

SUPPORTING STATEMENT
U.S. Department of Commerce
U.S. Census Bureau
Survey of Income and Program Participation Panel
OMB Control No. 0607-1000

B. Collection of Information Employing Statistical Methods

1. Universe and Respondent Selection

The SIPP respondent universe is the civilian, noninstitutionalized population based on the 2010 decennial census, which contains approximately 303.5 million individuals. The SIPP uses a two-stage stratified sample of this universe. The first stage involves the division of the United States into groups of counties called the Primary Selection Units (PSUs), which are assembled into homogeneous groups called strata. Two PSUs are then selected from each stratum. The second stage involves selection of units within the selected PSUs.

Within each PSU, living quarters (LQs) are systematically selected from lists of addresses prepared from the 2010 decennial census. Other sampling techniques are used to represent new construction and group quarters. Low-income households are over-sampled from the lists of addresses.

The SIPP sample comes from the 2010 sample redesign. The SIPP sample is currently planned for approximately 53,000 designated LQs, it will yield approximately 42,400 designated occupied LQs at the time of interview in 2021, of which approximately 33,600 will be interviewed. Each household contains an average of 2.1 eligible respondents (aged 15 years and older), therefore, the planned SIPP sample of approximately 53,000 designated LQs, and it should contain approximately 70,560 survey respondents.

The SIPP Sample Design

The SIPP respondent universe is the civilian, noninstitutionalized population based on the 2010 decennial census, which contains approximately 303.5 million individuals. The SIPP uses a multistage stratified sample of the U.S. civilian noninstitutionalized population. The first stage involves the definition and division of the United States into groups of counties called the PSUs, which are assembled into homogeneous groups called strata. The second stage involves selection of units within the PSU.

The SIPP overlapping panels began with the 2018 SIPP. Each year SIPP sends a list of approximately 53,000 housing units to be interviewed in the field, including new panel sample cases (see Table 1). The new panel size is 53,000 housing units minus the total of eligible housing units from earlier panels. There are 820 selected PSUs from the sample frame. Starting with 2018 SIPP, we re-stratified PSU's and reduced the number of sample areas by selecting only 686 from the original 820 PSU's in order to reduce field costs. The selected PSUs cover both urban and rural areas of the United States. PSUs are formed from one or more contiguous counties. Larger populated PSUs (>100,000 HUs) are identified as self-representing (SR) PSUs, while the remaining PSUs are identified as non-self-representing (NSR). SR PSUs are in the SIPP sample with certainty while the NSR PSUs are stratified and selected with a probability proportionate to their size. During the stratification process, NSR PSUs are grouped together within the same state to form strata. During the PSU selection process, two NSR PSUs are selected from each stratum. There are 252 SR PSUs and 434 NSR PSUs in sample for the 2018 SIPP.

Table 1. SIPP Overlapping Sample

PANEL	CY18	CY19	CY20	CY21	CY22	CY23	CY24	CY25	...
2018	Wave 1	Wave2	Wave3	Wave4					
2019		Wave 1	Wave2	Wave3	Wave4				
2020			Wave 1	Wave2	Wave3	Wave4			
2021				Wave 1	Wave2	Wave3	Wave4		
2022					Wave 1	Wave2	Wave3	Wave4	
2023						Wave 1	Wave2	Wave3	...
2024							Wave 1	Wave2	...
2025								Wave 1	...

Within each selected PSU, LQs are systematically selected from the list of addresses updated by the Census and listing operations. The universe of addresses within the sample PSU is divided into two strata, one with a higher concentration of low-income households and the other with a lower concentration of low-income households. In the sampling strata, low-income households are defined by the poverty thresholds at the national level for families of certain sizes and age compositions. For example, the low-income household cutoff for one person with no related individual is \$18,700, and the low-income household cutoff with three people is \$28,600. Cutoffs are adjusted by the Consumer Price Index, and applied to the sample frame by household size and by age of householder for one- and two-person households. Addresses are sorted by geographic and demographic variables and a systematic selection of units is taken from each stratum. A 28% higher sampling rate is used in the stratum with the higher concentration of low-income households, thereby resulting in an oversample of low-

income households. Oversampling occurs to the extent that the rise in the variance for the estimate for persons 55 and over is not increased by more than 5%.

The frame for the SIPP is the Master Address File (MAF), which is maintained by the U.S. Census Bureau and is the source of addresses for the decennial census, the American Community Survey, and other demographic surveys. The MAF is updated using the U.S. Postal Service's Delivery Sequence File and various automated, clerical, and field operations.

The addresses selected for the 2018 SIPP sample are not be eligible to be selected for another Census Bureau demographic survey (CPS, SCHIP, NCVS, CE, and AHS) before 5 years after the last SIPP interview.

SIPP Weighting

As mentioned previously, starting in 2018, the SIPP collects data for each calendar year using an overlapping panel design, as illustrated in Table 1. The Calendar Year 2019 (CY19) files, for example, contain records from Wave 2 of the 2018 Panel and Wave 1 of the 2019 Panel. Note that SIPP calendar years refer to the year of interview; for CY19, we interviewed households in 2019, and collected data for the reference months January through December of 2018.

