

# Redesign of the Sample for the Current Population Survey

Harland Shoemaker

**T**he Current Population Survey (CPS), sponsored by the Bureau of Labor Statistics (BLS) and conducted by the U.S. Census Bureau, is a monthly sample survey of households that produces reliable estimates of labor force characteristics of the working-age population.

Since the inception of the survey in 1940, there have been various changes in the design of the CPS sample. The last major sample redesign of the CPS took place in the early 1980s and involved a major effort to upgrade the quality and reliability of State-level data. Usually, a new sample is selected after each decennial census. The sample selected after the 1990 census, introduced between April 1994 and July 1995 and currently in use, is similar to the one designed in the early 1980s. Starting in April 2004, the 1990 sample began to be phased out through a series of changes that will be completed in July 2005. Again, the new sample is similar to the sample currently in use in terms of its size, although it does include some reallocation among the States to achieve the specified reliability requirements more efficiently. The principal reason for the redesign, however, is to maintain the efficiency of the sample design by updating the sampling frame with Census 2000 data.

This article discusses the scope and nature of the changes to the 1990 sample design and their effect upon the survey methods and the estimates. It also describes how the new design is being phased in.

## Design objectives

The CPS is a monthly probability sample based on a stratified sampling design that is intended to provide national and State estimates of labor force characteristics with a specified level of reliability. In the first stage of sampling, primary sampling units (PSUs), which correspond to substate areas, counties, or groups of counties that are geographically contiguous, are selected. These PSUs are then grouped into strata, and within each stratum, a single PSU is chosen for the sample. The probability of selection of this PSU, which represents the entire stratum from which it was selected, is proportional to its population as of the most recent decennial

census. In the case of strata consisting of only one PSU, the PSU is chosen with certainty.

The CPS sampling frame, which is based on a particular census, represents the actual population less efficiently over time. Replacing the 1990 sampling frame with one based on Census 2000 improves the efficiency of data collection by field representatives. In addition, the more recent and more accurate Census 2000 information was used to redefine the PSUs and to restratify them, incorporating variables that are more strongly related to the unemployment rate. Updated data from the BLS Quarterly Census of Employment and Wages, also known as the ES-202 program, also were used in the restratification as well. Finally, the updated Census 2000 information was utilized to select a more efficient sample within the selected PSUs.

The 2000 CPS sample design reliability requirements are the same as those for the 1990 design. The only difference is that, for the new design, the supplementary sample from the State Children's Health Insurance Program (SCHIP) survey, which was first used in the production of official CPS estimates in July 2001, was selected independently from the CPS sample. However, the data from both the CPS and the SCHIP survey will continue to be used to produce the official estimates each month. Reliability requirements for the sample are expressed in terms of coefficients of variation (CVs). All references to CVs in this article include an assumed unemployment rate of 6 percent.

*The Nation.* The required CV on the unemployment level for the Nation is 1.9 percent. The CV itself is based on the requirement that a difference of 0.2 percentage point in the unemployment rate for consecutive months be significant at the 90-percent confidence level.

*States and substate areas.* For all States, except California and New York, the CV on the annual average unemployment level is no larger than 8 percent. In New York and California, this CV applies separately to the substate areas of New York City (Bronx, Kings, New York, Queens, and Richmond counties), the balance of New York State, the Los Angeles-Long Beach-Glendale metropolitan division (Los Angeles County), and the balance of California.

---

Harland Shoemaker is Chief, Current Population Survey Branch, Demographic Statistical Methods Division, U.S. Census Bureau.

### Phase-in of the new design

The phase-in of the new CPS design will have two dimensions. The first will involve “continuing” areas—that is, those areas selected in both the old and new sample designs. The sampling frame in these areas began to be updated in April 2004 and will continue through July 2005. Sample households selected from Census 2000 and building permits for new residential construction since Census 2000 will replace households selected from the 1990 census lists and from building permits for new residential construction issued since the 1990 census. Ninety percent of the entire sample is in these “continuing” areas.

The second dimension of the phase-in involves changing the areas selected for sampling (or the sample PSUs). From August through November 2004, areas selected only for the new design gradually replaced those selected for the old design.

The redesigned sample will contain 824 of the 2,025 total PSUs, which cover the entire geographic area of the United States. In the 1990 design, 754 areas out of the 2,007 PSUs covering the Nation were in the sample. The numbers of PSUs in the respective samples are not directly comparable, however, because, in the 1990 design, some PSUs in New England and Hawaii were formed from minor civil divisions in addition to counties, while in the 2000 design, the PSUs are strictly county based. Also, the PSUs have been redefined to correspond to the new Core-Based Statistical Area (CBSA) definitions and to improve efficiency in field operations.

In both designs, the sample is divided into eight approximately equal rotation groups. The two designs also use the same rotation pattern, which has been a part of the CPS since 1953. According to this pattern, a rotation group is interviewed for 4 consecutive months, temporarily leaves the sample for 8 months, and then returns for 4 more consecutive months before retiring permanently from the CPS (after a total of eight interviews). The SCHIP sample shares these features with the CPS.

*Continuing areas.* From April 2004 through July 2005, the new CPS design is being phased in for continuing areas, one rotation group per month. The phase-in consists of gradually replacing the sample selected from the 1990 census-based frames with that selected from the Census 2000-based frames. The gradual aspect minimizes the impact on the continuity of the CPS data series. In the replacement process, sample sizes within individual areas may change, but the total monthly sample size for the survey will remain about the same. For each month from April through July 2004, an additional rotation group from the new sample was phased in. Since July 2004, half of the sample in continuing areas has been obtained from the new design. Due to the rotation pattern, this proportion will not change from July 2004 through March 2005. Then, from April through July 2005, the old-design sample will again be replaced by the new-design sample, one rotation group per month. The phase-in of the new sample

will be complete in July 2005. The phase-in pattern is depicted in table 1.

