



UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

February 6, 2020

MEMORANDUM FOR

Carolyn M. Pickering
Survey Director, Associate Director for Demographic Programs

From:

Anthony G. Tersine, Jr. *Anthony G. Tersine, Jr.*
Chief, Demographic Statistical Methods Division

Subject:

Survey of Income and Program Participation 2014 Panel: Source
and Accuracy Statement for Wave 1 Through Wave 4 Public Use
Files (S&A-22)

This memorandum documents the Source and Accuracy Statement for The Survey of Income and Program Participation 2014 Panel for Wave 1 Through Wave 4 Public Use Files.

The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20-POP001-0052.

If you have any questions about this document, please contact Mahdi Sundukchi at 301.763.4228 or mahdi.s.sundukchi@census.gov, Ashley Westra at 301.763.8536 or Ashley.m.westra@census.gov or Faith Nwaoha-Brown at 301.763.4696 faith.n.nwaoha.brown@census.gov.

cc

Elizabeth Sinclair (ADDP)
David Waddington (SEHSD)
Trudi Renwick
Stephanie Galvin
Alfred Gottschalck
Brian McKenzie
Kurt Bauman
Ashley Edwards
Rose Kreider
Jason Fields
Lindsay Monte
Holly Fee
Shelley Irving
Jonathan Rothbaum

Adam Smith
Sharon Stern
Nathaniel McKee (DSD)
Connie Bauer
James Farber (DSMD)
Tracy Mattingly
Mahdi Sundukchi
Ashley Westra
Julia Yang
Faith Nwaoha Brown

[This page is intentionally left blank]

SOURCE AND ACCURACY STATEMENT FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP) 2014 PANEL FOR WAVE 1 THROUGH WAVE 4 PUBLIC USE FILES¹.

DATA COLLECTION AND ESTIMATION

Source of Data: The data were collected in the 2014 Panel of the Survey of Income and Program Participation (SIPP). The population represented in the 2014 SIPP (the population universe) is the civilian noninstitutionalized population living in the United States. The institutionalized population, which is excluded from the universe, is composed primarily of the persons in correctional institutions and nursing homes (94 percent of the four million institutionalized people in the 2010 Census).

The SIPP 2014 Panel sample is located in 820 Primary Sampling Units (PSUs), each consisting of a county or a group of contiguous counties. Of these 820 PSUs, 344 are self-representing (SR) and 476 are non-self-representing (NSR). SR PSUs have a probability of selection of one. NSR PSUs have a probability of selection less than one. Within PSUs, housing units (HUs) were systematically selected from the Master Address File (MAF), which is the U.S. Census Bureau's official inventory of known housing units. The MAF was created using the decennial censuses, as well as the U.S. Postal Service's Delivery Sequence File (DSF). The Census Bureau continues to update the MAF using the DSF and various automated, clerical, and field operations.

Households were classified into two strata, such that one stratum had a higher concentration of low income households than the other. We oversampled the low income stratum by 24 percent to increase the accuracy of estimates for low income households and program participation. Analysts are strongly encouraged to use the SIPP weights when computing estimates since households are not selected with equal probability.

Each household in the sample was scheduled to be interviewed at yearly intervals over a period of roughly four years. The reference period for the interview questions is the preceding twelve-month calendar year. The most recent month, December, is designated reference month 12 and the earliest month, January, is reference month one. In general, one cycle of interviews covering the entire sample, using the same questionnaire, is called a wave.

Interviews for each wave of the 2014 SIPP are conducted from February through May each year. For Wave 1, interviews were conducted February through May of 2014, collecting data on January through December 2013. Similarly, respondents in Wave 2, Wave 3, and Wave 4 were interviewed in 2015, 2016, and 2017 respectively and provided data on the 2014, 2015, and 2016 calendar years. Data for up to 12 reference months are available for persons on the

¹ This source and accuracy statement can also be accessed through the U.S. Census Bureau website at <http://www.census.gov/programs-surveys/sipp/tech-documentation/source-accuracy-statements.html>

published files for each wave. Specific months available depend on a person's sample entry or exit date.

The SIPP 2014 Panel began with a designated sample of 53,000 HUs in Wave 1. About 10,500 of these HUs were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. Field Representatives (FRs) obtained interviews for 29,500 of the eligible HUs. FRs were unable to interview 12,500 eligible HUs in the panel because the occupants: (1) refused to be interviewed; (2) could not be found at home; (3) were temporarily absent; or (4) were otherwise unavailable. Thus, occupants of about 70 percent of all eligible HUs participated in the first interview of the panel.

For subsequent interviews, only original sample people (interviewed persons in Wave 1 sample households) and people living with them are eligible to be interviewed. Original sample people who move from their Wave 1 address to a new address in later waves are still included in the SIPP sample. However, FRs attempt telephone interviews in lieu of in-person interviews if their new address is more than 100 miles from a SIPP sample area. In Wave 2, 30,000 HUs were eligible for interview and interviews were obtained from 23,000 of these households. In Wave 3, 19,000 of 31,000 eligible HUs were interviewed. In Wave 4, the final wave of the 2014 Panel, FRs obtained interviews from 17,000 of 31,500 eligible HUs.

Since the SIPP follows all original sample members, those members that form new households are also included in the SIPP sample. This expansion of original households can be estimated within the interviewed sample, but is impossible to determine within the noninterviewed sample. Therefore, a growth factor based on the growth in the known sample is used to estimate the unknown expansion of the noninterviewed households.

Growth factors account for the additional nonresponse stemming from the expansion of noninterviewed households and are calculated for Wave 2 and later waves. They are used to get a more accurate estimate of the weighted number of noninterviewed HUs at each wave. There are two categories of noninterviewed households: Type A and Type D. Type A noninterviewed households are eligible households where the interviewer obtains no interview. Type D noninterviewed households are previously interviewed households who move to an unknown address or outside the SIPP universe; hence, Type D noninterviews only occur from Wave 2 onwards. To calculate this loss of sample, or "sample loss," we use Formula (1):

$$\text{Sample Loss} = \frac{(A_1 \times GF_c) + A_c + D_c}{I_c + (A_1 \times GF_c) + A_c + D_c} \quad (1)$$

where A_1 is the weighted number of Type A noninterviewed households in Wave 1, A_c is the weighted number of Type A noninterviewed households in the Current Wave, D_c is the weighted number of Type D noninterviewed households in the current wave, I_c is the weighted number of interviewed households in the current wave, and GF_c is the growth factor associated with the current wave.

The cumulative weighted sample loss at the end of each wave of the SIPP 2014 Panel was calculated using equation (1) and tabulated in Table A. The distribution of Type A nonrespondents by nonresponse reason is shown in Table B.

Table A. 2014 Panel Household Total Counts, Sample Loss, and Response Rates

Wave	Eligible HUs ²	Interviewed HUs	Type As		Type Ds		Growth Factor	Cumulative Weighted Response Rates (percent)	Weighted Sample Loss (percent)
			Total	Weighted Rate (percent)	Total	Weighted Rate (percent)			
1	42,500	29,500	12,500	31.2				69.8	31.2
2	30,000	23,000	6,400	21.8	700	2.1	1.0	52.3	47.7
3	31,000	19,000	10,000	34.1	1,500	4.6	1.0	42.2	57.8
4	31,500	17,000	12,500	40.4	1,900	5.9	1.1	36.9	63.1

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Table B. 2014 Panel Percent of Type A Nonresponse Status by Wave

Wave	Language Problem	Unable to Locate	No One Home	Temporarily Absent	Household Refused	Other
1	0.7	0.3	11.3	1.3	79.2	7.2
2	0.4	-	6.9	0.9	79.8	12.1
3	0.6	-	7.2	1.1	79.3	11.8
4	0.4	-	5.6	0.7	81.8	11.6

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Weights Produced: The SIPP produces weights for both cross-sectional and longitudinal analysis. Three weights are produced for each wave of the SIPP: monthly weights (*WPFINWGT*), calendar year weights (*CY2013*, *CY2014*, *CY2015*, *CY2016*), and panel weights (*FINPNL1*, *FINPNL2*, *FINPNL3*, *FINPNL4*).

Monthly weights are used to calculate estimates for each of the 12 months within a wave. Calendar year weights cover the reference period from January to December of a specified calendar year, and can be used to calculate estimates for any period within the year. For example, calendar year weights for Wave 1 of the 2014 SIPP can be used to compute monthly³, quarterly, and annual estimates for the time between January to December 2013 while calendar year weights for Wave 4 of the 2014 SIPP can be used to compute similar estimates from January 2016 through December 2016.

² Interviewed and Noninterviewed HUs may not sum up to Eligible HUs due to rounding.

³ We recommend analysts use monthly weights instead of calendar year weights to compute monthly estimates.

Calendar year weights for previous SIPP Panels were based on the SIPP survey universe in January of a specified year, and required respondents be interviewed for *all* months of the year to receive positive calendar year weights. However, the SIPP 2014 calendar year weights are based on the SIPP survey universe in December of a designated year and assign positive calendar year weights to all persons interviewed in December, regardless of their interview status in preceding months. As a result, calendar year weights are equal to the monthly December weights in each wave and separate calendar year weight files are not produced for the SIPP 2014 Panel.

Panel weights cover the reference period from the beginning of the panel to the end of the current wave, and can be used to calculate estimates in this interval. *FINPNL2* weights produced at the end of Wave 2 can be used to compute estimates for the reference period covering January 2013 through December 2014. *FINPNL3* weights produced at the end of Wave 3 can be used to calculate estimates between January 2013 and December 2015. *FINPNL4* weights produced at the end of Wave 4 can be used to compute estimates at any point in the reference period from January 2013 to December 2016.

Monthly and calendar year weights are cross-sectional weights because they are computed from single wave data. Panel weights, however, are computed by combining data from two or more waves and therefore considered longitudinal weights. Table C specifies the reference period for calendar and panel weights in the SIPP 2014 Panel.

Table C. Calendar Year and Panel Weights' Reference Periods for the 2014 Panel

Variable Name	Control Month	Beginning Wave	Beginning Month	Ending Wave	Ending Month
<i>Calendar year weights</i>					
<i>CY2013</i>	December 2013	Wave 1	January 2013	Wave 1	December 2013
<i>CY2014</i>	December 2014	Wave 2	January 2014	Wave 2	December 2014
<i>CY2015</i>	December 2015	Wave 3	January 2015	Wave 3	December 2015
<i>CY2016</i>	December 2016	Wave 4	January 2016	Wave 4	December 2016
<i>Panel weights</i>					
<i>FINPNL1</i>	December 2013	Wave 1	January 2013	Wave 1	December 2013
<i>FINPNL2</i>	December 2013	Wave 1	January 2013	Wave 2	December 2014
<i>FINPNL3</i>	December 2013	Wave 1	January 2013	Wave 3	December 2015
<i>FINPNL4</i>	December 2013	Wave 1	January 2013	Wave 4	December 2016

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

The eligible sample cohort for panel weights consists of persons who are in the SIPP sample universe and interviewed in December 2013. The monthly interview status of persons in this cohort are tracked from December 2013 to the end of the current wave. Eligible persons are then classified as interviewed (for panel weights) if they are interviewed in all subsequent months of the reference period, except for months in which they are *survey universe leavers*.

