

8. Using Sampling Weights on SIPP Files

This chapter describes the use of sampling weights in analyzing data from the Survey of Income and Program Participation (SIPP) for the current panels.¹ Each SIPP file contains a number of alternative sets of weights for use in data analysis. The several different sets of weights are needed to cater to the different possible units of analysis (persons, households, families, and subfamilies) and different time periods for which survey estimates may be required.

A common mistake in the analysis of a survey like SIPP is to ignore the weights entirely, that is, to perform an unweighted analysis. This chapter explains why an unweighted analysis is likely to produce biased estimates. It is important to understand the different sets of weights on the files and to use the set that is appropriate for a particular analysis. Topics covered in this chapter include:

- What weights are and why they should be used;
- Which weights are available in SIPP files;
- Which weights to use for a particular analysis;
- How weights are constructed;
- Using weights in the core wave files;
- Using weights in the topical module files;
- Using weights in the longitudinal file

What Weights Are and Why They Should Be Used

The weight for a responding unit in a survey data set is an estimate of the number of units in the target population that the responding unit represents. In general, since population units may be sampled with different selection probabilities and since response rates and coverage rates may vary across subpopulations, different responding units represent different numbers of units in the population. The use of weights in survey analysis compensates for this differential representation,

¹ Information on how to use sampling weights on panels prior to the 1996 Panel can be found in *SIPP Users' Guide*, 3rd Ed. [U.S. Census Bureau, 2001].

thus producing estimates that relate to the target population.

There are some variations in sampling rates in all SIPP panels and, more important, there are appreciable variations in response and coverage rates across subpopulations. As a result, there is nontrivial variation in SIPP weights. For example, in Wave 1 of the 2004 Panel, the final person lower quartile weight is 1,682 and the upper quartile weight is 3,429 (the maximum weight is 16,482). A respondent with a final person weight of 1,682 represents 1,682 people in the U.S. population for the reference month, whereas a respondent with a weight of 3,429 represents 3,429 people. Because weights in SIPP vary over a sufficiently large range of values, performing unweighted analyses may produce appreciably biased estimates for the U.S. population.

Table 8-1 illustrates the effects of weighting on a selection of estimates obtained from Wave 1 of the 2004 Panel. The 2004 Panel included an oversample of households with low income (less than 150% of the poverty threshold). Since the low income group is overrepresented in this sample, failure to use the weights would lead to overrepresentation in the population estimates based on that sample. For example, at the household level, the unweighted percentage of households headed by females with no spouse present and living with children is 13.4 percent, whereas the weighted estimate is 12.0 percent. At the person level, the magnitude of the differences between weighted and unweighted estimates is typically less, but still appreciable.

Table 8-1. Weighted and Unweighted Point-in-Time Estimates of Percentages Based on Core Wave 1 of the 2004 SIPP Panel for January 2004

Characteristics	Percentage	
	Weighted ^a	Unweighted
Household-Level		
Female-headed households with no spouse present, living with	12.0	13.4
Person-Level		
Female	51.1	51.9
Race/Ethnicity		
White Only	80.6	79.4
Black Only	12.5	13.0
Residual Race	6.8	7.8
Low Income (< 150%)	22.9	24.4
Age over 65 years	12.0	12.2
Receiving Food Stamps [RCUTYP27 (FOODSTMP)]	6.8	7.9
RCUTYP20 (AFDC or TANF)	1.1	1.3

^a Weighted by WPFINWGT- final weight for person - and WHFNWGT - final weight for households.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-2 illustrates the effect of weighting on a selection of estimates obtained from Wave 1 of the 1990 Panel.

Table 8-2. Weighted and Unweighted Point-in-Time Estimates of Percentages Based on Core Wave 1 of the 1990 SIPP Panel for January 1990

Characteristics	Percentage	
	Weighted ^a	Unweighted
Household-Level		
Female-headed households with no spouse present, living with relatives?	11.7	14.3
Person-Level		
Female	51.3	52.2
Race/Ethnicity		
White	84.2	82.1
Black	12.4	14.4
American Indian, Eskimo, or Aleut	0.6	0.6
Asian or Pacific Islanders	2.9	2.9
Age over 65 years	10.4	10.6
Receiving Food Stamps [RCUTYP27 (FOODSTMP)]	6.7	7.7
RCUTYP20 (AFDC)	3.8	4.6

^a Weighted by WPFINWGT (FNLWGT)-final weight for person-and WHFNWGT (HWGT) -final weight for households.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Weights Available in SIPP Files

Table 8-3 lists the weight variables in SIPP data files for the 1996, 2001 and 2004 Panels.

Table 8-3. Weight Variables in SIPP Files for the 1996, 2001 and 2004 Panels

Variable Name	Description
Core Wave Files	
WPFINWGT	Reference month, final weight of person
WHFNWGT	Reference month, final weight of household
WFFINWGT	Reference month, final weight of head of family
WSFINWGT	Reference month, final weight of related head of subfamily
Topical Module Files	
WPFINWGT	4th reference month, final weight of person
Longitudinal File^a	
LGTCY _x WT	Calendar year <i>x</i> , final weight of people in the calendar year cohort
LGTPNWT _x ^b	Panel weight based on the sample universe in waves 1 through <i>w(x)</i> ^c

^a The number of calendar year weights in the longitudinal file depends on the panel's duration. The 1996 longitudinal file has four calendar year weights (LGTCY1WT, LGTCY2WT, LGTCY3WT, LGTCY4WT). The 2001 longitudinal file has three calendar year weights (LGTCY1WT, LGTCY2WT, LGTCY3WT). The 2004 longitudinal file will have four calendar year weights when it is complete (LGTCY1WT, LGTCY2WT, LGTCY3WT, LGTCY4WT).

^b The final panel weight for the 1996 Panel was LGTPNLWT. Starting in the 2001 Panel, multiple panel weights are constructed. The 2001 Panel has three panel weights: LGTPNWT1 (waves 1-4), LGTPNWT2 (waves 1-7), and LGTPNWT3 (waves 1-10). The 2004 Panel will have four panel weights: LGTPNWT1 (waves 1-4), LGTPNWT2 (waves 1-7), LGTPNWT3 (waves 1-10), and LGTPNWT4 (waves 1-12).

