10. Using the Core Wave Files

This chapter discusses procedures for working with Survey of Income and Program Participation (SIPP) core wave data. Specifically, the focus is on SIPP documentation that accompanies the core wave public use files and the data files. The data file structure is described, and detailed explanations are provided about how to use the core wave files to perform common tasks, including (among others):

- Identifying persons, households, families, and program units;

- Understanding the effects of topcoding;

- Using imputation flags; and

- Identifying states and metropolitan areas

Before reading this chapter, users should read Chapter 9 for an introduction to Section II. Analysts using only one core wave file should also read about the use of sample weights (Chapter 8) and the computation of standard errors (Chapter 7). Users who merge data from multiple core wave files, from full panel files, or from topical module files should read Chapter 11 for information about topical module files, Chapter 12 for information about the full panel files for pre-1996 Panels, and Chapter 13 for information about linking SIPP public use files.

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

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

Using the Technical Documentation of the Core Wave Files

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

- The items booklet (for the 1996, 2001, and 2004 Panels);

- The paper survey instrument (for panels prior to the 1996 Panel);

- A glossary of selected terms;

- A cross-walk, mapping reference months into calendar months for each rotation group;

- A source and accuracy statement describing the sample weights and the computation of standard errors; and

- User Notes.

The items booklet shows all questions and their responses. Some questions employ skip patterns (Chapter 3), so users should pay particular attention to which questions were skipped for which respondents. The skip patterns are best understood by consulting the survey instruments. The data dictionary is another way to determine the universe. The instrument screens can be found on the SIPP web site, (http://www.sipp.census.gov/sipp/).

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

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

1.   The definition

2.   The sample universe of the corresponding survey question;

3.   The ranges for all legal values; and

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

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

The data dictionary is formatted for computer processing by user-written programs. As shown in Figure 10-1, a "D" in the first column signifies that the next few lines define the variable: (1) the

variable name; (2) the size (i.e., how many digits it contains); and (3) the starting position. A "U" in the first column signifies that the next words describe the universe.[1] A "V" in the first column indicates that the next number and phrase describe one of the values of the variable. An asterisk in the first column denotes a comment. A period (.) before a word denotes the start of the value label. In the dictionaries for files from the 1996+ Panels, lines beginning with a "T" contain short variable descriptions that can be used by many software packages as variable labels.

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

```
                        Wave 1 of the 1996+ Panels

D EENTAID    3     506
T PE: Address ID of hhld where person entered
      Sample Address ID of the household that this
      person belonged to at the time this
      person first became part of the sample
U All persons
V       11:129 .Entry address ID

D EPPPNUM    4     509
T PE: Person number
   Person number. This field differentiates
      persons within the sample unit. Person
      number is unique within the sample.
U All persons
V       101:1299 .Person number

D EPPINTVW   2     513
T PE: Person's interview status
U All persons
V         1 .Interview (self)
V         2 .Interview (proxy)
V         3 .Noninterview - Type Z
V         4 .Nonintrvw = pseudo Type Z.
V           .Left sample during the
V           .reference period
V         5 .Children under 15 during
V           .reference period
```

(figure continues)

---

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

Figure 10-1. Excerpt from a Data Dictionary for the Core Wave Files (*continued*)

```
               Wave 9 of the 1992 Panel (continued)

D ENTRY       2      457
Edited entry address ID
       Address ID of the household that this
       person belonged to at the time this
       person first became part of the sample
       Range=(11:99)
U All persons, including children

D PNUM 3      459
       Edited person number
       Range=(101:998)
U All persons, including children

D INTVW       1      462
       Person's interview status
       Range=(0:5)
U All persons, including children
V       0 .Not applicable (children under 15)
V       1 .Interview (self)
V       2 .Interview (proxy)
V       3 .Noninterview - Type Z refusal
V       4 .Noninterview - Type Z other
V       5 .Noninterview - left before
V         .interview month
```

Figure 10-2 shows sample SAS and FORTRAN syntax for reading the data described by the codebook fragment in Figure 10-1. Additional SAS program code could be used to associate value labels (SAS "formats") with the variables.

Relationship of the Core Wave Data Files to the SIPP Survey Instrument

Because the core wave data dictionary does not replicate the survey instrument, analysts should keep a few things in mind when using the data:

- The variables on the data files do not correspond one-to-one with the questionnaire items - the variables are listed in a different order, some variables are not included in the core wave files at all, and some variables are created from a combination of other variables;

- The range of possible values of the variables on the data files does not always correspond one-to-one with the response categories shown on the survey instrument or in the data

dictionary; [2]

- The variable name in the data dictionary may not readily indicate the variable's content;[3] and

- The complexity of the skip patterns will not be apparent by simply looking at the data dictionary.[4]

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

Structure of the Core Wave Files

Beginning with the 1990 Panel, the core wave files have been issued in person-month format, with one record per person for each month of the 4-month reference period the person is in the sample.[5] A person who was in the sample for all 4 months of the wave has four records. A person who was in the sample for 1 month has only one record. Records for persons interviewed by proxy are included in the files, as are records for persons for whom the data are imputed. The files also contain records for all children residing with original panel members.

As Table 10-1 illustrates, person number 0101 (101) was in the sample all 4 months, person number 0102 (102) was also in the sample all 4 months, person number 0201 (201) was in the sample for 2 months, and person number 0202 (202) was in the sample for 1 month. Users may find it helpful to review Figure 2-1 (pp. 2-10-2-14), which illustrates movement into and out of the sample.

---

[2] For example, in the 1996+ Panel the response categories on the instrument for CLWRK are (1) a government organization, (2) a private, for-profit company, (3) a nonprofit organization, (4) a family business or farm. The response categories for the corresponding edited variable ECLWRK in the data dictionary are 1 = private for-profit employee, 2 = private not-for-profit employee, 3 = local government worker, 4 = state government worker, 5 = federal government worker, 6 = family worker without pay.

[3] Although an attempt was made in the 1996 Panel to give all variables meaningful names, the eight-character limitation imposed by many software packages places severe constraints on the degree to which this can be done. Prior to the 1996 Panel, the situation was more pronounced since numeric sequencing was used to name variables (e.g., in the paper survey, SE22318 is the variable that indicates the total number of employees working for the second business; in CAI, that variable is TEMPB2). In the 1996+ Panels, variable names beginning with a "T" have been topcoded to protect respondent confidentiality.

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

[5] Prior to the 1990 Panel, core wave files had one record per person. Each record contained four occurrences of each monthly variable. For more information, see earlier editions of the SIPP Users' Guide.

**Figure 10-2. Corresponding SAS and FORTRAN Syntax to Read the Data
from the Core Wave Files (See Figure 10-1 for Data Dictionary)**

Wave 1 of the 1996+ Panels

SAS
| ```
INPUT
@506   EENTAID      3.
       EPPPNUM      4.
       EPPINTVW     2.;

LABEL  EENTAID = "Adrs ID where person entered sample"
       EPPPNUM = "Person number"
       EPPINTVW = "Person's interview status";
``` |

FORTRAN
| ```
       READ(infile,1000) EENTAID, EPPPNUM, EPPINTVW
1000   FORMAT(T506, I3, I4 I2))
``` |

Wave 9 of the 1992 Panel

SAS
| ```
INPUT
  @457 ENTRY  2.
       PNUM   3.
       INTVW  1.;

, ENTRY = "Edited Entry Address ID"
 PNUM =  "Edited Person Number"
 INTVW = "Person's Interview Status";
``` |

FORTRAN
| ```
READ (infile, 1000) ENTRY, PNUM, INTVW
1000   FORMAT(T457, I2, I3, I1)
``` |

Identifying Persons

There are many occasions when a user may need to identify which records belong to which individual in the SIPP data files. This need arises, for example, when:

- Merging data from topical module or full panel files to core wave files;

- Combining data from two or more core wave files;

- Linking husbands and wives;

- Linking parents and children; and

- Identifying which person received government transfer income on behalf of the family.