SIPP survey universe leavers for a given month are defined as sample persons who are known to have died or moved to an ineligible address including: institutions, military barracks, and non-US addresses. Eligible persons who are not interviewed in one or more months following December 2013 to the end of the current wave, and also not survey universe leavers in these months are categorized as noninterviewed.

Longitudinal response rates are computed at the person level as household composition may change over the duration of the panel. Table D shows the response rate associated with each panel weight. *FINPNL2* weighting procedure classified 52,500 of the 72,500 eligible persons as interviewed and had a weighted person response rate of 72.1 percent. *FINPNL3* and *FINPNL4* weighting procedure categorized 38,000 and 29,000 eligible persons respectively as interviewed with response rates of 52.3 percent and 39.8 percent respectively.

Table D. 2014 Panel Longitudinal Person Response Rates.

Panel Weight	Eligible Persons ⁴	Interviewed Persons	Noninterviewed Persons	Weighted Response Rate (percent)
FINPNL2	72,500	52,500	20,000	72.1
FINPNL3	72,500	38,000	34,500	52.3
FINPNL4	72,500	29,000	43,500	39.8

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Calendar year and panel weights have the same reference period for the first wave of the 2014 SIPP, January 2013 to December 2013. Both weights are produced based on the SIPP survey universe in the control month⁵ of December 2013, and hence are the same as the December 2013 monthly weights. This implies that a person must be in the SIPP sample universe and interviewed in December 2013 to receive positive calendar year (*CY2013*) and panel (*FINPNL1*) weights in Wave 1. Therefore, separate longitudinal weight files are not provided for Wave 1. Analysts should use the December 2013 weight and replicate weights provided with Wave 1 cross-sectional files for computing 2013 calendar year (*CY2013*) or first panel (*FINPNL1*) estimates.

All interviewed persons in the survey universe for a given reference month will receive a positive cross-sectional weight for that month, whereas those who are noninterviewed or out of the universe are assigned zero weights. Similarly, all persons classified as interviewed for the reference period of a longitudinal weight are assigned positive weights for that period, while

⁴ Interviewed and Noninterviewed Persons may not sum up to Eligible HUs due to rounding.

⁵ A control month is defined as a month during with the sample universe is under consideration. The SIPP weighting procedure adjusts both cross sectional and longitudinal weights to population estimates for a specific control month.

those classified as noninterviewed or ineligible are assigned zero weights. Longitudinal weights are produced at the completion of the second and later waves.

Estimation: The SIPP estimation procedure involves several stages of weight adjustments to derive the person level weights. For cross-sectional weights, i.e. monthly and calendar weights, each person is first given a base weight (BW) equal to the inverse of their household's probability of selection. Next, a Weighting Control Factor (WCF) is used to adjust for subsampling done in the field when the number of sample units is much larger than expected. Then a noninterview adjustment factor is applied to account for households that were eligible for the sample but which FRs could not interview in Wave 1 (FN_1). For subsequent waves i , the noninterview adjustment factor is (FN_i).

A Mover's Adjustment Factor (MAF) is applied in Waves 2+ to adjust for persons in the SIPP universe who move into sample households after Wave 1. The last adjustment is the Second Stage Adjustment Factor (F_{2S}). This adjusts estimates to population controls (benchmark population estimates) and equalizes married spouses' weights. The 2014 Panel adjusts weights to both national and state level controls for the corresponding control month, i.e. each of the twelve calendar months of interest for monthly weights, and December for calendar year weights.

The final cross-sectional weight for each month c , in Wave i , is $FW_c = BW * WCF * FN_1 * F_{2S}$ for Wave 1 and is $FW_c = BW * WCF * FN_1 * MAF * FN_i * F_{2S}$ for Waves 2+. Additional details of the weighting process are in *SIPP 2014: Weighting Specifications for Wave 1 and SIPP 2014: Cross-Sectional Weighting Specifications for the Second and Subsequent Waves*.

For longitudinal (panel) weights, eligible persons are given an initial weight (IW) equal to their cross-sectional household noninterview adjusted weight for the control month of December 2013. A noninterview adjustment factor (FP_{ni}) is then applied to account for person level nonresponse. Finally, a second stage adjustment (FP_{2S}) is applied to adjust the noninterview weights to independent population controls for December 2013. Spouse weights are not equalized in the panel weighting procedure.

The final panel person weight for Waves 2+⁶ is $PW_i = IW * FP_{ni} * FP_{2S}$. Additional details of the weighting process are in *Survey of Income and Program Participation 2014: Longitudinal Weighting Specifications for the Second and Subsequent Waves*.

Population Controls: The 2014 SIPP estimation procedure adjusts weighted sample estimates to agree with independently derived population estimates of the civilian noninstitutionalized population. This attempts to correct for undercoverage and thereby reduces the mean square error of the estimate. The national and state level population controls are obtained directly from the Population Division and are prepared each month to agree with the most current set

⁶ The final panel weight for Wave 1 is the same as Wave 1 calendar year weight, $FW_1 = BW * WCF * FN_1 * F_{2S}$

of population estimates released by the Census Bureau's population estimates and projections program.

The national level controls are distributed by demographic characteristics as follows:

- Age, Sex, and Race (White Alone, Black Alone, and all other groups combined)
- Age, Sex, and Hispanic Origin

The state level controls are distributed by demographic characteristics as follows:

- State by Age and Sex
- State by Hispanic origin
- State by Race (Black Alone, all other groups combined)

The estimates begin with the latest decennial census as the base and incorporate the latest available information on births and deaths along with the latest estimates of net international migration.

The net international migration component in the population estimates includes a combination of:

- Legal migration to the U.S.,
- Emigration of foreign born and native people from the U.S.,
- Net movement between the U.S. and Puerto Rico,
- Estimates of temporary migration, and
- Estimates of net residual foreign-born population, which include unauthorized migration.

Because the latest available information on these components lags the survey date, to develop the estimate for the survey date, it is necessary to make short-term projections of these components.

Use of Weights: The SIPP 2014 Panel monthly, calendar year, and panel weights are produced at the person level and intended for analyzing data at the person level. Every interviewed person in the SIPP universe for a given reference month has a person month weight for that reference month. Likewise, persons who provided data for the month of December during their interview have calendar year weights, and persons categorized as interviewed for a longitudinal reference period are assigned corresponding panel weights. Chapter 7 of the 2014 *SIPP User's Guide* provides additional information on how to use the weights.

In previous SIPP panels, public use files also contained household, family, and related subfamily monthly weights for analyzing the data at the appropriate household and family levels. These weights were set to be the person month weight of the household, family, or subfamily

reference person⁷ for that reference month⁸. For the SIPP 2014 Panel, the household structure of an interviewed unit is only set for the interview month. Up to five addresses are recorded for each person for the reference period, so interviewed persons can live in different households depending on the reference month. Therefore, for each reference month it is possible to tell which interviewed persons lived together and their relationships to each other, but the files do not specify a household ID or reference person for each of the reference months. The same is true for families. If a data user would like to conduct analysis at the household or family level, person weights can be used to specify a single household or family weight. One option is to use the average of the person month weights for all persons in the household or family. Another option is to specify a household or family reference person and use his or her person month weight as the household or family weight.

All estimates may be divided into two broad categories: longitudinal and cross-sectional. Longitudinal estimates require that data records for each person be linked across interviews, whereas cross-sectional estimates do not. For example, estimating the average duration spell of unemployment from January 2013 to December 2015 requires linking records from Wave 1 through Wave 3 and would be a longitudinal estimate. Because there is no linkage between interviews, cross-sectional estimates can combine data from different interviews only at the aggregate level.

Longitudinal person weights were developed for longitudinal estimation, but may be used for cross-sectional estimation as well. The panel weight can be used to form monthly, quarterly, annual, or multi-year estimates. For example, Wave 4 panel weights, *FINPNL4*, can be used in constructing estimates for time spans in the period between January 2013 and December 2016. However, note that SIPP 2014 wave files produced annually contain monthly (and hence calendar year) cross-sectional weights. Because of the larger sample size with positive weights available on the wave files, it is recommended that these files be used for cross-sectional estimation, if possible. To form an estimate for a particular month, use the monthly cross-sectional weight for the month of interest. Similarly, use calendar year weights to determine estimates for any time frame within a wave.

Data users are strongly encouraged to use the weights provided as instructed. Weights can vary due to the sample design, the level of nonresponse, and the time period of the estimate. Calculating estimates without using the weights as instructed will produce erroneous results.

Some basic types of estimates that can be constructed using the calendar year and panel weights are described below in terms of estimated numbers. More complex estimates, such as percentages, averages, ratios, etc., can be constructed from the estimated numbers.

⁷ During Wave 1 interviews, FRs identify a reference person who is usually the owner or renter of the residence. A household's reference person may change from wave to wave due to changes in the household composition overtime.

⁸ Only person-level calendar year and panel weights were available in previous SIPP panels.

1. The number of people who have ever experienced a characteristic during a given time period.

To construct such an estimate, use the person weight for the shortest time period which covers the entire time period of interest. Then sum the weights over all people who possessed the characteristic of interest at some point during the time period of interest. For example, to estimate the number of people who ever received food stamps during the last quarter of 2013, use Wave 1 calendar year weights, i.e. December weights (*WPFINWGT*, with *monthcode*=12 or *CY2013*), which cover all 12 months of 2013.

To estimate the number of persons who received any unemployment income in 2013 and 2014, sum second panel weights, *FINPNL2*, of all persons who received unemployment income anytime between January 2013 and December 2014.

To estimate the number of households who experienced poverty at any point during calendar years 2013 through 2016, sum the fourth household panel weights, *FINPNL4*, of all households whose total monthly income were less than the monthly household poverty threshold during this time frame. Analysts may define a household's *FINPNL4* weight as the household's reference person's *FINPNL4* weight, or the average *FINPNL4* weight for all members of the household.

2. The amount of a characteristic accumulated by people during a given time period.

To construct such an estimate, use the person weight for the shortest time period which covers the entire time period of interest. Then compute the product of the weight times the amount of the characteristic and sum this product over all appropriate people. For example, to estimate the aggregate 2015 annual income of people who were employed during all 12 months of the year, use Wave 3 calendar year weights, i.e. December weights (*WPFINWGT*, with *monthcode*=12 or *CY2015*). The same estimate could be generated using the *FINPNL3* weights but there may be fewer positively weighted people than those in the calendar year.

3. The average number of consecutive months of possession of a characteristic (i.e., the average spell length for a characteristic) during a given time period.

For example, one could estimate the average length of each spell of receiving food stamps during 2016. One could also estimate the average spell of unemployment that elapsed before a person found a new job. To construct such an estimate, first identify the people who possessed the characteristic at some point during the time period of interest. Then create two sums of these persons' appropriate weights: (1) sum the product of the weight times the number of months the spell lasted and (2) sum the weights only. Now, the estimated average spell length in months is given by (1) divided by (2). A person who experienced two spells during the time period of interest would be

treated as two people and appears twice in sums (1) and (2). An alternate method of calculating the average can be found in the section "Standard Error of a Mean."

4. The number of month-to-month changes in the status of a characteristic (i.e., number of transitions) summed over every set of two consecutive months during the time period of interest.

To construct such an estimate, sum the appropriate person weights each time a change is reported between two consecutive months during the time period of interest. For example, to estimate the number of people who changed from receiving food stamps in July 2015 to not receiving in August 2015, add together the Wave 3 calendar year weights of each person who had such a change.