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-4 lists the weight variables in SIPP data files for the 1990-1993 Panels. For earlier panels, the user should refer to the data dictionary for the particular file.

Table 8-4. Weight Variables in SIPP Files for the 1990-1993 Panels.

Variable Name	Description
Core Wave Files	
FNLWGT	Reference month, final weight of person
HWGT	Reference month, final weight of household
FWGT	Reference month, final weight of head of family
SWGTT	Reference month, final weight of related head of subfamily
P5WGT ^a	Interview (5 th) month, final weight of person
H5WGT ^a	Interview (5 th) month, final weight of household
Topical Module Files	
FINALWGT	Prior to 1996: interview month, final weight of person. 1996+: 4 th reference month, final weight of person
Full Panel Files^b	
FNLWGT _x	Calendar year <i>x</i> , final weight of people in the calendar year cohort
PNLWGT	Final weight for people in full panel cohort

^a Beginning with the 1996 Panel, SIPP files no longer include the interview month weights.

^b The number of calendar year weights in the full panel file depends on the panel's duration. The 1990 full panel file contains two calendar year weights: WPFINWGT90 (FNLWGT90) and WPFINWGT91 (FNLWGT91). The 1992 full panel file has three calendar year weights: WPFINWGT92 (FNLWGT92), WPFINWGT93 (FNLWGT93), and WPFINWGT94 (FNLWGT94).

Choosing a Weight

The decision of which weight to use for a given analysis depends on the population of interest. Useful guidance for choosing the correct set of weights is to consider to what population the results are intended to apply.

The weights in the SIPP files are constructed for sample cohorts defined by:

- Month (e.g., the reference month weights in the core wave files and the fourth reference month weights in the topical module files);
- Year (e.g., the calendar year weights in the longitudinal file); and
- Panel (e.g., the panel weights in the longitudinal file).

Users can choose to base their analyses on:

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

- A cross-sectional sample at a given month;
- A longitudinal sample that provides continuous monthly data over a year;
- A longitudinal sample that provides monthly data over the life of a panel (up to 32 months with the 2001 Panel, or up to 48 months with the 1996 and 2004 Panel); or
- A subset of the sample and/or the period in any of the above.

Monthly (cross-sectional) weights allow the use of all available data for a given month. For this type of analysis, users can choose among the following units of analysis:

- Person (e.g., WPFINWGT);
- Household (e.g., WHFNWGT);
- Family (e.g., WFFINWGT); and
- Related subfamily (e.g., WSFINWGT).

Analysts can use longitudinal samples to follow the same people over time and hence study such issues as the dynamics of program participation, lengths of poverty spells, and changes in other circumstances (e.g., household composition). The longitudinal weights allow the inclusion of all people for whom data were collected for every month of the period involved (calendar year or panel), including those who left the target population through death or because they moved to an ineligible address (institution, foreign living quarters, military barracks), as well as those for whom data were imputed for missing months. The Census Bureau makes nonresponse adjustments to the longitudinal weights to compensate for panel attrition and poststratification adjustments to make the weighted sample totals conform to population totals for key variables.

How Weights Are Constructed

This section describes how the weights are constructed. The basic components for all the different sets of weights are the same, namely:

- A base weight that reflects the probability of selection for a sample unit;
- An adjustment for subsampling within clusters;

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

- An adjustment for movers (in Waves 2 and beyond);
- A noninterview adjustment to compensate for sample nonresponse; and
- A poststratification (second-stage calibration) adjustment to correct for departures from known population totals.

Weights

Reference month final weights are provided on the SIPP core wave files for persons, households, families, and subfamilies. The special weights for persons are constructed first. The household, family, and related subfamily final weights are derived from the final person weights. This section summarizes the steps involved in constructing the various sets of weights, starting with the final person weights for a reference. Appendix C provides the technical details and reasons for some of the adjustments specifically for the 2004 Panel.

The reference month weights for people on the core wave files are computed (i.e., are nonzero) for all responding sample members who are "in scope" (i.e., a part of the survey's universe—the residential, noninstitutional population of the United States) in the specified month.² A number of factors lead to fluctuations in sample size from month to month. They include births, deaths, immigration, and emigration from the population (and therefore from the sample). In addition to those population dynamics, people move into and out of the sample as a result of the changing household composition of sample members. (Chapter 2 describes the SIPP "following rules.")

In Wave 1, the weight for each sample person per month is a product of four components:

1. **Wave 1 Base weight.** This weight is the inverse of the probability of a sample person's address being selected.
2. **Duplication-control Factor.** This factor adjusts for the occasional subsampling of clusters. Clusters are occasionally subsampled in the field when they turn out to be much larger than expected.³
3. **Wave 1 Household Noninterview Adjustment Factor (NAF).** This adjustment compensates for different rates of household noninterview within adjustment classes.

² Persons subjected to Type Z imputation receive weights, although they are not respondents.

³ This adjustment has been used since Wave 5 of the 1984 Panel.

More than 500 nonresponse adjustment classes are defined based on a cross-classification of characteristics. Those characteristics include within-PSU oversampling strata (poverty stratum, nonpoverty stratum); Census Region; race of reference person (black, nonblack); household tenure (owner, renter); Residence Status (MSA-central city, MSA-non-central city, other place); household size (1, 2, 3, 4+ people).

4. **Wave 1 Second-Stage Calibration Adjustment (SSCA).** This adjustment brings the sample estimates into agreement with independent monthly estimates of population totals. The characteristics used for calibration include age, race, sex, Hispanic origin, family relationship, household type and state.⁴ A raking procedure is used to ensure that the weights agree with all the control totals included for calibration. The adjustment is done by rotation group, with each group assigned one-fourth of the population total for the month.