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

1996+ Panels					
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Rotation Person Number (EPPPNUM)	Reference Group (SROTATION)	Calendar Month (SREFMON)	Month (RHCALMN)
123451000123	011	0101	2	1	2
123451000123	011	0101	2	2	3
123451000123	011	0101	2	3	4
123451000123	011	0101	2	4	5
123451000123	011	0102	2	1	2
123451000123	011	0102	2	2	3
123451000123	011	0102	2	3	4
123451000123	011	0102	2	4	5
123451000123	011	0201	2	1	2
123451000123	021	0201	2	2	3
123451000123	022	0202	2	4	5
Prior to the 1996 Panel					
Sample Unit ID (SUID)	Current Address ID (ADDID)	Person Number (PNUM)	Rotation Group (ROT)	Reference Month (REFMTH)	Calendar Month (MONTH)
123451000	11	101	2	1	2
123451000	11	101	2	2	3
123451000	11	101	2	3	4
123451000	11	101	2	4	5
123451000	11	102	2	1	2
123451000	11	102	2	2	3
123451000	11	102	2	3	4
123451000	11	102	2	4	5
123451000	21	201	2	1	2
123451000	21	201	2	2	3
123451000	22	202	2	4	5

To uniquely identify a person in the core wave files, analysts should employ the three variables shown in Table 10-2. Users should note that for Panels, 1996+, the entry address ID is no longer needed for unique identification. Its continued use will not create any problems; it is simply redundant information. That is a change from earlier panels in which the entry address ID was key to uniquely identifying persons.

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

Variable Name	Description
SSUID (SUID)	Sample unit ID
EENTAID (ENTRY)	Entry address ID (Not required for identification in the 1996 Panel)
EPPPNUM (PNUM)	Person number

The variables in Table 10-2 have the following characteristics:

- SSUID (SUID) uniquely identifies each initially sampled dwelling unit.[6] Every person in a core wave file was either a member of one of those units (an original sample member) or lives with someone who was a member of an initially sampled dwelling unit. A person's connection to that unit is an attribute of that person and does not change over time.[7] This means that as people move from address to address, their SSUID (SUID) stays the same. As new people join the homes of original sample members, they receive the SSUID (SUID) of the original sample members.

- EENTAID (ENTRY) identifies the address where the person lived at the time she or he was first interviewed. It does not change even if the person moves.[8] Prior to the 1996 Panel, it used in conjunction with the person number and sample unit ID to uniquely identify persons within the sampling unit. It is not needed to uniquely identify persons in the 1996 panel. Values for this variable are unique only within sample units. The entry address ID has two components. The first part of the ID number (two digits in the 1992 and 1996 Panels, and one digit in all others) identifies the wave in which SIPP interviews were first conducted at the address. The second part of the number (one digit in all panels) sequentially numbers addresses within a sample unit [SSUID (SUID)] that enter the sample in the same wave. See Chapter 9 for a more complete discussion.

- Prior to the 1996 Panel, PNUM uniquely identified a person within the sample unit and entry address ID. In the 1996+ Panel, EPPPNUM uniquely identifies a person within the sample unit. EPPPNUM (PNUM) does not change even if the person moves.[9] The first part of EPPPNUM (PNUM) (two digits in the 1992 and 1996 Panels, one digit in all others) indicates the wave in which the person was first interviewed.[10] The remaining two digits are sequentially assigned within the household. Thus, original sample members are assigned person numbers ranging from 100 to 199. Individuals who enter the SIPP sample in Wave 2 are assigned a person number ranging from 200 to 299. Those who enter in Wave 10 are assigned person numbers ranging from 1001 to 1099.

Table 10-3 illustrates how the combination of SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) uniquely identifies people and provides information about when they first entered the SIPP sample. In this example, there are eight individuals: five are original sample

---

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

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

[8] See footnote 6.

[9] See footnote 6.

[10] For Wave 10 of the 1992 Panel and for the 1996 Panel, the first two digits of PNUM instead of the first digit identify the wave in which the person entered sample.

members, one person joined the SIPP sample in Wave 3, one joined in Wave 4, and another joined in Wave 7. Note that the person who joined the sample in Wave 3 (pre-1996 Panel) was assigned a person number of 301, but an entry address ID of 21 (not 31). That is because the first part of the entry address ID indicates the wave in which that address was first occupied by any SIPP sample member, which is not necessarily the wave in which a given member entered the sample.

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

1996+ Panels			
SampleUnit ID (SSUID)	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	011	0101	Original sample member
123456789123	011	0102	Original sample member
123456789123	021	0301	Enters SIPP sample in Wave 3
123456789123	011	0401	Enters SIPP sample in Wave 4
123456789123	071	0701	Enters SIPP sample in Wave 7
321456789123	011	0101	Original sample member
321456789123	011	0102	Original sample member
321456789123	011	0103	Original sample member
Prior to the 1996 Panel			
Sample Unit ID (SSUID)	**Entry Address ID (EENTAID)**	**Person Number (EPPPNUM)**	**Notes**
123456789	11	101	Original sample member
123456789	11	102	Original sample member
123456789	21	301	Enters SIPP sample in Wave 3
123456789	11	401	Enters SIPP sample in Wave 4
123456789	71	701	Enters SIPP sample in Wave 7
321456789	11	101	Original sample member
321456789	11	102	Original sample member
321456789	11	103	Original sample member

Identifying Households

The term *household*, as used in Census Bureau publications, refers to a group of persons who occupy a housing unit. A house, an apartment or other group of rooms, or a single room is regarded as a housing unit if it is occupied or intended for occupancy as separate living quarters. That is, the occupants do not live and eat with any other persons in the structure and there is direct access from the outside or through a common hall. A group of friends sharing an apartment constitutes a household. Noninstitutional group quarters, such as rooming and boarding houses, college dormitories, convents, and monasteries, are classified as group quarters rather than households. To uniquely identify a household or group quarters in the core wave files, analysts should use the two variables shown in Table 10-4.

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

Variable Name	Description
SSUID (SUID)	Sample unit ID
SHHADID (ADDID)	Current address ID

People with the same SSUID (SUID) and SHHADID (ADDID) values live in the same household (or group quarters). The six individuals in Table 10-5 make up three households. The first household contains the first four individuals. The second household contains one person. The third household contains one person.

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

1996+ Panels			
Sample Unit ID (SSUID)	**Current Address ID (SHHADID)**	**Person Number (EPPPNUM)**	**Notes**
123456789123	071	0101	Four persons in this household
123456789123	071	0102	
123456789123	071	0401	
123456789123	071	0701	
321456789123	031	0101	One person in this household
321456789123	032	0102	One person in this household
Prior to the 1996 Panel			
Sample Unit ID (SUID)	**Current Address ID (ADDID)**	**Person Number (PNUM)**	**Notes**
123456789	71	101	
123456789	71	102	Four persons in this household
123456789	71	401	
123456789	71	701	
321456789	31	101	One person in this household
321456789	32	102	One person in this household

Each household contains one reference person. The household reference person is the person in whose name the home is owned or rented. If the house is owned or rented jointly by more than one person (such as a married couple or some roommate situations), any of those people may be listed as the "reference person." Users may find it helpful to refer to Figure 2-1 (pp. 2-10-2-14), which illustrates the concepts of household and changes in household composition.

Identifying Families

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

There are several types of families that the Census Bureau distinguishes:

- A *primary family* is a family containing the household reference person and all of his or her relatives. This means that a household composed of a husband and wife, their son, and their son's wife (i.e., the daughter-in-law) is classified as a primary family containing four people.

- A *related subfamily* is a nuclear family that is related to but does not include the household reference person. For example, the son and his wife (i.e., the daughter-in-law) in the preceding example are a related subfamily.

- An *unrelated subfamily* (sometimes called a secondary family) is a nuclear family that is not related to the household reference person. Thus, a husband and wife who live in a friend's house are classified as an unrelated subfamily. A mother and daughter who live in the mother's boyfriend's apartment are classified as an unrelated subfamily.

- A *primary individual* is a household reference person who lives alone or lives with only nonrelatives. Primary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.

- A *secondary individual* is not a household reference person and is not related to any other people in the household. Secondary individuals are sometimes treated by the Census Bureau as families with only one person and are referred to as pseudo-families.

To uniquely identify a family, analysts should use the variables shown in Table 10-6.

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

Variable Name	Description
SSUID (SUID)	Sample Unit ID
SHHADID (ADDID)	Current Address ID
and one of the following:	
RFID (FID)	Family ID
RFID2 (FID2)	Family ID, excluding related subfamily members
RSID (SID)	Family ID, for both related and unrelated subfamilies.

The Census Bureau has two principal methods for distinguishing families.

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

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

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

Table 10-7 illustrates the difference between the RFID (FID), RFID2 (FID2), and RSID (SID) variables. Those variables are set to new numbers in each month. For example, a mother, a father, and a child would be family 1 with RFID (FID) = 1 in month 1, RFID (FID) = 2 in month 2, RFID (FID) = 3 in month 3, and RFID (FID) = 4 in month 4, even though family composition remains the same. The first household in the table contains a primary family of five people. The primary family contains two related subfamilies. RFID (FID) and RFID2 (FID2) mask the fact that there are two related subfamilies; only RSID (SID) provides that information: RSID (SID) has nonzero values for those related subfamilies.

