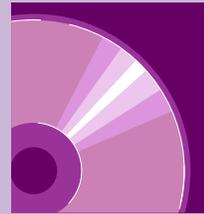


Using Core Wave Files

This section focuses on information specific to the core wave files.

- *File Structure*
- *Using the Data Dictionary*
 - 1996 Panel*
 - Pre-1996 Panels*
- *Identification/Description Variables*
 - Monthly Interview Status*
 - Identifying Persons*
 - Identifying Households*
 - Identifying Families*
 - Identifying Reference Persons*
 - Household Reference Person*
 - Family Reference Person*
 - Other Relationship Variables*
 - Program Units*
 - Movers & Household Composition Changes*
 - Identifying States & Metro Areas*
- *Family-Level Income Variables*
- *Topcoding*
- *Using Allocation Flags*
- *Weight Variables*



Structure of the Core Wave Files

In the first six SIPP panels, the core wave files were issued in person-record format. Beginning with the 1990 Panel, the core wave files have been issued in person-month format.

In the 1990–1996 Panels, one record per person exists for each month of the 4-month reference period that the person was in the sample. A person who was in the sample for all 4 months of the wave has four records.

If a person was not in the sample for the fourth month of the wave because he or she moved out of the country during the middle of the third month, for example, the file will contain three records. The third-month record for that person will contain information that was either imputed or collected by proxy from another household member.

The files also contain records for children under age 15 in sample households.

Using the Data Dictionary

The data dictionary is formatted to facilitate processing by user-written programs. The dictionaries in the 1996 Panel and earlier panels differ somewhat.

1996 Panel

- A “D” in the first column of the dictionary signifies that the line contains the variable name, size (i.e., the number of digits it contains), and the starting position.
- A “T” in the first column signifies that the line contains a short variable description that can be used by many software packages as a variable label.

```

SURVEY OF INCOME AND PROGRAM PARTICIPATION,
1996 PANEL WAVE 1 TOPICAL MODULE DATA DICTIONARY

DATA      SIZE  BEGIN
D  SSUSEQ   5    1
T  SU: Sequence Number of Sample Unit - Primary
    Sort Key
U  All persons
V    1:50000 .Sequence Number

D  SSUID    12    6
T  SU: Sample Unit Identifier
    Sample Unit identifier This identifier is
    created by scrambling together the PSU,
    Segment, Serial, Serial Suffix of the
    original sample address. It may be used
    in matching sample units from different
    waves.
U  All persons
V  000000000000:999999999999 .Scrambled Id

D  SPANEL   4    18
T  SU: Sample Code - Indicated Panel Year
U  All persons
V

```

- A “U” in the first column signifies that the next words describe the universe.
- 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 a comment.

Pre-1996 Panels

- A “D” in the first column of the dictionary signifies that the next few lines define the variable:
- The first line contains the variable name, size (i.e., the number of digits it contains), and the starting position.
- Succeeding lines contain a description of the variable.
- A “U” in the first column signifies that the next words describe the universe. *tip*
- 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.

Identification/Description Variables

Monthly Interview Status

All core wave files issued in person-month format (1990 and subsequent panels) contain records only for persons whose respondent interview status was equal to 1. Thus, the monthly interview status variable can be safely ignored.

In the six earlier panels, core wave files were issued in person-record format. Users should check each person’s monthly interview status variables in these files.

SIPP *tip*

The universe definitions included in the data dictionaries before the 1996 Panel were not always accurate. Users of those panels should check the skip patterns in the actual survey questionnaires to determine which subset of respondents was asked each question.

Identifying Persons

To uniquely identify persons in the core wave files, analysts should use the following variables:

Variable Description	Pre-1996 Panels	1996 Panel
Sample unit ID	SUID	SSUID
Entry address ID	ENTRY	EENTAID (optional)
Person number ID	PNUM	EPPNUM

Chapter 10 of the *SIPP Users' Guide* provides illustrations of how to use these variables to identify individuals and learn when they first entered the SIPP sample.

Identifying Households

To uniquely identify households and group quarters in the core wave files, analysts should use the following two variables:

Variable Description	Pre-1996 Panels	1996 Panel
Sample unit ID	SUID	SSUID
Current address ID	ADDID	SHHADID

People with the same sample unit ID and current address ID live in the same household.

Identifying Families

By using several core wave variables and their associated numbering schemes, analysts can uniquely identify the following family configurations.

Primary Family (family containing the household reference person and all relatives living with him or her)

Variable Description	Pre-1996 Panels	1996 Panel
Sample unit ID	SUID	SSUID
Current address ID	ADDID	SHHADID
Family ID	FID	RFID

Primary Family Excluding Related Subfamilies (related subfamily: a family unit within the primary family whose members are related to, but do not include, the household reference person)

Variable Description	Pre-1996 Panels	1996 Panel
Sample unit ID	SUID	SSUID
Current address ID	ADDID	SHHADID
Family ID (excluding related subfamilies)	FID2	RFID2

Related Subfamilies Only

Variable Description	Pre-1996 Panels	1996 Panel
Sample unit ID	SUID	SSUID
Current address ID	ADDID	SHHADID
Family ID (for related subfamilies)	SID	RSID
Type of family	FTYPE	ESTYPE

Multigenerational Families

Variable Description	Pre-1996 Panels	1996 Panel
Sample unit ID	SUID	SSUID
Current address ID	ADDID	SHHADID
Family ID (excluding related subfamilies)	FID2	RFID2
Family ID (for both related and unrelated subfamilies)	SID	RSID

Identifying Household and Family Reference Persons

Analysts can use the following variables in the core wave files to identify the household reference person (the owner or renter of record) and family reference persons.

