

## FILE INFORMATION

### Geographic Coverage

State codes are shown except for nine states which are collapsed into three groups. The three state groupings are as follows: Maine and Vermont; Iowa, North Dakota and South Dakota; and Alaska, Idaho, Montana, and Wyoming. The sample was not designed to produce state level estimates. State codes are primarily useful in relating a respondent's reciprocity of benefits to welfare reform thresholds and policies which may vary from state to state.

### Identification Number System/Match Key Variables

The SPD identification scheme uses match key variables designed to uniquely identify individuals, provide a means of linking data for the same individuals across files, and grouping individuals into households and families across files over time. The various components of the identification scheme are listed below:

SIPP Panel Number	SIPP_PNL
Sample Unit Identification Number	PP_ID
Entry ID	PP_ENTRY
Person Number	PP_PNUM
Address ID	ADDIDE(2,3,4,7,8)
SIPP_PNL, PP_ID, ADDID	IHHKEY(2,3,4,7,8)

### File Match Keys

This file includes match keys to merge with SPD and 1992/1993 SIPP data. These match keys link data for each person between the 1992 and 1993 SIPP Panels, the 1997 SPD Bridge, and the 1998 SPD files.

Use the following variables to match back at the person level:

SIPP Panel Number	SIPP_PNL
Sample Unit Identification Number	PP_ID
Person Number	PP_PNUM
Entry ID	PP_ENTRY

Use the following variables to match back at the household level:

SIPP Panel Number	SIPP_PNL
Sample Unit Identification Number	PP_ID
Address ID	ADDIDE(2, 3, 4, 7, 8)

The SIPP panel number identifies the panel in which the respondent participated. The sample person should either have an entry of 1992 or 1993 for their panel number.

The sample unit identification number was created by scrambling together the Primary Sampling Unit, segment, and serial numbers used for Census Bureau administrative purposes. These identifiers are constructed in the same manner as the 1992 and 1993 SIPP panel files, to enable matching to these files. To uniquely identify a household, you must use the sample unit identification (ID) number, the address ID, and the SIPP panel number. The sample unit identification number, the address ID, and the SIPP panel number can be used to link all households back to the original household.

The entry ID represents the address of the person at the time he/she was first interviewed and does not change even if the person moves. It is used in conjunction with the person number to uniquely identify people within the sample unit. This variable is the number 011 for all original sample people. For additional sample people, this variable can be 011 or greater depending on the address ID of the unit which the new sample person joined. For example, a person who moves into a household with an ADDIDE8 of 011 will receive an PP\_ENTRY of 011. Whereas, a person who enters a household spawned in 1998 (ADDIDE8=121) will have a PP\_ENTRY of 121.

The person numbers represent the wave the person entered the sample. Person numbers such as 0101 and 0102 are assigned in Wave 1 of the SIPP. Person numbers such as 0201 and 0202 are assigned to people added to the roster in Wave 2 of the SIPP. People added to the roster in the 1998 SPD have person numbers 1201 and 1202 and following sequentially as needed. People added to the roster in the 1997 SPD Bridge have person numbers 1101 and 1102 and following sequentially as needed.

The address ID is a three digit code that identifies the various household addresses associated with the same sample unit identification number. The first two digits of the address ID code indicate the wave or year in which that address was first interviewed. The third digit sequentially numbers that split into multiple households and have the same address ID. The address ID code is 011 for all sample addresses that are the same as in Wave 1 of the 1992 and 1993 SIPP panels. As the SIPP sample people move to new addresses, new address ID codes are assigned. For example, any new address to which sample unit members moved during the 1998 SPD is numbered from 121 to 129.

Households are defined at each cross sectional time point in this file: 1992, 1993, 1994, 1997, and 1998. If you would like to look at the household configuration in any given year, use the IHHKEY variable appropriate for that year. For example, to look at 1997 household structure, use IHHKEY97. The IHHKEY variables are a concatenation of SIPP\_PNL; PP\_ID and ADDID for the specific year. For example, IHHKEY97 is SIPP\_PNL; PP\_ID and ADDIDE7.

## **File Structure**

The file is a rectangular person-level file. Household and Family level variables are included on the record for every person in the household.

## Edits

The first longitudinal file is a fully edited and imputed data set.

## Topcoding of Variables

To protect against the possibility that a user may recognize the identity of a SPD respondent with a very high income, income from every source is topcoded so that no individual amounts above \$100,000 are revealed. This topcode amount is consistent with the topcoding for the 1992 and 1993 SIPP panels. Other economic variables are topcoded at the 97 percentile level, meaning the top 3 percent of values are not disclosed. Variables that have been topcoded will have a "T" in the second to the right position. NOTE: Aggregate amounts (PTOTVLR, PERNVLR, etc.) use topcoded amounts as input.

