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

February 7, 2019

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 2 Public Use Files (21)[1]

This memorandum documents the Source and Accuracy Statement for 2014 Panel Wave 2 Public Use Files of the Survey of Income and Program Participation.

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

Matthew Marlay	(ADDP)	Jonathan Rothbaum	
Holly Fee		Edward Welniak	
Shelley Irving		Gary Benedetto	
Nathaniel McKee	(DSD)	Sharon Stern	
Connie Bauer		James Farber	(DSMD)
David Waddington	(SEHSD)	Tracy Mattingly	
Trudi Renwick		Mahdi Sundukchi	
Stephanie Galvin		Ashley Westra	
Alfred Gottschalck		Ralph Culver III	
Rebecca Chenevert		Julia Yang	
Brian McKenzie		Faith Nwaoha Brown	
Kurt Bauman			
Ashley Edwards			
Rose Kreider			
Jason Fields			

[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*

SOURCE AND ACCURACY STATEMENT FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP) 2014 PANEL FOR WAVE 2 PUBLIC USE FILES[2]

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 4 million institutionalized people in Census 2010).

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 Census Bureau's official inventory of known housing units. The frame 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 statistics of low income households and program participation. Analysts are strongly encouraged to use the SIPP weights when creating 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 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 1. In general, one cycle of interviews covering the entire sample, using the same questionnaire, is called a wave. Interviews for each wave are conducted from February through May each year.

For Wave 1 of the SIPP 2014 panel, interviews were conducted February through May of 2014, collecting data on January through December 2013. Similarly, Wave 2 respondents were interviewed from February through May of 2015 and provided data on the 2014 calendar year, January through December 2014. For each wave, data for up to 12 reference months are

---

[2] For questions or further assistance with the information provided in this document contact: Tracy Mattingly of the Demographic Statistical Methods Division on 301-763-6445 or via email at Tracy.L.Mattingly@census.gov.

available for persons on the published file. Specific months available depend on a person's sample entry or exit date.

In Wave 1, the SIPP 2014 Panel began with a sample of 53,070 HUs. Of these HUs, 10,722 were found to be vacant, demolished, converted to nonresidential use, or otherwise ineligible for the survey. Field Representatives (FRs) were able to obtain interviews for 29,685 of the eligible HUs. FRs were unable to interview 12,663 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 (those in Wave 1 sample households and interviewed in Wave 1) and people living with them are eligible to be interviewed. In Wave 2, 30,109 HUs were eligible for interview and interviews were obtained from 23,036 of these households. The SIPP sample includes original sample people if they moved to a new address, unless the new address was more than 100 miles from a SIPP sample area. In this case, FRs attempt telephone interviews.

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 non-interviewed sample. Therefore, a growth factor based on the growth in the known sample is used to estimate the unknown expansion of the non-interviewed households.

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

$$Sample\ Loss = \frac{(A_1 \times GF) + A_c + D_c}{I_c + (A_1 \times GF) + A_c + D_c} \tag{1}$$

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

Based on the above equation, the weighted sample loss at each wave of the SIPP 2014 Panel was calculated and tabulated as shown in Table A below.

Table A. Sample Loss and Response Rates

Wave	Eligible HUs	Interviewed HUs	Type As		Type Ds		Growth Factor	Weighted Sample Loss
			Total	Weighted Rate	Total	Weighted Rate		
1	42,348	29,685	12,663	31.2%				31.2%
2	30,109	23,036	6,385	21.8%	688	2.1%	1.01	47.7%

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

Table B. Percent of Type As by Nonresponse Status

Wave	Language Problem	Unable to Locate	No One Home	Temporarily Absent	Household Refused	Other
1	0.8%	0.4%	11.0%	1.3%	79.2%	7.4%
2	0.4%		6.9%	0.9%	79.8%	12.1%

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, calendar year weights (CY2013, CY2014, CY2015, CY2016), and panel weights (PNL1, PNL2, PNL3, PNL4).

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 within the year. Calendar year weights for Wave 2 of the 2014 SIPP can be used to compute monthly, quarterly, and annual estimates for the time between January and December 2014. 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 weight is based on the SIPP survey universe in December of a designated year and assigns 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.

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. Therefore, PNL2 weights can be used to compute estimates for any time frame between January 2013 and December 2014. Table C specifies the reference period for calendar and panel weights.