In subsequent waves, each person receives an initial weight that is carried over from the preceding wave. This weight is adjusted to compensate for changes in the sample between waves resulting from movers and nonresponse, and then it is realigned to match the population totals for the reference or interview month:

- **Wave 2+ Initial Weight.** This is the weight from the previous wave before the second-stage calibration for each original sample person who is a reference person or is in group quarters for the current wave.
- **Wave 2+ Mover's Adjustment.** This adjustment is made to compensate for including people who were not in the original sample but were in the SIPP universe in Wave 1 and who moved into a sample household after Wave 1. For people in housing units that contain adult members who were not part of the original sample but were in the SIPP universe at Wave 1, the weights are decreased. For example, if a third adult moves into a household occupied by two original sample persons, all three adults would receive the initial weight of the original sample persons multiplied by a factor of two-thirds.
- **Wave 2+ Noninterview Adjustment.** The noninterview adjustment for Waves 2 and beyond is used to compensate for household nonresponse after the first interview. The noninterview adjustment classes are defined on the basis of sample unit characteristics and personal demographic characteristics from the most recent wave. The information used consists of household characteristics. Reference person

⁴ State was added for the 2004 Panel.

characteristics are used to define some of the household characteristics. Tenure (owner/renter occupied), household type (female householder, no spouse present; 65+; other), race and Hispanic origin, and education level are defined at the household level by using reference person data. Other household characteristics include size, poverty status, type of income, type of financial assets, census division, and number of imputed items. Poverty threshold, census division, and number of imputed items were new starting in the 1996 Panel. Some adjustment classes are combined to ensure that the adjustment for each class does not exceed a factor of 2, and each class contains at least 30 unweighted sample households.

- **Wave 2+ Second-Stage Calibration.** To derive this adjustment, use the same procedure as in Wave 1; that is, use the appropriate population control totals by reference month.

The reference month final weights for households, families, and subfamilies are derived from the person weights:

- The household weight is the person weight of the household reference person (renter/owner of housing unit).
- The family weight is the person weight of the family reference person.
- The subfamily weight for a related subfamily is the person weight of the related subfamily reference person (Chapter 10 explains how to identify households, families, and subfamilies).

Final Panel and Calendar Year Weights

Final panel and final calendar year weights are provided on the longitudinal weight file for eligible sample members. Prior to the 2001 Panel there is one set of final panel weights and generally more than one set of calendar year weights, one for each calendar year covered by the panel. The 1996 longitudinal weight file has four sets of calendar year weights. Starting with the 2001 Panel there are more than one set of final panel weights and more than one set of calendar year weights. For example, the 2001 longitudinal weight file has three sets of panel weights and three sets of calendar year weights.

Final panel weights are computed only for people who are in the sample at Wave 1 of the panel and for whom data are obtained (either reported or imputed) for every month up to $w(x)$ ⁵ for which they

⁵ $w(x)$ for LGTPNWT1 is Wave 4, $w(x)$ for LGTPNWT2 is Wave 7, $w(x)$ for LGTPNWT3 is Wave 10 and $w(x)$ for

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

were in scope for the survey. Other people in the longitudinal weight file are assigned final panel weights of zero. Most people with nonzero final panel weights have provided data for all months of the panel. However, people who missed a wave and whose missing wave data were imputed and people who provided data up to the point that they left the survey (through death or because they moved to an ineligible address) are also assigned nonzero final panel weights. (In core panels, it also includes those missing up to two consecutive waves, if the waves are bounded.)

Final calendar year weights are computed only for people who had an interview covering the control date⁶ and for whom data are obtained (either reported or imputed) for every month of the calendar year for which they were in scope for the survey. Other people are assigned final calendar year weights of zero. Some people who joined the household of an original sample person after the start of the panel are assigned nonzero calendar year weights for the second calendar year, if data are obtained for that period.

The panel weighting scheme does not assign weights to people who enter the sample universe after Wave 1. Similarly, the calendar year weighting scheme does not assign weights to people who do not have an interview covering the control date. This group consists of (a) people who enter the sample universe after the first wave of interviewing for the calendar year and (b) people who were in the sample universe in the first wave of interviewing in the calendar year but did not have an interview covering the control date. For example, newborn infants and people leaving institutions who are entering the sample universe after Wave 1 are assigned panel and calendar year 1 weights of zero. Note that the same people will receive positive calendar year 2 (CY2) weights if they are in the sample universe in the first wave of interviewing for CY2 and have an interview covering the control date for CY2.

The final panel and final calendar year weights are constructed from the following three components:

- 1. Initial Weight.** This weight is constructed from the components of the cross-sectional weights at the start of the panel and calendar year weighting periods before the second-stage calibration adjustment.
- 2. Noninterview Adjustment Factors.** These factors account for noninterviewed eligible sample persons not already accounted for in the noninterview adjustment component of the initial weight. The adjustment classes are similar to those used in the Wave 2+ nonresponse adjustment factors.

LGTPNWT4 is Wave 12.

⁶ The calendar year control dates for January 1 for the given calendar year. The exception is calendar year 1996 for the 1996 Panel. Its control date is March 1, 1996

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

- 3. Second-Stage Calibration Factors.** These factors are determined by a process similar to that used for reference month weighting. The control totals used for the calendar year weights are the population estimates for the control date of the relevant year. Those for the panel weights are the population estimates for a designated date in the first wave of the panel.

Using Weights in the Core Wave Files

Each core wave file contains reference month weights for persons, households, families, and subfamilies. This section describes those weights and indicates how they should be used for different types of analysis.

Reference Month Weights

To understand the format of the reference month weights, analysts may find it useful to recall the SIPP survey design and the file structure for the core wave file. The full SIPP sample consists of four rotation groups; for each wave, interviewing is spread over 4 months. One rotation group is interviewed per month, with the reference months for each rotation group being the 4 months preceding the interview month. As successive rotation groups are interviewed, the 4-month reference periods advance by 1 month. Therefore, there are 4 interview months and 4 reference months per rotation group for each wave.

There are four final person reference month weights per sample person, one for each month in the reference period. Beginning with the 1990 Panel, the reference month weights are provided as one variable - that is, WPFINWGT for persons - in four separate person-month records per person. The reference month weight on each record refers to the specific month to which the data relate.