We topcode age by bottom coding year of birth. For the First Longitudinal SPD file no age will be older than 88.

## Weighting

For the first longitudinal file, the longitudinal panel final weights are on the file for all sample people (including children). The longitudinal panel final weight is referred to as "the final weight." The sample people who meet the following definition have a positive final weight:

1. Lived in a 1992/1993 SIPP panel household during Wave 1 interview.
2. Were interviewed (self, proxy or imputed) in each and every reference month in SIPP.
3. Were interviewed (self, proxy or imputed) in 1997 SPD Bridge and 1998 SPD.

All the other sample people included on the file have zero final weights with the exception for the sample children aged 6 or less from the SIPP Panel 1992 and aged 5 or less from the SIPP Panel 1993 who were born to original sample persons. For each child in this group, if her/his designated parent is an original sample person then the weight of the child was assigned the same as her/his designated parent's weight, else the child's weight was assigned as zero. A designated parent of a child is either a biological parent or an adopted parent or a blood-related guardian or a not blood related guardian. The procedure used for calculating the longitudinal panel final weights of the sample people on the first longitudinal file is described in the eight steps provided below. Refer to section 4 (Glossary) of the Technical Documentation for definitions for the key terms used in the eight steps of weighting described on pages 3-3 through 3-6.

**STEP 1** Choose March 1993 as the reference point in time (control date) for the SPD panel universe used for measuring the effect of the Welfare Reform. Namely, the SPD panel universe consists of persons (including children) residing in the United States households and persons living in group quarters in March 1993. (Note that persons living in military barracks and institutions, such as prisons and nursing homes, are excluded.) The SPD panel universe is represented by combining the SIPP Panels 1992 and 1993 into one sample. The effect of the Welfare Reform can be assessed based on the various longitudinal (time dependent) characteristics of the people in the SPD panel universe.

**STEP 2** Assign an initial weight (IW) to each original sample person (including children). An original sample person is a person who at the time of the Wave 1 interview resided in an interviewed sample household or group quarters. The inverse of this initial weight represents the probability of an original sample person residing in an interviewed Wave 1 sample household in either the SIPP Panel 1992 or

1993, depending on which SIPP panel he/she originally belonged. Thus, by definition, the initial weights can be expressed as follows:

The SIPP Panel 1992 original sample persons receive their initial weights of the SIPP Panel 1992 base weight the SIPP Panel 1992 Wave 1 household non-interview adjustment factor.

The SIPP Panel 1993 original sample persons receive their initial weights of the SIPP Panel 1993 base weight the SIPP Panel 1993 Wave 1 household non-interview adjustment factor.

For every sample person who is not an original sample person, assign an SPD 1998 longitudinal panel final weight of zero and exclude this person from further processing in the weighting calculation. Note that this group of sample persons nominally represents the children and adults born prior to the inception of the SPD panel universe.

For every sample person aged six or less if spawned from the SIPP Panel 1992 and aged five or less if spawned from the SIPP Panel 1993 (as of March 1998) who is not an original sample person, assign an SPD 1998 longitudinal panel final weight according to the procedure provided in Step 8. Note that this group of sample persons nominally represents the children born after the inception of the SPD panel universe.

**STEP 3** Since each of the SIPP Panels 1992 and 1993 (samples) is a nationally representative sample by itself, combining them into one sample reduces the weight of each panel sample person proportionately to their sample sizes. The sample sizes of the SIPP Panels 1992 and 1993 are approximately the same. Therefore, assign a combined panel factor (CPF) of 0.5 to each of the original sample person. Namely, the weight (CPAW) of each original person after accounting the panel combining effect can be expressed in terms of IW and CPF as  $CPAW = IW \times CPF$ .

**STEP 4** Divide the original sample persons into two cohorts as follows: the first cohort consists of the original sample persons who are qualified as longitudinally interviewed between Wave 1 of the SIPP Panels 1992 and 1993 up to the SPD Bridge. The second cohort consists of the original sample persons who are not qualified as longitudinally interviewed between Wave 1 of the SIPP Panels 1992 and 1993 up to the SPD Bridge. Hereinafter, each person in the first cohort will be referred to as a SPD Bridge longitudinally interviewed person. Likewise, each person in the second cohort will be referred to as a SPD Bridge longitudinally non-interviewed person.

Calculate the SPD Bridge non-interview adjustment factors  $\frac{1}{F_{1c}}$  ( $F_{1c}$ ).  $F_{1c}$  serves as a means to appropriately transfer the weights from the SPD Bridge longitudinally non-interviewed persons to the SPD Bridge longitudinally interviewed persons with similar demographic and economic characteristics. Herein, the similarity in demographic and economic characteristics among the sample persons are classified based on the following variables: age, race, ethnicity, education, labor force status, employment status, income types, assets, and income level. Then assign the SPD Bridge non-interview adjusted weights (NIAWB) in terms of  $F_{1c}$  and CPAW in the following manner.

