

ATTACHMENT F

Overview of CPS Sample Design and Methodology

(Note: A detailed description of the CPS sample design and methodology is available in *Current Population Survey, Design and Methodology, Technical Paper 63RV, March 2002*, available from the Census Bureau Web site:<http://www.bls.census.gov/cps/tp/tp63.htm>.)



Sampling

Since the inception of the survey, there have been various changes in the design of the CPS sample. The sample is traditionally redesigned and a new sample selected after each decennial census. Also, the number of sample areas and the number of sample persons are changed occasionally. Most of these changes are made in order to improve the efficiency of the sample design, increase the reliability of the sample estimates, or control cost.

Changes in this regard since 1960 are as follows: When Alaska and Hawaii received statehood in 1959 and 1960, respectively, three sample areas were added to the existing sample to account for the population of these States. In January 1978, a supplemental sample of 9,000 housing units, selected in 24 States and the District of Columbia, was designed to provide more reliable annual average estimates for States. In October 1978, a coverage improvement sample of approximately 450 sample household units representing 237,000 occupied mobile homes and 600,000 new construction housing units was added. In January 1980, another supplemental sample of 9,000 households selected in 32 States and the District of Columbia was added. A sample reduction of about 6,000 units was implemented in May 1981. In January 1982, the sample was expanded by 100 households to provide additional coverage in counties added to the Standard Metropolitan Statistical Areas (SMSA's), which were redefined in 1973. In January 1985, a new State-based CPS sample was selected based on 1980 census information. A sample reduction of about 4,000 households was implemented in April 1988; they were reinstated during the 8-month period, April-November 1989. A redesigned CPS sample based on the 1990 decennial census was selected for use during the 1990's. Households from this new sample were phased into the CPS between April 1994 and July 1995. The July 1995 sample was the first monthly sample based entirely on the 1990 census. For further information on the 1990 sample redesign, see "Redesign of the Sample for the Current Population Survey" in the May 1994 issue of this publication.

The original 1990 census-based sample design included about 66,000 housing units per month located in 792 selected geographic areas called primary sampling units (PSU's). The sample was initially selected to meet specific reliability criteria for the Nation, for each of the 50 States and the District of Columbia, and for the sub-State areas of New York City and the Los Angeles-Long Beach metropolitan area. In 1996, the original sample design reliability criteria were modified to reduce costs. The current criteria, given below, are based on the coefficient of variation (CV) of the unemployment level, where the CV is defined as the standard error of the estimate divided by the estimate, expressed as a percentage. These CV controls assume a 6-percent unemployment rate to establish a consistent specification of sampling error.

The current sample design, introduced in January 1996, includes about 59,000 households from 754 sample areas and maintains a 1.9-percent CV on national monthly estimates of unemployment level. This translates into a change of 0.2 percentage point in the unemployment rate being significant at a 90-percent confidence level. For each of the fifty States and for the District of Columbia, the design maintains a CV of at most 8-percent on the annual average estimate of unemployment level, assuming a 6-percent unemployment rate. Due to the national reliability criterion, estimates for several large States are substantially more reliable than the State design criterion requires. Annual average unemployment

estimates for California, Florida, New York, and Texas, for example, carry a CV of less than 4-percent.

In the first stage of sampling, the 754 sample areas are chosen. In the second stage, ultimate sampling unit clusters composed of about four housing units each are selected. Each month, about 59,000 housing units are assigned for data collection, of which about 50,000 are occupied and thus eligible for interview. The remainder are units found to be destroyed, vacant, converted to nonresidential use, containing persons whose usual place of residence is elsewhere, or ineligible for other reasons. Of the 50,000 housing units, about 6.5 percent are not interviewed in a given month due to temporary absence (vacation, etc.), other failures to make contact after repeated attempts, inability of persons contacted to respond, unavailability for other reasons, and refusals to cooperate (about half of the noninterviews). Information is obtained each month for about 94,000 persons 16 years of age or older.

Selection of sample areas. The entire area of the United States, consisting of 3,141 counties and independent cities, is divided into 2,007 sample units (PSU's). In most States, a PSU consists of a county or a number of contiguous counties. In New England and Hawaii, minor civil divisions are used instead of counties. Metropolitan areas within a State are used as a basis for forming PSU's. Outside of metropolitan areas, counties normally are combined except when the geographic area of an individual county is too large. Combining counties to form PSU's provides greater heterogeneity; a typical PSU includes urban and rural residents of both high and low economic levels and encompasses, to the extent feasible, diverse occupations and industries. Another important consideration is that the PSU be sufficiently compact so that, with a small sample spread throughout, it can be efficiently canvassed without undue travel cost.

