

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 of the United States, which contains approximately 329.5 million individuals at the time of sampling for the 2024 SIPP. The SIPP uses a two-stage stratified sample of this universe. The first stage involves grouping counties within each state into Primary Selection Units (PSUs), which are assembled into homogeneous groups called strata. One or two PSUs are then selected from each stratum. The second stage involves selection of units within the selected PSUs.

Within each PSU, housing units (HUs) are annually and systematically selected from lists of addresses prepared from the 2020 decennial census and updated every six months for new construction. Other techniques are used to sample housing unit equivalents from group quarters. Low-income households are oversampled from the lists of addresses.

The annual 2024 SIPP sample is currently planned for approximately 35,000 designated HUs, spread across four overlapping panels. It will yield approximately 33,000 eligible HUs at the time of interview in 2024, of which approximately 20,000 (60% response rate) or more should be interviewed to meet the U.S. Census Bureau Statistical Quality Standards. Each household contains an average of 2.0 eligible respondents (aged 15 years and older), therefore, the planned SIPP sample should contain approximately 40,000 person-level survey respondents.

The SIPP Sample Design

The SIPP respondent universe is the civilian noninstitutionalized population, which contains approximately 329.5 million individuals at the time of sampling for the 2024 SIPP, which uses the list of addresses updated in January 2023. The SIPP uses a multistage stratified sample of the U.S. civilian noninstitutionalized population. The first stage involves the definition of PSUs, which are assembled into homogeneous

groups called strata, from which one or two PSUs are selected to be in sample. The second stage involves selection of units within each selected PSU.

The SIPP overlapping panels began with the 2018 SIPP. From 2018 through 2022, SIPP sent a list of approximately 53,000 designated housing units to be interviewed in the field each year, including new panel sample cases (see Table 1). Starting in 2023, SIPP reduced that number from 53,000 to approximately 35,000 housing units in the annual sample. The new 2024 panel sample size is approximately 14,000 housing units, which is 35,000 housing units minus the total of eligible housing units from earlier panels. There were originally 820 selected PSUs from the sample frame in the 2010 sample redesign. Starting with 2018 SIPP, we re-stratified PSUs and reduced the number of sample areas by selecting only 686 from the original 820 PSUs 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. Meanwhile, each SR PSU is defined as a stratum on its own, with one PSU selected from the stratum. There are 252 SR PSUs and 434 NSR PSUs in sample for the 2018 SIPP and subsequent panels. This two-stage clustered sample design means that most SIPP estimates will have greater standard errors than those obtained from a simple random sample. More information on the SIPP design effect and guidance for calculating standard errors can be found in the annual Source and Accuracy Statement released by the U.S. Census Bureau.

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, HUs are systematically selected from the list of addresses updated by the decennial censuses 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 \$21,870, and the low-income household cutoff with three people is \$37,290, where the low-income cutoff is 150% of the poverty threshold. 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 32% 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 twice a year using the U.S. Postal Service's Delivery Sequence File and various automated, clerical, and field operations.

The addresses selected for the SIPP sample are not 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 2024 (CY24) files will contain records from Wave 1 of the 2024 Panel, Wave 2 of the 2023 Panel, Wave 3 of the 2022 Panel, and Wave 4 of the 2021 Panel. Note that SIPP calendar years refer to the year of interview; for CY24, we interview households in 2024 and collect data for the reference months January through December of 2023.

Each year, cross-sectional person-level weights are created for each of the 12 reference months and the entire calendar year. Longitudinal weights are also produced covering reference periods of two to four years. All weights are calculated as the product of three components: a base or initial 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 =

$$\sum_{\square} (interviewed\ weights + noninterview\ weights) / \sum_{\square} (interviewed\ weights) \text{ for each cell.}$$

These factors are applied to the base weights in Wave 1, and to the Wave 1 noninterview weights adjusted for movers 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 4,000 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 longitudinal weight is based on December of the last reference year, so the controls for second stage raking for the panel weights are those for December of the last reference year.

As a continued effort to evaluate the quality of the SIPP data, research into estimations is an ongoing endeavor. 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). We publish SIPP survey key estimates in the annual Source and Accuracy Statement.

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 themselves. 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 35,000 household sample, we estimate that each household contains 2.0 people aged 15 and above, yielding approximately 20,000 household-level interviews and 40,000 person-level interviews in 2024. We estimate that completing the SIPP interview takes approximately 50 minutes per adult on average, and consequently the total annual burden for SIPP interviews will be 33,330 hours per year.

3. Methods to Maximize Response, Ensure Geographical Coverage, and Minimize Bias

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, households may 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.). Households may 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 provides a SIPP Fact Sheet (Attachment E) which gives examples of how SIPP data are used in various government programs. For households that 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.

During the data collection period, adaptive design measures are taken to ensure the SIPP respondents provide sufficient geographical coverage of the sampled PSUs and accurately represent population characteristics. Case prioritization is used to guide FRs to households that are in either geographical areas with low response or reflect demographic characteristics that are underrepresented among the respondents.

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

Ashley Westra Supervisory Mathematical Statistician (SIPP)
Demographic Statistical Methods Division
301-763-8536

Data Content, Collection, and Tabulation

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

Attachments

- A. SIPP_2024_Instrument_Booklet
- B. SIPP_2024_Flashcard_Booklet
- C. SIPP-05(L1)W1 – Wave1 Advance Letter
- D. SIPP-05(L3)W2+ – Wave2+ Advance Letter
- E. SIPP-23(L)(C)(GQ)(AT) – Group Quarters Letter
- F. SIPP-06(L1)2023 – Thank You Letter
- G. SIPP Fact Sheet
- H. SIPP Brochure: Health Coverage and Disability
- I. SIPP Brochure: Families with Children
- J. SIPP Brochure: Labor Force and Employment
- K. SIPP in the News Postcard
- L. 60-Day Federal Register Notice for 2024 SIPP
- M. SIPP Response to OMB 11.10.2022