchild of the reference person in the 1996+ Panels (ERRP = 5), but as another relative of the household reference person in earlier panels (RRP = 5, where the same value has a different meaning from that of the 1996+ Panels variable).

²¹ Prior to the 1996 Panel, the RRPV variable, available in the core wave files, provides additional detail not contained in the RRP variable. When needed, RRPV can be merged to data from the topical module files (Chapters 10 and 13).

If the daughter is the reference person, her son is listed as a child of the household reference person (RRP = 4) and her mother is listed as the parent of the reference person in the 1996+ Panels (ERRP = 6), but as another relative of the household reference person in earlier panels (RRP = 5).²² Users should note that the identity of the household reference person could change from one month to the next; thus, the household description could also change.

**Table 11 -13. Relationship to the Household Reference Person
in the Topical Module Files**

1996+ Panels	
ERRP	Description
1	Reference person w/related people in household
2	Reference person w/out related people in household
3	Spouse of reference person
4	Child of reference person
5	Grandchild of reference person
6	Parent of reference person
7	Brother or sister of reference person
8	Other relative of reference person
9	Foster child of reference person
10	Unmarried partner of reference person
11	Housemate or roommate
12	Roomer or boarder
13	Other nonrelative of reference person
Panels Prior to 1996	
Revised Relationship to the Household Reference Person (RRP)	Description
1	Household reference person, living with relatives
2	Household reference person, living alone or with nonrelatives
3	Spouse of household reference person
4	Child of household reference person
5	Other relative of household reference person
6	Nonrelative of household reference person, but related to other members of the household
7	Nonrelative of all members of the household

²² Because it is impossible to anticipate all of the different living arrangements found in SIPP sample households, and in some cases more than one rule for identifying a reference person may apply, some interviewer discretion in identifying the reference person is inevitable. For that reason, the resulting choices can sometimes appear somewhat arbitrary to the analyst.

Table 11 -14. ERRP (RRP) Coding for the Same Three-Generation Household When Two Different People Are Designated as the Reference Person in the Topical Module Files

Designated Reference Person	Relationship to the Household Reference Person [ERRP (RRP)]	Meaning of ERRP (RRP) Value
Mother as Household Reference Person		
Mother	1 (1)	Reference person (Reference person)
Daughter	4 (4)	Child of reference person (Child of reference person)
Daughter's son	5 (5)	Grandchild of reference person (Other relative of reference person)
Daughter as Household Reference Person		
Mother	6 (5)	Parent of reference person (Other relative of reference person)
Daughter	1 (1)	Reference person (Reference person)
Daughter's son	4 (4)	Child of reference person (Child of reference person)

Identifying a Person's Spouse, Parent, or Guardian

Four other variables on the topical module files from the 1996+ Panels can be used to describe household and family composition. They are EPNSPOUS, EPNDAD or EPNMOM, and EPNGUARD. These variables identify the person number of the spouse, the father or mother (just one parent is identified in files from Pre-1996 Panels), and guardian of the person, respectively. On the topical module files from Pre-1996 Panels, only two variables are found: PNPT and PNSP, the person numbers of the person's parent and spouse, respectively. In each case, the relative is identified only if she or he is living at the same address as the person.

By building from these variables, the analyst can identify a variety of family configurations. For example, these variables can be used to identify households containing three generations. Table 11-15 displays one household containing a mother and her two children. One child, EPPPNUM = 0102 (PNUM = 102), has a son; the other child, EPPPNUM = 0104 (PNUM = 104), has a spouse.