The 2,007 PSU's are grouped into strata within each State. Then one PSU is selected from each stratum with the probability of selection proportional to the population of the PSU. Nationally, there are a total of 428 PSU's in strata by themselves. These strata are self-representing and are generally the most populous PSU's in each State. The 326 remaining strata are formed by combining PSU's that are similar in such characteristics as unemployment, proportion of housing units with three or more persons, number of persons employed in various industries, and average monthly wages for various industries. The single PSU randomly selected from each of these strata is non-self-representing because it represents not only itself but the entire stratum. The probability of selecting a particular PSU in a non-self-representing stratum is proportional to its 1990 population. For example, within a stratum, the chance that a PSU with a population of 50,000 would be selected for the sample is twice that for a PSU having a population of 25,000.

Selection of sample households. Because the sample design is State based, the sampling ratio differs by State and depends on State population size as well as both national and State reliability requirements. The State sampling ratios range roughly from 1 in every 100 households to 1 in every 3,000 households. The sampling ratio occasionally is modified slightly to hold the size of the sample relatively constant given the overall growth of the population. The sampling ratio used within a sample PSU depends on the probability of selection of the PSU and the sampling ratio for the State. In a sample PSU with a probability of selection of 1 in 10 and a State sampling ratio of 3,000, a within-PSU sampling ratio of 1 in 300 achieves the desired ratio of 1 in 3,000 for the stratum.

The 1990 within-PSU sample design was developed using block-level data from the 1990 census. (The 1990 census was the first decennial census that produced data at the block level for the entire country.) Normally, census blocks are bounded by streets and other prominent physical features such as rivers or railroad tracks. County, Minor Civil Division, and census place limits also serve as block boundaries. In cities, blocks can be bounded by four streets and be quite small in land area. In rural areas, blocks can be several square miles in size.

For the purpose of sample selection, census blocks were grouped into three strata: Unit, group quarters, and area. (Occasionally, units within a block were split between the unit and group quarters strata.) The unit stratum contained regular housing units with addresses that were easy to locate (e.g. most single family homes, townhouses, condominiums, apartment units, and mobil homes). The group quarters stratum contained housing units where residents shared common facilities or received formal or authorized care or custody. Unit and group quarters blocks exist primarily in urban areas. The area stratum contains blocks with addresses that are more difficult to locate. Area blocks exist primarily in rural areas.

To reduce the variability of the survey estimates and to ensure that the within-PSU sample would reflect the demographic and socioeconomic characteristics of the PSU, blocks within the unit, group quarters, and area strata were sorted using geographic and block-level data from the census. Examples of the census variables used for sorting include proportion of minority renter-occupied housing units, proportion of housing units with female householders, and proportion of owner-occupied housing units. The specific sorting variables used differed by type of PSU (urban or rural) and stratum.

Within each block, housing units were sorted geographically and grouped into clusters of approximately four units. A systematic sample of these clusters was then selected independently from each stratum using the appropriate within-PSU sampling ratio. The geographic clustering of the sample units reduces field representative travel costs. Prior to interviewing, special listing procedures are used to locate the particular sample addresses in the group quarters and area blocks.

Units in the three strata described above all existed at the time of the 1990 decennial census. Through a series of additional procedures, a sample of building permits is included in the CPS to represent housing units built after the decennial census. Adding these newly built units keeps the sample up-to-date and representative of the population. It also helps to keep the sample size stable: over the life of the sample, the addition of newly built housing units compensates for the loss of "old" units which may be abandoned, demolished, or converted to nonresidential use.

Rotation of sample. Part of the sample is changed each month. Each monthly sample is divided into eight representative subsamples or rotation groups. A given rotation group is interviewed for a total of 8 months, divided into two equal periods. It is in the sample for 4 consecutive months, leaves the sample during the following 8 months, and then returns for another 4 consecutive months. In each monthly sample, one of the eight rotation groups is in the first month of enumeration, another rotation group is in the second month, and so on. Under this system, 75 percent of the sample is common from month to month and 50 percent from year to year for the same month. This procedure provides a substantial amount of month-to-month and year-to-year overlap in the sample, thus providing better estimates of change and reducing discontinuities in the series of data without burdening any specific group of households with an unduly long period of inquiry.

CPS Sample, 1947 to present. Table 1-A provides a description of some aspects of the CPS sample designs in use since 1947. A more detailed account of the history of the CPS sample design appears in *The Current Population Survey: Design and Methodology*, Technical Paper No. 40, Bureau of the Census, or *Concepts and Methods Used in Labor Force Statistics Derived from the Current Population Survey*, Report 463, Bureau of Labor Statistics. A description of the 1990 census-based sample design appears in "Redesign of the Sample for the Current Population Survey," in the May 1994 issue of this publication.

