

**Attachment F**

**Overview of CPS Sample Design and Methodology (updated 2016)**

## OVERVIEW OF 2010 CPS SAMPLE DESIGN AND METHODOLOGY

### 1. CPS Sample Design and Selection

The Current Population Survey (CPS) is a monthly survey designed primarily to produce national and state estimates of labor force characteristics of the civilian noninstitutional population (CNP) 16 years of age and older. It is conducted in approximately 59,000 eligible housing units throughout the United States. (Note: 'Eligible' can be simplistically defined as an occupied housing unit having at least one person in the CNP.) This sample includes 10,000 eligible housing units from the monthly supplementary sample to improve state-level estimates of health insurance coverage for low-income children, also known as the CHIP expansion. This supplementary sample has been part of the official CPS since July 2001. Thirty-two states plus the District of Columbia contain this supplementary sample each month.

The CPS sample has been redesigned based on information from the 2010 Decennial Census, in accordance with usual practice. Historically, the CPS sample has been redesigned after each Decennial Census.

The CPS sample is a probability sample based on a stratified two-stage sampling scheme: selection of sample primary sampling units (PSUs) and selection of sample housing units within those PSUs. In general, the CPS sample is selected from lists of addresses obtained from the Master Address File (MAF) with updates from the United States Postal Service (USPS) twice a year. The MAF is the Census Bureau's permanent list of addresses, including their geographic locations, for individual living quarters. It is continuously maintained through partnerships with the USPS; with Federal, State, regional, and local agencies; and with the private sector, and it is used as a sample frame by many Census Bureau demographic surveys.

The CHIP sample selection methodology is similar to that used for the CPS.

#### a. State-Based Design

In the first stage of sampling, PSUs are selected. These PSUs consist of counties or groups of contiguous counties in the United States, and are grouped into strata. The CPS is a state-based design. Therefore, all PSUs and strata are defined within state boundaries and the sample is allocated among the states to produce state and national estimates with the required reliability, while keeping total sample size to a minimum. The specified coefficient of variation (CV) requirement for the monthly unemployment level for the nation, given a 6 percent unemployment rate, is 1.9 percent or less. (Note: The CV of an estimate is the estimate itself divided by its standard error, usually expressed as a percent.) This CV is based on the requirement that a difference of 0.2 percentage points in the unemployment

rate between two consecutive months be statistically significant at the 0.10 level. Additionally, the required CV on the annual average unemployment level for each state and the District of Columbia, given a 6 percent unemployment rate, is 8 percent or less. For New York and California, the state reliability requirement applies to the following substate areas: New York City (five boroughs only), the balance of New York State, Los Angeles County, and the balance of California.

b. First Stage of the Sample Design: PSU Stratification and Selection

The variables chosen for grouping PSUs in each state into strata reflect the primary interest of the CPS in maximizing the reliability of estimates of labor force characteristics. Basically, the same set of stratification variables, from the 2010 Decennial Census and the American Community Survey (ACS), are used for each state: unemployment statistics by gender; number of families maintained by a woman; and the proportion of occupied housing units with three or more people. In addition, the number of persons employed in selected industries and the average monthly wage for selected industries are used as stratification variables in some states. The industry-specific data are averages over the period 2000 through 2008 and are obtained from the Quarterly Census of Employment and Wages program of the BLS.

Thus, each stratum consists of one or more PSUs. Within each stratum, a single PSU is chosen for the sample, with probability proportional to its population as of the 2010 Census. Some strata have only one PSU, and each is included in the sample as a self-representing PSU; these strata generally include the most populous counties within each state. The remaining PSUs are grouped into non-self-representing strata within state boundaries. In each of these strata, one PSU is selected to represent all of the PSUs in that stratum.

The PSUs, strata, and sample PSUs are the same for CPS and CHIP. This differs from the 2000 sample design, which had three states with different designs. In total, 852 PSUs (1,385 counties) from a total of 1,987 PSUs (3,143 counties) in the United States are in sample for either just the basic CPS or for both the basic CPS and the CHIP expansion.

c. Second Stage of the Sample Design: Selection of Housing Units

- 1) The 2010 sample design comprises three frames: unit, coverage improvement (CI) and group quarters (GQ). The unit frame consists of housing units in Census blocks that contain a very high proportion of complete addresses. It covers most of the population and accounts for approximately 95% of the CPS sample. It is updated every six months with new growth records and will be sampled from annually. The CI frame is intended to improve the coverage of the unit frame. It is feasible to target blocks (in 13 targeted states) and then list them to efficiently capture most of the undercoverage. The CI frame is updated annually with

information from July MAF extracts. There is a single GQ frame in the 2010 sample design and its sample is selected in a three-year cycle.