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

The needs of the analysis will help to determine which family classification to use. The following guide may prove helpful:

- To group people into families in the same way that the Census Bureau does, use SSUID (SUID), SHHADID (ADDID), and RFID (FID).

- To analyze people in related subfamilies, include only those records with RSID (SID) greater than zero and ESFTYPE (FTYPE) equal to 2.

- To analyze all families and to keep subfamilies separate from primary families, use SSUID (SUID), SHHADID (ADDID), RFID2 (FID2), and RSID (SID) to uniquely identify each family.

Table 10-7. Uniquely Identifying Families in the Core Wave Files

1996+ Panels								
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Related Person Number (EPPPNUM)	Family ID, Including Related Subfamily (RFID)	Family ID, Excluding Related Subfamily (RFID2)	Related Subfamily ID (RSID)	Family Type (EFTYPE)[a]	Subfamily Type (ESFTYPE)	Notes
110011111123	011	0101	1	1	0	1	0	This household contains a primary family of five people. The primary family contains two subfamilies.
110011111123	011	0102	1	0	2	1	2	
110011111123	011	0103	1	0	2	1	2	
110011111123	011	0104	1	0	3	1	2	
110011111123	011	0105	1	0	3	1	2	
1100777777723	011	101	1	1	0	1	0	Three households formed by people who were originally members of the same originally sampled household (SSUID of 11007777723). Two subfamilies split off from the original household to become two new primary families at addresses 21 and 22.
1100777777723	021	102	1	1	0	1	0	
1100777777723	021	103	1	1	0	1	0	
1100777777723	022	104	1	1	0	1	0	
1100777777723	022	105	1	1	0	1	0	
12210000123	011	0101	1	1	0	1	0	This household contains a primary family and two unrelated subfamilies.
12210000123	011	0104	1	1	0	1	0	
12210000123	011	0305	2	2	0	3	0	
12210000123	011	0306	2	2	0	3	0	
12210000123	011	0307	3	3	0	3	0	
12210000123	011	0308	3	3	0	3	0	
55555555123	021	101	1	1	0	4	0	This household contains a primary individual and an unrelated subfamily
55555555123	021	201	2	2	0	3	0	
55555555123	021	202	2	2	0	3	0	
55555555123	021	203	2	2	0	3	0	
61000000123	032	0101	1	1	0	4	0	Primary Individual
897454644123	011	0101	1	1	0	5	0	Group Quarters with two secondary individuals.
897454644123	011	0102	2	2	0	5	0	

[a] EFTYPE = 1 means the person belongs to a primary family (including related subfamily members). EFTYPE = 3 means the person belongs to an unrelated subfamily. EFTYPE = 4 means the person is a primary individual. EFTYPE = 5 means the person is a secondary individual.

Table 10-7. Uniquely Identifying Families in the Core Wave Files (continued)

Sample Unit ID (SUID)	Current Address ID (ADDID)	Related Person Number (PNUM)	Family ID, Including Related Subfamily (RFID)	Family ID, Excluding Related Subfamily (RFID2)	Related Subfamily ID (RSID)	Family Type (FAMTYP)[b]	Subfamily Type (ESFTYPE)	Notes
Pre -1996 Panels								
110011111	11	101	1	1	0	1		This household contains a primary family of five people. The primary family contains two subfamilies.
110011111	11	102	1	0	2	1		
110011111	11	103	1	0	2	1		
110011111	11	104	1	0	3	1		
110011111	11	105	1	0	3	1		
110077777	11	101	1	1	0	1	0	Three households formed by people who ere originally members of the same originally sampled household (SSUID of 11007777723). Two subfamilies split off from the original household to become two new primary families at addresses 21 and 22.
110077777	21	102	1	1	0	1	0	
110077777	21	103	1	1	0	1	0	
110077777	22	104	1	1	0	1	0	
110077777	22	105	1	1	0	1	0	
122100000	11	101	1	1	0	1		This household contains a primary family and two unrelated subfamilies.
122100000	11	104	1	1	0	1		
122100000	11	305	2	2	0	3		
122100000	11	306	2	2	0	3		
122100000	11	307	3	3	0	3		
122100000	11	308	3	3	0	3		
55555555	21	101	1	1	0	4		This household contains a primary individual and an unrelated subfamily
55555555	21	201	2	2	0	3		
55555555	21	202	2	2	0	3		
55555555	21	203	2	2	0	3		
61000000	32	101	1	1	0	4		Primary Individual
89745464	11	101	1	1	0	5		Group Quarters with two secondary individuals.
89745464	11	102	2	2	0	5		

[b] FAMTYP = 1 means the person belongs to a primary family (including related subfamily members). FAMTYP = 3 means the person belongs to an unrelated subfamily. FAMTYP = 4 means the person is a primary individual. FAMTYP = 5 means the person is a secondary individual.

Other Variables Describing Household and Family Composition

Table 10-8 shows the primary core wave variables summarizing household and family composition.[11]

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

Variable Name		Description
1996+ Panels	**Prior to the 1996 Panel**	
RHNF	HNF	Number of families subfamilies and pseudo-families in household
RHNFAM	HNFAM	Number of families and pseudo-families but excluding related subfamilies in household
RHNSF	HNSF	Number of related subfamilies in household
EHREFPER	HREFPER	Household reference person (ENTRY concatenated with PNUM)
EHHNUMPP	HNP	Number of persons in household
RHTYPE	HTYPE	Type of household (e.g. married-couple family, male householder family, etc.)
EFREFPER	FREFPER	Family reference person (ENTRY concatenated with PNUM)
EFTYPE	FTYPE	Type of family (e.g. primary family, unrelated subfamily, etc.)
EFKIND	FKIND	Head of family (e.g. husband and wife, male reference person, etc.)
ESFT	FAMTYP	Type of family to which this person belongs (e.g., primary family, related subfamily, etc.)
ESFR[a]	FAMREL	Family relationship (e.g. reference person spouse of family reference person, child of family reference person, etc.)
ERRP	RRP	Recoded relationship to the household reference person (e.g. household reference person living with relatives, child of household reference person, etc.)
Not a variable for the 1996 Panel	RRPU	Unedited relationship to the household reference person (e.g. stepchild of household reference person, grandchild of household reference person, etc.)
EPNSPOUS	PNSP	Person number of spouse
EPNGUARD	PNGDU	Person number of guardian
EPNMOM		Person number of mother
EPNDAD		Person number of father
	PNPT	Person number of parent

[a] ESFR (edited subfamily relationship) is defined the same as FAMREL, but it applies only to subfamilies (both related and unrelated).

---

[11] Detailed information about the relationships between members is collected in the Household Relationships topical module (see Chapter 3 for a discussion of topical module content). See those data for extensive information about household composition.

Identifying Household and Family Reference Persons

The EHREFPER (HREFPER) variable's value identifies the household reference person. As explained in Chapter 2, the household reference person is the owner or renter of record. Prior to the 1996 Panel, the variable identified the household reference person by concatenating ENTRY with PNUM. For the 1996+ Panels, the variable simply contains the person number of the household reference person (EHREFPER = EPPPNUM). Prior to the 1996 Panel, the household reference person was the one for whom:

- HREFPER = ENTRY * 1000 + PNUM (for Waves 1-9) or

- HREFPER = ENTRY * 10000 + PNUM (for Wave 10 of the 1992 Panel)

The EFREFPER (FREFPER) variable identifies the family reference person. For the 1996 Panel, the variable simply contains the person number of the family reference person (EFREFPER = EPPPNUM). Prior to the 1996 Panel, the family reference person was the one for whom:

- FREFPER = ENTRY * 1000 + PNUM (for Waves 1-9) or

- REFPER = ENTRY * 10000 + PNUM (for Wave 10 of the 1992 Panel)

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

For the 1996+ Panels, ERRP describes how each person is related to the household reference person. As seen in Table 10-9, the new variable provides information about several household relationship categories that were not available from earlier panels. However, as in earlier panels, this variable summarizes the relationship to the household reference person, not to the family reference person.

Prior to the 1996 Panel, both edited and unedited versions of the RRP variable were included on the core wave files. As shown in Table 10-10, RRP (the edited version of the variable) summarized the values of RRPU (the unedited variable). The RRPU variable can distinguish whether someone is a grandchild, stepchild, foster child, or natural/adopted child of the *household* reference person. What it cannot do, however, is distinguish the type of child within each family: RRPU is the relationship to the *household* reference person, not the relationship to the *family* reference person. For example, using records with RRPU = 6 will *not* identify all foster children, because some could be in an unrelated subfamily. The variable FAMREL summarizes the relationship of the person to the family reference person (as reference person of family, spouse, or child).