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 non-interviewed. PNL2 weighting procedure classified 52,486 people as interviewed and had a weighted person response rate of 72.1%[3].

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 month[4] of December 2013, and hence are the same as the December 2013 monthly weights. As a result, separate longitudinal weight files are not provided for Wave 1. This implies that a person must be in the SIPP sample universe and interviewed in December 2013 to receive a positive calendar year (CY2013) and panel (PNL1) weights in Wave 1.

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 non-interviewed 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 those classified as non-interviewed or ineligible are assigned zero weights. Longitudinal weights are produced at the completion of each wave.

Table C. Reference Periods for Calendar Year and Panel Weights

Variable Name	Control Month	Beginning Wave	Beginning Month	Ending Wave	Ending Month
			Calendar year weights		
CY2013	December 2013	Wave 1	January 2013	Wave 1	December 2013
CY2014	December 2014	Wave 2	January 2014	Wave 2	December 2014
CY2015	December 2015	Wave 3	January 2015	Wave 3	December 2015
CY2016	December 2016	Wave 4	January 2016	Wave 4	December 2016
			Panel weights		
PNL1	December 2013	Wave 1	January 2013	Wave 1	December 2013
PNL2	December 2013	Wave 1	January 2013	Wave 2	December 2014

[3] 72,678 persons were eligible for PNL2 weights
[4] A control month is defined as a month during with the sample universe is under consideration. The SIPP weighing procedure adjusts both cross sectional and longitudinal weights to population estimates for a specific control month.

Variable Name	Control Month	Beginning Wave	Beginning Month	Ending Wave	Ending Month
PNL3	December 2013	Wave 1	January 2013	Wave 3	December 2015
PNL4	December 2013	Wave 1	January 2013	Wave 4	December 2016

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

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 the probability of selection of a their household. 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 non-interview adjustment factor is applied to account for households that were eligible for the sample but which FRs could not interview in Wave 1(F_{N1})**.** Similarly for subsequent waves *i*, the non-interview adjustment factor is (F_{Ni}). 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 is $FW_c = BW * WCF * FN_1 * F_{2S}$ for Wave 1 and is $FW_c = BW * WCF * FN_1 * MAF * FN_2 * F_{2S}$ 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 non-interview adjusted weight for the control month, December 2013. A non-interview 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 non-interview weights to independent population controls for December 2013. Spouse weights are not equalized in the panel weighting procedure.

The final panel weight for Waves 2+[5] is $PW_c = IW * FP_{ni} * FP_{2S}$. 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 results 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

---

[5] The final panel weight for Wave 1 is same as Wave 1 calendar year weight, $FW_c = BW * WCF * FN_1 * F_{2S}$

from the Population Division and are prepared each month to agree with the most current set of population estimates released by the U.S. 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. Likewise, person interviewed in December have calendar year weights, and persons categorized as interviewed for a longitudinal reference period are assigned panel weights. Chapter 7 of the 2014 *SIPP User's Guide* provides additional information on how to use the weights.

In historic 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 person for that reference month[6]. 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, the person weights can be used to specify a single household or family weight. One option is to take 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 take 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 (e.g., the panel weights for Wave 2 can be used for constructing estimates at any time spans in the period between January 2013 and December 2014). The calendar year weight can be used to form monthly, quarterly, or annual estimates within a specific calendar year. However, note that wave files with cross-sectional weights are also produced for the SIPP. 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 reference month cross-sectional weight for the month of interest. Similarly, use calendar year weights to determine estimates for any time frame within a wave.

Users should be forewarned to apply the appropriate weights given on weighting files before attempting to calculate estimates. The weights vary with demographic and time units of analysis (person, family, and household, monthly in 2013, quarterly in 2013, annually between 2013 to 2014, etc.) due to differences in control months, longitudinal reference periods, interview-refusal and unlocated-mover nonresponses, sample reduction effects if there is a sample reduction, etc. that are factored in the weighting adjustments. If an analysis/estimate is done for a cohort of people or families or households (in the survey universe) without applying the appropriate weights, the results will be erroneous.

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

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.

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 six months 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. The same estimate could be generated using the panel weights, but there may be fewer positively weighted people than those in the calendar year. To estimate the number of persons who received any unemployment income in 2013 and 2014, use panel weights, i.e. *PNL2*.