More About Using the SIPP ID Variables: Identifying Movers

Most of the SIPP topical modules collect information that pertains to a single month generally month four of the wave's core reference period in the 1996+ Panels, and month five (the interview month) for prior panels. However, some topical modules collect information about longer reference periods, most commonly either the previous 4 months (the same period as the core questions but often not

**Table 11 - 15. Identifying Households Containing Three Generations
in the Topical Module Files**

1996+ Panels					
Household Member	Person Number (EPPPNUM)	Recoded Relationship to Household Reference Person (ERRP)	Spouse (EPNSPOUS)	Parent (EPNMOM)	Notes
Mother	0101	1	9999	9999	Mother
Daughter #1	0102	4	9999	0101	Child
Daughter #1's Son	0103	5	9999	0102	Grandchild
Daughter #2	0104	4	0105	0101	Child
Spouse of Daughter #2	0105	8	0104	9999	Spouse of child
Panels Prior to 1996					
Household Member	Person Number (EPPPNUM)	Recoded Relationship to Household Reference Person (ERRP)	Spouse (EPNSPOUS)	Parent (EPNMOM)	Notes
Mother	101	1	999	999	Mother
Daughter #1	102	4	999	101	Child
Daughter #1's Son	103	5	999	102	Grandchild
Daughter #2	104	4	105	101	Child
Spouse of Daughter #2	105	5	104	999	Spouse of child

Note: Value of 999 or 9999 means not applicable.

with monthly resolution), the year prior to the interview (e.g., some items in the child and adult well-being topical modules), or the prior calendar year (e.g., the annual income and retirement accounts topical module of the 1996+ Panels). In instances such as these, it is sometimes useful to know something about household composition during the reference period of the topical module.²³ This section of the Users' Guide is primarily for users who need to know how to access that kind of information. This section may also be helpful to those who wish to gain a better understanding of the SIPP ID variables for other reasons.

When a person moves, the current address field, SHHADID (ADDID), changes. The SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) values remain the same. The first part (two digits in the 1992 Panel and the 1996 Panel, one digit in all others) of SHHADID (ADDID) indicates the wave in which a household is first interviewed at that new address. The remaining digit

²³ For example, a person who joined the SIPP sample in Wave 4 of the 1996+ Panels could not have contributed to the household income (at least not as a household member) of the prior calendar year.

sequentially numbers the households that split into two or more households, as a result of a move to a different location by original sample members. Thus, new addresses in Wave 2 are numbered 021 (21), 022 (22), and so on. New addresses in Wave 3 are numbered 031 (31), 032 (32), and so on.

Table 11-16 shows that persons 0101 (101) and 0102 (102) in the first household are original sample members. Person 0401 (401) moved into the home of persons 0101 (101) and 0102 (102) in Wave 4. In Wave 7, all three of them moved to a new location and were joined by person 0701 (701). In the second household, person 101 is an original sample member who moved to a new location in Wave 3. In the third household, person 0102 (102) is also an original sample member who used to live with persons 0101 (101) and 0103 (103) of the same sample unit ID, but moved to a new location in Wave 3 [to a different location from person 0101 (101)]. In the fourth household, person number 0103 (103) is an original sample member who used to live with persons 0101 (101) and 0102 (102) of the same sample unit ID number. All but two people moved from their original location [i.e., only two people have SHHADID (ADDID) equal to EENTAID (ENTRY)].

The next example (Table 11-17) further illustrates how the ID system works as people move to new addresses, additional people move in with them, and households split. (Users may also find it helpful to review Figure 2-1, which illustrates changes in household composition.)

- In Wave 1, there is a five-person household consisting of a husband, a wife, a daughter, a son, and a cousin. Since this is the first wave, the current address number is 011 (11), indicating address 1 of Wave 1, and the entry address number for each member of the household is the same as the current address number. Since they are assigned in Wave 1, the person numbers are in the 0100 (100) series and numbered sequentially, beginning with 0101 (101).
- During Wave 2, the son joins the Army, moves into the military barracks, and therefore leaves the SIPP sample. For the son's record, person number 0104 (104), the person-month file will contain a Wave 1 record for him and a Wave 2 record containing information (either imputed or provided by proxy) on his characteristics in the months of Wave 2 that he was still in the sample. If he does not return to the sample during the remainder of the panel, there will be no records for him beyond Wave 2.
- During Wave 3, the daughter marries and her husband moves into the household. The current address number where the mother, father, cousin, daughter, and son-in-law live remains the same since it is the same address. The son-in-law's entry address number is 011 (11), since he first enters the SIPP sample at an address coded 011 (11). The person number for the son-in-law is in the 0300 (300) series [0301 (301)] since he joins the SIPP sample in Wave 3.