To estimate the number of changes in monthly salary income during the 2016 fiscal year (September 2015 to September 2016), use *FINPNL4* weights and sum together the estimate of the weighted number of people who had a change between September and October, between October and November, between November and December, ..., and between August and September.

Note that spell and transition estimates should be used with caution because of the biases that are associated with them. Sample people tend to report the same status of a characteristic for all months of a reference period. This tendency also affects transition estimates in that, for many characteristics, the number of characteristics, the number of month-to-month transitions reported between the last month of one reference period and the first month of the next reference period are much greater than the number of reported transitions between any two months within a reference period. Additionally, spells extending before or after the time period of interest are cut off (censored) at the boundaries of the time period. If they are used in estimating average spell length, a downward bias will result.

5. Monthly estimates of a characteristic averaged over a number of consecutive months.

For example, one could estimate the monthly average number of Temporary Assistance for Needy Families (TANF) recipients over the months July 2016 through December 2016. To construct such an estimate, first form an estimate for each month in the time period of interest. Sum the fourth calendar year weight, *CY2016*, of all persons who possessed the characteristic of interest during each of the six months of interest. Then sum the monthly estimates and divide by six, the number of months.

ACCURACY OF ESTIMATES

SIPP estimates are based on a sample; they may differ somewhat from the figures that would have been obtained from a complete census using the same questionnaire, instructions, and

enumerators. There are two types of errors possible in an estimate based on a sample survey: sampling and nonsampling errors. For a given estimator, the difference between an estimate based on a sample and the estimate that would result if the sample were to include the entire population is known as sampling error. For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. We are able to provide estimates of the magnitude of SIPP sampling error, but this is not true of nonsampling error.

Nonsampling Error: Nonsampling errors can be attributed to many sources:

- inability to obtain information about all cases in the sample
- definitional difficulties
- differences in the interpretation of questions
- inability or unwillingness on the part of the respondents to provide correct information
- errors made in the following: collection such as in recording or coding the data, processing the data, estimating values for missing data
- biases resulting from the differing recall periods caused by the interviewing pattern used and undercoverage.

Quality control and edit procedures were used to reduce errors made by respondents, coders and interviewers. More detailed discussions of the existence and control of nonsampling errors in the SIPP can be found in the *SIPP Quality Profile, 1998 SIPP Working Paper Number 230*, issued June 1998 (Kalton, 1998).

Undercoverage in SIPP results from missed HUs and missed persons within sample HUs. It is known that undercoverage varies with age, race, and sex. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks. Ratio estimation to independent age-race-sex population controls partially corrects for the bias due to survey undercoverage. However, biases exist in the estimates to the extent that persons in missed households or missed persons in interviewed households have characteristics different from those of interviewed persons in the same age-race-sex group.

A common measure of survey coverage is the coverage ratio, the estimated population before ratio adjustment divided by the independent population control. Tables E, F, G, and H show SIPP 2014 coverage ratios for age-sex-race groups in December of years 2013, 2014, 2015, and 2016 respectively using calendar year cross-sectional weights prior to the ratio adjustment.

Table I shows the coverage ratios for December 2013 using *FINPNL4* weights prior to post stratification ratio adjustment. Note that Table E also corresponds to the coverage ratios using *FINPNL1* since *CY2013* and *FINPNL1* weights have the same reference period and control months. Therefore, it can be assumed that *FINPNL2* and *FINPNL3* coverage ratios for each age-sex-race-group will be between the values of those in Table E (*CY2013/FINPNL1*) and Table I (*FINPNL4*). The SIPP coverage ratios exhibit some variability from month to month, but these

are a typical set of coverage ratios. Other Census Bureau household surveys (e.g. the Current Population Survey) experience similar coverage.

Table E. Coverage Ratios for December 2013 for *CY2013* Weights by Age, Race and Sex

Age	White Only		Black Only		Residual	
	Male	Female	Male	Female	Male	Female
<15	0.8844	0.8999	0.8025	0.7863	0.8836	0.8164
15	0.8628	0.8244	0.8780	0.8681	0.8846	0.8714
16-17	0.8366	0.8534	0.8371	0.8584	0.8878	0.8851
18-19	0.8463	0.7947	0.8351	0.8699	0.8894	0.8979
20-21	0.8630	0.8451	0.7104	0.6541	0.8532	0.8882
22-24	0.8188	0.7519	0.7137	0.6583	0.8608	0.8938
25-29	0.8113	0.8130	0.6405	0.7262	0.8614	0.7530
30-34	0.8257	0.8698	0.6982	0.7226	0.8394	0.7512
35-39	0.8696	0.9011	0.7681	0.8387	0.7980	0.7813
40-44	0.8504	0.8622	0.7490	0.8440	0.7855	0.7845
45-49	0.8196	0.8856	0.7802	0.7987	0.8465	0.9196
50-54	0.8715	0.9040	0.8066	0.8211	0.8667	0.9129
55-59	0.9286	0.9644	0.7458	0.8467	0.9092	0.9350
60-61	0.9755	1.034	0.7450	0.8193	0.9689	0.9332
62-64	1.003	0.9570	0.7449	0.8568	0.9169	0.9266
65-69	0.9952	1.015	0.9908	0.8958	0.8697	0.9402
70-74	1.013	0.9943	0.9932	0.9185	0.8697	0.9269
75-79	1.035	1.025	0.9937	0.9086	0.8695	0.9325
80-84	1.021	1.064	0.9976	0.9085	0.8724	0.9012
85+	0.9369	0.8719	1.062	0.9047	0.8996	0.9487

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Table F. Coverage Ratios for December 2014 for CY2014 Weights by Age, Race, and Sex

Age	White Only		Black Only		Residual	
	Male	Female	Male	Female	Male	Female
<15	0.8263	0.8443	0.7157	0.6992	0.8599	0.7984
15	0.7775	0.8210	0.7559	0.8194	0.7575	0.7995
16-17	0.7960	0.8226	0.7652	0.8171	0.7766	0.7750
18-19	0.7896	0.7026	0.7557	0.8524	0.7938	0.7899
20-21	0.7018	0.7471	0.6448	0.5472	0.7595	0.7733
22-24	0.6895	0.6293	0.6503	0.5454	0.7575	0.7915
25-29	0.7370	0.7274	0.5307	0.5831	0.7751	0.7442
30-34	0.7423	0.7909	0.6341	0.6509	0.7879	0.7450
35-39	0.8041	0.8746	0.6601	0.7749	0.7820	0.7532
40-44	0.7898	0.8199	0.7155	0.7572	0.7750	0.7519
45-49	0.7922	0.8315	0.8082	0.7256	0.7716	0.8357
50-54	0.8099	0.8461	0.7383	0.8208	0.7807	0.8538
55-59	0.8889	0.9306	0.7284	0.8514	0.8913	0.9820
60-61	0.9818	1.018	0.7531	0.8475	0.9344	1.0190
62-64	0.9783	0.9728	0.7417	0.8440	0.9163	0.9939
65-69	0.9970	1.0180	1.0720	0.9341	0.9868	0.9419
70-74	0.9967	1.0100	1.0980	0.9835	0.9305	0.9687
75-79	1.0820	1.0380	1.0720	0.9998	0.9232	0.9740
80-84	1.1770	1.1070	1.0740	0.9780	0.9586	0.9626
85+	0.9985	0.9428	1.0900	0.9887	0.9768	0.9099

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Table G. Coverage Ratios for December 2015 for CY2015 Weights by Age, Race, and Sex

Age	White Only		Black Only		Residual	
	Male	Female	Male	Female	Male	Female
<15	0.8185	0.7878	0.6988	0.6989	0.7998	0.7821
15	0.8882	0.8672	0.6898	0.7122	0.7354	0.6670
16-17	0.7896	0.7563	0.6925	0.6969	0.7287	0.6849
18-19	0.6997	0.7235	0.6671	0.7041	0.7479	0.7307
20-21	0.7334	0.6377	0.5807	0.5307	0.7264	0.7278
22-24	0.6131	0.6317	0.5944	0.5152	0.6954	0.6821
25-29	0.6563	0.6352	0.4562	0.4895	0.7144	0.6419
30-34	0.7196	0.7734	0.5861	0.6087	0.7190	0.6491
35-39	0.7783	0.8113	0.6283	0.7325	0.7512	0.6923
40-44	0.7701	0.8086	0.6418	0.6519	0.7641	0.7071
45-49	0.7325	0.7755	0.7033	0.6665	0.7710	0.7656
50-54	0.8118	0.8633	0.6473	0.7450	0.7909	0.8042
55-59	0.8513	0.9116	0.7671	0.9373	0.8659	0.9893
60-61	0.9290	1.0110	0.7550	0.9969	0.9278	1.0700
62-64	1.0270	1.0460	0.8101	0.9184	0.9082	0.9856
65-69	1.0140	0.9797	1.1170	1.0230	0.9192	0.9202
70-74	0.9843	1.066	1.1000	1.0570	0.8623	0.9136
75-79	1.1200	1.0920	1.1010	0.9836	0.8608	0.9390
80-84	1.1830	1.2170	1.1670	1.0760	0.9111	0.9646
85+	1.1660	1.0000	1.0890	1.012	0.8725	0.8810

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Table H. Coverage Ratios for December 2016 for CY2016 Weights by Age, Race, and Sex

Age	White Only		Black Only		Residual	
	Male	Female	Male	Female	Male	Female
<15	0.7588	0.7632	0.6238	0.6757	0.8440	0.8176
15	0.7840	0.8406	0.7146	0.7102	0.7239	0.7714
16-17	0.7675	0.7508	0.7136	0.7310	0.7401	0.7329
18-19	0.6854	0.7241	0.7214	0.6922	0.7299	0.7909
20-21	0.6610	0.5958	0.5895	0.5120	0.7546	0.7821
22-24	0.5698	0.6161	0.5797	0.5095	0.7017	0.7220
25-29	0.6485	0.6056	0.5055	0.4745	0.6357	0.6090
30-34	0.6768	0.7073	0.4953	0.5497	0.6484	0.6608
35-39	0.7224	0.7785	0.5086	0.6652	0.8746	0.7620
40-44	0.7068	0.7262	0.7468	0.7307	0.9017	0.7729
45-49	0.7249	0.8213	0.6744	0.6783	0.7694	0.7428
50-54	0.7773	0.7945	0.6830	0.7115	0.8450	0.7635
55-59	0.7994	0.8855	0.8093	0.8883	0.8771	1.0150
60-61	0.9569	1.0170	0.8114	0.9357	0.9153	1.0080
62-64	1.0130	1.0070	0.8585	0.9124	0.9381	1.0300
65-69	1.0010	1.0220	1.1830	1.0700	0.9761	0.9735
70-74	1.0540	1.1350	1.1390	1.0760	0.8615	0.9883
75-79	1.1520	1.0520	1.0970	1.0760	0.9408	0.9545
80-84	1.1640	1.2430	1.1360	1.0810	0.9199	1.0050
85+	1.2210	1.0400	1.1530	1.1450	0.9222	0.9760

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Table I. Coverage Ratios for December 2013 for *FINPNL4* Weights by Age, Race, and Sex