Table 10-9. The ERRP Variable in the 1996+ Core Wave Files

Edited Relationship to the Household Reference Person (ERRP)	Description
1	Household reference person, living with relatives
2	Household reference person, living alone or with nonrelatives
3	Spouse of household reference person
4	Child of household reference person
5	Grandchild of household reference person
6	Parent of household reference person
7	Brother or sister of household reference person
8	Other relative of household reference person
9	Foster child of household reference person
10	Unmarried partner of household reference person
11	Housemate or roommate
12	Roomer or boarder
13	Other nonrelative of household reference person

The ERRP (RRP) variable contains summary information about each person's relationship to the household reference person. Analysts should keep in mind that the household description depends upon the identity of the household reference person. For example, the household in Table 10-11 contains a mother, her daughter, and her daughter's son. If the mother is the household reference person [ERRP=1 (RRP=1)], her daughter is listed as a child of the household reference person [ERRP=4 (RRP=4)], and the daughter's son is listed as a grandchild of the reference person in the 1996 Panel (ERRP=5), but as another relative of the household reference person in earlier panels (RRP=5, but the same value has a different meaning from that of the 1996 Panel variable). If the daughter is the reference person, her son is listed as a child of the household reference person (RRP=4), and her mother is listed as the parent of the reference person in the 1996 Panel (ERRP= 6), but as another relative of the household reference person in earlier panels (RRP=5).[12] Users should note that the identity of the household reference person could change from one month to the next; thus, the household description could also change.

---

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

**Table 10-10. Comparison of RRP and RRPU Variables of the
Core Wave Files Prior to the 1996+ Panels**

Edited Relationship to the Household Reference Person (RRP)	Description	Relationship to the Household Reference Person (RRPU)	Notes
1	Household reference person, living with relatives	1	Same as code 1 under RRP
2	Household reference person, living alone or with nonrelatives	2	Same as code 2 under RRP
3	Spouse of household reference person	3	Same as code 3 under RRP
4	Child of household reference person	4	Natural/adopted child of household reference person
		5	Stepchild of household reference person
5	Other relative of household reference person	7	Grandchild of household reference person
		8	Parent of household reference person
		9	Brother/sister of household reference person
		10	Other relative of household reference person
6	Nonrelative of household reference person, but related to other members of the household	11	Same as code 6 under RRP
7	Nonrelative of all members of the household	6	Foster child of household reference person
		12	Partner/roommate of household reference person
		13	Other type of nonrelative of household reference person

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

1996+ Panels		
Household Member	**Relationship to Household Reference Person (ERRP)**	**Notes**
Mother as Household Reference Person		
Mother	1	Reference person
Daughter	4	Child of reference person
Daughter's son	5	Grandchild of reference person
Daughter as Household Reference Person		
Daughter	1	Reference person
Daughter's son	4	Child of reference person
Mother	6	Parent of reference person
Panels Prior to 1996		
Household Member	**Relationship to Household Reference Person (RRP)**	**Notes**
Mother as Household Reference Person		
Mother	1	Reference person
Daughter	4	Child of reference person
Daughter's son	5	Other relative of reference person
Daughter as Household Reference Person		
Daughter	1	Reference person
Daughter's son	4	Child of reference person
Mother	5	Other relative of reference person

Identifying a Person's Spouse, Parent, or Guardian

Four other variables on the core wave files (three prior to the 1996 Panel) can also be used to describe household and family composition. They are EPNSPOUS (PNSP), EPNDAD or EPNMOM (PNPT), and EPNGUARD (PNGDU). These variables identify the person number of the spouse, the father or mother (just one parent is identified in files from panels prior to 1996), and guardian of the person, respectively. In each case, the relative is identified only if she or he is living at the same address as the person. By building from these variables, analysts can identify a variety of family configurations. For example, these variables can be used to identify households containing three generations. Table 10-12 displays one household containing a mother and her two children. One child, EPPPNUM=0102 (PNUM=0102), has a son, and the other child, EPPPNUM=0104 (PNUM=0104), has a spouse.
.

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

1996+ Panels					
Household Member	**Person Number (EPPPNUM)**	**Recoded Relationship to Household 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 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.

Using Family-Level Income Variables

The core wave files contain a number of family-level income variables. The family income variables on these files include the income of all related subfamily members. *In other words, primary family members, including related subfamily members, are treated as one family by the Census Bureau in calculating family-level income amounts.* The core wave files also contain related subfamily income variables. These variables pool the income of all persons who are members of the same related subfamily.

Table 10-13 illustrates how the family income variables on the core wave files include the income of related subfamily members. From the previous example of a primary family of five people, the primary family contains two related subfamilies. Total family income, TFTOTINC (FTOTINC), is $4,200. The first related subfamily has a total income, TSTOTINC (STOTINC), of $1,000. The second related subfamily has $2,000 in total income.

More About Using the SIPP ID Variables:
Identifying Movers

When a person moves, the current address field, SHHADID (ADDID), changes. The SSUID (SUID), EENTAID (ENTRY), and EPPPNUM (PNUM) values remain the same. The first part (two digits in the 1992 Panel and the 1996+ Panels, one digit in all others) of SHHADID (ADDID) indicate(s) the wave in which a household is first interviewed at that new address. The remaining digits sequentially number the households that split into two or more households, as a result of a move to a different location by original sample members. Thus, new addresses in Wave 2 are numbered 021 (21), 022 (22), and so on. New addresses in Wave 3 are numbered 031 (31), 032 (32), and so on.

Table 10-14 shows that persons 0101 (101) and 0102 (102) in the first household are original sample members. Person 0401 (401) moved into the home of persons 0101 (101) and 0102 (102) in Wave 4. In Wave 7, all three of them moved to a new location and were joined by person 0701 (701). In the second household, person 101 is an original sample member who moved to a new location in Wave 3. In the third household, person 0102 (102) is an original sample member who used to live with persons 0101 (101) and 0103 (103) of the same sample unit ID, but moved to a new location in Wave 3 [to a different location from person 0101 (101)]. In the fourth household, person number 0103 (103) is an original sample member who used to live with persons 0101 (101) and 0102 (102) of the same sample unit ID number. All but two people moved from their original location [i.e., only two people have SHHADID (ADDID) equal to EENTAID (ENTRY)].

The next example (Table 10-15) further illustrates how the ID system works as people move to new addresses, additional people move in with them, and households split. A review of Figure 2-1 may help in understanding the various household changes.

- In Wave 1, there is a five-person household consisting of a husband, wife, daughter, son, and cousin. Since this is the first wave, the current address number is 011 (11), indicating address 1 of Wave 1, and the entry address number for each member of the household is the same as the current address number. Since they are assigned in Wave 1, the person numbers are in the 0100 (100) series and are numbered sequentially, beginning with 0101(101).

- 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.

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

1996+ Panels									
Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Person Number (EPPPNUM)	Family ID, Including Subfamily (RFID)	Subfamily ID (RSID)	Number of Persons in Family (EFNP)	Total Family Income (TFTOTINC)	Number of Persons in Related Subfamily (EFNP)	Total Related Subfamily Income (TSTOTINC)	Total Primary Family Income Net or Related Subfamily
110011111123	11	0101	2	0	5	$4,200	0	$0	$1,200
110011111123	11	0102	2	2	5	$4,200	2	$1,000	NA
110011111123	11	0103	2	2	5	$4,200	2	$1,000	NA
110011111123	11	0104	2	3	5	$4,200	2	$2,000	NA
110011111123	11	0105	2	3	5	$4,200	2	$2,000	NA
Prior to the 1996 Panel									
Sample Unit ID (SUID)	Current Address ID (ADDID)	Person Number (PNUM)	Family ID, Including Subfamily (FID)	Subfamily ID (SID)	Number of Persons in Family (FNP)	Total Family Income (FTOTINC)	Number of Persons in Related Subfamily (SNP)	Total Related Subfamily Income (STOTINC)	Total Primary Family Income Net or Related Subfamily
11001111	11	101	2	0	5	$4,200	0	$0	$1,200
11001111	11	102	2	2	5	$4,200	2	$1,000	NA
11001111	11	103	2	2	5	$4,200	2	$1,000	NA
11001111	11	104	2	3	5	$4,200	2	$2,000	NA
11001111	11	105	2	3	5	$4,200	2	$2,000	NA

Note: NA equals not applicable.