Person Reference Month Weights

For person-level analyses, the weight available in the core wave file is WPFINWGT (the reference month weight). WPFINWGT is the estimated number of people in the population that the sample person represents in a specific reference month. The reference month is given by the variables RHCALMN (calendar month) and RHCALYR (calendar year), which are derived based on SROTATON (rotation group) and SREFMON (reference month).

Table 8-5 shows the reference months weights for two hypothetical sample persons in Wave 1 of the 2004 Panel, based on the person-month format. The persons can be identified by the variables SSUID, EENTAID, and EPPPNUM (Chapter 10 describes how to identify a person). There are four records per person, one for each reference month. The first four records are for the first person, who

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

is from rotation group 2: SROTATON = 2. Reference month 1, SREFMON = 1, corresponds to November 2003 (RHCALMN and RHCALYR). WPFINWGT for SREFMON = 1 is 2,000, meaning that this person represents 2,000 people in the population in November 2003. The values of WPFINWGT in subsequent months are slightly different because of adjustments to the weight resulting from fluctuations in the population and in the sample. The second person is from rotation group 3. Since the month of interview for this person is different from that of the first person, the reference months for this person are also different. The variables RHCALMN and RHCALYR can be used to select records with data for a particular month.

Table 8-5. Final Person Weights for Four Reference Months in Wave 1 of the 2004 Panel

SSUID	ENTAID	EPPPNUM	SROTATON	SREFMON	RHCALM	RHCALYR	WPFINWGT
123456789	11	101	2	1	11	03	2,000
123456789	11	101	2	2	12	03	2,005
123456789	11	101	2	3	01	04	2,010
123456789	11	101	2	4	02	04	2,020
123456789	11	101	3	1	12	03	3,500
123456789	11	101	3	2	01	04	3,510
123456789	11	101	3	3	02	04	3,520
123456789	11	101	3	4	03	04	3,530

Table 8-6 lists the weight variables in SIPP data files for the 1991 Panel.

Table 8-6. Final Person Weights for Four Reference Months and one Interview Month in Wave 1 of the 1991 Panel

SUID	ENTRY	PNUM	ROT	REFMON	MONTH	YEAR	FNLWGT	P5WGT
123456789	11	101	2	1	10	90	5,000	5,025
123456789	11	101	2	2	11	90	5,005	5,025
123456789	11	101	2	3	02	90	5,010	5,025
123456789	11	101	2	4	01	90	5,020	5,025
123456789	11	101	3	1	12	90	6,500	6,525
123456789	11	101	3	2	01	90	6,510	6,525
123456789	11	101	3	3	01	91	6,520	6,525
123456789	11	101	3	4	02	91	6,530	6,525

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Household Reference Month Weights

Households in the core wave file refer to a group of people who occupy a housing unit in a specific calendar month. For each household, the household weight WHFNWGT is the weight of the reference person (the renter/owner of a housing unit) of the household. WHFNWGT is the number of households in the population that the sample household represents in that reference month. Note that the household reference person can change from one month to the next, resulting in a change of WHFNWGT. WHFNWGT is assigned to all household members.

Table 8-7 shows WHFNWGT for five members of a household and their person weights. The variables SSUID and SHHADID identify the household (Chapter 10 describes how to identify households). The WHFNWGTs for all members of a household are equal to the WPFINWGTs of the reference person in the household. In this case, the household reference person is the father. The user should note that weights for husbands and wives are equalized in the weight process. Therefore, couples (e.g., father and mother, daughter and son-in-law) have the same person weights.

Table 8-7. Household, Reference Month Weights for Members of a Household for a Given Month in Wave 1 of the 2004 Panel

Household Member	SSUID	SHHADID	EENTAID	EPPNUM	WHFNWGT	WPFINWGT
Father ^a	101111103	11	11	101	2,000	2,000
Mother	101111103	11	11	102	2,000	2,000
Daughter	101111103	11	11	103	2,000	1,800
Son-in-law	101111103	11	11	104	2,000	1,800
Grandchild	101111103	11	11	105	2,000	500

^a Reference person of household.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-8 shows WHFNWGT (HWGT) and WHFNWGT (H5WGT) for five members of a household and their person weights for Wave 1 of the 1990 Panel.

Table 8-8. Household, Reference Month Weights for Members of a Household for a Given Month in Wave 1 of the 1990 Panel

Household Member	SUID	ADDID	ENTRY	PNUM	HWGT	H5WGT	FNLWGT	P5WGT
Father ^a	101111103	11	11	101	5,000	5,050	5,000	5,050
Mother	101111103	11	11	102	5,000	5,050	5,000	5,050
Daughter	101111103	11	11	103	5,000	5,050	4,800	4,865
Son-in-law	101111103	11	11	104	5,000	5,050	4,800	4,865
Grandchild	101111103	11	11	105	5,000	5,050	3,000	3,035

^a Reference person of household.

Family and Related Subfamily Reference Month Weights

All sample persons in a core wave file are assigned a family type, EFTYPE, consisting of the following categories: primary families, unrelated subfamilies, primary individuals, and secondary individuals. A family is defined as a group of two or more persons related by birth, marriage, or adoption who reside together. A primary family is a family containing the household reference person and all of his or her relatives. An unrelated subfamily is a family in a household that is not related to the household reference person. A primary individual is a household reference person who lives alone or lives with only nonrelatives. A secondary individual is not a household reference person and is not related to any other people in the household.

Related subfamily units within primary families are identified by ESFTYPE = 2. Related subfamilies are families that are related to, but do not include, the household reference person. For example, the daughter, son-in-law, and grandchild in Table 8-9 constitute a related subfamily within a primary family. They are members of the father and mother's primary family unit, as well as members of their own subfamily.