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 2014 annual income of people who were employed during all 12 months of the year, use Wave 2 calendar year weights. The same estimate could be generated using the panel 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 2014. Also, one could 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 or Aggregate."

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 weight 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 2013 to not receiving in August 2013, add together the Wave 1 calendar year weights of each person who had such a change. To estimate the number of changes in monthly salary income during the 2014 fiscal year (September 2013 to September 2014), use PNL2 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 December 2013 through December 2014. To construct such an estimate, first form an estimate for each month in the time period of interest. Use the panel weight, *PNL2*, summing over all people who possessed the characteristic of interest during the twelve months of interest. Then sum the monthly estimates and divide by 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 if a complete census had been taken 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. 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. Table D and E below show SIPP 2014 coverage ratios for age-sex-race groups in December 2013 and December 2014 respectively using calendar year cross-sectional weights prior to the ratio adjustment. 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.

Comparability with Other Estimates. Caution should be exercised when comparing this data with data from other SIPP products or with data from other surveys. The comparability problems are caused by such sources 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 would include the average result of all possible samples.

2. Approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

3. Approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average result of all possible samples.

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

Table D. 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	0.88	0.90	0.80	0.79	0.88	0.82
15	0.86	0.82	0.88	0.87	0.88	0.87
16-17	0.84	0.85	0.84	0.86	0.89	0.89
18-19	0.85	0.79	0.84	0.87	0.89	0.90
20-21	0.86	0.85	0.71	0.65	0.85	0.89
22-24	0.82	0.75	0.71	0.66	0.86	0.89
25-29	0.81	0.81	0.64	0.73	0.86	0.75
30-34	0.83	0.87	0.70	0.72	0.84	0.75
35-39	0.87	0.90	0.77	0.84	0.80	0.78
40-44	0.85	0.86	0.75	0.84	0.79	0.78
45-49	0.82	0.89	0.78	0.80	0.85	0.92
50-54	0.87	0.90	0.81	0.82	0.87	0.91
55-59	0.93	0.96	0.75	0.85	0.91	0.94
60-61	0.98	1.03	0.75	0.82	0.97	0.93
62-64	1.00	0.96	0.74	0.86	0.92	0.93
65-69	1.00	1.02	0.99	0.90	0.87	0.94
70-74	1.01	0.99	0.99	0.92	0.87	0.93
75-79	1.03	1.02	0.99	0.91	0.87	0.93
80-84	1.02	1.06	1.00	0.91	0.87	0.90
85+	0.94	0.87	1.06	0.90	0.90	0.95

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

Table E. 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	0.83	0.84	0.72	0.70	0.86	0.80
15	0.78	0.82	0.76	0.82	0.76	0.80
16-17	0.80	0.82	0.77	0.82	0.78	0.78
18-19	0.79	0.70	0.76	0.85	0.79	0.79
20-21	0.70	0.75	0.64	0.55	0.76	0.77
22-24	0.69	0.63	0.65	0.55	0.76	0.79
25-29	0.74	0.73	0.53	0.58	0.78	0.74
30-34	0.74	0.79	0.63	0.65	0.79	0.74
35-39	0.80	0.87	0.66	0.77	0.78	0.75
40-44	0.79	0.82	0.72	0.76	0.77	0.75
45-49	0.79	0.83	0.81	0.73	0.77	0.84
50-54	0.81	0.85	0.74	0.82	0.78	0.85
55-59	0.89	0.93	0.73	0.85	0.89	0.98
60-61	0.98	1.02	0.75	0.85	0.93	1.02
62-64	0.98	0.97	0.74	0.84	0.92	0.99
65-69	1.00	1.02	1.07	0.93	0.99	0.94
70-74	1.00	1.01	1.10	0.98	0.93	0.97
75-79	1.08	1.04	1.07	1.00	0.92	0.97
80-84	1.18	1.11	1.07	0.98	0.96	0.96
85+	1.00	0.94	1.09	0.99	0.98	0.91

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

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.

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}. 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 level. In this event, it is commonly accepted practice to say that the characteristics are different. We recommend that users report only those differences that are significant at the 10 percent 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 done to account for this potential problem that consists of 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 weighted base smaller than 150,000. Also, nonsampling error in one or more of the small number of cases providing the estimation can 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;
- Generalized variance function parameters (denoted as a and b); and
- Simplified tables of SEs based on the a and b parameters.