Table 11 -16. Identifying Movers in the Core Wave Files

1996+ Panels				
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Entry Address ID (EENTAID)	Person Number (EPPPNUM)	Notes
123456789123	071	011	0101	Persons 101 and 102 are the original sample members. Person 401 begins to live with them in Wave 4. All three people move in Wave 7 and person 701 joins them.
123456789123	071	011	0102	
123456789123	071	011	0401	
123456789123	071	071	0701	
321456789123	031	11	0101	Person 101 is an original sample member who moved in Wave 3.
321456789123	032	011	0102	Person 102 is an original sample member who moved in WAVE 3 to a different location from person 0101.
Panels Prior to 1996				
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Entry Address ID (EENTAID)	Person Number (EPPPNUM)	Notes
123456789	71	11	101	Persons 101 and 102 are the original sample members. Person 401 begins to live with them in Wave 4. All three people move in Wave 7 and person 701 joins them.
123456789	71	11	102	
123456789	71	11	401	
123456789	71	71	701	
321456789	31	11	101	Person 101 is an original sample who moved in Wave 3.
321456789	32	11	102	Person 102 is an original sample who moved in Wave 3 to a different location from person 101.

- During Wave 4, the daughter and son-in-law move into a new house. Their current address number changes to 041 (41) to indicate that a new address has been established in Wave 4. Meanwhile, the cousin, who is over age 15, moves in with an uncle.²⁴ The cousin's current address number changes to 042 (42) (i.e., the second new household formed in the fourth wave from this sample unit). The assignment of address number 041 (41) to the daughter and 042 (42) to the cousin is arbitrary-it could be the other way around. The uncle enters the SIPP sample and receives an address number of 042 (42) and an entry address number of 042 (42). The uncle's person number is in the 0400 (400) series [0401 (401)] because he joins the survey in Wave 4.

²⁴ In the 1993 Panel, all original sample members were followed, regardless of age. In all other panels (including the 1996 Panel), only those aged 15 or older were followed when they moved to new addresses.

Table 11-17. Example of Household Changes and Their Effects on the ID Variables in the Core Wave Files

1996+ Panels				
Household Member	Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Entry Address ID (EENTAID)	Person Number (EPPNUM)
Wave 1				
Father	10111103123	011	011	0101
Mother	10111103123	011	011	0102
Daughter	10111103123	011	011	0103
Son	10111103123	011	011	0104
Cousin	10111103123	011	011	0105
Wave 2				
Father	10111103123	011	011	0101
Mother	10111103123	011	011	0102
Daughter	10111103123	011	011	0103
Son	10111103123	011	011	0104
Cousin	10111103123	011	011	0105
Wave 3				
Father	10111103123	011	011	0101
Mother	10111101233	011	011	0102
Daughter	10111103123	011	011	0103
Son-in-Law	10111103123	011	011	0301
Cousin	10111103123	011	011	0105
Wave 4				
Parent's Household				
Father	10111103123	011	011	0101
Mother	10111103123	011	011	0102
Daughter's Household				
Daughter	10111103123	041	011	0103
Son-in-Law	10111103123	041	011	0301
Cousin's Household				
Cousin	10111103123	042	011	0105
Uncle	10111103123	042	042	0401
Wave 10				
Parent's Household				
Father	10111103123	011	011	0101
Mother	10111103123	011	011	0102
Daughter's Household				
Daughter	10111103123	101	011	0103
Son-in-Law	10111103123	101	011	0301
Newborn	10111103123	101	041	1001

(table continues)

Table 11-17. Example of Household Changes and Their Effects on the ID Variables in the Core Wave Files