The SIPP core wave files provide reference month weights for families and related subfamilies. The family reference month weight WFFINWGT is equal to the person weight of the family reference person in that month; it is assigned to all family members. The subfamily reference month weight WSFINWGT is equal to the person weight of the related subfamily reference person; it is assigned to all subfamily members and is set equal to zero for people not in related subfamilies. Primary individuals are the household reference persons and the family reference persons. For a primary individual, WFFINWGT = WPFINWGT = WHFNWGT. Secondary individuals are classified as family reference persons who are not household reference persons. Therefore, for secondary

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

individuals, WFFINWGT = WPFINWGT

Table 8-9 shows the weights for the 2004 Panel for the different analysis units by type of household, RHTYPE, and by type of family, EFTYPE. Three households are shown. The first household is a married couple family household, RHTYPE = 1, consisting of a primary family and a related subfamily, ESFTYPE = 2. The WHFNWGT for each member of this household is equal to the person weight of the household reference person (i.e., the father in this case). Members of this household belong to one primary family. Therefore, the WFFINWGT for each member is equal to the person weight of the family reference person (who is also the father). Some members of this primary family belong to a related subfamily unit (i.e., daughter, son-in-law, and grandchild). The subfamily weight WSFINWGT for each member of the subfamily is equal to the person weight of the subfamily reference person (e.g., the daughter). WSFINWGT is zero for the father and mother who are not part of the subfamily.

The second household is a male-householder nonfamily household, RHTYPE = 4, with three unrelated individuals. The household reference person is the primary individual, EFTYPE = 4, and the others are secondary individuals, EFTYPE = 5. The WHFNWGT for this household is the person weight of the household reference person, and the weight is the same for all individuals. The WFFINWGT is different for each individual because each one is treated as his or her own family reference person.

The third household is a group-quarters household, RHTYPE = 6. Because there is no household reference person based on the typical definition of renter or owner, both individuals are classified as secondary individuals, EFTYPE = 5. The first secondary individual in a group quarters is labeled as the household reference person, and the WHFNWGT for each person in group quarters is the weight of that individual. The WFFINWGT for each individual is different because each forms an individual family.

Table 8-10 shows the weights for the 1990 Panel for different analysis units by type of household, HTYPE, and by type of family, FTYPE.

Calendar Month Estimation: Using a Single Core Wave File

Each core wave file consists of data from 7 calendar months covered by the reference month periods for the four rotation groups. There is only 1 calendar month with complete data from all four rotation groups. As an illustration, Table 8-11 shows the calendar months within the reference periods for Wave 1 of the 2004 Panel and the number of rotation groups available per month. The table shows that data from all four rotation groups are available for only January 2004. Data are available from three rotation groups for December 2003 and February 2004, for two rotation groups for November 2003 and March 2004, and for one rotation group for October 2003 and April 2004.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

The reference month weights for each rotation group are designed to represent a quarter of the population at the month of reference. The weights for each rotation group can be inflated to represent the full population. For every month, the inflation adjustment equals four divided by the number of rotation groups available. For example, the adjustment for October 2003 is 4/1 because there is only one rotation group in this month. For January 2004, the adjustment factor is 1 because all four rotation groups are available for this month.

Users are strongly encouraged to use the full sample of all four rotation groups whenever possible. The core wave files are designed to support analysis using the full sample of all four rotation groups (discussed below). While the weights can be modified to compensate for a smaller sample, estimates based on a subset of rotation groups will be less reliable than those based on the full sample.

Table 8 - 9. Family and Subfamily Reference Months Weights, by RHTYPE, EFTYPE, and ESFTYPE in Wave 1 of the 2004 Panel

Household Member	SSUID	SHH ADID	RFID	RFID2	RSID	EENTAID	EPPP NUM	WPFIN WGT	WHFN WGT	WFFIN WGT	WSFIN WGT	EFTYPE	ESFTYPE
RHTYPE = 1 Married-couple family household													
Father ^{a,b}	101111103	11	1	1	0	11	101	2,000	2,000	2,000	0	1	-1
Mother	101111103	11	1	1	0	11	102	2,000	2,000	2,000	0	1	-1
Daughter ^c	101111103	11	1	0	1	11	103	1,800	2,000	2,000	1,800	1	2
Son-in-law	101111103	11	1	0	1	11	104	1,800	2,000	2,000	1800	1	2
Grandchild	101111103	11	1	0	1	11	105	500	2,000	2,000	1800	1	2
RHTYPE = 4—Male-householder nonfamily													
Male 1 ^{a,b}	122210000	11	1	1	0	11	101	3,000	3,000	3,000	0	4	-1
Person 2 ^b	122210000	11	1	1	0	11	102	1,500	3,000	1,500	0	5	-1
Person 3	122210000	11	1	1	0	11	103	2,500	3,000	2,500	0	5	-1
RHTYPE = 6 Group quarters													
Individual 1 ^a	222210000	11	1	1	0	11	101	1,500	1,500	1,500	0	5	-1
Individual 2	222210000	11	1	1	0	11	102	1,000	1,500	1,000	0	5	-1

Notes: RHTYPE—type of household: 1 = married couple family household, 2 = male householder family household, 3 = female householder family household, 4 = male householder nonfamily household, 5 = female householder nonfamily household, 6 = group quarters; EFTYPE—type of family: 1 = primary family, 2 = related subfamily, 3 = unrelated subfamily, 4 = primary individual, 5 = secondary individual.

^a Household reference person—see text.

^b Family reference person.

^c Related subfamily reference person.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-10. Family and Subfamily Reference Months Weights, by HTYPE, FTYPE, and SFTYPE in Wave 1 of the 1990 Panel

Household Member	SUID	ADDID	FID	FID2	SID	ENTRY	PNUM	FNLWGT	HWGT	FWGT	SWGTT	FTYPE	STYPE
RHTYPE = 1 Married-couple family household													
Father ^{a,b}	101111103	11	1	1	0	11	101	5,000	5,000	5,000	0	1	0
Mother	101111103	11	1	1	0	11	102	5,000	5,000	5,000	0	1	0
Daughter ^c	101111103	11	1	0	1	11	103	4,800	5,000	5,000	4,800	1	1
Son-in-law	101111103	11	1	0	1	11	104	4,800	5,000	5,000	4,800	1	1
Grandchild	101111103	11	1	0	1	11	105	3,000	5,000	5,000	4,800	1	1
RHTYPE = 4—Male-householder nonfamily													
Male 1 ^{a,b}	122210000	11	1	1	0	11	101	6,000	6,000	6,000	0	4	0
Person 2 ^b	122210000	11	1	1	0	11	102	4,500	6,000	4,500	0	5	0
Person 3	122210000	11	1	1	0	11	103	5,500	6,000	5,500	0	5	0
RHTYPE = 6 Group quarters													
Individual 1 ^a	222210000	11	1	1	0	11	101	4,500	4,500	4,500	0	5	0
Individual 2	222210000	11	1	1	0	11	102	3,500	4,500	3,500	0	5	0

Notes: month = 01; Year = 1990. RHTYPE—type of household: 1 = married couple family household, 2 = male householder family household, 3 = female householder family household, 4 = male householder nonfamily household, 5 = female householder nonfamily household, 6 = group quarters; EFTYPE—type of family: 1 = primary family, 2 = related subfamily, 3 = unrelated subfamily,

4 = primary individual, 5 = secondary individual.