While the replicate weight methods provide the most accurate variance estimates, this approach requires more computing resources and more 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 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 cluster 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 $\boldsymbol{a}$ and $\boldsymbol{b}$) 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 Wave 2 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 taken directly from Tables 1 and 7. The choice of parameter depends on the weights used, on the characteristic of interest, and on 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 and 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 and 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:

$$g = \frac{\sum_{i=1}^{n} m_i^2}{\sum_{i=1}^{n} m_i} \tag{2}$$

   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 and 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 those 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 and 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):

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

where f is the appropriate f factor from Tables 1 and 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):

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

Here x is the size of the estimate and a and b are the appropriate parameters from Tables 1 and 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).

<u>Illustration 1.</u>

Suppose SIPP estimates based on Wave 1 of the 2014 panel show that there were 2,000,000 females aged 25 to 44 with a monthly income of greater than $6,000 in September 2013. The appropriate parameters and factor from Table 1a and the appropriate general standard error from Table 3 are:

$$a = -0.00004570 \qquad b = 5{,}925 \qquad f = 1.057 \qquad s = 102{,}771$$

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

$$s_x = 1.057 \times 102{,}771 = 108{,}629$$

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

$$s_x = \sqrt{(-0.00004570 \times 2{,}000{,}000^2) + (5{,}925 * 2{,}000{,}000)} = 108{,}015\ females.$$

Using the standard error based on Formula (4), the approximate 90 percent confidence interval as shown by the data is from 1,822,315 to 2,177,685 females ($i.e.$, $2{,}000{,}000 \pm 1.645 \times 108{,}015$). Therefore, a conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90% of all samples.

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 of females aged 25 to 34. 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:

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

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:

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

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

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:

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

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

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

<u>Illustration 2.</u>
<u>Method 1</u>

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 \$38,703.4 Applying Formula (6), the approximate population variance, s^2, is:

$$s^2 = \left(\frac{370}{23{,}527}\right)(2{,}500)^2 + \cdots + \left(\frac{2{,}138}{23{,}527}\right)(105{,}000)^2 - (38{,}703.4)^2 = 649{,}411{,}468\,.$$

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

$$s_{\bar{x}} = \sqrt{\frac{7{,}880}{23{,}527{,}000} \times 649{,}411{,}468} = \$466.38$$

Thus, the approximate 90 percent confidence interval as shown by the data ranges from \$37,936.21 to \$39,470.60.

Method 2

Suppose that we are interested in estimating the average length of spells of food stamp recipiency during the calendar year 2014 for a given 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 their weights could be substantially different from those given). The number of consecutive months of food stamp recipiency during 2014 and the calendar year 2014 weights are given in the table below for each sample person:

Sample Person	Spell Length in Months	Calendar Year 2014 Weight (CY2014)
1	4, 3	5,300
2	5	7,100
3	9	4,900
4	3, 3, 2	6,500
5	12	9,200
6	12	5,900
7	4, 1	7,600
8	7	4,200
9	6	5,500
10	4	5,700

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

$$\bar{x} = \frac{(5300)(4) + (5300)(3) + \ldots + (5700)(4)}{5300 + 5300 + \ldots 5700} = 5.4$$

The standard error will be computed by Formula (6). First, the estimated population variance can be obtained by Formula (7):

$$s^2 = \frac{(5300)(4)^2 + (5300)(3)^2 + \ldots + (5700)(4)^2}{5300 + 5300 + \ldots + 5700} - (5.4)^2$$

$$= 12.4\ (months)^2$$

Next, the base ***b*** parameter of 6,861 is taken from Table 1b and multiplied by the factor computed from Formula (2):

$$g = \frac{2^2+1+1+3^2+1+1+2^2+1+1+1}{2+1+1+3+1+1+2+1+1+1}$$
$$= 1.71$$

Therefore, the final ***b*** parameter is 1.71 $\times$ 6,861 = 11,732, and the standard error of the mean from Formula (5) is:

$$s_{\bar{x}} = \sqrt{\frac{(11{,}732)(12.4)}{87{,}800}} = 1.29\ months$$

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:

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

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:

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

where f is the appropriate f factor from Tables 1 and 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:

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

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) above and should be used when data from less than four rotations are used to estimate p.

