

## OVERVIEW OF CPS SAMPLE DESIGN AND METHODOLOGY (For the Sample Design Based on Census 2000)

### 1. CPS Sample Design and Selection

The Current Population Survey (CPS) is a monthly survey conducted in approximately 60,000 occupied households throughout the United States, including approximately 10,000 households from the monthly supplementary sample to improve state-level estimates of health insurance coverage for low-income children, also known as the SCHIP expansion. This supplementary sample has been part of the official CPS since July 2001. Twenty-six states plus the District of Columbia contain this supplementary sample each month.

The CPS is a probability sample based on a stratified sampling scheme. In general, the CPS sample is selected from lists of addresses obtained from the most recent decennial census and updated for new construction. The SCHJP sample selection methodology is generally similar to that used for the CPS.

#### a. State-Based Design

In the first stage of sampling, primary sampling units (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 national reliability requirement is a coefficient of variation (CV) of 1.9 percent or less on the monthly estimate of unemployed, assuming an unemployment rate of 6 percent. The state reliability requirement is a CV of 8 percent or less on the annual average estimate of unemployed, assuming an unemployment rate of 6 percent. For New York and California, the state reliability requirement applies to the following substate areas: New York City (the five boroughs only), the balance of New York state, Los Angeles County, and the balance of California.

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 most recent decennial census (in this case, Census 2000). This PSU represents the entire stratum from which it was selected. In the case of strata consisting of only one PSU, the PSU is termed “self-representing.” Sample PSUs from strata with more than one PSU are called “non self-representing.” The sample PSUs are selected from each stratum using a statistical method which increases the probability of keeping the same areas in the sample from one sample redesign to the next.

In total, 824 geographic areas from a total of 2,025 geographic areas in the United States are in sample for either the basic CPS, the SCRIP expansion, or for both the basic CPS and the SCHIP expansion; four of these areas contain sample for the SCHIP expansion only.

A sample of addresses within the sample PSUs is obtained in a second step. Most of the sample addresses are selected from census lists in a single stage of sampling within the selected PSU; for a relatively small proportion, an additional stage of selection within the PSU is necessary.

b. PSU Stratification

The variables chosen for grouping the non self-representing 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 Census 2000) were used for each state: unemployment statistics by male, female; number of families with a female head; 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 were used as stratification variables in some states. The industry-specific data are averages over the period 1990 through 1998 and were obtained from the Bureau of Labor Statistics.

In states with SCHIP sample, the self-representing PSUs are the same for both CPS and SCHIP. In all but three SCHIP states, the same non self-representing sample PSUs are in sample for both surveys. In the three remaining states, in order to improve the reliability of the SCHIP estimates in those states, the SCHIP non self-representing PSUs were selected independently from the CPS sample PSUs, with replacement.

c. Rotation System

Each sample is divided into eight approximately equal 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. The end result of this rotation pattern is an improvement in the reliability of estimates of month-to-month change as well as estimates of year-to-year change.

2. CPS Estimation Procedure

Under the estimating methods used in the CPS, all of the results for a given month become available simultaneously and are based on returns for the entire panel of respondents. The CPS estimation procedure involves weighting the data from each sample person. The unbiased weight, 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 unbiased weight. The unbiased weights are then adjusted for noninterview, and a ratio adjustment procedure is applied.

a. Noninterview Adjustment

The weights for all interviewed households are adjusted to account for occupied sample housing units for which no information was obtained. Reasons for a noninterviewed household 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 2000 population) are grouped in 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 sex. Since these characteristics are correlated closely with labor force participation and other principal measurements made from the sample, the latter estimates can be improved substantially when weighted appropriately by the known distribution of these population characteristics. This is accomplished through four adjustments as follows:

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 requiring sample areas to represent areas not in sample (i.e., 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 were introduced in January 2003 in an effort to improve the national and state estimates by race, Hispanic origin, sex, and age. The national coverage adjustment is done by Black alone, White alone, Asian alone, and All Other Race for non-Hispanics and by White alone, and All Other Race for Hispanics. The All-Other-Race category includes respondents who indicate they belong to more than one race. These race/ethnicity categories are further divided into cells representing various combinations of age and sex.

The 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 sex. 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-sex-race-Hispanic origin groups to independently derived census-based estimates of the civilian noninstitutional population (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 carried out 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 sex. There are three age cells by sex (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/sex 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, All Other Race). The All-Other- Race category includes respondents who indicate they belong to more than one race. The cell division is by age/race/sex. 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-sex-race-Hispanic origin categories will be virtually equal to the independent population controls.

3. Specialized Sampling

An unusual problem occurs in the CPS sample that requires specialized sampling procedures. This problem occurs in rural areas in the state of Alaska because of their sparse population. Special methods of defining the ultimate sampling units are used in these areas.

4. Periodic Data Collection Cycle

Collection of CPS data on a monthly basis is mandated in Paragraph 2 of Title 29, United States Code. Less frequent collection would place the Bureau of Labor Statistics, which is the co-sponsor of the CPS, in violation of this code.

5. 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 Supervisory Field Representative 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 Paragraph 2.a. above. Individual item nonresponse is allocated using a procedure in which the missing data are assigned from individuals whose data are complete and have similar characteristics. The CPS household noninterview rate ranges between 6.5-7.5 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 7 percent of all CPS enumerators have their assignments reinterviewed for quality control purposes. Errors uncovered during the reinterview are discussed with the original interviewer and remedial action is taken. An additional 7 percent of CPS enumerators have their assignments reinterviewed monthly to measure response error.

6. Methodological Testing

The basic CPS program is not used as a vehicle for testing procedures or methods. There was, however, an extensive program of testing conducted on CPS methods, procedures, and content. This program, the Methods Development Survey (MDS), and its predecessor, the Methods Test Panel (MTP), was conducted between May 1978 and December 1993.

The result of this testing is the new CPS Labor Force Instrument that is collected in a completely automated CATI/CAPI environment. For further information on the testing procedures and results see the February 1994 Employment and Earnings article titled, "Revisions to the Current Population Survey Effective January 1994."

7. CPS Statistical Contacts

Individuals consulted on the statistical aspects of the CPS are Alan R. Tupek (301/763-4287), Larry Cahoon (301/763-4203), and Harland Shoemaker (301/763-4275) of the Demographic Statistical Methods Division, U.S. Census Bureau. Chester E. Bowie (301/457-3773), Enrique Lamas (301/457-3811), and Maria Reed (301/763-3806), of the Demographic Surveys Division, U.S. Census Bureau, have responsibility for data collection and processing. Shail Butani (202/691-6347) and Edwin Robison (202/691-6363) are the contacts for statistical aspects of the CPS at the Bureau of Labor Statistics; Tom Nardone (202/691-6379) is responsible for data analysis.