Table 10-14. 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 0101 and 0102 are the original sample members. Person 0401 begins to live with them in Wave 4. All three people move in Wave 7 and person
123456789123	071	011	0102	
123456789123	071	011	0401	
123456789123	071	071	0701	
321456789123	031	011	0101	Person 0101 is an original sample member who moved in Wave 3.
321456789123	032	011	0102	Person 0102 is an original sample member who moved in Wave 3 to a different location from person 0101.
Prior to the 1996 Panel				
Sample Unit ID (SUID)	Current Address ID (ADDID)	Entry Address ID (ENTRY)	Person Number (PNUM)	Notes
123456789	71	11	101	Persons 101 and 102 are the original sample members. Person 401 begins to live with them in Wave 4. All three people move in Wave 7 and person 701 joins them.
123456789	71	11	102	
123456789	71	11	401	
123456789	71	71	701	
321456789	31	11	101	Person 101 is an original sample member who moved in Wave 3
321456789	32	11	102	Person 102 is an original sample member who moved in Wave 3 to a different location from person 101.

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

Household Members	Sample Unit ID (SSUID)	Current Address ID (SHHADID)	Entry Address ID (EENTAID)	Person Number (EPPPNUM)
1996+ Panels				
Wave 1				
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
Daughter	101111103123	011	011	0103
Son	101111103123	011	011	0104
Cousin	101111103123	011	011	0105
Wave 2				
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
Daughter	101111103123	011	011	0103
Son	101111103123	011	011	0104
Cousin	101111103123	011	011	0105
Wave 3				
Father	101111103123	011	011	0101
Mother	101111101233	011	011	0102
Daughter	101111103123	011	011	0103
Son-in-Law	101111103123	011	011	0301
Cousin	101111103123	011	011	0105
Wave 4	**Parent's Household**			
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
	Daughter's Household			
Daughter	101111103123	041	011	0103
Son-in-Law	101111103123	041	011	0301
	Cousin's Household			
Cousin	101111103123	042	011	0105
Uncle	101111103123	042	042	0401
Wave 10	**Parent's Household**			
Father	101111103123	011	011	0101
Mother	101111103123	011	011	0102
	Daughter's Household			
Daughter	101111103123	101	011	0103
Son-in-Law	101111103123	101	011	0301
Newborn	101111103123	101	041	1001

(table continues)

Table 10-15. Example of Household Changes and Their Effects on the ID Variables of the Core Wave Files (*continued*)

Household Members	Sample Unit ID (SUID)	Current Address ID (ADDID)	Entry Address ID (ENTRY)	Person Number (PNUM)
1996+ Panels				
Wave 1				
Father	101111103	011	011	0101
Mother	101111103	011	011	0102
Daughter	101111103	011	011	0103
Son	101111103	011	011	0104
Cousin	101111103	011	011	0105
Wave 2				
Father	101111103	011	011	0101
Mother	101111103	011	011	0102
Daughter	101111103	011	011	0103
Son	101111103	011	011	0104
Cousin	101111103	011	011	0105
Wave 3				
Father	101111103	011	011	0101
Mother	101111103	011	011	0102
Daughter	101111103	011	011	0103
Son-in-Law	101111103	011	011	0301
Cousin	101111103	011	011	0105
Wave 4	**Parent's Household**			
Father	101111103	011	011	0101
Mother	101111103	011	011	0102
	Daughter's Household			
Daughter	101111103	041	011	0103
Son-in-Law	101111103	041	011	0301
	Cousin's Household			
Cousin	101111103	042	011	0105
Uncle	101111103	042	042	0401
Wave 10[a]	**Parent's Household**			
Father	101111103	011	011	0101
Mother	101111103	011	011	0102
	Daughter's Household			
Daughter	101111103	041	011	0103
Son-in-Law	101111103	041	011	0301
Newborn	101111103	041	041	01001

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

- During Wave 3, the daughter marries and her husband moves into the household. The current address number where the mother, father, cousin, daughter, and son-in-law live remains the same since it is the same address. The son-in-law's entry address number is 011 (11), since he first enters the SIPP sample at an address coded 011 (11). The person number for the son-in-law is in the 0300 (300) series [0301 (301)] since he joins the SIPP sample in Wave 3.

- During Wave 4, the daughter and son-in-law move into a new house. Their current address number changes to 041 (41) to indicate that a new address has been established in Wave 4. Meanwhile, the cousin, who is over age 15, moves in with an uncle.[13] The cousin's current address number changes to 042 (42) (i.e., the second new household formed in the fourth wave from this sample unit). The assignment of address number 041 (41) to the daughter and 2 (42) to the cousin is arbitrary-it could be the other way around. The uncle enters the SIPP sample and receives an address number of 042 (42) and an entry address number of 042 (42). The uncle's person number is in the 0400 (400) series [0401 (401)], since he joins the survey in Wave 4.

- No changes in household composition are observed during Waves 5-9.

- During Wave 10,[14] the daughter and son-in-law have a baby. This new sample member is assigned the sample unit ID of the daughter and son-in-law. The newborn's entry address is 041 (41) because that is the current address ID of the daughter and son-in-law at the time of birth. The newborn's person number is 1001, reflecting the fact that the newborn came into the SIPP sample in Wave 10. Meanwhile, the cousin moves to Europe and therefore leaves the SIPP sample. The uncle, even though he did not move to Europe with the cousin, also leaves the SIPP sample because he no longer resides with an original SIPP sample member. Their records are no longer listed.

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

1. The first occasion was when two separate sampling units, each containing original sample members, were merged, perhaps because of a marriage. In this situation, one of the original sets of SUID and ENTRY values was retained and the other set was changed to agree with that retained set. The person-number values (PNUM) of the changed set were modified further to be between 180 and 199, inclusive.

---

[13] In the 1993 Panel, all original sample members were followed, no matter what their age. In all other panels, only those age 15 or older were followed when they moved to new addresses.

[14] Prior to the 1996 Panel, only the 1992 Panel had 10 or more waves.

2.   The second occasion was when a household split into two new households (in which each new household gained a new sample person) and later the households recombined. For example, suppose that a married couple separated in Wave 3, each moving in with a sibling. Both siblings were assigned a person number of 301 because they entered the sample in Wave 3 at different addresses (thus, ADDID = 31 and 32). If the husband and wife reunited in Wave 6, bringing the siblings with them, one sibling's person number would have been changed. In this case, one of the siblings would have a person number of 301 and the other would have a person number of 680 (or some number between 680 and 699, inclusive).

Those two occasions were the only times when SUID, ENTRY, and PNUM changed. When it did occur, the old ID variables were stored in the previous wave variables (PWSUID, PWENTRY, and PWPNUM).[15]

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

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

Identifying Program Units

Besides household and family composition, the core wave files contain detailed information about participation in health insurance and various government transfer programs. For most programs, three characteristics are recorded (Table 10-16):

1. Whether the person is covered;

2. Who received the income or benefit; and

3. The amount of the income or benefit.

---

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

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

Program	Coverage	Authorized Recipient	Recipiency	Amount
Panels 1996+				
Program	**Coverage**	**Authorized Recipient**	**Recipiency**	**Amount**
Social Security-Adults	RCUTYP01	RCUOWN01	ER01A	T01AMTA
Social Security-Children			ER01K	T01AMTK
Railroad Retirement-Adults			ER02	T02AMT
Federal Supplemental Security Income	RCUTYP03	RCUOWN03	ER03	T03AMT
Veteran's Benefits	RCUTYP08	RCUOWN08	ER08	T08AMT
Aid to Families with Dependent Children/ Temporary Assistance for Needy Families[a]	RCUTYP20	RCUOWN20	ER20	T20AMT
General Assistance	RCUTYP21	RCUOWN21	ER21	T21AMT
Foster Child Care	RCUTYP23	RCUOWN23	ER23	T23AMT
Other Welfare	RCUTYP24	RCUOWN24	ER24	T24AMT
Women, Infants and Children (WIC)	RCUTYP25	RCUOWN25	ER25	T25AMT
Food Stamps	RCUTYP27	RCUOWN27	ER27	T27AMT
Medicare		ECRMTH		
Medicaid	RCUTYP57	RCUOWN57	ER57	
CHAMPUS		RCHAMPM		
Other Health Insurance	RCUTYP58	RCUOWN58	ER58	
Panels Prior to 1996				
Program	**Coverage**	**Authorized Recipient**	**Recipiency**	**Amount**
Social Security-Adults	SOCSEC	SSPNUM	R01A	S01AMTA
Social Security-Children			R01K	S01AMTK
Railroad Retirement-Adults	RAILRD	RRPNUM	R02A	S02AMTA
Railroad Retirement-Children			R02K	S02AMTK
Federal Supplemental Security Income	SSICOVRG[b]		R03	S03AMT
Veteran's Benefits	VETS	VETNUM	R08	S08AMT
Aid to Families with Dependent Children	AFDC	AFDCPNUM	R20	S20AMT
General Assistance	GENASST	GAPNUM	R21	S21AMT
Foster Child Care	FOSTKID	FKPNUM	R23	S23AMT
Other Welfare	OTHWELF	OWPNUM	R24	S24AMT
Women, Infants and Children (WIC)	WICCOV	WICPNUM	R25	WICVAL
Food Stamps	FOODSTMP	FSPNUM	R27	S27AMT
Medicare	CARECOV			
Medicaid	CAIDCOV	MCDPNUM		
CHAMPUS	CHAMP	CHPNUM		
Other Health Insurance	HIIND	HIPNUM		