Illustration 3.

Suppose that using the second panel weight, PNL2, it was estimated that 59,355,000 males were employed in December 2013 and an estimated 2.4 percent of them became unemployed in January 2014. Using Formula (10), with a b parameter of 8,170 from Table 7, the approximate standard error is:

$$s_{(x,p)} = \sqrt{\frac{8{,}170}{59{,}355{,}000} \times 2.4 \times (100 - 2.4)} = 0.18\, percent$$

Consequently, the 90 percent confidence interval as shown by these data is from 2.10 percent to 2.70 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:

$$p_I = 100\left(\frac{x_A}{x_N}\right),$$

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

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

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

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

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 $\bar{x}_N$. To calculate s_p, use Formula (10). The standard errors of $\bar{x}_N$ and $\bar{x}_A$ may be calculated using Formula (5).

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

Illustration 4.

Suppose that in September 2013, 9.8 percent of the households own rental property, the mean value of rental property is $72,121, the mean value of assets is $78,734, and the corresponding standard errors are 0.18 percent, $5,468, and $2,703, respectively. In total there are 125,906,141 households. Then, the percent of all household assets held in rental property is:

$$100\left(0.098 \times \frac{72{,}121}{78{,}734}\right) = 9.0\, percent$$

Using Formula (11), the appropriate standard error is:

$$s_I = \sqrt{\left(\frac{0.098 \times 72{,}121}{78{,}734}\right)^2 \left[\left(\frac{0.0018}{0.098}\right)^2 + \left(\frac{5{,}468}{72{,}121}\right)^2 + \left(\frac{2{,}703}{78{,}734}\right)^2\right]} = 0.77\, percent.$$

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.
Suppose that for September 2014, SIPP estimates show the number of persons aged 35-44 years with annual cash income of $50,000 to $59,999 was 3,186,000 and the number of persons aged 25-34 years with annual cash income of $50,000 to $59,999 in the same time period was 2,619,000. Then, using the parameters $a = -0.00003112 \ and \ b = 7{,}880$ from Table 1b and Formula (4), the standard errors of these numbers are approximately 157,448 and 142,913 respectively. The difference in sample estimates is 567,000 and using Formula (12), the approximate standard error of the difference is:

$$\sqrt{157{,}448^{\,2} + 142{,}913^2} = 212{,}636\,.$$

Suppose that it is desired to test at the 10 percent significance level whether the number of persons with monthly cash income of $50,000 to $59,999 was different for people age 35-44 years than for people age 25-34 years. To perform the test, compare the difference of 567,000 to the product $1.645 \times 212{,}636 \; = 349{,}786$. Since the difference is greater than 1.645 times the standard error of the difference, the data show that the two age groups are significantly different at the 10 percent significance level.

Standard Error of a Median. The median quantity of some items such as income for a given group of people is that quantity such that at least half the group have as much or more and at least half the group have as much or less. 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 $p = 0.5$. 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 50 percent of the subpopulation falls at or below that value and 50 percent 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 Formulas (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 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:

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

if Pareto Interpolation is indicated and:

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

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

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 0.93 percentage points.

2. Following step 2, the two percentages of interest are 49.07 and 50.93.

3. By examining Table 6, we see that the percentage 49.07 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 49.07 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

$$\$30{,}000 \times \exp\left[\left(\frac{\ln\left(\frac{0.4907 \times 23{,}527{,}000}{12{,}881{,}000}\right)}{\ln\left(\frac{8{,}285{,}000}{12{,}881{,}000}\right)}\right) \times \ln\left(\frac{40{,}000}{30{,}000}\right)\right] = \$32{,}221.$$

Also by examining Table 6, we see that 50.93 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

$$\$30{,}000 \times \exp\left[\left(\frac{\ln\left(\frac{0.5093 \times 23{,}527{,}000}{12{,}881{,}000}\right)}{\ln\left(\frac{8{,}285{,}000}{12{,}881{,}000}\right)}\right) \times \ln\left(\frac{40{,}000}{30{,}000}\right)\right] = \$31{,}448.$$

Thus, the 68 percent confidence interval on the estimated median is from $31,448 to $32,221.

4. Then the approximate standard error of the median is

$$\frac{\$32{,}221 - \$31{,}448}{2} = \$386.74$$

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

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

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:

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

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 replicate weights in SAS include the VARMETHOD=BRR(FAY=0.5) option in the PROC statement and specify the replicate weights with a REPWEIGHTS. 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.