For each SPD Bridge longitudinally interviewed person, assign  $NIAWB = CPAW \times F_{1c} = IW \times CPF \times F_{1c}$ .

For each SPD Bridge longitudinally non-interviewed person, assign a longitudinal panel final weight of zero, and exclude this person from further processing of the weighting calculation.

**STEP 5** Due to budget constraints, some sample household units were cut<sup>2</sup> in the SPD 1998. Therefore, identify all the SPD Bridge longitudinally interviewed persons belonging to the sample households selected into the SPD 1998 sample, and call this cohort of persons "the selected SPD Bridge longitudinally interviewed persons". Similarly, identify all the SPD Bridge longitudinally interviewed persons belonging to the sample households not selected into SPD 1998 sample, and call this cohort of persons "the not-selected SPD Bridge longitudinally interviewed persons".

For each selected SPD Bridge longitudinally interviewed person, assign a sample cut factor, SCF (sampling rate) in accordance with the household demographic characteristics provided in the document referred to in Footnote 2. The SCFs jointly serve as a means to appropriately transfer the weights (NIAWB) from the not-selected SPD Bridge longitudinally interviewed persons to the selected SPD Bridge longitudinally interviewed persons in the following manner.

For each selected SPD Bridge longitudinally interviewed person, let SCAW denote her/his weight after the adjustment for the sample cut using the SCF. By definition, SCAW can be expressed as  $SCAW = NIAWB \times SCF = IW \times CPF \times F_{1c} \times SCF$ .

For each not-selected SPD Bridge longitudinally interviewed person, assign a SPD 1998 longitudinal panel final weight of zero, and exclude this person from further processing of the weighting calculation.

**STEP 6** Divide the selected SPD Bridge longitudinally interviewed persons into two cohorts as follows: the first cohort consists of the selected SPD Bridge longitudinally interviewed persons who are also qualified as interviewed in the SPD 1998 survey. The second cohort consists of the selected SPD Bridge longitudinally interviewed persons who are not qualified as interviewed in the SPD 1998 survey. Hereinafter, each person in the first cohort will be referred to as a SPD 1998 longitudinally interviewed person. Likewise, each person in the second cohort will be referred to as a SPD 1998 longitudinally non-interviewed person.

Calculate the SPD 1998 non-interview adjustment factors ( $F_{2c}$ ) in the same manner as calculating  $F_{1c}$  in Step 4. Namely,  $F_{2c}$  serves as a means to appropriately transfer the weights from the SPD 1998 longitudinally non-interviewed persons to the SPD 1998 longitudinally interviewed persons with similar demographic and economic characteristics as classified in Step 4. Thus assign the SPD 1998 non-interview adjusted weights (NIAW98) in the following manner.

For each SPD 1998 longitudinally interviewed person, assign  $NIAW98 = SCAW \times F_{2c} = IW \times CPF \times F_{1c} \times SCF \times F_{2c}$ .

For each SIPP to SPD 1998 longitudinally non-interviewed person, assign a longitudinal panel final weight of zero, and exclude this person from further processing of the weighting calculation.

**STEP 7** For the SPD 1998 longitudinally interviewed persons, perform second stage ratio estimation and then compute their second stage weights. The second stage ratio estimation<sup>3</sup> principally involves raking to match a set of SPD population estimates with a corresponding set of control (benchmark) population estimates at a representative control date. The control population estimates are available for the classifications based on the following demographic variables: age, sex, race, ethnicity, householder living with or not living with relative, not-householder related to or not related to householder. As described in Step 1, for the SPD 1998, the representative control date is March 1993. The control population estimates are from the CPS population estimates produced for March 1993.

The second stage ratio adjustment serves as a means to improve the population coverage of the SPD 1998 sample. The classification of the sample persons into groups for the second stage ratio estimation also serves as a post-sampling stratification. Therefore, as a byproduct of the second stage ratio estimation, the variance estimates of the estimates based on the SPD sample will be generally improved as well.

The weights obtained after the second stage adjustment are the final weights for the SPD 1998 longitudinally interviewed persons.

**STEP 8** Assign the final weight for each sample person (child) aged 6 or less (as of March 1998) if spawned from the SIPP Panel 1992 and aged 5 (as of March 1998) or less if spawned from the SIPP Panel 1993 as follows. If the designated parent of the child is an original sample person then assign the final weight of the child to be the same as the weight of the child's designated parent, else if the designated parent of the child is not an original sample person then assign the final weight of zero to the child. A designated parent of a child can be either a biological parent or an adopted parent or a blood-related guardian or a not blood related guardian.