Prior to 1996 Panel				
Household Member	Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Entry Address ID (EENTAID)	Person Number (EPPPNUM)
Wave 1				
Father	101111103	11	11	0101
Mother	101111103	11	11	0102
Daughter	101111103	11	11	0103
Son	101111103	11	11	0104
Cousin	101111103	11	11	0105
Wave 2				
Father	101111103	11	11	0101
Mother	101111103	11	11	0102
Daughter	101111103	11	11	0103
Son	101111103	11	11	0104
Cousin	101111103	11	11	0105
Wave 3				
Father	101111103	11	11	0101
Mother	101111103	11	11	0102
Daughter	101111103	11	11	0103
Son-in-Law	101111103	11	11	0301
Cousin	101111103	11	11	0105
Wave 4				
Parent's Household				
Father	101111103	11	11	0101
Mother	101111103	11	11	0102
Daughter's Household				
Daughter	101111103	41	11	0103
Son-in-Law	101111103	41	11	0301
Cousin's Household				
Cousin	101111103	42	11	0105
Uncle	101111103	42	11	0401
Wave 10				
Parent's Household				
Father	101111103	11	11	0101
Mother	101111103	11	42	0102
Daughter's Household				
Daughter	101111103	41	11	0103
Son-in-Law	101111103	41	11	0301
Newborn	101111103	41	41	1001

^a Prior to the 1996 Panel, only the 1992 Panel had 10 or more waves. Wave 2 of the 1992 Panel of the core wave files has expanded address and person ID fields (3 and 4 digits, respectively) to accommodate Wave 10 of the 1992 panel.

- No changes in household composition are observed during Waves 5 through 9.
- During Wave 10,²⁵ the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. The newborn's entry address is 041 (41), since that is the current address ID of the daughter and son-in-law at the time of birth. The newborn's person number is 1001, reflecting the fact that the newborn came into the SIPP sample in Wave 10. Meanwhile, the cousin moves to Europe and therefore leaves the SIPP sample. The uncle, even though he did not move to Europe with the cousin, also leaves the SIPP sample because he no longer resides with an original SIPP sample member. Their records are no longer listed.

Prior to the 1996 Panel, there were two extremely rare occasions when the original ID, ENTRY, and PNUM values were modified by the Census Bureau:

1. The first occasion was when two separate sampling units, each containing original sample members, were merged, perhaps because of a marriage. In this situation, one of the original sets of ID and ENTRY values was retained and the other set was changed to agree with that retained set. The person-number values (PNUM) of the changed set were modified further to be between 180 and 199, inclusive.
2. The second occasion was when a household split into two new households (in which each new household gained a new sample person) and later the households recombined. For example, suppose that a married couple separated in Wave 3, each moving in with a sibling. Both siblings were assigned a person number of 301 because they entered the sample in Wave 3 at different addresses (thus, ADDID = 31 and 32). If the husband and wife reunited in Wave 6, and brought the siblings with them, one of the sibling's person numbers would have been changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699, inclusive).

Those two occasions were the only times when ID, ENTRY, and PNUM changed. When it did occur, the old ID variables were stored in the previous wave variables (PWSUID, PWENTRY, and PWPNUM), found only on the core wave files.²⁶

When the merge occurred after the first month of a reference period, the members of the merged household (whose ID variables were modified) were assigned two sets of monthly records in the core wave file. The first set of records contained the original ID information and identified the person as having exited the sample at the time of the merge. The second set contained the new ID information and identified the person as having entered the sample at the time of the merge. When

²⁵ Prior to the 1996 Panel, only the 1992 Panel had more than nine waves.

²⁶ In the 1993 Panel, merged households are identified with the variables PWSUID, PWENTRY, and PWPNUM. Before the 1993 Panel, they were identified with the variables PREV-ID, SC0064, and SC0066.

the merge occurred at the start of the reference period, only the second set of records was retained in the core wave files.

Because merged households were very rare prior to the 1996 Panel, information about them will no longer be carried on the topical module files from the 1996+ Panels. When either of those two kinds of events occur in the 1996+ Panels, one or more original sample members will appear to leave the sample when the merge takes place, and new people will appear to enter the sample when the merged household forms. There is no indication in the data files that the “new” sample members were previously members of the SIPP sample with different ID values.