<u>Illustration 7.</u>

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 of the total household income (THTOTINC) using SIPP replicate weights is:

```
proc surveymeans data=pu2014w2 mean varmethod=brr(Fay=0.5) mean;
  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(repwgt1-repwgt240) fay(.5) vce(brr) mse
svy: mean thtotinc
```

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., and 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. and Tekansik, S. (May 6, 2011). Calculating Standard Errors and Confidence Intervals of Medians. Internal Memorandum for Documentation. U.S. Census Bureau.

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 (2016). Chapter 7: Nonsampling Error, Sampling Error and Weighting. *Survey of Income and Program Participation: 2014 Panel Users' Guide,* 1st Ed. Washington, D.C: U.S. Census Bureau. Retrieved from https://www.census.gov/content/dam/Census/programs-surveys/sipp/methodology/2014-SIPP-Panel-Users-Guide.pdf

U.S. Census Bureau (2017). SIPP 2014: Weighting Specifications for Wave 1 (WGT-29). Internal Memorandum from James B. Treat to Barry F. Sessamen, June 27, 2017.

U.S. Census Bureau (2018). SIPP 2014: Cross-Sectional Weighting Specifications for the Second and Subsequent Waves (WGT-DRAFT-12182018). Internal Memorandum from Anthony G. Tersine Jr. to Kimberly D. Wortman, December 18, 2018. Not finalized

U.S. Census Bureau (2019). Survey of Income and Program Participation 2014: Longitudinal Weighting Specifications for the Second and Subsequent Waves (WGT-30). Internal Memorandum from to Anthony G. Tersine Jr. to Nathaniel McKee, February 5, 2019. Not finalized.

Wolter, K. M. (2007). Chapter 7: Generalized Variance Functions, *Introduction to Variance Estimation,* 2nd Ed. New York: Springer, pp. 272-297.

Woodruff, Ralph S (1952). Confidence Intervals for Medians and Other Position Measures, *Journal of the American Statistical Association*, 47(260), 635-646.

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

Table 1a. Generalized Variance Parameters for Wave 1

Domain	Parameters		Design Effect[7]	f
	a	b		
Poverty and Program Participation, Persons 15+				
Total	-0.00002111	5,295	2.05	0.998
Male	-0.00004368	5,295		
Female	-0.00004085	5,295		
Income and Labor Force Participation, Persons 15+				
Total	-0.00002361	5,925	2.30	1.057
Male	-0.00004886	5,925		
Female	-0.00004570	5,925		
Other, Persons 0+				
Total (or White)	-0.00001704	5,315	2.06	1.000
Male	-0.00003487	5,315		
Female	-0.00003332	5,315		
Black, Persons 0+	-0.00012014	4,857	1.88	0.955
Male	-0.00025752	4,857		
Female	-0.00022520	4,857		
Hispanic, Persons 0+	-0.00010122	5,455	2.11	1.012
Male	-0.00020095	5,455		
Female	-0.00020397	5,455		
Households				
Total (or White)	-0.00003751	4,723	1.83	1.00
Black	-0.00028286	4,723		
Hispanic	-0.00028981	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.

[7] Design Effect=b/sample interval where sample interval=2,580

Table 1b. Generalized Variance Parameters for Wave 2

Domain	Parameters a	Parameters b	Design Effect[8]	f
Poverty and Program Participation, Persons 15+				
Total	-0.00002710	6,861	2.66	1.136
Male	-0.00005604	6,861		
Female	-0.00005247	6,861		
Income and Labor Force Participation, Persons 15+				
Total	-0.00003112	7,880	3.05	1.217
Male	-0.00006436	7,880		
Female	-0.00006026	7,880		
Other, Persons 0+				
Total (or White)	-0.00002206	6,931	2.69	1.143
Male	-0.00004513	6,931		
Female	-0.00004316	6,931		
Black, Persons 0+	-0.00015851	6,474	2.51	1.104
Male	-0.00033947	6,474		
Female	-0.00029736	6,474		
Hispanic, Persons 0+	-0.00013268	7,296	2.83	1.172
Male	-0.00026330	7,296		
Female	-0.00026743	7,296		
Households				
Total (or White)	-0.00004940	6,289	2.44	1.155
Black	-0.00037262	6,289		
Hispanic	-0.00038517	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.