[a] In August 1996, the Personal Responsibility and Work Opportunity Reconciliation Act was signed into law. This legislation replaced the old welfare system, Aid to Families with Dependent Children (AFDC), with a new program, Temporary Assistance for Needy Families (TANF). In the 1996 Panel, the questions for income type 20 referred to the AFDC program prior to Wave 4 and to the TANF program beginning in Wave 4. In Wave 9, the questions were expanded somewhat to capture the larger array of program types that could exist under TANF.
[b] During the 1990s, SSI was extended to children with disabilities. Consequently, beginning with the 1992 Panel, SSICOVRG was added to the core wave data files.

The coverage variables identify whether the income or benefit covers that person. In other words, when a person is flagged as covered by food stamps, RCUTYP27 (FOODSTMP) = 1, the person received the benefits either directly (because he or she was the authorized food stamp recipient) or indirectly (because he or she was in the same food stamp unit as the authorized recipient). The coverage variables also allow users to determine situations in which the program unit is a subset of the family or household.[16]

The authorized recipient variables identify the people who actually received the income or benefit for the people in their program units. In the 1996+ Panels, the variables identifying the authorized recipient use only the person number, EPPPNUM. Prior to the 1996 Panel, the variables identifying the authorized recipient were constructed by concatenating the entry address, ENTRY, with the person number, PNUM.

Individuals who are members of a common program unit can be identified by using the sample unit ID, SSUID (SUID), and the authorized recipient variable. For example, members of a common food stamp unit are those with common values of SSUID (SUID) and RCUOWN27 (FSPNUM). Identifying members of common units is often necessary because most programs allow more than one program unit in a household. Medicare, however, is a person-based program in which each participant is an authorized recipient, so no additional authorized recipient variable for that program is included on the files. Prior to the 1996 Panel, there was also no authorized recipient variable for SSI on the core wave files.

There are some exceptions to these rules:

* Social Security, Railroad Retirement (prior to 1996), WIC, AFDC, and Medicaid can offer benefits solely to children. When that happens, an adult receives the income on behalf of the children. The adult, therefore, is flagged as the authorized recipient but is not flagged as covered by the program. The children are flagged as covered and have nonzero benefits.

* Most SSI recipients are elderly and disabled adults, but they can also be disabled children. In the 1990s, the definition of qualifying disabling conditions was expanded. That change in definition resulted in a rapid expansion of the child SSI caseload. Consequently, the SSICOVRG variable was included (beginning with the 1992 Panel). This variable indicates on the recipient's (the adult's) record whether the children, the adults, or both, within a family are covered by the income. Prior to the 1996 Panel, however, SSICOVRG did not flag each person individually, like the other coverage variables. Only the recipient will have had a nonzero SSI income. Beginning with the 1996 Panel, two new variables identify each individual covered by federally administered SSI (RCUTYP03) or state-administered SSI (RCUTYP04).

---

[16] In the 1984 and 1985 Panels, WIC coverage was imputed to children under 6 years old if a mother reported participation in the WIC program. Beginning with the 1986 Panel, WIC coverage is assessed directly for all sample members.

- The medical insurance variables simply reflect who is enrolled in which type of program. There are no associated amount variables.

These rules and exceptions are illustrated in Table 10-17. The household contains one AFDC unit and two food stamp units. The mother is covered by Social Security and SSI. The mother of the disabled child receives WIC benefits and SSI on behalf of her child, but she did not receive WIC or SSI for herself. Everyone in the household is enrolled in Medicaid. The coverage variables are set to '2' whenever a family member is not covered by a particular program; the one exception (for panels prior to 1996) is SSI coverage-a value of 2 means that only the children are covered.

Users should note that, except for WIC, no amounts of income or benefit from government transfer and health insurance programs are listed in the records of children under age 15. Thus, in the case of WIC, users need to sum the amounts over all persons, including children, to get the proper WIC unit total. For all other programs, users will find the unit total benefit in the recipient's record.

Income Topcoding in the 1996+ Panels

To protect the confidentiality of SIPP respondents, the Census Bureau topcodes very high incomes on the SIPP public use data files. New income topcoding procedures were instituted with the 1996 Panel. As in the past, summary income variables for persons, families, and households are the sums of the component variables after they have been topcoded. The summary variables are not independently topcoded. Thus, a person, family, or household with high income from several sources (multiple jobs, businesses, property) could have aggregate monthly income well over the topcode threshold for each source.

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

	Mother	Daughter #1	Daughter #1's Son	Daughter #2	Spouse of Daughter #2	Daughter #2's Pregnant Daughter
1996+ Panels						
EPPPNUM	0101	0102	0103	0104	0105	0106
TAGE	70	21	4	35	36	16
AFDC/TANF						
RCUTYP20	2	1	1	2	2	2
RCUOWN20	0	0102	0102	0	0	0
ER20	0	1	0	0	0	0
T20AMT	0	123	0	0	0	0
Food Stamps						
RCUTYP27	2	1	1	1	1	1
RCUOWN27	0	0102	0102	0104	0104	0104
ER27	0	1	0	1	0	0
T27AMT	0	160	0	130	0	0
SSI						
RCUTYP03	1	2	1	0	0	0
ER03	1	1	0	0	0	0
T03AMT	188	122	0	0	0	0
WIC						
RCUTYP25	2	2	1	2	2	1
RCUOWN25	0	0	0102	0	0	0106
ER25	0	1	0	0	0	1
WICVAL	0	30.12	0	0	0	27.50
Medicaid						
RCUTYP57	1	1	1	1	1	1
RCUOWN57	0101	0102	0102	0104	0104	0106
Social Security						
RCUTYP01A	1	2	2	2	2	2
RCUOWN01A	0101	0	0	0	0	0
R01A	1	0	0	0	0	0
T01AMTA	470	0	0	0	0	0

(*table continues*)

**Table 10-17. Example of Program Units, Coverage, and
Recipiency in the Core Wave Files (*continued*)**

Panels Prior to 1996						
	Mother	**Daughter #1**	**Daughter #1's Son**	**Daughter #2**	**Spouse of Daughter #2**	**Daughter #2's Pregnant Daughter**
PNUM	101	102	103	104	105	106
AGE	70	21	4	35	36	16
AFDC						
AFDCCOV	2	1	1	2	2	2
AFDCPNUM	0	11102	11102	0	0	0
R20	0	1	0	0	0	0
S20AMT	0	123	0	0	0	0
Food Stamps						
FOODSTMP	2	1	1	1	1	1
FSPNUM	0	11102	11102	11104	11104	11104
R27	0	1	0	1	0	0
S27AMT	0	160	0	130	0	0
SSI						
SSICOVRG	1	2	1	0	0	0
R03	1	1	0	0	0	0
S03AMT	188	122	0	0	0	0
WIC						
WICCOV	2	2	1	2	2	1
WICPNUM	0	0	11102	0	0	11106
R25	0	1	0	0	0	1
WICVAL	0	30.12	0	0	0	27.50
Medicaid						
CAIDCOV	1	1	1	1	1	1
MCDPNUM	11101	11102	11102	11104	11104	11106
Social Security						
SOCSEC	1	2	2	2	2	2
SSPNUM	11101	0	0	0	0	0
R01A	1	0	0	0	0	0
R01K	0	0	0	0	0	0
S01AMTA	470	0	0	0	0	0
S01AMTK	0	0	0	0	0	0

Topcoding Unearned Income in the 1996 Panel

When the total amount of asset income or of certain types of general income for a wave exceeds the established ceiling, the monthly amounts in excess of the monthly threshold are replaced by monthly topcode values. For example:

- When the amount of interest on joint municipal/corporate bonds exceeds $10,000 for the wave, each monthly amount in excess of $2,500 is recoded to $2,500.

- When the amount of interest on self-owned municipal/corporate bonds exceeds $12,800 for the wave, each monthly amount in excess of $3,200 is recoded to $3,200.