Age	White Only		Black Only		Residual	
	Male	Female	Male	Female	Male	Female
<15	0.8668	0.8643	0.6096	0.6303	0.8996	0.7691
15	0.8169	0.7893	0.6029	0.4891	0.8111	0.6591
16-17	0.7281	0.6623	0.5436	0.6108	0.7644	0.7336
18-19	0.6032	0.5253	0.6128	0.5850	0.7848	0.6864
20-21	0.5478	0.5996	0.4260	0.3722	0.6371	0.7145
22-24	0.6390	0.5427	0.4523	0.4061	0.6402	0.6707
25-29	0.6906	0.6649	0.3882	0.5682	0.9480	0.7276
30-34	0.7351	0.8159	0.5237	0.5861	0.9131	0.6829
35-39	0.7970	0.8025	0.7243	0.6709	0.8048	0.8161
40-44	0.8175	0.7927	0.6100	0.6378	0.7737	0.7949
45-49	0.7810	0.8629	0.7738	0.7469	1.0330	0.9897
50-54	0.9198	0.9301	0.8862	0.8485	0.9955	0.9820
55-59	1.0080	1.1270	0.8845	1.2170	1.0080	1.1230
60-61	1.2100	1.2180	0.9504	1.1200	0.9477	1.1170
62-64	1.2080	1.1570	0.8535	1.0950	0.9419	1.1110
65-69	1.1200	1.2530	1.5260	1.2620	1.0820	1.1790
70-74	1.2550	1.2140	1.4950	1.2170	1.1290	1.0710
75-79	1.3110	1.2970	1.3850	1.2380	1.1130	1.1490
80-84	1.3400	1.3250	1.5060	1.3490	1.0960	1.1180
85+	1.2290	0.9309	1.5380	1.2700	1.2080	1.0930

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Comparability with Other Estimates: Caution should be exercised when comparing the SIPP 2014 data with data from other SIPP products⁹ or with data from other surveys. The comparability problems are caused by sources such as the seasonal patterns for many characteristics, different nonsampling errors, and different concepts and procedures. Refer to the *SIPP Quality Profile* for known differences with data from other sources and further discussions.

Sampling Variability: Standard errors indicate the magnitude of the sampling error. They also partially measure the effect of some nonsampling errors in response and enumeration, but do not measure any systematic biases in the data. The standard errors for the most part measure the variations that occurred by chance because a sample rather than the entire population was surveyed.

USES AND COMPUTATION OF STANDARD ERRORS

Confidence Intervals: The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range about a given estimate that has a known probability of including the result of a complete enumeration. For example, if all possible samples were selected, each of these being surveyed under essentially the same conditions and using the same sample design, and if an estimate and its standard error were calculated from each sample, then:

1. Approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate of all possible samples would include the actual value of the population parameter.
2. Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate of all possible samples would include the actual value of the population parameter.
3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate of all possible samples would include the actual value of the population parameter.

The estimate derived from all possible samples may or may not be contained in any particular computed interval. However, for a particular sample, one can say with a specified confidence

⁹The reference period covered by the SIPP 2008 (May 2008 to November 2013) and SIPP 2014 (January 2013 to December 2016) Panels overlap from January 2013 to November 2013. Analysts should be especially cautious when comparing estimates from this interval between the two panels as the SIPP was redesigned following the SIPP 2008 Panel. The questionnaire, interview frequency, collection format, and sample size of the SIPP 2014 Panel differ from those of previous SIPP panels. Detailed information on the reengineered SIPP are provided in the 2014 SIPP Users' Guide <https://www2.census.gov/programs-surveys/sipp/tech-documentation/methodology/2014-SIPP-Panel-Users-Guide.pdf>

that the estimate derived from all possible samples is included in the confidence interval.

Hypothesis Testing: Standard errors may also be used for hypothesis testing, a procedure for distinguishing between population characteristics using sample estimates. The most common types of hypotheses tested are 1) the population characteristics are identical versus 2) they are different. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical. The Census Bureau's uses 90 percent confidence level for hypothesis testing but stricter confidence levels – for example 95 percent or higher – may also be used.

To perform the most common test, compute the difference $X_A - X_B$, where X_A and X_B are sample estimates of the characteristics of interest. A later section explains how to derive an estimate of the standard error of the difference $X_A - X_B$. Let that standard error be S_{DIFF} . Calculate a test statistic $|z|$, as $|z| = |X_A - X_B| / S_{DIFF}$. For large sample sizes, one can conclude the difference is significant at the 90 percent confidence level if $|z|$ is larger than the critical value of 1.645.

Confidence intervals can also be used to test for significant difference between two sample estimates. If $X_A - X_B$ is between $(-1.645 \times S_{DIFF})$ and $(+1.645 \times S_{DIFF})$, no conclusion about the characteristics is justified at the 10 percent significance level. If, on the other hand $X_A - X_B$, is smaller than $(-1.645 \times S_{DIFF})$ or larger than $(+1.645 \times S_{DIFF})$, the observed difference is significant at the 10 percent significance level. In this event, it is commonly accepted practice to say that the characteristics are significantly different. In accordance with the Census Bureau's Statistical Quality Standards, we recommend that users report only those differences that are significant at the 10 percent significance level or better. Of course, sometimes this conclusion will be wrong. When the characteristics are the same, there is a 10 percent chance of concluding that they are different.

Note that as more tests are performed, more erroneous significant differences will occur. For example, at the 10 percent significance level, if 100 independent hypothesis tests are performed in which there are no real differences, it is likely that about 10 erroneous differences will occur. Therefore, the significance of any single test should be interpreted cautiously. A Bonferroni correction can be used to account for this potential problem. The Bonferroni method involves dividing your stated level of significance by the number of tests you are performing (Sedgwick, 2014; Stoline, 1981). This correction results in a conservative test of significance.

Note Concerning Small Estimates and Small Differences: Because of the large standard errors involved, there is little chance that estimates will reveal useful information when computed on a small number of sample cases. We recommend minimum weighted estimates – i.e. the weighted totals of the sample cases included in the analysis – of 220,000 and 320,000 for computing cross-sectional and longitudinal estimates respectively. Nonsampling error in one or more of the small number of cases providing the estimation can also cause large relative error in that particular estimate. Care must be taken in the interpretation of small differences since

even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Calculating Standard Errors for SIPP Estimates: There are three main ways we calculate the Standard Errors (SEs) for SIPP Estimates. They are as follows:

- Direct estimates using replicate weight methods (highly recommended);
- Generalized variance function parameters (denoted as a and b); and
- Simplified tables of SEs based on the a and b parameters.

Replicate weight methods provide the most accurate variance estimates but may require more computing resources and expertise on the part of the user. The Generalized Variance Function (GVF) parameters provide a method of balancing accuracy with resource usage as well as having a smoothing effect on SE estimates across time. SIPP uses the Replicate Weighting Method to produce GVF parameters (see K. Wolter, *Introduction to Variance Estimation*, for more information). The GVF parameters are used to create the simplified tables of SEs.

Standard Error Parameters and Tables and Their Use: Most SIPP estimates have greater standard errors than those obtained through a simple random sample because of its two-stage clustered sample design. To derive standard errors that would be applicable to a wide variety of estimates and could be prepared at a moderate cost, a number of approximations were required.

Estimates with similar standard error behavior were grouped together and two parameters (denoted as a and b) were developed to approximate the standard error behavior of each group of estimates. Because the actual standard error behavior was not identical for all estimates within a group, the standard errors computed from these parameters provide an indication of the order of magnitude of the standard error for any specific estimate. These a and b parameters vary by characteristic and by demographic subgroup to which the estimate applies. Tables 1 and 7 provide a and b parameters for the core domains to be used for the 2014 Panel cross-sectional (monthly and calendar year) and longitudinal (panel) estimates respectively.

The creation of appropriate a and b parameters for the previously discussed types estimates are described below.

1. The number of people who have ever experienced a characteristic during a given time period.

The appropriate a and b parameters are obtained directly from Tables 1 or 7. The choice of parameter depends on the weights used, the characteristic of interest, and the demographic subgroup of interest.

2. Amount of a characteristic accumulated by people during a given time period.

The appropriate ***b*** parameters are also taken directly from Tables 1 or 7.

3. The average number of consecutive months of possession of a characteristic per spell (i.e., the average spell length for a characteristic) during a given time period.

Start with the appropriate base ***a*** and ***b*** parameters from Tables 1 or 7. The parameters are then inflated by an additional factor, *g*, to account for people who experience multiple spells during the time period of interest. This factor is computed by:

$$g = \frac{\sum_{i=1}^n m_i^2}{\sum_{i=1}^n m_i} \quad (2)$$

where there are *n* people with at least one spell and *m_i* is the number of spells experienced by person *i* during the time period of interest.

4. The number of month-to-month changes in the status of a characteristic (i.e., number of transitions) summed over every set of two consecutive months during the time period of interest.

Obtain a set of adjusted ***a*** and ***b*** parameters exactly as just described in 3, then multiply these parameters by an additional factor. Use 1.0 if the time period of interest is two months and 2.0 for a longer time period. (The factor of 2.0 is based on the conservative assumption that each spell produces two transitions within the time period of interest.)

5. Monthly estimates of a characteristic averaged over a number of consecutive months.

Appropriate base ***a*** and ***b*** parameters are taken from Tables 1 or 7. If more than one longitudinal weight has been used in the monthly average (i.e., when Wave 2+ files are available), then there is a choice of parameters. Choose the table which gives the largest parameter.

For users who wish further simplification, we have also provided base standard errors for estimates of totals and percentages in Tables 2 through 5. Note that these base standard errors must be adjusted by a *f* factor provided in Tables 1 or 7 depending on the domain and type of estimate being calculated i.e. cross-sectional or longitudinal estimates. The standard errors resulting from this simplified approach are less accurate. Methods for using these parameters and tables for computation of standard errors for different estimates are given in the following sections. Later, we will describe how to use software packages to directly compute standard errors using replicate weights.

Standard Errors of Estimated Numbers: The approximate standard error, s_x , of an estimated number of persons, households, families, unrelated individuals and so forth, can be obtained in two ways. Note that neither method should be applied to dollar values.

The standard error may be obtained by the use of Formula (3):

$$s_x = f \times s, \quad (3)$$

where f is the appropriate f factor from Tables 1 or 7, and s is the base standard error on the estimate obtained by interpolation from Tables 2 or 3.

Alternatively, s_x may be approximated by Formula (4):

$$s_x = \sqrt{ax^2 + bx} \quad (4)$$

Here x is the size of the estimate and a and b are the appropriate parameters from Tables 1 or 7 associated with the characteristic being estimated (and the wave which applies). This formula was used to calculate the base standard errors in Tables 2 and 3. Use of Formula (4) will generally provide more accurate results than the use of Formula (3).

Illustration 1.

Suppose SIPP estimates based on Wave 1 of the 2014 panel show that there were 4,000,000 females aged 25 to 44 with a monthly earned income greater than \$6,000 in September 2013. The appropriate parameters and factor from Table 1a are:

$$a = -0.00004570 \quad b = 5,925 \quad f = 1.056$$

Table 3 does not contain an entry for a sample size of 4,000,000; therefore its corresponding base standard error s , in Formula (3) will be computed via linear interpolation.