Variable Description	Pre-1996 Panels	1996 Panel
Household reference person	HREFPER	EHREFPER
Family reference person	FREFPER	EFREFPER

Describing Relationship to Household Reference Person

Analysts should note that there are two variables in the pre-1996 core wave files that describe how each person is related to the household reference person. One is an edited version of the other. The unedited version allows the analyst to describe more household relationships.

Variable Description	Pre-1996 Panels	1996 Panel
Relationship to household reference person	RRP RRPU (unedited)	ERRP

Chapter 10 of the *SIPP Users' Guide* contains tables that provide the values and value descriptions for these variables.

Describing Relationship to Family Reference Person

In the pre-1996 core wave files, analysts can use the variable FAMREL to identify the relationship of a person to the family reference person (such as spouse or child of family reference person).

The 1996 core wave files do not contain a variable that corresponds exactly to FAMREL. They do contain the variable ESFR (edited subfamily relationship), which is defined the same as FAMREL but applies only to related and unrelated subfamilies.

Identifying Other Relationship Variables

The core wave files contain many variables that describe household and family composition. [Link to a table from the SIPP Users' Guide that lists these variables.](#) Other material in Chapter 10 of the Guide provides more detail on these topics.

Note that in the following list of four of the relationship variables, just one parent is identified in files from panels before 1996.

Variable Description	Pre-1996 Panels	1996 Panel
Spouse	PNSP	EPNSPOUS
Parent	PNPT	
Father		EPNDAD
Mother		EPNMOM
Guardian	PNGDU	EPNGUARD

Identifying Program Units

Users will quickly note that the variable names for program units in the 1996 Panel are quite different from those in earlier panels.

[Link to a table from the SIPP Users' Guide that contains variable names for government transfer programs and health insurance programs in the core wave files.](#)

Questions about program units in the 1996 Panel were expanded in Waves 4 and 9 in response to replacement of the Aid to Families with Dependent Children (AFDC) program by a new program, Temporary Assistance for Needy Families (TANF). TANF provides a broader array of program types.

Identifying Movers and Household Composition Changes

Tables 10-14 and 10-15 in the *SIPP Users' Guide* provide examples of how to identify movers and changes in household composition in the core wave files.

In the rare cases of persons in merged households who were assigned new ID values, two records exist in the pre-1996 Panel core wave files for those persons when the move occurred after the first reference month. When the move occurred in the first reference month, only one record exists. Merged households cannot be identified in the 1996 Panel core wave files.

Identifying States and Metropolitan Areas

The purpose of including variables to identify states in the core wave files is to allow analysts to examine how state-level characteristics affect national estimates. As noted earlier, because SIPP data do not identify all states or uniquely identify nonmetropolitan residences, they should not be used to produce state-level or nonmetropolitan population estimates.

Variable Description	Pre-1996 Panels	1996 Panel
41 states, DC, and 3 combinations of 9 states	HSTATE	
45 states, DC, and 2 combinations of 5 states		TFIPSST
Metropolitan residences	HMETRO	METRO
93 MSAs and CMSAs	HMSA	TMSA

Family-Level Income Variables

Family-level income variables in the core wave files include the income of all related subfamily members. In other words, the Census Bureau treats primary family members, including related subfamily members, as one family when calculating family-level income amounts. The core wave files, however, also contain related subfamily income variables that aggregate the income of members of the same related subfamily.

Variable Description	Pre-1996 Panels	1996 Panel
Family income	FTOTINC	TFTOTINC
Related subfamily income	STOTINC	TSTOTINC

Analysts should keep these variable distinctions in mind when examining family income.

Topcoding

To protect respondents' confidentiality, the Census Bureau topcodes income and age-related variables in the public use files. See the information on topcoding income in the tutorial section SIPP Public Use Files.

Appendix B of the *SIPP Users' Guide* describes the Census Bureau's topcoding specifications for SIPP.

Using Allocation (Imputation) Flags

Almost all imputed person-level variables in the core wave files have allocation (imputation) flags.

In panels prior to 1996, the entire record was imputed if

- (1) MIS5 = 2 and MIS_j = 1 for j = 1, 2, 3, or 4 or
- (2) INTVW = 3 or 4.

The whole record was imputed in the 1996 Panel if EPPINTVW = 3 or 4.

EPPINTVW and INTVW describe the type of interview or noninterview that occurred with a person. 

Weight Variables

The core wave files include alternative reference month weights. Beginning with the 1996 Panel, SIPP files no longer include interview month weights.

Variable Description	1990–1993 Panels	1996 Panel
Reference month—final weight		
Person	FNLWGT	WPFINWGT
Household	HWGT	WHFNWGT
Family	FWGT	WFFINWGT
Related subfamily	SWGTT	WSFINWGT
Interview month—final weight		
Person	P5WGT	
Household	H5WGT	

SIPP tip

Users should note that the codes for EPPINTVW and INTVW differ. Also, the method for identifying persons who were in the sample early in the wave but not at the time of the interview changed for the 1990–1993 Panels.