Not all income sources are topcoded. For example, the amount of food stamp income is not topcoded. For a complete list of topcoded income variables with the topcode amounts for the 1996 Panel, users should refer to Appendix B (Topcoding).

Topcoding Employment Income in the 1996 Panel

Three different sources of monthly employment income are identified in the SIPP public use files: (1) wage and salary income, (2) self-employed earnings, and (3) other worker arrangements. Each of these three sources is topcoded separately. For each source, monthly amounts over $12,500 (one-twelfth of the $150,000 annual benchmark) are topcoded if the total income from those sources from all 4 months in the wave is greater than $50,000 (one-third of $150,000). Table 10-18 provides examples of employment income amounts that require topcoding.

Table 10-18. Topcoding Criteria for the 1996 Panel

Example	Reported Monthly Earned Income Amounts				Sum for the Wave	Is the Sum Greater than $50,000?	Topcoding Procedure
	Month 1	Month 2	Month 3	Month 4			
1	$3,000	$4,000	$5,000	$5,000	$17,000	No	None
2	$0	$0	$0	$55,000	$55,000	Yes	Topcode month 4
3	$15,000	$10,000	$10,000	$12,000	$52,000	Yes	Topcode month 1
4	$12,000	$15,000	$15,000	$15,000	$60,000	Yes	Topcode months 2, 3, and 4
5	$0	$0	$0	$49,000	$49,000	No	None
6	$15,000	$15,000	$15,000	$15,000	$60,000	Yes	Topcode all 4

When topcoding is required because the reported value exceeds the acceptable threshold, the value assigned to the variable can be determined in one of two ways: it can be set equal to the threshold, or it can be set equal to the mean of the reported amounts above the threshold. In the second case, the topcode value that is assigned is based on the respondent's gender, race/ethnic origin, and

employment status (full or part year, full or part time). Table 10-19 illustrates the procedure. It shows the topcodes used in Wave 1 of the 1996 Panel for employment income. Those Wave-1-based topcodes are adjusted for inflation and real growth in earned income (see Box 10-1) and then used for all later waves of the panel.

Because of the way in which the topcode values were computed (explained in the next paragraph), the values listed for each cell are greater than the monthly value that is tested ($12,500). This method of computation may result in instances in which use of the topcode values results in total amounts for the wave (summed across all 4 months) that are greater than $50,000.

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

1996 Panel				
Example	**Sex**	**Race**	**Worker Status**	**Earned Income Topcode**
1	Male	Nonblack, non-Hispanic	Full year; full time	$29,660
2	Male	Nonblack, non-Hispanic	Not full year; full time	$38,270
3	Male	Black, non-Hispanic	Full year; full time	$17,530
4	Male	Black, non-Hispanic	Not full year; full time	$24,015
5	Male	Hispanic, any race	Full year; full time	$26,250
6	Male	Hispanic, any race	Not full year; full time	$24,015
7	Female	Nonblack, non-Hispanic	Full year; full time	$21,990
8	Female	Nonblack, non-Hispanic	Not full year; full time	$49,450
9	Female	Black, non-Hispanic	Full year, full time	$24,015
10	Female	Black, non-Hispanic	Not full year; full time	$24,015
11	Female	Hispanic, any race	Full year; full time	$24,015
12	Female	Hispanic, any race	Not full year; full time	$24,015

Box 10-1. Computing Earned Income Topcode Amounts for Waves 2-12 in the 1996 Panel

The topcode amount for wave k is computed as:

$Topcode_{Wave\ k} = Topcode_{Wave\ 1} * 1019^{k-1}$

Example: Nonblack, non-Hispanic male employed full year, full time.

Wave 1 Topcode (from Table 10-19) = $29,660

Wave 7 Topcode = $29,660 * 1.019^{(7-1)} = $29,660 * 1.120 = $33,206

The topcode values were computed from data collected in Wave 1 of the 1996 Panel. The topcode values are the unweighted mean amounts from records identified for topcoding in Wave 1 of the 1996 Panel. A separate topcode value was computed for each of the 12 cells of Table 10-19. Each topcode value is based on amounts from all three employment income sources, and the same topcode is used

for all three employment income sources. The algorithm used to calculate the assigned topcode amount is as follows:

1.   Add the four monthly amounts of wage and salary income. If the sum is greater than $50,000, store the monthly amounts greater than $12,500 in the 12-cell matrix.

2.   Add the four monthly amounts of self-employed earnings. If the sum is greater than $50,000, store the monthly amounts greater than $12,500 in the 12-cell matrix.

3.   Add the four monthly amounts of contingent worker earnings. If the sum is greater than $50,000, then store the monthly amounts greater than $12,500 in the 12-cell matrix.

On the basis of the amounts accumulated, compute a mean amount within each of the 12 cells of the matrix. That mean amount is the topcode value shown in Table 10-19.

The amounts shown in Table 10-19 were computed with data from Wave 1. Current plans call for using these amounts, adjusted for inflation and real growth in earned income by 1.019 percent per wave for all remaining waves of the 1996 Panel. This is equivalent to an annual increase of 5.8 percent. The mean amounts will not be recomputed from microdata for later waves. The formula to compute the topcode amounts for earned income in later waves is shown in Box 10-1.

The following three examples and Table 10-20 illustrate employment income topcoding:

• A black male software consultant works full time for the entire year and reports an annual salary of $196,600. His salary income varies from month to month, however, sometimes dramatically. For this wave, it is $57,100, above the first test of $50,000. The earned income topcode value for black males who work full time, full year is $17,530 (see Table 10-19: example 3, last column). That value will be used instead of the consultant's reported monthly earned income for the 1 month in which his earned income exceeded $12,500.

• A Hispanic female attorney normally works full time, the full year, with an annual income of about $300,000. In the middle of this wave, she has returned from a 6-month maternity leave; for the first 2 months of the wave, she has no earned income. Her income for the wave in question is $51,000, just over the threshold value of $50,000. The earned income topcode value for Hispanic women who work full time, full year is $24,015 (see Table 10-19: example 11, last column). That is the value that will be used as the attorney's monthly earned income for the months in which her income exceeds $12,500.

• A white male psychiatrist spends the month of August at his beach house. While on vacation, he has no earned income. When he returns to the city in September his income returns to its usual level of $20,000 for the next 3 months. His income for the wave is $60,000, exceeding the $50,000 threshold. The earned income topcode for nonblack, non-Hispanic males is $38,270 (see Table 10-19: example 2, last column). That value is used for the 3 months the psychiatrist reported income over $12,500, resulting in a total earned income for the wave of $114,810.

That total, after topcoding, is substantially higher than $50,000.

- A white television actress does not work during her series' hiatus. When the series is in production, she works full time. Her annual earned income is $880,000; her income for the wave in question is $160,000. She has earned nothing in the first 3 months of the wave, and $160,000 for the fourth month. The SIPP matrix topcode for nonblack, non-Hispanic women who work full time but less than full year is $49,450 for each month (see Table 10-19: example 8, last column). That value will be assigned for the 1 month of the wave in which the actress reported earned income.

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

Worker Characteristics	Income	Reported Monthly Income Amounts				Sum for the Wave
		Month 1	Month 2	Month 3	Month 4	
Black, non-Hispanic male, working full time, full year	Reported	$10,000	$10,000	$12,300	$ 24,800	$ 57,100
	Topcoded	$10,000	$10,000	$12,300	$ 17,530	$ 49,830
Hispanic female, working full year, full year	Reported	$0	$0	$25,000	$ 26,000	$ 51,000
	Topcoded	$0	$0	$24,015	$ 24,015	$ 48,030
Nonblack, non-Hispanic male working full time, party year	Reported	$0	$20,000	$20,000	$ 20,000	$ 60,000
	Topcoded	$0	$38,270	$38,270	$ 38,270	$114,810
Nonblack, female, not full year	Reported	$0	$0	$0	$160,000	$160,000
	Topcoded	$0	$0	$0	$ 49,450	$ 49,450

Topcoding Prior to the 1996 Panel

Prior to the 1996 Panel, the data dictionary indicates a topcode of $33,332 for monthly income; that is also the income topcode for the wave. That topcode is, therefore, rarely used for a single month. In most cases, the monthly income is topcoded at $8,333 (one-fourth of $33,332), which actually represents $8,333 or more. Individual amounts above $8,333 may occasionally be shown if the respondent's income varied considerably from month to month. For example, if a respondent's income from a single job was concentrated in only 1 of the 4 reference months, SIPP could show a figure as high as $33,332.