^a Household reference person—see text.

^b Family reference person.

^c Related subfamily reference person.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-11. Calendar Month Estimation: Using a Single Core Wave File in Wave 1 of the 2004 Panel

Rotation Group	Interview Month	Reference Months—Wave 1, 2004 Panel						
		2003 Oct.	2003 Nov.	2003 Dec.	2004 Jan.	2004 Feb.	2004 Mar.	2004 Apr.
1	Feb. 2004	1	2	3	4			
2	Mar. 2004		1	2	3	4		
3	Apr. 2004			1	2	3	4	
4	May 2004				1	2	3	4
Rotation Group Adjustment		4	2	4/3	1	4/3	2	4

Table 8-12. Calendar Month Estimation: Using a Single Core Wave File in Wave 1 of the 1991 and 1996 Panels

Rotation Group	Interview Month	Reference Months—Wave 1, 1991 Panel						
		1990 Oct.	1990 Nov.	1990 Dec.	1991 Jan.	1991 Feb.	1991 Mar.	1991 Apr.
1	Feb. 1991	1	2	3	4			
2	Mar. 1991		1	2	3	4		
3	Apr. 1991			1	2	3	4	
4	May 1991				1	2	3	4
Rotation Group Adjustment		4	2	4/3	1	4/3	2	4
Rotation Group	Interview Month	Reference Months—Wave 1, 1996 Panel						
		1995 Dec.	1996 Jan.	1996 Feb.	1996 Mar.	1996 Apr.	1996 May	1991 June
1	Apr. 1996	1	2	3	4			
2	May 1996		1	2	3	4		
3	June 1996			1	2	3	4	
4	July 1996				1	2	3	4
Rotation Group Adjustment		4	2	4/3	1	4/3	2	4

Calendar Month and Quarterly Estimation: Using Two or More Core Wave Files

Combining data from two or more core wave files can increase the data available for making estimates for calendar months or continuations of calendar months such as quarters of the year. As an example, Table 8-13 shows the effects of cumulating calendar month data across two waves:

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Waves 1 and 2 of the 2004 Panel. By combining Waves 1 and 2, there are now four rotation groups for calendar month estimations from January through April 2004. To calculate calendar month estimates for each of those months, the user can simply select the person-month records for the month of interest from a file that pools records from Waves 1 and 2 and apply the WPFINWGT associated with each record to obtain the full sample estimate.

Quarterly estimates in the form of average month estimates also can be computed based on a combined file. For example, to calculate the percentage of people receiving food stamps in the first quarter of 2004, users can obtain the weighted number of people receiving food stamps and the weighted number of the total population in each month of the quarter. Then the percentage of people receiving food stamps is the sum across months of the weighted number of people receiving food stamps divided by the sum of the weighted number of total population in the quarter. In deriving quarterly estimates, or estimates for any time interval, from data in the core wave files, users need to include all four rotation groups in each month of the estimation. The quarterly estimates derived by this method are cross-sectional estimates, based on the samples in each month of the quarter.

Table 8-13. Calendar Month Estimation: Using Two Core Wave Files from Waves 1 and 2 of the 2004 Panel

Rotation Group	Interview Month	2003 Oct.	2003 Nov.	2003 Dec.	2004 Jan.	2004 Feb.	2004 Mar.	2004 Apr.
Wave 1, 2004 Panel								
1	February	1	2	3	4			
2	March		1	2	3	4		
3	April			1	2	3	4	
4	May				1	2	3	4
Wave 2, 2004 Panel ^a								
1	June					1	2	3
2	July						1	2
3	August							1
3	September							
Rotation Group Adjustment		4	2	4/3	1	1	1	1

^a Not all data from Wave 2 are shown in the table.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-14. Calendar Month Estimation: Using Two Core Wave Files from Waves 1 and 2 of the 1991 and 1996 Panels

Rotation Group	Interview Month	Reference Months						
		1990 Oct.	1990 Nov.	1990 Dec.	1991 Jan.	1991 Feb.	1991 Mar.	1991 Apr.
Wave 1, 1991 Panel								
1	February	1	2	3	4			
2	March		1	2	3	4		
3	April			1	2	3	4	
4	May				1	2	3	4
Wave 2, 1991 Panel ^a								
1	June					1	2	3
2	July						1	2
3	August							1
3	September							
Rotation Group Adjustment		4	2	4/3	1	1	1	1
Rotation Group	Interview Month	Reference Months						
		1995 Dec.	1996 Jan.	1996 Feb.	1996 Mar.	1996 Apr.	1996 May	1996 June
Wave 1, 1996 Panel								
1	Apr. 1996	1	2	3	4			
2	May		1	2	3	4		
3	June			1	2	3	4	
4	July				1	2	3	4
Rotation Group Adjustment		4	2	4/3	1	1	1	1
Wave 2, 1996 Panel ^a								
1	August					1	2	3
2	September						1	2
3	October							1
4	November							
Rotation Group Adjustment		4	2	4/3	1	1	1	1

^a Not all data from Wave 2 are shown in the table.

Using Weights in the Topical Module Files

The topical module files contain one weight variable—WPFINWGT. Starting with the 1996 Panel, this weight is the person cross-sectional weight for the fourth reference month. It shows the number of people in the population represented by the sample person in the interview month.