First determine the sample size, x_1 , (and its base standard error, y_1) on Table 3 that is closest to and also less than 4,000,000. Next determine the sample size, x_2 , (and its base standard error, y_2) on Table 3 that is closest to and also greater than 4,000,000. The linear interpolated base standard error s for a sample size x is defined as:

$$s = y_1 + \frac{(y_2 - y_1) \times (x - x_1)}{(x_2 - x_1)}$$

In our example, $x_1 = 3,000,000$, $y_1 = 125,700$, $x_2 = 5,000,000$, $y_2 = 161,700$ and $x = 4,000,000$

$$s = 125,700 + \frac{(161,700 - 125,700) \times (4,000,000 - 3,000,000)}{(5,000,000 - 3,000,000)} = 143,700$$

Using Formula (3), the approximate standard error is:

$$s_x = 1.056 \times 143,700 = 151,747$$

Using Formula (4), the approximate standard error is:

$$s_x = \sqrt{(-0.00004570 \times 4,000,000^2) + (5,925 \times 4,000,000)} = 151,555 \text{ females.}$$

Using the standard error based on Formula (4), the approximate 90 percent confidence interval as shown by the data is from 3,750,692 to 4,249,308 females (*i. e.*, $4,000,000 \pm 1.645 \times 151,555$). Therefore, 90 percent of confidence intervals from all possible samples will contain the actual number of females aged 25 to 44 with a monthly earned income greater than \$6,000 in September 2013 in the population.

Standard Error of a Mean: A mean is defined here to be the average quantity of some item (other than persons, families, or households) per person, family or household. For example, it could be the average monthly household income for a specified demographic. The standard error of a mean can be approximated by Formula (5) below. Because of the approximations used in developing Formula (5), an estimate of the standard error of the mean obtained from this formula will generally underestimate the true standard error. The formula used to estimate the standard error of a mean \bar{x} is:

$$s_{\bar{x}} = \sqrt{\left(\frac{b}{y}\right) s^2}, \quad (5)$$

where y is the size of the base, s^2 is the estimated population variance of the item and b is the parameter associated with the particular type of item.

The population variance s^2 may be estimated by one of two methods. In both methods, we assume x_i is the value of the item for i^{th} unit. (A unit may be person, family, or household). To use the first method, the range of values for the item is divided into c intervals. The lower and upper boundaries of interval j are Z_{j-1} and Z_j , respectively. Each unit, x_i , is placed into one of c intervals such that $Z_{j-1} < x_i \leq Z_j$. The estimated population mean, \bar{x} , and variance, s^2 , are given by the formulas:

$$\bar{x} = \sum_{j=1}^c p_j m_j$$

$$s^2 = \sum_{j=1}^c p_j m_j^2 - \bar{x}^2 \quad (6)$$

where $m_j = (Z_{j-1} + Z_j)/2$, and p_j is the estimated proportion of units in the interval j . The most representative value of the item in the interval j is assumed to be m_j . If the interval c is open-ended, or no upper interval boundary exists, then an approximate value for m_c is

$$m_c = \frac{3}{2} Z_{c-1}.$$

In the second method, the estimated population mean, \bar{x} , and variance, s^2 are given by:

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

$$s^2 = \frac{\sum_{i=1}^n w_i x_i^2}{\sum_{i=1}^n w_i} - \bar{x}^2 \quad (7)$$

where there are n units with the item of interest and w_i is the final weight for i^{th} unit. (Note that $\sum w_i = y$.)

Illustration 2.

Method 1

Suppose that based on Wave 2 data, the distribution of annual income for persons aged 25 to 34 who were employed for all 12 months of 2014 is given in Table 6. Using these data, the mean monthly cash income for persons aged 25 to 34 is \$38,703. Applying Formula (6), the approximate population variance, s^2 , is:

$$s^2 = \left(\frac{370}{23,527}\right)(2,500)^2 + \dots + \left(\frac{2,138}{23,527}\right)(105,000)^2 - (38,703)^2 = 649,442,787$$

Using Formula (5) and a b parameter of 7,880 from Table 1b, the estimated standard error of a mean \bar{x} is:

$$s_{\bar{x}} = \sqrt{\frac{7,880}{23,527,000} \times 649,442,787} = \$466$$

Thus, the approximate 90 percent confidence interval as shown by the data ranges from \$37,936 to \$39,470.

Method 2

Suppose that we are interested in estimating the average length of spells of food stamp reciprocity during the entire period covered by the SIPP 2014 Panel, i.e. from January 2013 to December 2016, for Black subpopulation. Also, suppose there are only 10 sample people in the subpopulation who were food stamp recipients. (This example is a hypothetical situation used for illustrative purposes only; actually, 10 sample cases would be too few for a reliable estimate and the sum of weights is lower than the recommended size of 320,000). The number of consecutive months of food stamp reciprocity during 2014 and final panel weights (*FINPNL4*) are given in Table J below for each sample person:

Table J. Hypothetical Food Stamp Reciprocity Spells Among SIPP Sample Persons
(Not Actual Data, Only Use for Calculation Illustrations).

Sample Person	Spell Length in Months	Panel Four Weight (<i>FINPNL4</i>)
1	4, 3	12,190
2	5	9,500
3	9	8,063
4	3, 3, 2	9,298
5	12	10,070
6	12	2,819
7	4, 1	7,441
8	7	14,060
9	6	5,829
10	4	11,590

Using formula (7), the average spell of food stamp reciprocity is estimated to be:

$$\bar{x} = \frac{(12,190)(4) + (12,190)(3) + \dots + (11,590)(4)}{12,190 + 12,190 + \dots + 11,590} = \frac{651,408}{129,087} = 5.046$$

The standard error will be computed by Formula (5). First, the estimated population variance can be obtained by Formula (7):

$$s^2 = \frac{(12,190)(4)^2 + (12,190)(3)^2 + \dots + (11,590)(4)^2}{12,190 + 12,190 + \dots + 11,590} - 5.046^2 = 9.140(\text{months})^2$$

Next, the base *b* parameter of 14,030 is taken from Table 7c and multiplied by the factor computed from Formula (2):

$$g = \frac{2^2 + 1 + 1 + 3^2 + 1 + 1 + 2^2 + 1 + 1 + 1}{2 + 1 + 1 + 3 + 1 + 1 + 2 + 1 + 1 + 1} = 1.714$$

Therefore, the final **b** parameter is $1.714 \times 14,030 = 24,047$ and the standard error of the mean from Formula (5) is:

$$s_{\bar{x}} = \sqrt{\frac{(24,047)(9.140)}{129,087}} = 1.305 \text{ months}$$

Standard Error of an Aggregate: An aggregate is defined to be the total quantity of an item summed over all the units in a group. The standard error of an aggregate can be approximated using Formula (8). As with the estimate of the standard error of a mean, the estimate of the standard error of an aggregate will generally underestimate the true standard error. Let y be the size of the base, s^2 be the estimated population variance of the item obtained using Formula (6) or Formula (7) and b be the parameter associated with the particular type of item. The standard error of an aggregate is:

$$s_x = \sqrt{b \times y \times s^2}. \quad (8)$$

Standard Errors of Estimated Percentages: The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total upon which the percentage is based. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. For example, the percent of people employed is more reliable than the estimated number of people employed. When the numerator and denominator of the percentage have different parameters, use the parameter (and appropriate factor) of the numerator. If proportions are presented instead of percentages, note that the standard error of a proportion is equal to the standard error of the corresponding percentage divided by 100.

There are two types of percentages commonly estimated. The first is the percentage of people sharing a particular characteristic such as the percent of people owning their own home. The second type is the percentage of money or some similar concept held by a particular group of people or held in a particular form. Examples are the percent of total wealth held by people with high income and the percent of total income received by people on welfare.

For the percentage of people, the approximate standard error, $s_{(x,p)}$, of the estimated percentage p can be obtained by the formula:

$$s_{(x,p)} = f \times s, \quad (9)$$

where f is the appropriate f factor from Tables 1 or 7 (for the appropriate reference period) and s is the base standard error of the estimate from Tables 4 or 5.

Alternatively, it may be approximated by the formula:

$$s_{(x,p)} = \sqrt{\frac{b}{x}(p)(100 - p)}, \quad (10)$$

from which the standard errors in Tables 4 and 5 were calculated. Here x is the size of the subclass of social units which is the base of the percentage, p is the percentage ($0 < p < 100$), and b is the parameter associated with the characteristic in the numerator. Use of Formula (10) will give more accurate results than use of Formula (9).

Illustration 3.

Suppose that using the third panel weight, *FINPNL3*, it was estimated that 21,610,000 males were in poverty in December 2014 and an estimated 5.482 percent of them exited poverty in January 2015. Using Formula (10), with a b parameter of 10,310 from Table 7b, the approximate standard error is:

$$s_{(x,p)} = \sqrt{\frac{10,310}{21,610,000} \times 5.482 \times (100 - 5.482)} = 0.4972 \text{ percent}$$

Consequently, the 90 percent confidence interval as shown by these data is from 4.664 percent to 6.300 percent.

For percentages of money, a more complicated formula is required. A percentage of money will usually be estimated in one of two ways. It may be the ratio of two aggregates:

$$p_R = 100 \left(\frac{x_A}{x_B} \right),$$

or it may be the ratio of two means with an adjustment for different bases:

$$p_R = 100 \left(\hat{p}_A \left(\frac{\bar{x}_A}{\bar{x}_B} \right) \right),$$

where x_A and x_B are aggregate money figures, \bar{x}_A and \bar{x}_B are mean money figures, and \hat{p}_A is the estimated number in group A divided by the estimated number in group B. In either case, we estimate the standard error as

$$s_I = \sqrt{\left(\frac{\hat{p}_A \bar{x}_A}{\bar{x}_B} \right)^2 \left[\left(\frac{s_p}{\hat{p}_A} \right)^2 + \left(\frac{s_A}{\bar{x}_A} \right)^2 + \left(\frac{s_B}{\bar{x}_B} \right)^2 \right]}, \quad (11)$$

where s_p is the standard error of \hat{p}_A , s_A is the standard error of \bar{x}_A and s_B is the standard error of \bar{x}_B . To calculate s_p , use Formula (10). The standard errors of \bar{x}_B and \bar{x}_A may be calculated using Formula (5).

It should be noted that there is frequently some correlation between \hat{p}_A , \bar{x}_B , and \bar{x}_A . Depending on the magnitude and sign of the correlations, the standard error will be over or underestimated.

Illustration 4.

In October 2016, 10.89 percent of males 16 years or older, who were employed for at least one week out of the month were Black. The average monthly earnings of these Black males was \$4,885 and the average monthly earnings of all males 16 years and older who were employed for at least one week in the same month was \$7,862. The corresponding standard errors were 0.20 percent, \$573, and \$2,499 respectively. The percent of all male earning in this age group by Blacks in October 2016 is

$$100 \left(0.1089 \times \frac{4,885}{7,862} \right) = 6.766 \text{ percent}$$

Using formula (11), the corresponding standard error is

$$s_i = \sqrt{\left(\frac{0.1089 \times 4,885}{7,862} \right)^2 \left[\left(\frac{0.0020}{0.1089} \right)^2 + \left(\frac{573}{4,885} \right)^2 + \left(\frac{2,499}{7,862} \right)^2 \right]} = 2.296 \text{ percent}$$

Standard Error of a Difference: The standard error of a difference between two sample estimates is approximately equal to

$$s_{(x-y)} = \sqrt{s_x^2 + s_y^2 - r s_x s_y} \quad (12)$$

where s_x and s_y are the standard errors of the estimates x and y .