[8] 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	35,438	30,000,000	379,750
300,000	43,385	40,000,000	415,355
500,000	55,966	50,000,000	436,978
750,000	68,476	60,000,000	446,654
1,000,000	78,991	70,000,000	445,163
2,000,000	111,267	80,000,000	432,389
3,000,000	135,729	90,000,000	407,271
5,000,000	173,810	95,000,000	389,384
7,500,000	210,686	99,500,000	369,707
10,000,000	240,728	105,000,000	340,162
15,000,000	288,479	110,000,000	306,676
25,000,000	355,457	117,610,000	237,367

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 factor from Tables 1 and 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=1.155

Table 3. Base Standard Errors of Estimated Numbers of Persons

Size of Estimate	Standard Error	Size of Estimate	Standard Error
200,000	32,593	110,000,000	615,196
300,000	39,912	120,000,000	626,438
500,000	51,510	130,000,000	634,802
750,000	63,061	140,000,000	640,403
1,000,000	72,787	150,000,000	643,312
2,000,000	102,771	160,000,000	643,565
3,000,000	125,665	170,000,000	641,166
5,000,000	161,707	180,000,000	636,085
7,500,000	197,241	190,000,000	628,256
10,000,000	226,817	200,000,000	617,576
15,000,000	275,483	210,000,000	603,892
25,000,000	349,607	220,000,000	586,996
30,000,000	379,623	230,000,000	566,599
40,000,000	430,507	240,000,000	542,306
50,000,000	472,388	250,000,000	513,566
60,000,000	507,500	260,000,000	479,579
70,000,000	537,172	270,000,000	439,129
80,000,000	562,267	275,000,000	415,903
90,000,000	583,375	280,000,000	390,210
100,000,000	600,916	299,340,000	253,244

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 and 7. For example, to calculate standard error for Wave 2 cross-sectional estimates related to labor force characteristics, multiply the appropriate standard error (based on the size of the estimate) by f=1.217

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	1.53%	2.15%	3.35%	4.61%	6.65%	7.68%
300,000	1.25%	1.76%	2.73%	3.76%	5.43%	6.27%
500,000	0.97%	1.36%	2.12%	2.92%	4.21%	4.86%
750,000	0.79%	1.11%	1.73%	2.38%	3.44%	3.97%
1,000,000	0.68%	0.96%	1.50%	2.06%	2.98%	3.44%
2,000,000	0.48%	0.68%	1.06%	1.46%	2.10%	2.43%
3,000,000	0.39%	0.56%	0.86%	1.19%	1.72%	1.98%
5,000,000	0.31%	0.43%	0.67%	0.92%	1.33%	1.54%
7,500,000	0.25%	0.35%	0.55%	0.75%	1.09%	1.25%
10,000,000	0.22%	0.30%	0.47%	0.65%	0.94%	1.09%
15,000,000	0.18%	0.25%	0.39%	0.53%	0.77%	0.89%
25,000,000	0.14%	0.19%	0.30%	0.41%	0.60%	0.69%
30,000,000	0.12%	0.18%	0.27%	0.38%	0.54%	0.63%
40,000,000	0.11%	0.15%	0.24%	0.33%	0.47%	0.54%
50,000,000	0.10%	0.14%	0.21%	0.29%	0.42%	0.49%
60,000,000	0.09%	0.12%	0.19%	0.27%	0.38%	0.44%
70,000,000	0.08%	0.11%	0.18%	0.25%	0.36%	0.41%
80,000,000	0.08%	0.11%	0.17%	0.23%	0.33%	0.38%
90,000,000	0.07%	0.10%	0.16%	0.22%	0.31%	0.36%
105,000,000	0.07%	0.09%	0.15%	0.20%	0.29%	0.34%
110,000,000	0.07%	0.09%	0.14%	0.20%	0.28%	0.33%
117,610,000	0.06%	0.09%	0.14%	0.19%	0.27%	0.32%

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 factor from Tables 1 and 7. For example to estimate standard errors for Wave 2 longitudinal household estimates, multiply by f=1.171