The sample weights on the topical module files are defined in the same manner as the sample weights on the core wave files. The WPFINWGT for each rotation group is defined to represent a quarter of the population at the interview month.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Using Weights in the Full Panel File - Pre-96

The weight variables in the full panel file are the calendar year weights, WPFINWGT (FNLWGT), and the full panel weight (PNLWGT). The number of calendar year weights on the file depends on the duration of the panel. Most panels before the 1996 Panel have two calendar year weights. The exceptions are the 1989 Panel, which has one calendar year weight - WPFINWGT89 (FNLWGT89) - and the 1992 Panel, which has three calendar year weights - WPFINWGT92 (FNLWGT92), WPFINWGT93 (FNLWGT93), and WPFINWGT94 (FNLWGT94). When the 1996 full panel file is complete, it will have four calendar year weights.

The weight variables are defined for sample persons who are in the sample for different periods of time. The calendar year weights apply to sample persons who had interviews covering the control date of the corresponding calendar year and who have complete data (either reported or imputed) for every month of the year (excluding months of ineligibility). The panel weight applies to sample persons who are in the sample in Wave 1 of the panel and who have complete data (either reported or imputed) for every month of a panel (excluding months of ineligibility). People are assigned calendar year weights equal to zero when they do not have interviews covering the control date, have missing data for one or more months of the year, or both.

Similarly, people are assigned panel weights equal to zero if they were not in sample in Wave 1, have missing data for one or more months of the panel, or both.

The population of inference for each of these weights is the population of survivors of the January (or Wave 1, depending on the weight) population. Infants born after the beginning of the panel are assigned a PNLWGT of zero. Similarly, infants born after the control date are assigned a calendar year weight of zero for that year. This weighting can have important implications for those studying young children when infants are a sizable fraction of the population. For example, the WIC program serves children under 5 years of age. Infants in their first year constitute 20 percent of that population.

The SIPP full panel file contains records for every person who was ever part of a responding SIPP household. There is one record for each such person, excluding people who may have been in the sample for only 1 month. The first number in PP-EENTAID (PP-ENTRY) and in PP-EPPPNUM (PP-NUM) indicate the wave in which the person entered the sample. Each record contains month-by-month data collected at every wave. However, records with incomplete data for a given period (year or full period of the panel) are assigned weights of zero. As discussed in Chapter 4, beginning with the 1991 Panel, a new imputation procedure was put into place to allow more people to have positive weights in the full panel files. All people with one or more missing waves, each of which was bounded on both sides by interviewed waves, have their data imputed for the bounded missing waves. With this procedure, a significant portion of the panel nonrespondent records became usable.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

records for longitudinal analysis.

The variables PPID (PP-ID), PP-EENTAID (PP-ENTRY), and PP-EPPNUM (PP-PNUM) identify people in the full panel files (Chapter 12). Table 8-15 provides examples of the weights in the 1990-1993 full panel file. The 1990-1993 Panels provided three weights: WPFINWGT (FNLWGT90), WPFINWGT91 (FNLWGT91), and PNLWGT. The person on the first row is a complete panel member, with all three weights greater than zero. The second person has positive calendar year weights but zero PNLWGT, which probably indicates that this person provided data for the first 2 calendar years but left before Wave 8. The third person had complete (reported or imputed) data for the first calendar year, but probably left before the end of the second calendar year. The fourth person entered the panel at Wave 4 and probably remained in sample until the end of the panel. He was eligible for only a calendar year 2 weight. The last person entered at Wave 7 and was assigned a weight of zero for all three weights on the panel file (however, this person would have had reference month and interview month weights on the Wave 7 and 8 core files).

Using Weights in the Longitudinal File

There was no missing data imputations for Panels, 1984 through 1990, however missing wave imputations were implemented for Panels, 1991 through 1993. For the 1996 and 2001 Panels, SIPP performed longitudinally edited cross-sectional files.

The weight variables in the longitudinal file are the calendar year weights and the panel weights. The number of calendar year weights on the file depends on the duration of the panel. Starting with the 1996 Panel, the number of calendar years increased. The 1996 Panel has four calendar year weights, the 2001 Panel has three calendar year weights and the 2004 Panel has four calendar year weights. The 1996 Panel has one panel weight, but starting in the 2001 Panel the number of panel weights increased. The 2001 Panel has three panel weights and the 2004 Panel will have four panel weights.

The weight variables are defined for sample persons who are in the sample for different periods of time. The calendar year weights apply to sample persons who had interviews covering the control date of the corresponding calendar year and who have complete data (either reported or imputed) for every month of the year (excluding months of ineligibility). The panel weights apply to sample persons who are in the sample in Wave 1 of the panel and who have complete data (either reported or imputed) for every month up through $w(x)_7$ (excluding months of ineligibility). People are assigned calendar year weights equal to zero when they do not have interviews covering the control date, have missing data for one or more months of the year, or both. Similarly, people are assigned

7 See Table 8-3 for definition of $w(x)$.

panel weights equal to zero if they were not in sample in Wave 1, have missing data for one or more months of the panel, or both.

The population of inference for each of these weights is the population of survivors of the January (or Wave 1, depending on the weight) population. Infants born after the beginning of the panel are assigned a LGTPNWTx of zero. Similarly, infants born after the control date are assigned a calendar year weight of zero for that year. This weighting can have important implications for those studying young children when infants are a sizable fraction of the population. For example, the WIC program serves children under 5 years of age. Infants in their first year constitute 20 percent of that population.

The SIPP longitudinal file contains records for every person who was ever part of a responding SIPP household. There is one record for each such person, excluding people who may have been in the sample for only 1 month. The first number in EPPNUM indicate the wave in which the person entered the sample. Each record contains month-by-month data collected at every wave. However, records with incomplete data for a given period (year or panel) are assigned weights of zero. As discussed in Chapter 4, beginning with the 1991 Panel, a new imputation procedure was put into place to allow more people to have positive weights in the longitudinal file. All people with one or more missing waves, each of which was bounded on both sides by interviewed waves, have their data imputed for the bounded missing waves. With this procedure, a significant portion of the panel nonrespondent records became usable records for longitudinal analysis.