*New sample areas.* In August 2004, the new sample areas began to replace the outgoing areas. The effects of this change are expected to be minor, because the sample in continuing areas accounts for about 90 percent of the national estimate. During the phase-in, only 10 percent of the sample at any one time involves transitional areas (incoming new areas or outgoing old areas). As shown in table 1, from August 2004 through November 2004, the new areas were introduced, two rotation groups per month. One of these replaced a retiring rotation group from the old design, and the other replaced a rotation group from the old design that would otherwise have returned to the sample for its fifth interview. For this reason, the normal rotation pattern in these areas was not preserved during the August through November 2004 period. The replacement of outgoing areas with new areas was completed in November 2004, 8 months before the phase-in of the new sample in continuing areas is scheduled to be completed.

As the new areas are phased in, the base weights and weighting factors used in computing the estimates will be updated to account for the change in the areas that make up the sample. The factors include the noninterview adjustment and the first-stage estimation procedure. (See the section titled “Explanatory Notes and Estimates of Error” in this publication for descriptions of these procedures.) These updated factors will be introduced into the estimation process at the same rate as the new sample areas are introduced into the sample.

Table 1. Rotation group redesign scheme

Year and month	Continuing areas (90 percent)		Replacement areas (10 percent)		Overall percent of households on 2000 design
	New sample rotations	Old sample rotations	New sample rotations	Old sample rotations	
2004:					
April .....	1	7	0	8	11
May .....	2	6	0	8	23
June .....	3	5	0	8	34
July .....	4	4	0	8	45
August .....	4	4	2	6	48
September .....	4	4	4	4	50
October .....	4	4	6	2	53
November .....	4	4	8	0	55
December .....	4	4	8	0	55
2005:					
January .....	4	4	8	0	55
February .....	4	4	8	0	55
March .....	4	4	8	0	55
April .....	5	3	8	0	66
May .....	6	2	8	0	78
June .....	7	1	8	0	89
July .....	8	0	8	0	100

## Effects of the changeover

*Field operations.* Because new areas will be phased into the sample and some old areas will be dropped, increased hiring and training of field representatives, as well as some dismissals, will be required during this period. To minimize the possible influence of inexperienced field representatives on the CPS program, trainees will have half the normal assignments the first month they interview (July 2004). In addition, all of these interviews will be personal visits, and the initial interview results will not be used for estimates. During their first month of interviewing, new field representatives also will have 3 days of classroom study, 20 hours of self-study, and 2 days of on-the-job training with a senior field representative. Their workload will increase by one rotation group per month in each of the 4 months after July 2004; this additional sample will be included in the official estimates. Beginning in November 2004, all data collected by these new field representatives will be incorporated into the official estimates.

Efforts also will be made to maintain the experienced staff of field representatives who will be affected by the phaseout of certain sample areas, by offering them work in adjacent, continuing areas or on other programs. In any event, they will be notified in advance of the phaseout of their sample areas. Their work during this period will be closely monitored to avoid any adverse impact that their anticipation of termination may have on the data.

*Continuity of data series.* Three features of the new CPS design have the potential to affect published estimates: (1) the temporary disruption of the rotation pattern from August 2004 through June 2005 for a comparatively small portion of the overall sample, (2) the change in sample areas, and (3) the introduction of the new CBSAs (formerly called metropolitan areas).

Labor force estimates in the CPS are affected by changes in the rotation pattern because of the relationship between the number of times a rotation group has been interviewed and the expected value of the estimate obtained from the group. Normally, the eight rotation groups contributing to the published estimate are evenly distributed in terms of the

number of times they have been in the sample (that is, one rotation group in the sample for the first time, one for the second time, ..., and one for the eighth time). However, when the new areas are phased in at the rate of two rotation groups per month, this even distribution by time spent in the sample is lost. For example, in August 2004, two rotation groups in new PSUs were in the sample for the first time and none for the fifth time. Because published CPS estimates represent averages of the eight rotation group estimates, a change in the distribution of the sample rotation groups' number of months in sample may result in a difference in the overall estimates for those months, even though the actual value of the characteristic being estimated may not have changed at all. Conversely, this effect may mask real changes in the estimates. In either case, the disruption to the rotation pattern is expected to have a negligible effect on the published estimates.

In addition to the temporary phase-in effect just discussed, the new design may produce some small permanent differences, particularly in the State estimates. These differences are the result of restratifying the PSUs within each State for the 2000 design, selecting a different set of sample PSUs to represent each State, and adjusting certain weighting factors to reflect this change. The differences may appear to be discontinuities in the data series, but in fact are the result of variability arising from selecting only a sample of PSUs to represent entire States.

The introduction of new sample areas also has an effect on data series, the upshot of which is that there will be slightly less overlap of the sample used to produce estimates of year-to-year change between 2004 and 2005. However, the increase in variability due to the smaller overlap is expected to be negligible, because the proportion of overlap in the samples will drop to a minimum of 45 percent from the usual 50 percent.

Finally, changes in the definitions of metropolitan areas (now called Core-Based Statistical Areas) could affect the CPS estimates of metropolitan populations at both the State and national levels. However, preliminary research has shown that the national-level estimates of the metropolitan and nonmetropolitan populations should not change appreciably because of the new sample design.