- 2) Within these sampling frames, housing units are sorted based on characteristics of the ACS and geography. Then, from each frame, a systematic sample of addresses within the sample PSUs is obtained. Most of the sample addresses are selected in a single stage of sampling within the selected PSUs; for a relatively small proportion, an additional stage of selection within the PSU is necessary.

d. Rotation System

Each sample is divided into eight approximately equal panels, called rotation groups. A rotation group is interviewed for four consecutive months, temporarily leaves the sample for eight months, and then returns for four more consecutive months before retiring permanently from the CPS (after a total of eight interviews). This rotation scheme has been in use since July 1953. When compared to the previous rotation pattern, the implementation of this rotation pattern resulted in an improvement in the reliability of estimates of month-to-month change as well as estimates of year-to-year change.

e. Major Differences from the 2000 CPS Sample Design

The 2010 sample design differs from that of 2000 in a variety of ways. These changes have resulted after consideration of numerous factors, including improving reliability of the estimates, minimizing costs, and maximizing comparability of estimates across time. Major changes include the following:

- 1) Sample is now selected from the continually updated MAF, with sample phase-in beginning in 2014, and ACS data is used to sort and stratify the housing units on the MAF. Previously, sample was selected from Decennial Census address lists and stratification was done using information also from the Decennial Census.
- 2) In the past, the CPS sample universe was distributed across four frames: unit, permit, GQ, and area, with approximately 80% of the CPS sample coming from the unit frame. As mentioned in Paragraph 1.c.1, the 2010 sample design comprises three frames: unit (updated with new growth records), CI, and GQ. As the result of improved flexibility and reduced complexity of block listing via the CI frame, an area frame no longer exists. Instead, the block listing process will enable a flexible workload that can change as often as annually, depending on budget resources and on the need for coverage improvement. Rather than having GQs split between the GQ frame and the area frame as in past designs, there is now a single GQ frame. An additional change is the exclusion of military

GQs from the sampling universe since research showed that they are extremely unlikely to convert to a non-institutional GQ.

- 3) In past designs, the CPS had selected a decade of sample housing units all at once, occurring just after the Decennial Census, with periodic supplementation of new construction through sampling of building permits and area listing results. The selected housing units were then parsed into monthly samples throughout the decade. This approach was the most cost effective and sensible method of sampling in the context of once-a-decade operations.

In the 2010 sample design, sampling occurs annually for the unit frame. This changes the second-stage sample selection of housing units from once-a-decade sampling to annual sampling. The benefits of selecting a fully representative sample of housing units on an annual basis include:

- Better control of survey sample size.
- More accurate addresses due to twice-a-year updates of valid/invalid status, geocoding errors, and geography changes of previously existing records that are eligible for selection.
- Ability to modify or select new samples more quickly in response to population shifts in order to meet reliability criteria.
- More flexibility in accommodating sample expansions and contractions in response to changes in budget or data requirements.
- Ability to implement methodological changes and process improvements more quickly and easily than before.
- Potential to reduce variances on annual average estimates with annual sampling; this is a potential for cost saving because less sample is needed.

Note that annual sampling does not apply to the GQ frame, where sample is selected three years at a time, or to the first-stage sample selection of the PSUs. Also, a housing unit selected by any demographic survey will not be available for selection by subsequent surveys until five years after its last interview.

## 2. CPS Estimation Procedure

Under the estimating methods used in the CPS, initial second-stage results for a given month are based on responses obtained from the monthly sample of eight panels. It involves weighting the data from each sample person. The baseweight, which is the inverse of the probability of the person being in the sample, is a rough measure of the number of actual persons that the sample person represents. Almost all sample persons within the same state have the same baseweight, and every person in the same housing unit receives the same baseweight. These weights are then adjusted for noninterview, and a ratio adjustment procedure is applied.

a. Noninterview Adjustment

The baseweights for all interviewed housing units are adjusted to account for occupied sample housing units for which no information was obtained. Reasons for a noninterviewed housing unit include absence of the occupants, impassable roads, refusal of the occupant to participate in the survey, or unavailability of the occupant for other reasons. The noninterview adjustment is performed by noninterview cluster. Noninterview clusters are classified as either metropolitan or non-metropolitan. PSUs classified as metropolitan are assigned to metropolitan clusters. PSUs representing metropolitan areas of the same or similar size (based on Census 2010 population) are grouped into the same noninterview cluster. Each metropolitan cluster is further divided into two cells: central city and balance of the metropolitan area. Likewise, non-metropolitan PSUs are assigned to non-metropolitan clusters. All non-metropolitan areas in a state are placed within the same noninterview cluster. Due to small sample sizes, a few non-metropolitan noninterview clusters contain PSUs from more than one state.

b. Adjusting Estimates to Population Controls

The distribution of the population selected in the sample may differ somewhat, by chance, from that of the population as a whole in such characteristics as age, race, Hispanic origin, and gender. Since these characteristics are correlated closely with labor force participation and other principal measurements made from the sample, survey estimates are substantially improved when weighted appropriately by the known distribution of these population characteristics. This is accomplished through four adjustments:

1) First-stage ratio adjustment

In the CPS, some of the sample areas are chosen to represent both themselves and other areas in the same state, but not in the sample; the remainder of the sample areas represent only themselves. The first-stage ratio estimation procedure is designed to reduce that portion of the variance resulting from non-self-representing PSUs. Therefore, this adjustment procedure is applied only to sample areas that represent other areas and is done by Black alone / not Black alone cells at a state level. Each race cell is further divided into two age cells: age 0-15, and age 16 and older.