The 2001 Panel provides six weights: LGTCY1WT, LGTCY2WT, LGTCY3WT, LGTPNWT1, LGTPNWT2, and LGTPNWT3. Table 8-16 provides examples of some of the weights in the longitudinal weight file. The person on the first row is a complete panel member, with all weights greater than zero. The second person has positive calendar year weights but zero LGTPNWT3, which probably indicates that this person provided data for the first 2 calendar years but left before Wave 8. The third person had complete (reported or imputed) data for the first calendar year, but probably left before the end of the second calendar year. The fourth person entered the panel at Wave 4 and probably remained in sample until the end of the panel. He was eligible for only a calendar year 2 weight. The last person entered at Wave 7 and was assigned a weight of zero for all three weights on the longitudinal weight file (however, this person would have had reference month weights on the Wave 7 and 8 core files).

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Table 8-15. Selected Calendar Year and Panel Weights, 1990-1993

PP-ID	PP-ENTRY	PP-PNUM	FNLWGT90	FNLWGT91	PNLWGT
123456789	11	101	5,500	6000	6,500
123456789	11	102	5,500	6,000	0
123456789	11	101	7,200	0	0
221456789	41	401	0	6,500	0
567891211	71	701	0	0	0

Table 8-16. Selected Calendar Year and Panel Weights, 2001-2003

SSUID	EPPPNUM	LGTCY1WT	LGTCY2WT	LGTPN3WT
123456789	101	2,500	3,000	3,500
123456789	102	2,500	3,000	0
123456789	101	4,200	0	0
221456789	401	0	3,500	0
567891211	701	0	0	0

Calendar Year Estimation: Using the Longitudinal File

Although the SIPP collects most core content with monthly resolution, users may need to construct calendar year estimates of quantities such as total annual income. One way to construct such estimates is to work with the longitudinal weight file, extracting those records with positive calendar year weights. For example, to estimate average annual wages in 2001 for people over age 25 on January 1, 2001, one could identify records from the 2001 Panel with positive values on the calendar year weight, LGTCY1WT. The annual income amount for each sample person is the sum of the amounts received during each month of the calendar year. The aggregate income estimates for the population can be derived by multiplying each person's annual income by LGTCY1WT and summing the products across all people. An estimate of average income is this weighted total income divided by the sum of the weights (summed across the same subsample of the population).⁸

⁸ For purposes of exposition, this discussion has neglected the complication that not all persons with positive calendar year weights will have 12 months of data. For example, any person who was in the population January 1 but who spent at least 1 month during that year in an institution would have fewer than 12 months of data. If that person had complete data for the months when he or she was *not* in the institution, the person would have a positive value for LGTCY1WT. This issue is particularly pertinent for studies of the elderly, since a noneligible portion of that population spend some time in a nursing home or some other type of extended care facility.

Throughout this chapter, pre-1996 variable names appear in parentheses following 1996 variable names.

Annual estimates computed with this method are based on monthly data from the same person collected at three or four points in time (depending on the rotation group of the respondent). The shorter recall period used by SIPP is generally believed to provide estimates of annual measures with less nonsampling error than other surveys that collect annual income measured only once during a year. Chapter 6 and the *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a), provide a more detailed discussion of nonsampling error in SIPP.

Spell Estimation: Using the Longitudinal File

Analysis of SIPP data that takes full advantage of the longitudinal nature of the survey can take a number of forms. In studies of the dynamics of household composition, labor force activity, and welfare reciprocity, analysts have applied a set of methods that fall under the general headings of survival analysis (see Kalbfleisch and Prentice, 1980) and event-history analysis (see Tuma and Hannan, 1984). Among many other topics, researchers have studied the length of time that a woman remains single, a person remains unemployed, or a person receives food stamps before marrying, getting a job, or moving off the Food Stamp program. A spell of being single, unemployed, or receiving food stamps is a period of time during which a person's status did not change, and it is the duration of those spells that is often of interest.

In these studies, the unit of analysis is the spell. A file of spells is built from the person records in the longitudinal files, scanning across months to find a transition into and out of the state of interest. An example of the approach is provided by Shea (1995b). She constructed spells from the records of people with positive panel weights (PNLWGT⁹ greater than zero), restricting her analysis to spells starting after the beginning of the panel, as is commonly done. Methods have been proposed that allow for the use of spells in progress at the start of the panel when the beginning dates of those spells are known (see Guo, 1993).

An alternative approach is to use all people in the longitudinal file. Spells can be constructed whenever a transition into the state of interest is observed (e.g., the birth of a child to a single woman). There are three possible outcomes that might be of interest: (1) a transition out of "single parenthood" is observed when the woman marries; (2) the spell is right-censored because the woman is lost through attrition from the sample before the end of the panel and before she marries; and (3) the spell is right-censored because the panel ends before she marries. If modeled in that way, the appropriate weight would be the woman's calendar month weight associated with the month that the spell of single parenthood began. Calendar month weights are not on the longitudinal file, but can be merged into that file from the appropriate core wave files.

⁹ Prior to the 2001 Panel, there was only one panel weight and prior to the 1996 Panel the final full panel weight was PNLWGT.

During the course of a SIPP panel, some panel members can experience multiple spells (e.g., of participation in a given program). There are two approaches to handling this situation: (1) select only the first spells that started during the life of the panel (Ruggles and Williams, 1989), or (2) use all spells starting during the life of the panel (Kalton et al., 1992).

The length of spells that can be fully observed depends on the duration of a panel. The 1996 Panel has 48 months of data. The 2001 Panel lasted 36 months. The 2004 Panel lasted 48 months.

A note for users of spell analysis is that, in SIPP, as in other panel surveys, people tend to report a change in reciprocity more often between waves than within waves (the seam effect). This suggests that it may not be possible to pinpoint changes to a specific month. More detailed discussions of the seam effect are provided in Chapter 6 and in the *SIPP Quality Profile*, 3rd Ed. (U.S. Census Bureau, 1998a).