The estimates can be numbers, percent, ratios, etc. The correlation between x and y is represented by r . The above formula assumes that the correlation coefficient between the characteristics estimated by x and y is non-zero. If no correlations have been provided for a given set of x and y estimates, assume $r = 0$. However, if the correlation is really positive (negative), then this assumption will tend to cause overestimates (underestimates) of the true standard error.

Illustration 5.

Supposed that SIPP estimates show 1,369,000 persons under 18 years old and 2,221,000 persons age 65 or older received Federal and/or State administered Supplemental Security Income (SSI) in December 2015. Then, using the parameters $a = -0.00003492$ and $b = 8,945$ from Table 1c and Formula (4), the standard errors of these numbers are approximately 110,364 and 140,337 respectively. The difference in sample estimates is 852,000 and using Formula (12), the approximate standard error of the difference is:

$$\sqrt{110,364^2 + 140,337^2} = 178,535.$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons receiving SSI was different for persons under age 18 compared to persons age 65 or older. To perform the test, compute the Z statistic and compare it to the critical value of 1.645 for the 10 percent significance level.

$$Z = \frac{852,000}{178,535} = 4.77$$

Since 4.77 is greater than 1.645, we can conclude that the number of SSI recipients in the two age groups are significantly different at the 10 percent significance level.

We can also compare the difference of 852,000 to the product $1.645 \times 178,535 = 293,690$. Since the difference is greater than 1.645 times the standard error of the difference, we come to the same conclusion that the number of SSI recipients in both age groups are significantly different at the 10 percent significance level.

Standard Error of a Median: The median quantity of some item such as income for a given group of people is that quantity such that at most 50 percent of the group have less and at most 50 percent of the group have more. The sampling variability of an estimated median depends upon the form of the distribution of the item as well as the size of the group. To calculate standard errors on medians, the procedure described below may be used.

The median, like the mean, can be estimated using either data which have been grouped into intervals or ungrouped data. If grouped data are used, the median is estimated using Formulas (13) or (14) with $q = 0.5$, where q is the proportion of a group possessing characteristics of interest. If ungrouped data are used, the data records are ordered based on the value of the characteristic, then the estimated median is the value of the characteristic such that the weighted estimate of half – i.e. 50 percent – of the subpopulation falls at or below that value and half is at or above that value. Note that the method of standard error computation which is presented here requires the use of grouped data. Therefore, it should be easier to compute the median by grouping the data and using Formula (13) or (14).

An approximate method for measuring the reliability of an estimated median is to determine a confidence interval about it. (See the section on sampling variability for a general discussion of confidence intervals.) The following procedure may be used to estimate the 68 percent confidence limits and hence the standard error of a median based on sample data.

1. Determine, using either Formula (9) or $p = 50$ in Formula (10), the standard error of an estimate of 50 percent of the group.
2. Add to and subtract from 50 percent, the standard error determined in step 1.

3. Using the distribution of the item within the group, calculate the quantity of the item such that the percent of the group with more of the item is equal to the smaller percentage found in step 2. This quantity will be the upper limit for the 68 percent confidence interval. In a similar fashion, calculate the quantity of the item such that the percent of the group with more of the item is equal to the larger percentage found in step 2. This quantity will be the lower limit for the 68 percent confidence interval.
4. Divide the difference between the two quantities determined in step 3 by two to obtain the standard error of the median.

To perform step 3, it will be necessary to interpolate. Different methods of interpolation may be used. The most common are simple linear interpolation and Pareto interpolation. The appropriateness of the method depends on the form of the distribution around the median. If density is declining in the area, then we recommend Pareto interpolation. If density is fairly constant in the area, then we recommend linear interpolation. Note, however, that Pareto interpolation can never be used if the interval contains zero or negative measures of the item of interest. Interpolation is used as follows.

The quantity of the item such that p percent have more of the item is:

$$X_{pN} = A_1 \times \exp \left[\left(\frac{\ln \left(\frac{qN}{N_1} \right)}{\ln \left(\frac{N_2}{N_1} \right)} \right) \ln \left(\frac{A_2}{A_1} \right) \right] \quad (13)$$

if Pareto Interpolation is indicated and:

$$X_{pN} = \left[A_1 + \left(\frac{qN - N_1}{N_2 - N_1} \right) (A_2 - A_1) \right], \quad (14)$$

if linear interpolation is indicated, where:

N	is the size of the group,
A_1 and A_2	are the lower and upper bounds, respectively, of the interval in which X_{pN} falls
N_1 and N_2	are the estimated number of group members owning more than A_1 and A_2 , respectively
exp	refers to the exponential function and
ln	refers to the natural logarithm function

q is the corresponding proportion for p i.e., $q = \frac{p}{100}$

Illustration 6.

To illustrate the calculations for the sampling error on a median, we return to Table 6. The median annual income for this group using Formula (13) is \$31,828. The size of the group is 23,527,000.

1. Using Formula (10), the standard error of 50 percent on a base of 23,527,000 is about 0.92 percentage points.
2. Following step 2, the two percentages of interest are 49.08 and 50.92.
3. By examining Table 6, we see that the percentage 49.08 falls in the income interval from \$30,000 to \$39,999. (Since 54.7 percent receive more than \$30,000 per annum, the dollar value corresponding to 49.08 must be between \$30,000 and \$40,000.) Thus, $A_1 = \$30,000$, $A_2 = \$40,000$, $N_1 = 12,881,000$ and $N_2 = 8,285,000$.

In this case, we decided to use Pareto interpolation. Therefore, using Formula (13), the upper bound of a 68 percent confidence interval for the median is

$$\$30,000 \times \exp \left[\left(\frac{\ln \left(\frac{0.4908 \times 23,527,000}{12,881,000} \right)}{\ln \left(\frac{8,285,000}{12,881,000} \right)} \right) \times \ln \left(\frac{40,000}{30,000} \right) \right] = \$32,216.$$

Also by examining Table 6, we see that 50.92 falls in the same income interval. Thus, A_1, A_2, N_1 and N_2 are the same. We also use Pareto interpolation for this case. So the lower bound of a 68 percent confidence interval for the median is

$$\$30,000 \times \exp \left[\left(\frac{\ln \left(\frac{0.5092 \times 23,527,000}{12,881,000} \right)}{\ln \left(\frac{8,285,000}{12,881,000} \right)} \right) \times \ln \left(\frac{40,000}{30,000} \right) \right] = \$31,452.$$

Thus, the 68 percent confidence interval on the estimated median is from \$31,452 to \$32,216.

4. Then the approximate standard error of the median is

$$\frac{\$32,216 - \$31,452}{2} = \$382$$

Standard Errors of Ratios of Means and Medians: The standard error for a ratio of means or medians is approximated by:

$$s_{\frac{x}{y}} = \sqrt{\left(\frac{x}{y}\right)^2 \left[\left(\frac{s_y}{y}\right)^2 + \left(\frac{s_x}{x}\right)^2\right]}, \quad (15)$$

where x and y are the means or medians, and s_x and s_y are their associated standard errors. Formula (15) assumes that the means are not correlated. If the correlation between the population means estimated by x and y are actually positive (negative), then this procedure will tend to produce overestimates (underestimates) of the true standard error for the ratio of means.

Standard Errors Using Software Packages: Standard errors and their associated variance, calculated by statistical software packages such as SAS or Stata, do not accurately reflect the SIPP's complex sample design. Erroneous conclusions will result if these standard errors are used directly. We provide adjustment factors by characteristics that should be used to correctly compensate for likely under-estimates. The factors called design effects (DEFF), available in Tables 1 and 7, must be applied to SAS or Stata generated variances. The square root of DEFF can be directly applied to similarly generated standard errors. These factors approximate design effects which adjust statistical measures for sample designs more complex than simple random sample.

Replicate weights for SIPP are also provided and can be used to estimate more accurate standard errors and variances. While replicate weighting methods require more computing resources, many statistical software packages, including SAS, have procedures that simplify the use of replicate weights for users. To calculate variances using replicate weights use the formula:

$$Var(\theta_0) = \frac{1}{G(0.5)^2} \times \sum_{i=1}^G (\theta_i - \theta_0)^2 \quad (16)$$

where G is the number of replicates, θ_0 is the estimate using full sample weights, and θ_i is the estimate using the replicate weights. For the 2014 panel, $G=240$ for the number of replicate weights provided in the public use files. Replicate weights are created using Fay's method, with a Fay coefficient of 0.5 (Chakrabarty, 1993; Fay, 1984).

Instead of direct computation, various SAS procedures include options to use replicate weights when estimating standard errors or variances. To use Balanced Repeated Replication (BRR) method using replicate weights in SAS include the VARMETHOD=BRR(FAY=0.5) option in the PROC statement and specify the replicate weights with a REPWEIGHTS statement. Other computer packages have similar methods.

Formula (16) produces variance estimates close to zero for the median when multiple observations have value equal to the median. In this case, two methods can be used to estimate the variance of the median. The first technique incorporates replicate weights in

Woodruff's method for estimating variability (Woodruff, 1952). Gossett et al. (2002) documents the procedure for combining Woodruff's method with Jackknife replication and provides sample codes adapted by Mack and Tekansik (2011) for Fay's BRR. The second method uses VARMETHOD=TAYLOR option, a direct application of Woodruff's method, along with the cluster and strata statements instead of replicate weights to account for SIPP's complex design.

Illustration 7.

In SAS, the SURVEYMEANS procedure is used to estimate statistics such as means, totals, proportions, quantiles, and ratios for a survey sample. An example syntax for estimating the mean and median of the total household income, *THTOTINC*, for any month in 2016 using SIPP replicate weights is:

```
proc surveymeans data=pu2014w410 varmethod=brr(fay=0.5) mean median11;
  var THTOTINC;
  weight WPFINWGT;
  repweights REPWGT1-REPWGT240;
run;
```

Similarly, replicate weights can be used to estimate standard errors in the SURVEYFREQ (for frequency tables and cross-tabulations), SURVEYREG (for regression analysis), SURVEYLOGISTIC (for logistic regression analysis), and SURVEYPHREG (for proportional hazards regression analysis) SAS procedures by using the same VARMETHOD = BRR(FAY=0.5) option and REPWEIGHTS statement.

In Stata, the SVY command is used to fit a statistical model to a complex survey dataset. SVYSET is used to determine the survey design and provide information about the variance estimation. The following Stata syntax is equivalent to using SURVEYMEANS by SAS:

```
use pu2014w2.dta
svyset [pweight=wpfinwgt], brrweight(repwt1-repwt240) fay(.5) vce(brr) mse
svy: mean thtotinc
```

Illustration 8.

Illustration 7 showed how to compute cross-sectional estimates and their standard errors using final and replicate weights. The same method can also be applied to compute longitudinal estimates and their standard errors using the longitudinal weighting files published with Wave 2+ cross-sectional files.