2) National and state coverage adjustments

The national and state coverage adjustments are intended to improve the national and state estimates by race, Hispanic origin, gender, and age. The national coverage adjustment is done by Black alone, White alone, Asian alone, and the residual of all other race categories for non-Hispanics, and White alone and not White alone for Hispanics. (Note that respondents

who indicate that they belong to more than one race are included in the Residual race category.) These race/ethnicity categories are further divided into cells representing various combinations of age and gender. This national adjustment is performed by month-in-sample pair (1,5; 2,6; 3,7; and 4,8).

The cells used in the state coverage adjustment are defined by race category (Black alone, not Black alone), age, and gender. The adjustment is performed either for each month-in-sample pair or for all eight month-in-sample groups combined. The actual cells used vary by state and race category.

3) Second-stage ratio adjustment

The second-stage ratio adjustment modifies sample estimates in a number of age-gender-race-Hispanic origin groups to independently derived Census-based estimates of the CNP in each of these groups. This adjustment reduces mean square error of sample estimates by reducing bias due to differential coverage of the sampling frame. The adjustment is executed in three steps and each set of three steps is referred to as a “rake.” There are 10 cycles (or iterations) of raking. Each step in each rake is done by month-in-sample pair.

In the first step, the sample estimates are adjusted for each state and the District of Columbia to independent controls for the CNP by age and gender. There are three age cells by gender (0-15, 16-44, 45 and over). The second step of the adjustment is done at the national level by Hispanic origin status. Hispanic and non-Hispanic each have 13 age/gender cells, which are adjusted to nationwide independent controls. The third and final step of the second-stage adjustment is performed by race (Black alone, White alone, Residual race). The cell division is by age/race/gender. Each of these cells is adjusted to national independent population controls as in the previous step.

The entire second-stage adjustment procedure is iterated through 10 rakes. This iteration ensures that the sample estimates of state and national population by the various age-race-gender-Hispanic origin categories will be virtually equal to the independent population controls.

c. Composite Estimation and Weighting

The last step in the preparation of most CPS estimates makes use of a composite estimation procedure. A basic composite estimate is a weighted average of 1) a second-stage estimate based solely on current month responses and 2) a composite estimate from the previous month that is updated to the current month with an estimate of month-to-month change based on six sample panels that are

common to both months. Estimates of month-to-month change in employment and unemployment that are computed using composite estimates generally have lower sampling errors than comparable change estimates using second-stage estimates. A composite weighting procedure computes a weight for each person. Using these weights, it is then unnecessary to recompute composite estimates of labor force each time a table is produced.

### 3. Nonresponse in the CPS

If a respondent is reluctant to participate in the CPS, the interviewer immediately informs the regional office staff. The regional office sends a follow-up letter to the household explaining CPS in greater detail and urging cooperation. The interviewer then recontacts the household and attempts the interview again. If this procedure fails, a field supervisor then contacts the household in an attempt to convert the reluctant respondent. Methods used to interview reluctant households include conducting telephone or personal interviews with the household, if so requested, and interviewing a designated individual within the household. The CPS estimation procedure adjusts for household nonresponse in its noninterview adjustment procedure, detailed in the preceding Paragraph 2.a. Three imputation methods for individual item nonresponse are used: relational imputation, hot-deck imputation, and longitudinal assignments. As appropriate, longitudinal assignments are used in most of the labor force edits. The CPS household noninterview rate ranges between 9 and 10 percent monthly. Accuracy of the CPS data is maintained through interviewer training and monthly home studies, monitoring of error and noninterview rates, and systematic reinterviewing of CPS households. Each month about 10 percent of all CPS enumerators have a portion of their assignments reinterviewed for quality control purposes. Depending on the interviewer's experience level and position, they can be selected as many as three times every 15 months. Errors uncovered during the reinterview are discussed with the original interviewer and remedial action is taken. Also, 1 percent of cases are reinterviewed to measure response error.

### 4. CPS Contact Persons

At the Census Bureau, individuals consulted on the statistical aspects of the CPS are Yang Cheng, CPS Lead Scientist of the DSMD at (301) 763-3287; CPS Survey Design Lead of the DSMD at (301) 763-3714. Lisa Clement, CPS Survey Director of the Associate Director for Demographic Programs Division (ADDP) at (301) 763-5482 and Gregory Weyland of the ADDP at (301) 763-3790 can be contacted for survey design, data collection, and processing issues.

At the Bureau of Labor Statistics, Ed Robison (202-691-6363) is the contact for statistical aspects of the CPS, and Dorinda Allard (202-691-6470) is responsible for data analysis.