This weight will then be identified as "SPD Longitudinal Annual Weight" and its application is described in the next section (see below).

### **Longitudinal Weight with Non-original Sample children**

SPD Longitudinal Annual Weight (ANNUALWT) - This weight is for calculating annual or calendar year estimates. This variable is nonzero for original sample people who provided a self or proxy, or whose data were imputed in each and every reference month for SIPP, SPD 1997 Bridge, and SPD 1998 and for non-original sample children (age six or less) born of or adopted by original sample people after the first interview. When calculating annual estimates, include only weighted non-original sample children born during or in prior calendar years. The non-original sample children weights, for those age six or less, are derived by assigning the weight of child's designated parent (blood-related or adopted) to the child and is intended only to approximate growth in the universe. Some caution should be given when interpreting results for children 6 or less because this strategy has resulted in general overestimate of 2.2 percent based on 1998 benchmarks. By race, the results are a 3.6 percent overestimate for non-Black and a 5.4 underestimate for Black children six or less by 1998.

### **Estimation of Person Characteristics**

Some basic types of longitudinal estimates that can be constructed using the longitudinal weight are described below in terms of estimated numbers. Of course, more complex estimates, such as percents, averages, and ratios can be constructed from the estimated numbers. The SPD longitudinal weights can be used to construct the following types of longitudinal estimates:

1. The number of people who have ever experienced a characteristic or situation during a given time. (For example, the number of people who experience unemployment during 1997.) To construct such an estimate, sum the weights over all people who possessed the characteristic of interest at some point during the time period of interest.
2. The amount of a characteristic accumulated by people during a given time. (For example, the amount of unemployment compensation received by unemployed people during 1998.) To construct such an estimate, compute the product of the weight times the amount of the characteristic and sum this product over all appropriate people.

## Longitudinal Research Using This File

The SPD is designed exclusively to support longitudinal analysis of the impact of welfare reform. The First Longitudinal SPD data can be linked to the 1992 and 1993 SIPP Panel and Cross-sectional files, the 1997 SPD Bridge, and the 1998 SPD file using the following variables:

SIPP Panel Number	SIPP_PNL
Sample Unit Identification Number	PP_ID
Entry ID	PP_ENTRY
Person Number	PP_PNUM
Address ID	ADDIDE(2,3,4,7,8)

A Longitudinal weight is assigned to 100-level persons with full panel weights in the 1992/1993 SIPP file who were successfully interviewed in 1998. Note the full panel weights on the SIPP files were assigned to 100-level people who were interviewed for the entire time they remained in the SIPP universe or who had at most 1 missing interview bounded by successful interviews.

The SPD data represent the behavior and characteristics of people in two fixed cohorts. One cohort represents the population as it existed in March 1992 from the 1992 panel of the SIPP and the other population as of March 1993 from the 1993 panel. This is not a traditional longitudinal survey in that it does not repeat the same measure throughout the period. Each round of the SPD interviewing, beginning with the Bridge in 1997, does not represent cross-sectional snapshots of the U.S. population. It does offer insight into what the current condition is of the people in the U. S. population in the early 1990s just prior to welfare reform.

The core information common throughout the data collection (although reference periods and question phrasing vary) consists of basic demographics, labor force activity, income, and program participation. The longitudinal file consists of data collected using three different instruments, each with variations in wording and context.

### Obtaining Access to SAQ Data

The SAQ data (adult and adolescent) will only be available through the Census Bureau's Research Data Centers. Contact the Research Data Center staff for the requirements for reviewing the SAQ data.

---

## FOOTNOTES

<sup>1</sup> A full detailed specification for calculating the SPD Bridge non-interview adjustment factors ( $F_{1c}$ ) can be obtained from the memorandum dated July 15, 1999 on "Specifications for the longitudinal Weighting of Sample Persons in the Period Spanning 1992 through 1997 of the Survey of Program Dynamics" for C. E. Bowie from A. R. Tupek.

<sup>2</sup> The sample cut procedure and the sampling rates (sample cut factor) for each of the six household strata of the SPD Bridge sample can be obtained from the memorandum dated January 22, 1999 on "1998 SPD: Sampling Specifications" for C. E. Bowie from L. S. Cahoon.

<sup>3</sup> A full detailed specification for calculating the second stage adjustment ratio can be obtained from the memorandum dated July 15, 1999 on "Specifications for the longitudinal Weighting of Sample Persons in the Period Spanning 1992 through 1997 of the Survey of Program Dynamics" for C. E. Bowie from A. R. Tupek.