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	1.62%	2.28%	3.55%	4.89%	7.06%	8.15%
300,000	1.32%	1.86%	2.90%	3.99%	5.76%	6.66%
500,000	1.03%	1.44%	2.25%	3.09%	4.46%	5.16%
750,000	0.84%	1.18%	1.83%	2.53%	3.65%	4.21%
1,000,000	0.73%	1.02%	1.59%	2.19%	3.16%	3.65%
2,000,000	0.51%	0.72%	1.12%	1.55%	2.23%	2.58%
3,000,000	0.42%	0.59%	0.92%	1.26%	1.82%	2.10%
5,000,000	0.32%	0.46%	0.71%	0.98%	1.41%	1.63%
7,500,000	0.26%	0.37%	0.58%	0.80%	1.15%	1.33%
10,000,000	0.23%	0.32%	0.50%	0.69%	1.00%	1.15%
15,000,000	0.19%	0.26%	0.41%	0.56%	0.82%	0.94%
25,000,000	0.15%	0.20%	0.32%	0.44%	0.63%	0.73%
30,000,000	0.13%	0.19%	0.29%	0.40%	0.58%	0.67%
40,000,000	0.11%	0.16%	0.25%	0.35%	0.50%	0.58%
50,000,000	0.10%	0.14%	0.22%	0.31%	0.45%	0.52%
60,000,000	0.09%	0.13%	0.21%	0.28%	0.41%	0.47%
70,000,000	0.09%	0.12%	0.19%	0.26%	0.38%	0.44%
100,000,000	0.07%	0.10%	0.16%	0.22%	0.32%	0.36%
110,000,000	0.07%	0.10%	0.15%	0.21%	0.30%	0.35%
120,000,000	0.07%	0.09%	0.15%	0.20%	0.29%	0.33%
130,000,000	0.06%	0.09%	0.14%	0.19%	0.28%	0.32%
140,000,000	0.06%	0.09%	0.13%	0.18%	0.27%	0.31%
150,000,000	0.06%	0.08%	0.13%	0.18%	0.26%	0.30%
160,000,000	0.06%	0.08%	0.13%	0.17%	0.25%	0.29%
170,000,000	0.06%	0.08%	0.12%	0.17%	0.24%	0.28%
180,000,000	0.05%	0.08%	0.12%	0.16%	0.24%	0.27%
190,000,000	0.05%	0.07%	0.12%	0.16%	0.23%	0.26%
200,000,000	0.05%	0.07%	0.11%	0.15%	0.22%	0.26%
210,000,000	0.05%	0.07%	0.11%	0.15%	0.22%	0.25%
220,000,000	0.05%	0.07%	0.11%	0.15%	0.21%	0.25%
230,000,000	0.05%	0.07%	0.10%	0.14%	0.21%	0.24%
240,000,000	0.05%	0.07%	0.10%	0.14%	0.20%	0.24%
250,000,000	0.05%	0.06%	0.10%	0.14%	0.20%	0.23%
280,000,000	0.04%	0.06%	0.09%	0.13%	0.19%	0.22%
299,340,000	0.04%	0.06%	0.09%	0.13%	0.18%	0.21%

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 and 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 7: Longitudinal Generalized Variance Parameters

Table 7. Generalized Variance Parameters for PNL2[9]

Domain	Parameters *a*	Parameters *b*	Design[10] Effect	*f*
Poverty and Program Participation, Persons 15+				
Total	-0.00002877	7,217	2.80	1.166
Male	-0.00005953	7,217		
Female	-0.00005567	7,217		
Income and Labor Force Participation, Persons 15+				
Total	-0.00003257	8,170	3.17	1.240
Male	-0.00006739	8,170		
Female	-0.00006302	8,170		
Other, Persons 0+				
Total (or White)	-0.00002312	7,211	2.79	1.164
Male	-0.00004730	7,211		
Female	-0.00004521	7,211		
Black, Persons 0+	-0.00016460	6,655	2.58	1.119
Male	-0.00035281	6,655		
Female	-0.00030854	6,655		
Hispanic, Persons 0+	-0.00014327	7,721	2.99	1.205
Male	-0.00028442	7,721		
Female	-0.00028869	7,721		
Households				
Total (or White)	-0.00005101	6,477	2.51	1.171
Black	-0.00038494	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 7:

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.

9 PNL1 parameters are same as those in Table 1a since it covers the same reference period as CY2013

10 Design effect=b/sample interval, where sample interval=2,580