For example, to estimate the proportion of persons who received Medicaid at any point between January 2013 and December 2016, first examine Wave 1 through Wave 4 person

¹⁰ This snippet of code requires the analytic, final weight, and replicate weights variables – *THTOTINC*, *WPFINWGT*, and *REPWGT1-REPWGT240* respectively – are included in the dataset *pu2014w4*

¹¹ The documentation for the Surveymeans procedure provides a list of “statistic keywords” that can be supplied to compute estimates of different statistics.

monthly datasets (*pu2014w1*, *pu2014w2*, *pu2014w3*, and *pu2014w4*) and determined which persons were enrolled in Medicaid during the period of interest, i.e., persons who have *EMDMTH*=1 for one or more months in the cross-sectional wave files.

Merge the resulting dataset with the longitudinal weight (*lgtwgt2014pnl4*) and longitudinal replicate weight (*lrw2014pnl4*) files to obtain final (*FINPNL4*) and replicate (*repwgt1-repwgt240*) panel weights for the reference period. The syntax in SAS would be:

```
proc surveymeans data=<final data> mean varmethod=brr(Fay=0.5) mean;
  var MEDICAID;
  weight FINPNL4;
  repweights REPWGT1-REPWGT240;
run;
```

where <final data> is the result of combining final panel weight and replicate weights with person level file Medicaid receipt variables obtained from Wave 1 through Wave 4 cross-sectional files variables including *MEDICAID*, an indicator variable equal to one if a person was enrolled in Medicaid for at least one month in the reference period and equal to zero otherwise.

REFERENCES

- Chakrabarty, R. P. (1993). Variance estimation by users of SIPP micro-data files [Working Paper]. Retrieved from <https://www.census.gov/content/dam/Census/library/working-papers/1993/adrm/rr93-6.pdf>
- Fay, R. E. (1984). Some properties of estimates of variance based on replication methods. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 495-500.
- Gossett, J. M., Simpson, P., Parker, J.G., & Simon, W.L. (2002). How complex can complex survey analysis be with SAS? *Proceedings of the 27th Annual SAS Users Group International Conference*. Retrieved from <https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p266-27.pdf>
- Kalton, G. (1998). *SIPP quality profile* [Working Paper]. U.S. Census Bureau. Retrieved from <https://www.census.gov/content/dam/Census/library/working-papers/1998/demo/SEHSD-WP1998-11.pdf>
- Mack, S. & Tekansik, S. (2011). "Calculating Standard Errors and Confidence Intervals of Medians." U.S. Census Bureau Memorandum for Documentation dated May 6, 2011.

- Sedgwick, P. (2014). Multiple hypothesis testing and Bonferroni's correction. *BMJ: British Medical Journal*, 349. Retrieved from <https://www.jstor.org/stable/26517649>
- Stoline, M. (1981). The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. *The American Statistician*, 35(3), 134-141. Retrieved from www.jstor.org/stable/2683979
- U.S. Census Bureau. (2019). Chapter 7: Nonsampling error, sampling error and weighting. *Survey of Income and Program Participation: 2014 Panel Users' Guide*. Washington, D.C. Retrieved from <https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/2014-SIPP-Panel-Users-Guide.pdf>
- Treat, J.B. (2017). "SIPP 2014: Weighting Specifications for Wave 1 (WGT-29)." U.S. Census Bureau Memorandum from James B. Treat to Barry F. Sessamen dated June 27, 2017.
- Tersine Jr., A.G. (2018). "SIPP 2014: Cross-Sectional Weighting Specifications for the Second and Subsequent Waves (WGT-DRAFT-12182018)." U.S. Census Bureau Memorandum from Anthony G. Tersine, Jr. to Kimberly D. Wortman, December 2018. Not yet finalized
- Tersine Jr, A.G. (2019). "Survey of Income and Program Participation 2014: Longitudinal Weighting Specifications for the Second and Subsequent Waves (WGT-30)." U.S. Census Bureau Memorandum from Anthony G. Tersine, Jr. to Nathaniel McKee, December, 2019. Not yet finalized.
- Wolter, K. M. (2007). Chapter 7: Generalized variance functions, *Introduction to Variance Estimation*, 2nd Ed. New York: Springer, pp. 272-297.
- Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, 47(260), 635-646.

Tables 1a-1d: Cross-Sectional Generalized Variance Parameters

Table 1a. Generalized Variance Parameters for Wave 1

Domain	Parameters		Design Effect ¹²	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation,				
Persons 15+				
Total	-0.00002111	5,295	2.052	0.9981
Male	-0.00004368	5,295		
Female	-0.00004085	5295		
Income and Labor Force Participation,				
Persons 15+				
Total	-0.00002361	5,925	2.297	1.056
Male	-0.00004886	5,925		
Female	-0.00004570	5,925		
Other, Persons 0+				
Total (or White)	-0.00001704	5,315	2.060	1
Male	-0.00003487	5,315		
Female	-0.00003332	5,315		
Black, Persons 0+				
Male	-0.00012010	4,857	1.883	0.9559
Female	-0.00025750	4,857		
	-0.00022520	4,857		
Hispanic, Persons 0+				
Male	-0.00010120	5,455	2.114	1.013
Female	-0.00020090	5,455		
	-0.0002040	5,455		
Households				
Total (or White)	-0.00003751	4,723	1.831	1
Black	-0.00028290	4,723		
Hispanic	-0.00028980	4,723		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 1a

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

¹² Design Effect=b/sample interval where sample interval=2,580

Table 1b. Generalized Variance Parameters for Wave 2

Domain	Parameters		Design Effect ¹³	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation, Persons 15+				
Total	-0.00002710	6,861	2.659	1.136
Male	-0.00005604	6,861		
Female	-0.00005247	6,861		
Income and Labor Force Participation, Persons 15+				
Total	-0.00003112	7,880	3.054	1.218
Male	-0.00006436	7,880		
Female	-0.00006026	7,880		
Other, Persons 0+				
Total (or White)	-0.00002206	6,931	2.686	1.142
Male	-0.00004513	6,931		
Female	-0.00004316	6,931		
Black, Persons 0+	-0.00015850	6,474	2.509	1.104
Male	-0.00033950	6,474		
Female	-0.00029740	6,474		
Hispanic, Persons 0+	-0.00013270	7,296	2.828	1.172
Male	-0.00026330	7,296		
Female	-0.00026740	7,296		
Households				
Total (or White)	-0.00004940	6,289	2.438	1.154
Black	-0.00037260	6,289		
Hispanic	-0.00038520	6,289		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 1b

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

¹³ Design Effect=b/sample interval where sample interval=2,580

Table 1c. Generalized Variance Parameters for Wave 3

Domain	Parameters		Design Effect ¹⁴	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation,				
Persons 15+				
Total	-0.00003492	8,945	3.467	1.297
Male	-0.00007224	8,945		
Female	-0.00006760	8,945		
Income and Labor Force Participation,				
Persons 15+				
Total	-0.00003905	10,000	3.876	1.372
Male	-0.00008079	10,000		
Female	-0.00007559	10,000		
Other, Persons 0+				
Total (or White)	-0.00002813	8,921	3.458	1.296
Male	-0.00005758	8,921		
Female	-0.00005501	8,921		
Black, Persons 0+				
Male	-0.00019970	8,272	3.206	1.248
Female	-0.00042740	8,272		
	-0.00037490	8,272		
Hispanic, Persons 0+				
Male	-0.00016450	9,278	3.596	1.321
Female	-0.00032810	9,278		
	-0.00033010	9,278		
Households				
Total (or White)	-0.00006117	7,892	3.059	1.293
Black	-0.00046090	7,892		
Hispanic	-0.00046980	7,892		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 1c

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

¹⁴ Design Effect=b/sample interval where sample interval=2,580

Table 1d. Generalized Variance Parameters for Wave 4

Domain	Parameters		Design Effect ¹⁵	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation,				
Persons 15+				
Total	-0.00004111	10,640	4.124	1.415
Male	-0.00008500	10,640		
Female	-0.00007962	10640		
Income and Labor Force Participation,				
Persons 15+				
Total	-0.00004546	11,770	4.562	1.488
Male	-0.00009399	11,770		
Female	-0.00008803	11,770		
Other, Persons 0+				
Total (or White)	-0.00003297	10,550	4.089	1.409
Male	-0.00006744	10,550		
Female	-0.00006448	10,550		
Black, Persons 0+	-0.00023280	9,758	3.782	1.355
Male	-0.00049760	9,758		
Female	-0.00043730	9,758		
Hispanic, Persons 0+	-0.00020310	11,710	4.539	1.484
Male	-0.00040520	11,710		
Female	-0.0004070	11,710		
Households				
Total (or White)	-0.00007149	9,330	3.616	1.406
Black	-0.00054650	9,330		
Hispanic	-0.00052960	9,330		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 1d

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

¹⁵ Design Effect=b/sample interval where sample interval=2,580

Tables 2-5: Simplified Base Standard Errors for Estimated Numbers and Percentages of Households and Persons

Table 2. Base Standard Errors of Estimated Numbers of Households or Families

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200,000	30,710	30,000,000	328,500
300,000	37,600	40,000,000	359,000
500,000	48,500	50,000,000	377,300
750,000	59,340	60,000,000	385,200
1,000,000	68,450	70,000,000	383,200
2,000,000	96,420	80,000,000	371,200
3,000,000	117,600	90,000,000	348,200
5,000,000	150,600	95,000,000	331,900
7,500,000	182,500	99,500,000	314,000
10,000,000	208,500	105,000,000	287,000
15,000,000	249,800	110,000,000	256,200
25,000,000	307,600	117,610,000	191,400

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

- Notes: (1). These estimates are calculations using the Household Total (or White) a and b parameters from Table 1a and Formula (4).
- (2). To estimate household standard errors, multiply the standard error from this table by appropriate f factor from Tables 1 or 7. For example to estimate standard errors for Wave 2 cross-sectional household estimates, multiply the appropriate standard error (based on the size of the estimate) by $f = 1.154$

Table 3. Base Standard Errors of Estimated Numbers of Persons

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200,000	32,590	110,000,000	615,200
300,000	39,910	120,000,000	626,400
500,000	51,510	130,000,000	634,800
750,000	63,060	140,000,000	640,400
1,000,000	72,790	150,000,000	643,300
2,000,000	102,800	160,000,000	643,600
3,000,000	125,700	170,000,000	641,200
5,000,000	161,700	180,000,000	636,100
7,500,000	197,200	190,000,000	628,300
10,000,000	226,800	200,000,000	617,600
15,000,000	275,500	210,000,000	603,900
25,000,000	349,600	220,000,000	587,000
30,000,000	379,600	230,000,000	566,600
40,000,000	430,500	240,000,000	542,300
50,000,000	472,400	250,000,000	513,600
60,000,000	507,500	260,000,000	479,600
70,000,000	537,200	270,000,000	439,100
80,000,000	562,300	275,000,000	415,900
90,000,000	583,400	280,000,000	390,200
100,000,000	600,900	299,340,000	253,200

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

- Notes: (1) These estimates are calculations using the Other Persons 0+ a and b parameters from Table 1a and Formula (4).
- (2) To calculate the standard error for another domain and/or reference period, multiply the standard error from this table by the appropriate f factor from Tables 1 or 7. For example, to calculate standard error for Wave 4 cross-sectional estimates related to labor force characteristics, multiply the appropriate standard error (based on the size of the estimate) by $f = 1.488$