Each year, person-level weights are created for both every reference month and the whole calendar year. Starting in the second year of a panel, monthly and calendar year panel weights are also produced. All weights are calculated as the product of three components: a base weight, a household noninterview adjustment factor, and a second stage adjustment factor. The noninterview adjustment is calculated for each noninterview cell based on the following formula: Adjustment factor = $\sum (interviewed\ weights + noninterview\ weights) / \sum (interviewed\ weights)$ for each cell.

These factors are applied to the base weights in Wave 1, and to the noninterview weights in Wave 2+. There are 512 noninterview household cells in Wave 1, and 70 cells in Wave 2+. The second stage weights are calculated as a ratio adjustment of the sum of noninterview weights to the population controls and applied to each cell in five dimensions. The second stage weighting procedure consists of raking (with 400 maximum iterations and a tolerance level of 250), cell collapsing, and spouse equalization. This procedure is done for each reference month separately.

Due to the overlapping panels within the CYyy files, where yy represents the last two digits of the calendar year, some weighting steps are done separately by panel, and

others are done by combining the overlapping panels. Due to different response rates across panels, the Wave 1 Noninterview Adjustment, Movers' Adjustment, and Wave 2+ Noninterview Adjustment steps must be done separately for each panel. However, because all overlapping panels are used to produce the cross-sectional estimates for a calendar year, the Second Stage Adjustment procedure is done for all panels combined.

Each monthly weight for the SIPP CYyy is produced based on the SIPP survey universe corresponding to that month. Therefore, the controls (benchmark population estimates) for second stage raking for each monthly weight are those for the corresponding reference month. Each calendar year weight is produced based on the SIPP survey universe in December of that year. Therefore, the controls for second stage raking for the calendar year weights are those for December of that calendar year. Meanwhile, the panel weight is based on December of the reference year, so the controls for second stage raking for the panel weights are those for December of the reference panel years.

As a continued effort to evaluate the quality of the SIPP data, research into estimations is an ongoing endeavor. For instance, a report comparing retirement reporting to administrative records will be released in 2021. There are plans for future research to further investigate reporting of other program topics. We currently use administrative records to correct confusion between Supplemental Security Income (SSI) and Social Security (SS).

2. Procedures for Collecting Information

Within sampled households, all people 15 years old and over are interviewed using regular proxy-respondent rules. Each household member 15 years old or older who is present at the time of interview should respond for himself/herself. If a 15+ person is unable to respond, or is absent and not expected to be able to respond, proxy respondents may be selected to complete the interviews in their stead. Any knowledgeable household member who is 15 years or older may serve as a proxy.

For a 53,000 household sample, we estimate that each household contains 2.1 people aged 15 and above, yielding approximately 33,600 person-level interviews in 2021. We estimate that completing the SIPP interview takes approximately 63 minutes per adult on average, and consequently the total annual burden for SIPP interviews will be 74,088 hours per year.

3. Methods to Maximize Response

In all SIPP collections, we make special efforts to minimize non-interviews. Every household in the active sample receives an advance letter that explains the purpose of the survey and why the household's cooperation is important. In SIPP, each household will be given one of three brochures (Attachments F-H) that contains specific information about the SIPP and how it specifically relates to them in one of the following four areas: (1) Health Coverage and Disability (insurance, expenditures, etc.); (2) Families with Children (child care, child well-being, etc.); and (3) Labor Force (retirement, employment, unemployment, etc.). Each household will also receive a glossy "SIPP in the News" postcard (Attachment I) which highlights when SIPP appeared in major news sources such as the Washington Post. SIPP also distributes a SIPP Fact Sheet (Attachment E) which gives examples of how SIPP data are used in various government programs. For households that are identified but a response cannot be obtained (called Type A refusals), standard procedures include additional visits to the household by another FR or if needed, a Supervisory Field Representative (SFR) to obtain the household response.

4. Contacts for Statistical Aspects and Data Collection

The Census Bureau will collect and process these data. Within the Census Bureau, please consult the following individuals for further information:

Sample Design

Tracy Mattingly Lead Scientist (SIPP)
Demographic Statistical Methods Division
301-763-6445

Mahdi S. Sundukchi SIPP Survey Design Lead
Demographic Statistical Methods Division
301-763-4228

Data Content, Collection, and Tabulation

Hyon B. Shin Assistant Division Chief
Social, Economic, and Housing Statistics Division
301-763-6169

Attachments

- A. SIPP Instrument Items Booklet
- B. SIPP-(W1) – Wave1 Advance Letter
- C. SIPP-(W2) – Wave2 Advance Letter
- D. SIPP-(L1) – Thank You Letter
- E. SIPP Fact Sheet
- F. SIPP Brochure: Health Coverage and Disability
- G. SIPP Brochure: Families with Children
- H. SIPP Brochure: Labor Force and Employment
- I. SIPP in the News postcard