Topcoding

To protect the confidentiality of SIPP respondents, the Census Bureau topcodes characteristics available on the topical module files that might allow a user to recognize the identity of a SIPP respondent. The topcoding procedures used in the topical module files are similar to those used in the core wave files.²⁷ Generally, topcodes for continuous variables that apply to the total universe include at least 1/2 of 1 percent of all cases. For income variables that apply to subpopulations, topcodes include either 3 percent of the appropriate cases or 1/2 of 1 percent of all cases, whichever is the higher topcode. Any discrete information that is topcoded in the core wave files is topcoded in a consistent manner in the topical module files.

Characteristics that are frequently topcoded in SIPP topical module files include income and expense values, including those for a broad range of assets and liabilities. For example, the following groups of topical module variables appear in Wave 3 of the 1996+ Panels: assets and liabilities, interest earnings, medical expenses, mortgage amounts, other financial assets, real estate, rental properties, stocks and mutual funds, value of business, and work-related expenses and child support paid. The documentation for the variables included in these groups indicates whether the values are topcoded and the value ranges for the variables.

Users should refer to Chapter 10, “*Earnings Topcoding for the 2001 Panel*” and “*Earnings Topcoding for the 2004 Panel*” for more information on topcoding for Panels, 2001 and 2004.

Using Allocation (Imputation) Flags

As described in Chapter 4, the Census Bureau often imputes information when a person does not respond to the survey or to a particular question. A variable of interest may be imputed. In the topical module files prior to the 1996 Panel, there is an allocation (imputation) flag for almost all of

²⁷ Chapter 10 contains a discussion of both the new income topcoding procedures used in the 1996+ Panels core wave files and the income topcoding procedures used in the pre-1996 core wave files. See also Appendix B: SIPP Topcoding Specifications.

the person-level variables. Beginning with the 1996 Panel, there is an allocation (imputation) flag associated with every variable subject to imputation. For example, AEDUCATE is the allocation (imputation) variable that identifies whether EEDUCATE is imputed.

Variables are imputed and the allocation (imputation) flags are set before composite variables are created. For example, if income is imputed for one member of a household, that person's allocation (imputation) flag is set. However, total household income is computed after that imputation; if any household member had any income imputed, total household income is based, in part, on imputed information. There is no direct indication on the records of other household members that any information has been imputed.

Using Weights

The topical module files contain one weight variable, WPFINWGT (FINALWGT). For the 1996+ Panels, this weight is the person cross-sectional weight for the fourth reference month. Prior to 1996, this weight was the person interview month weight for people who provided data for a topical module. It shows the number of people in the population represented by the sample person in the interview month. Chapter 8 of this Guide contains a full discussion of how to use weights in SIPP data files.

The source and accuracy statements that accompany all SIPP topical module files ordered from the Census Bureau provide suggestions on how to use the topical module weight variable. The Source and Accuracy Statement can be found on the SIPP home page under "*Publications*".

Identifying States

For the 1996+ Panels, the variable TFIPSST identifies 45 states and the District of Columbia. The remaining five states are combined as follows:

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

The topical module files from panels prior to the 1996 Panel contain a variable STATE that identifies the state in which the household resides. The variable identifies 41 individual states and the District of Columbia; the nine other states are combined into three groups:

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

3. Alaska, Idaho, Montana, Wyoming.

Even though it is possible to identify most states, SIPP, prior to the 2004 Panel, was not designed to be representative at the state level and should not be used to produce state-level estimates. The state variable is included on the public use files to allow examination of how state-level characteristics affect national estimates. For example, a user could apply the state-specific eligibility criteria for a means-tested program in order to arrive at a national estimate of the number of eligible participants. Because some states are not uniquely identified, some method of allocating the state-specific eligibility rules to sample people in those states would need to be devised.

The 2004 SIPP Panel can be used to produce state estimates. It was designed to produce reliable low-income estimates for the 33 largest states.

Identifying Metropolitan Areas

The topical module files do not contain any variables identifying metropolitan areas. Those needing that information should merge it from the core wave files or the full panel files. Analysts should see Chapters 10 and 12 for discussions of the core wave files and the full panel files, respectively. Chapter 13 discusses how to merge multiple SIPP public use files.