Table 4. Base Standard Errors for Percentages of Households or Families

Base of Estimated Percentages	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200,000	1.529	2.151	3.349	4.610	6.654	7.684
300,000	1.248	1.757	2.735	3.764	5.433	6.274
500,000	0.967	1.361	2.118	2.916	4.208	4.860
750,000	0.7896	1.111	1.730	2.381	3.436	3.968
1,000,000	0.6838	0.9621	1.498	2.062	2.976	3.436
2,000,000	0.4835	0.6803	1.059	1.458	2.104	2.430
3,000,000	0.3948	0.5555	0.8648	1.190	1.718	1.984
5,000,000	0.3058	0.4303	0.6698	0.922	1.331	1.537
7,500,000	0.2497	0.3513	0.5469	0.7528	1.087	1.255
10,000,000	0.2162	0.3043	0.4736	0.652	0.9410	1.087
15,000,000	0.1766	0.2484	0.3867	0.5323	0.7684	0.8872
25,000,000	0.1368	0.1924	0.2996	0.4123	0.5952	0.6872
30,000,000	0.1248	0.1757	0.2735	0.3764	0.5433	0.6274
40,000,000	0.1081	0.1521	0.2368	0.3206	0.4705	0.5433
50,000,000	0.0967	0.1361	0.2118	0.2916	0.4208	0.4860
60,000,000	0.08828	0.1242	0.1934	0.2662	0.3842	0.4436
70,000,000	0.08173	0.1150	0.1790	0.2464	0.3557	0.4107
80,000,000	0.07645	0.1076	0.1675	0.2305	0.3327	0.3842
90,000,000	0.07208	0.1014	0.1579	0.2173	0.3137	0.3622
105,000,000	0.06673	0.09389	0.1462	0.2012	0.2904	0.3353
110,000,000	0.06520	0.09174	0.1428	0.1966	0.2837	0.3276
117,610,000	0.06305	0.08872	0.1381	0.1901	0.2744	0.3169

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

- Note: (1). These estimates are calculations using the Households Total (or White) b parameter from Table 1a and Formula (10).
- (2). To estimate Household standard errors, multiply the standard errors from this table by appropriate f factor from Tables 1 or 7. For example to estimate standard errors for Wave 4 longitudinal household estimates, multiply by $f = 1.575$

Table 5. Base Standard Errors for Percentages of Persons

Base of Estimated Percentages	Estimated Percentages					
	≤ 1 or ≥ 99	2 or 98	5 or 95	10 or 90	25 or 75	50
200,000	1.622	2.282	3.553	4.891	7.059	8.151
300,000	1.324	1.863	2.901	3.993	5.764	6.655
500,000	1.026	1.443	2.247	3.093	4.464	5.155
750,000	0.8376	1.179	1.835	2.525	3.645	4.209
1,000,000	0.7254	1.021	1.589	2.187	3.157	3.645
2,000,000	0.5129	0.7217	1.124	1.547	2.232	2.578
3,000,000	0.4188	0.5893	0.9174	1.263	1.823	2.105
5,000,000	0.3244	0.4565	0.7106	0.9781	1.412	1.630
7,500,000	0.2649	0.3727	0.5802	0.7986	1.153	1.331
10,000,000	0.2294	0.3228	0.5025	0.6916	0.9983	1.153
15,000,000	0.1873	0.2635	0.4103	0.5647	0.8151	0.9412
25,000,000	0.1451	0.2041	0.3178	0.4374	0.6314	0.7290
30,000,000	0.1324	0.1863	0.2901	0.3993	0.5764	0.6655
40,000,000	0.1147	0.1614	0.2512	0.3458	0.4991	0.5764
50,000,000	0.1026	0.1443	0.2247	0.3093	0.4464	0.5155
60,000,000	0.09365	0.1318	0.2051	0.2824	0.4075	0.4706
70,000,000	0.08670	0.1220	0.1899	0.2614	0.3773	0.4357
100,000,000	0.07254	0.1021	0.1589	0.2187	0.3157	0.3645
110,000,000	0.06916	0.09732	0.1515	0.2085	0.3010	0.3476
120,000,000	0.06622	0.09317	0.1450	0.1997	0.2882	0.3328
130,000,000	0.06362	0.08952	0.1394	0.1918	0.2769	0.3197
140,000,000	0.06131	0.08626	0.1343	0.1848	0.2668	0.3081
150,000,000	0.05923	0.08334	0.1297	0.1786	0.2578	0.2976
160,000,000	0.05735	0.08069	0.1256	0.1729	0.2496	0.2882
170,000,000	0.05563	0.07828	0.1219	0.1677	0.2421	0.2796
180,000,000	0.05407	0.07608	0.1184	0.1630	0.2353	0.2717
190,000,000	0.05263	0.07405	0.1153	0.1587	0.2290	0.2645
200,000,000	0.05129	0.07217	0.1124	0.1547	0.2232	0.2578
210,000,000	0.05006	0.07043	0.1096	0.1509	0.2178	0.2515
220,000,000	0.04891	0.06881	0.1071	0.1475	0.2128	0.2458
230,000,000	0.04783	0.0673	0.1048	0.1442	0.2082	0.2404
240,000,000	0.04682	0.06588	0.1026	0.1412	0.2038	0.2353
250,000,000	0.04588	0.06455	0.1005	0.1383	0.1997	0.2305
280,000,000	0.04335	0.0610	0.09496	0.1307	0.1887	0.2178
299,340,000	0.04193	0.05899	0.09184	0.1264	0.1825	0.2107

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

- Notes: (1) These estimates are calculations using the Other Persons 0+ a and b parameter from Table 1a and Formula (10).
- (2) To calculate the standard for another domain and/or reference period multiply the standard error from this table by the appropriate f factor from Tables 1 or 7.

Table 6. Hypothetical Distribution of Annual Cash Income Among People 25 to 34 Years Old
(Not Actual Data, Only Use for Calculation Illustrations)

	Interval of Annual Cash Income												
	under \$5000	\$5000 to \$7499	\$7500 to \$9999	\$10000 to \$12,499	\$12,500 to \$14,999	\$15,000 to \$17,499	\$17,500 to \$19,999	\$20,000 to \$29,999	\$30,000 to \$39,999	\$40,000 to \$49,999	\$50,000 to \$59,999	\$60,000 to \$69,999	\$70,000 and over
Number of People in Each Interval (in thousands)	370	302	447	685	935	1,113	1,298	5,496	4,596	3,121	1,902	1,124	2,138
Cumulative Number of People with at Least as Much as Lower Bound of Each Interval (in thousands)	23,527 (Total People)	23,158	22,856	22,409	21,724	20,789	19,675	18,377	12,881	8,285	5,164	3,262	2,138
Percent of People with at Least as Much as Lower Bound of Each Interval	100.0	98.4	97.1	95.2	92.3	88.4	83.6	78.1	54.7	35.2	21.9	13.9	9.1

Table 7a-7c: Longitudinal Generalized Variance ParametersTable 7a. Generalized Variance Parameters for *FINPNL2*¹⁶

Domain	Parameters		Design Effect ¹⁷	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation,				
Persons 15+				
Total	-0.00002877	7,217	2.797	1.165
Male	-0.00005953	7217		
Female	-0.00005567	7217		
Income and Labor Force Participation,				
Persons 15+				
Total	-0.00003257	8,170	3.167	1.240
Male	-0.00006739	8,170		
Female	-0.00006302	8,170		
Other, Persons 0+				
Total (or White)	-0.00002312	7,211	2.795	1.165
Male	-0.00004730	7,211		
Female	-0.00004521	7,211		
Black, Persons 0+				
Male	-0.00016460	6,655	2.579	1.119
Female	-0.00035280	6,655		
Female	-0.00030850	6,655		
Hispanic, Persons 0+				
Male	-0.00014330	7,721	2.993	1.205
Male	-0.00028440	7,721		
Female	-0.00028870	7,721		
Households				
Total (or White)	-0.00005101	6,477	2.510	1.171
Black	-0.00038490	6,477		
Hispanic	-0.00039740	6,477		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 7a

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

Table 7b. Generalized Variance Parameters for *FINPNL3*¹⁶ *FINPNL1* parameters are same as those in Table 1a since it covers the same reference period as *CY2013*¹⁷ Design Effect=b/sample interval where sample interval=2,580

Domain	Parameters		Design Effect ¹⁸	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation,				
Persons 15+				
Total	-0.00004109	10,310	3.996	1.393
Male	-0.00008503	10,310		
Female	-0.00007952	10,310		
Income and Labor Force Participation,				
Persons 15+				
Total	-0.00004619	11,590	4.492	1.477
Male	-0.00009557	11,590		
Female	-0.00008938	11,590		
Other, Persons 0+				
Total (or White)	-0.00003261	10,170	3.942	1.383
Male	-0.00006674	10,170		
Female	-0.00006378	10,170		
Black, Persons 0+	-0.00024810	10,030	3.888	1.374
Male	-0.00053190	10,030		
Female	-0.00046510	10,030		
Hispanic, Persons 0+	-0.00019580	10,550	4.089	1.409
Male	-0.00038860	10,550		
Female	-0.00039450	10,550		
Households				
Total (or White)	-0.00006899	8,847	3.429	1.369
Black	-0.00052400	8,847		
Hispanic	-0.00054410	8,847		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 7b

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

Table 7c. Generalized Variance Parameters for *FINPNL4*

¹⁸ Design Effect=b/sample interval where sample interval=2,580

Domain	Parameters		Design Effect ¹⁹	<i>f</i>
	<i>a</i>	<i>b</i>		
Poverty and Program Participation,				
Persons 15+				
Total	-0.00005715	14,340	5.558	1.643
Male	-0.00011820	14,340		
Female	-0.00011060	14,340		
Income and Labor Force Participation,				
Persons 15+				
Total	-0.00006221	15,610	6.050	1.714
Male	-0.00012870	15,610		
Female	-0.00012040	15,610		
Other, Persons 0+				
Total (or White)	-0.00004470	13,940	5.403	1.619
Male	-0.00009146	13,940		
Female	-0.00008741	13,940		
Black, Persons 0+	-0.00034700	14,030	5.438	1.625
Male	-0.00074390	14,030		
Female	-0.00065050	14,030		
Hispanic, Persons 0+	-0.00027560	14,850	5.756	1.672
Male	-0.00054720	14,850		
Female	-0.00055540	14,850		
Households				
Total (or White)	-0.00009114	11,720	4.543	1.575
Black	-0.00070030	11,720		
Hispanic	-0.00071910	11,720		

Source: U.S. Census Bureau, 2014 Survey of Income and Program Participation

Notes on Domain Usage for Table 7c

Poverty and Program Participation	Use these parameters for estimates concerning poverty rates, welfare program participation (e.g. Supplemental Security Income, SSI), and other programs for adults with low incomes.
Income and Labor Force	These parameters are for estimates concerning income, sources of income, labor force participation, economic well-being other than poverty, employment related estimates (e.g. occupation, hours worked a week), and other income, job, or employment related estimates.
Other Persons	Use the "Other Persons" parameters for estimates of total (or white) persons aged 0+ in the labor force, and all other characteristics not specified in this table, for the total or white population.
Black/Hispanic Persons	Use these parameters for estimates of Black and Hispanic persons 0+.
Households	Use these parameters for all household level estimates.

¹⁹ Design Effect=b/sample interval where sample interval=2,580