SIPP *Survey of Income and Program Participation*

Using Topical Module Files

This section focuses on information specific to the topical module files.

■ *File Structure & Content*

  File Structure

  General Content

  Topical Module Vs. Core Wave Files

■ *Variable Names, Reference Periods, & Sample Universe*

  Variable Names

  Reference Periods & Sample Universe

■ *Using the Data Dictionary*

■ *Identification/Description Variables*

  Monthly Interview Status

  Identifying Persons & Households

  Identifying Families

  Household & Family Composition

  Relationship to Household Reference Person
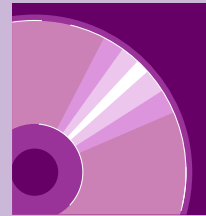
  Movers & Household Composition Changes

  Identifying States & Metro Areas

■ *Topcoding*

■ *Using Allocation Flags*

■ *Weight Variables*

File Structure and Content

Structure of the Topical Module Files

1996 Panel. The topical module files for the 1996 Panel contain one record for each person who was in the sample with a completed or imputed interview in the fourth month of the wave's reference period (the month before the interview).

Pre-1996 Panels. The topical module files for panels before 1996 contain one record for each sample member who was interviewed or for whom an interview was attempted during the interview month (month 5), not the fourth month of the reference period.

General Content of the Topical Module Files

Each topical module file contains data for all topical module subject areas administered during a given wave. The files also contain selected identification and demographic information from the SIPP core, making it possible to do some analysis of those files independently from core wave and full panel files.

If more detailed demographic information is necessary for an analysis, users can acquire that information by merging topical module files with core wave or full panel files, as discussed in the tutorial section Linking Files and in Chapter 13 of the *SIPP Users' Guide*.

Topical Module Vs. Core Wave Files

Topical module files differ from core wave files in key ways:

- The core wave files contain up to four records for each sample person in each wave (one record for each month of the wave the person was in the sample). The topical module files contain only one record for each SIPP sample member in each wave.

- For panels before 1996, topical module files include records for people whose entire households were not interviewed. Those people are excluded from the pre-1996 core wave files.

- As noted, the topical module files contain identification and demographic data also present in the core wave files. In the 1996 Panel, the values of those data correspond to month-4 data in the core wave file for the same wave.

- Prior to the 1996 Panel, however, the identification and demographic data in the topical module files correspond to data collected in the interview month (month 5), not to data in any of the 4 reference months. If any changes in those variables occurred between months 4 and 5, the values for the variables could differ between the core wave and topical module files.

Variable Names, Reference Periods, and Sample Universe

Variable Name

Prior to the 1996 Panel, some variable names were used in different topical module files for different variables, so the variable might change meaning depending on context.

Reference Periods and Sample Universe

Sample definitions and reference periods in topical modules vary across panels, across topical modules within panels, and even within topical modules. Analysts therefore need to pay close attention to those details in the topical module files they use.

1996 Panel. As noted above, most topical module questions were asked only of people in the sample during the fourth month of the wave's reference period. People who were members of SIPP households only at the time of the interview were not asked the topical module questions.

In addition, many questions applied only to the month of the reference period (month 4). However, some topical module questions—even entire topical modules—refer to longer periods of time.

Pre-1996 Panels. Most topical module questions were asked of people in the sample at the time of the interview (month 5). Thus questions were asked of some people who were not in the sample during the 4 previous months, the reference period for core questions in that wave. Consequently, to obtain core data that correspond to the topical module data, analysts must often merge core data from the subsequent wave.

Many topical module questions referred to "current" (interview month) conditions, although some asked about longer periods of time.

Using the Data Dictionary

The data dictionaries in the core wave and topical module files share the same format. The changes in format that occurred in the 1996 files apply to both core wave and topical module files. See the previous tutorial section, *Using Core Wave Files*.

Identification/Description Variables

Monthly Interview Status

Analysts should use data only for months in which the interview status variable has a value of 1.

1996 Panel. Only one interview status variable appears in the 1996 topical module files (EPPMIS4). Because those files contain records only for people who were in the sample, EPPMIS4 always equals 1 and can be safely ignored.

Pre-1996 Panels. The topical module files for these panels are different. There are five interview status variables (PP-MISx), one for each of the reference months and one for the interview month itself (PP-MIS5). Questions were asked only of sample members whose interview status in the interview month was equal to 1.

```
        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
```

Identifying Persons and Households

The variables analysts should use to track persons and households are the same in both the core wave and topical module files except that the variable name of the sample unit ID in the pre-1996 topical module files is ID (see previous tutorial section, *Using Core Wave Files*).

Identifying Families

The variables analysts should use to track families in the topical module files are also the same as those used in the core wave files, except that the topical module files for the 1996 Panel do not contain the variable needed to determine whether all subfamily members are members of the same subfamily. To determine that, an analyst must merge the RSID variable from the month-4 records in the core wave file.

Describing Household and Family Composition

The topical module files contain fewer variables describing house-hold and family composition than do the core wave files. *Link to a table with the topical module variables.*

Analysts wanting more details can merge additional variables from the core wave or full panel files.

Describing Relationship to Household Reference Person

The 1996 Panel core wave and topical module files contain the ERRP variable, which analysts can use to describe relationships to the household reference person. The pre-1996 topical module files contain only RRP, the edited version of the variable used to describe relationships to the household reference person. When a fuller description is needed, analysts can merge the unedited variable (RRPU) from the core wave files.

Identifying Movers and Household Composition Changes

The procedures for identifying movers and household changes are the same in the topical module files and the core wave files. Chapter 11 of the *SIPP Users' Guide* describes and illustrates the procedures in text and tables.

In the rare cases of merged households where persons may have two sets of ID values, the pre-1996 topical module files contain records for those persons only after the move. Analysts must use the core wave file records to identify those persons before the move. Persons in merged households cannot be identified in the 1996 Panel files.

Identifying States and Metropolitan Areas

The same caveat that applies to the core wave files also applies to the topical module files regarding state identification: SIPP was not designed to be representative at the state level and should not be used to produce state-level estimates.

The following variables for identifying states were included in the topical module files only to allow analysts to examine how state-level characteristics affect national estimates.

Variable Description	Pre-1996 Panels	1996 Panel
41 states, DC, and 3 combinations of 9 states	State	
45 states, DC, and 2 combinations of 5 states		TFIPSST

The topical module files do not contain any variables identifying metropolitan areas. Analysts needing that information must merge it from core wave files.

Topcoding

The topcoding procedures used in the topical module files are similar to those used in the core wave files. In general, topcodes for continuous variables, such as income, that apply to the total population include at least $1/2$ of 1 percent of all cases. For income variables that apply to subpopulations, topcodes include either 3 percent of the appropriate cases or $1/2$ of 1 percent of all cases, whichever is higher.

Characteristics frequently topcoded in the topical module files include income and expense values, including those for a broad range of assets and liabilities. The documentation for these variables indicates whether the values are topcoded and the value ranges for the variables.

Using Allocation (Imputation) Flags

As in the core wave files, there is an allocation (imputation) flag for almost all of the person-level variables that are imputed.

There are two ways to identify cases with edited or imputed data in panels prior to 1996: The entire record was imputed if:

(1) PP-MIS5 = 2 and PP-MISj = 1 for j = 1, 2, 3, or 4 or

(2) INTVW = 3 or 4.

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

Weight Variables

The topical module files contain one weight variable:

- WPFINWGT in the 1996 Panel—the person cross-sectional weight for the fourth reference month
- FINALWGT in the pre-1996 Panels—the person interview month weight for people who provided data for a topical module