Summary income variables on the person, family, and household records are simply the sums of the component variables after they have been topcoded. The summary variables are not independently topcoded. Thus, a person with high income from several sources (multiple jobs, businesses, property) could have aggregate monthly income well over the topcode for each source and yet SIPP could still be greatly understating the person's true income.

As shown in Table 10-21, person 101 has wages topcoded. The person received considerably more money in December than in the other months. In addition, total family income and total household income are the sum of the income amounts (in this case, WS1AMT+S01AMT) after they have been topcoded.

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

Person Number (PNUM)	Calendar Month (MONTH)	Household Total Income (HTOTINC)	Family Total Income (FTOTINC)	Topcoded Wages (WS1AMT)	Social Security (S01AMT)	Actual Wages
101	10	$9,333	$9,333	$8,333	$1,000	$ 8,333
101	11	$9,333	$9,333	$8,333	$1,000	$ 8,333
101	12	$9,333	$9,333	$8,333	$1,000	$12,123
101	01	$9,583	$9,583	$8,333	$1,250	$ 9,456

Earnings Topcoding in the 2001 Panel

Table 10-22 contains the topcode amounts for earnings for the 2001 Panel.

Table 10-22. 2001 Panel Earnings Topcodes

2004 Panel Earnings Topcodes				
Cell	Sex	Race	Worker Status	Topcode
1	Sex = 1 (Male)	Non-black, not Hispanic	Full Year, Full Time	$29,057
2	Sex = 1 (Male)	Non-black, not Hispanic	Not Full Year, Full Time	$24,956
3	Sex = 1 (Male)	Black, not Hispanic	Full Year, Full Time	$20,769
4	Sex = 1 (Male)	Black, not Hispanic	Not Full Year, Full Time	$20,769
5	Sex = 1 (Male)	Hispanic, any Race	Full Year, Full Time	$24,283
6	Sex = 1 (Male)	Hispanic, any Race	Not Full Year, Full Time	$36,866
7	Sex = 2 (Female)	Non-black, not Hispanic	Full Year, Full Time	$23,420
8	Sex = 2 (Female)	Non-black, not Hispanic	Not Full Year, Full Time	$25,973
9	Sex = 2 (Female)	Black, not Hispanic	Full Year, Full Time	$26,841
10	Sex = 2 (Female)	Black, not Hispanic	Not Full Year, Full Time	$26,841
11	Sex = 2 (Female)	Hispanic, any Race	Full Year, Full Time	$31,909
12	Sex = 2 (Female)	Hispanic, any Race	Not Full Year, Full Time	$31,909

Earnings Topcoding in the 2004 Panel

Table 10-23 contains the topcode amount for earnings for the 2004 Panel.

Table 10-23. 2004 Panel Earnings Topcodes

2004 Panel Earnings Topcodes				
Cell	Sex	Race	Worker Status	Topcode Amount
1	Sex = 1 (Male)	Non-Black, not Hispanic	Full Year, Full Time	$37,750
2	Sex = 1 (Male)	Non-Black, not Hispanic	Not Full Year, Full Time	$38,900
3	Sex = 1 (Male)	Black, not Hispanic	Full Year, Full Time	$51,400
4	Sex = 1 (Male)	Black, not Hispanic	Not Full Year, Full Time	$51,400
5	Sex = 1 (Male)	Hispanic, any race	Full Year, Full Time	$33,600
6	Sex = 1 (Male)	Hispanic, any race	Not Full Year, Full Time	$33,600
7	Sex = 2 (Female)	Non-Black, not Hispanic	Full Year, Full Time	$30,000
8	Sex = 2 (Female)	Non-Black, not Hispanic	Not Full Year, Full Time	$43,500
9	Sex = 2 (Female)	Black, not Hispanic	Full Year, Full Time	$51,400
10	Sex = 2 (Female)	Black, not Hispanic	Not Full Year, Full Time	$51,400
11	Sex = 2 (Female)	Hispanic, any race	Full Year, Full Time	$33,600
12	Sex = 2 (Female)	Hispanic, any race	Not Full Year, Full Time	$33,600

Using Allocation (Imputation) Flags

As described in Chapter 4, the Census Bureau often imputes information when a person does not respond to the survey or to a particular question.

1.   Prior to the 1996 Panel, the whole record may have been imputed because the person refused to be interviewed (and no proxy interview was obtained) or because the person left the sample in the middle of the wave and no interview was conducted. If that happened, INTVW will be 3 or 4.[17]

2.   A variable of interest may be imputed. In the core wave files prior to the 1996 Panel, there is an allocation (imputation) flag for almost all of the person-level variables. Beginning with the 1996 Panel, there is an allocation (imputation) flag associated with every variable subject to imputation. For example, AEDUCATE is the allocation (imputation) variable that identifies whether EEDUCATE is imputed.

---

[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.

For labor force items, the Census Bureau uses the following special imputation procedures when a person has no current wave information indicating whether or not he or she worked during the reference period.[18] If the Census Bureau can infer from what it knows about the previous reference period whether the person had a job or business at the start of the current period, the Census Bureau carries out the following procedure:

1. If the person was working at the end of the prior wave, then labor force participation is imputed from a single donor for the complete current wave.

2. The Census Bureau then projects job characteristics for the person from the person's prior wave through the current wave.

3. Finally, the Census Bureau edits the job characteristics for consistency with the imputed labor force participation variables.

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

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

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

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

Because the edit and imputation procedures used in the core wave files and in the full panel longitudinal research files are different, data from the two sources will not always agree. See Chapter 4 for a more detailed discussion of the SIPP edit and imputation procedures.

---

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

Using Weights

The core wave files include a number of alternative reference month weights for use in data analysis. Table 10-24 includes examples of the weights for the 1996+ Panels. Table 10-25 includes examples of the weights for the 1990-1993 Panel core wave files. The choice of the appropriate weight for a given analysis depends on the population of interest for that analysis-person, household, family, or related subfamily. Chapter 8 of the Guide contains a full discussion of how to use weights in the core wave files.

Table 10-24. Weight Variables in SIPP Core Wave Files for the 1996+ Panels

1996+ Panels	
Variable Name	**Description**
WPFINWGT [a]	Reference month, final weight of person
WHFNWGT [a]	Reference month, final weight of household
WFFINWGT	Reference month, final weight of family
WSFINWGT	Reference month, final weight of related subfamily

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

Table 10-25. Weight Variables in SIPP Core Wave Files for the 1990-1993

1990-1993 Panels	
Variable Name	**Description**
FNLWGT	Reference month, final weight of person
HWGT	Reference month, final weight of household
FWGT	Reference month, final weight of family
SWGT	Reference month, final weight of related subfamily
P5WGT	Interview (5th) month, final weight of person
H5WGT	Interview (5th) month, final weight of household

Identifying States

For the 2004 and 2008 Panel, the variable, TFIPSST, which is based on the Federal Information Processing (FIPS) Standards State Code, identifies the 50 states and District of Columbia. The 2004 and 2008 SIPP Panels can be used to produce state estimates. The survey was designed to produce reliable low-income estimates for the 33 largest states.

For the 1996 Panel and the 2001 Panel, the same variable, TFIPSST, identifies 45 states and the District of Columbia. To help protect the confidentiality of respondents, the Census Bureau combined the remaining five states as follows:

1. Maine, Vermont; and

2. North Dakota, South Dakota, Wyoming.

For Pre-96 Panels, the core wave files contain the variable HSTATE, which identifies 41 individual states and the District of Columbia; the nine other states are combined into three groups:

1. Maine, Vermont;

2. Iowa, North Dakota, South Dakota; and

3. Alaska, Idaho, Montana, Wyoming.

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

Identifying Metropolitan Areas

Panels, 2001 and 2004

The 2001 and 2004 Panels include a variable, TMETRO, that identifies residences located in metropolitan areas. It can be used to produce national estimates of the metropolitan population. However, it cannot be used to produce estimates of the nonmetropolitan population. To protect respondent confidentiality, the Census Bureau recoded and identified a small random sample of metropolitan households in the public use files as nonmetropolitan. The remaining metropolitan sample should still produce (approximately) unbiased estimates of the metropolitan population. However, the procedure "contaminates" the nonmetropolitan sample, and estimates of nonmetropolitan characteristics based on that sample will be biased (the magnitude of the bias depends on the specific analysis being performed).

Panels, 1990 – 1996

For the Pre-96 Panels, the variable, TMETRO, is named, HMETRO. For the 1996 panel, the variable is named TMETRO. In addition to TMETRO (HMETRO), the 1990 – 1996 Panels include a variable, TMSA (HMSA), that identifies MSAs (Metropolitan Statistical Areas) and CMSAs (Consolidated Metropolitan Statistical Areas), as defined by the Office of Management and Budget.