Table 1-A. Characteristics of the CPS sample, 1947 to present

Time period	Number of sample areas	Households eligible	
		Interviewed	Not interviewe
Aug. 1947 to Jan. 1954	68	21,000	500-1,000
Feb. 1954 to Apr. 1956	230	21,000	500-1,000
May 1956 to Dec. 1959	1/330	33,500	1,500
Jan. 1960 to Feb. 1963	2/333	33,500	1,500
Mar. 1963 to Dec. 1966	357	33,500	1,500
Jan. 1967 to July 1971	449	48,000	2,000
Aug. 1971 to July 1972	449	45,000	2,000
Aug. 1972 to Dec. 1977	461	45,000	2,000
Jan. 1978 to Dec. 1979	614	53,500	2,500
Jan. 1980 to Apr. 1981	629	62,200	2,800
May 1981 to Dec. 1984	629	57,800	2,500
Jan. 1985 to Mar. 1988	729	57,000	2,500
Apr. 1988 to Mar. 1989.	729	53,200	2,600
Apr. 1989 to Sept. 1994 3/....	729	57,400	2,600
Oct. 1994 to Aug. 1995 4/.....	792	54,500	3,500
Sept. 1995 to Dec. 1995.....	792	52,900	3,400
Jan. 1996 to present.....	754	46,800	3,200

1 Beginning in May 1956, these areas were chosen to provide coverage in each State a Columbia.

2 Three sample areas were added in 1960 to represent Alaska and Hawaii after stateho

3 The sample was increased incrementally during the 8-month period, April-November 1

4 Includes 2,000 additional assigned housing units for Georgia and Virginia that wer during the 10-month period, October 1994-August 1995.

Interviewer's Manual, Pt. A1E, Description of the Sample



Basic Monthly Survey Methodology and Documentation Page



CPS Main Page

Source: CPS Main

Author: Greg Weyland-Census/DSD/CPSB

Contact: (cpshelp@info.census.gov) CPS Help-Census/DSD/CPSB

Last revised: August 01, 1997

URL: <http://www.bls.census.gov/cps/bsampdes.htm>



Weighting

For all CPS data files a final weight is prepared and used to compute the monthly labor force status estimates. An additional weight is prepared for the earnings universe which roughly corresponds to wage and salary workers in the two outgoing rotations. This is explained below in the section on earnings data. In this section we briefly describe the construction and use of these weights. Chapter 5 of Technical paper 40, *The Current Population Survey: Design and Methodology* provides documentation of the weighting procedures for the CPS.

The final weight, which is the product of several adjustments, is used to produce estimates for the various characteristics covered in the full monthly CPS. This weight is constructed from the basic weight for each person, which represents the probability of selection for the survey. The basic weight is adjusted for special sampling situations and failure to obtain interviews from eligible households (noninterview adjustment). A two-stage ratio estimation procedure adjusts the sample population to the known distribution of the entire population. This two-stage ratio estimation process produces factors which are applied to the basic weight (after the special weighting and noninterview adjustments are made) and results in the final weight associated with each record. In summary, the final weight is the product of: (1) the basic weight, (2) adjustments for special weighting, (3) noninterview adjustment, (4) first stage ratio adjustment factor, and (5) second stage ratio adjustment factor. This final weight should be used when producing estimates from the basic CPS data.

Earnings Data

Beginning in January, 1992 usual hourly and weekly earnings data began appearing on the monthly CPS File for that portion of the population roughly corresponding to wage and salary workers (self-employed persons in incorporated businesses are excluded, although they are normally included with the wage and salary population). These data are collected on a monthly basis in the two outgoing rotation groups as part of the basic CPS labor force interview.

Since the intent of the regular collection of earnings data was to initiate a family earnings data series, all persons in the two outgoing rotations receive an "earnings weight," even if they are not eligible for the earnings item. The earnings weight is based on a simple ratio estimator using the person's labor force status by age, race, sex and class of worker. When tabulating estimates of earnings based on basic CPS data, use the earnings weight.

Further information on this earnings series is contained in *Technical Description of the Quarterly Data on Weekly Earnings from the Current Population Survey*, BLS Report #601, July 1980. This report is available on request from the U.S. Department of Labor, Bureau of Labor Statistics, Washington, D.C. 20212. Attn: Office of Inquiries and Correspondence.
