

United States 2000: Survey Information

The information below is taken from the web site of the Current Population Survey (<http://www.bls.census.gov/cps/cpsmain.htm>).

Table of contents:

- A. General Characteristics
- B. Population, sample size and sampling methods
- C. Data collection and acquisition
- D. Definition of the survey units
- E. Contents
- F. Quality of data

A. General characteristics

Official name of the survey/data source:

Current Population Survey (CPS)

Administrative Unit responsible for the survey:

Joint project between the Bureau of Labor Statistics and the Bureau of the Census
e-mail: cpshelp@info.census.gov

Basic Monthly Survey

The Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey has been conducted for more than 50 years. The CPS is the primary source of information on the labor force characteristics of the U.S. population. The sample is scientifically selected to represent the civilian non-institutional population. Respondents are interviewed to obtain information about the employment status of each member of the household 15 years of age and older. However, published data focus on those ages 16 and over. The sample provides estimates for the nation as a whole and serves as part of model-based estimates for individual states and other geographic areas. Estimates obtained from the CPS include employment, unemployment, earnings, hours of work, and other indicators. They are available by a variety of demographic characteristics including age, sex, race, marital status, and educational attainment. They are also available by occupation, industry, and class of worker. Supplemental questions to produce estimates on a variety of topics including school enrolment, income, previous work experience, health,

employee benefits, and work schedules are also often added to the regular CPS questionnaire.

CPS data are used by government policymakers and legislators as important indicators of our nation's economic situation and for planning and evaluating many government programs. They are also used by the press, students, academics, and the general public.

Annual Demographic Survey or March CPS supplement

The Annual Demographic Survey or March CPS supplement is the primary source of detailed information on income and work experience in the United States. Numerous publications based on this survey are issued each year by the Bureau of Labor Statistics and Census. A public-use microdata file is available for private researchers, who also produce many academic and policy-related documents based on these data.

The Annual Demographic Survey is used to generate the annual Population Profile of the United States, reports on geographical mobility and educational attainment, and detailed analysis of money income and poverty status. The labor force and work experience data from this survey are used to profile the U.S. labor market and to make employment projections. To allow for the same type of in-depth analysis of Hispanics, additional Hispanic sample units are added to the basic CPS sample in March each year. Additional weighting is also performed so that estimates can be made for households and families, in addition to persons.

LIS used data from the March Supplement for the LIS files (us00hh, us00pp and us00kd) and for the integrated LES file (us00ll), and data from the Basic Monthly for the true LES file (us00lfs). Hereafter, the March Supplement is described in further detail.

B. Population, sampling size and sampling methods

Coverage:

Population coverage includes the civilian population of the United States plus approximately 820,000 members of the Armed Forces in the United States living off post or with their families on post but excludes all other members of the Armed Forces. This file excludes inmates of institutions. The labor force and work experience data are not collected for Armed Forces members.

Sample size:

About 50,000 households.

Sampling design:

The sample for the March CPS consists of the basic CPS sample and an additional sample of Hispanic households. The basic CPS sample is selected from multiple frames

using multiple stages of selection. Each unit is selected with a known probability to represent similar units in the universe. The sample design is a state-based design with the sample in each state being independent of the others. The sample size is sufficient to produce a 1.9 percent coefficient of variation (CV) on the unemployment level for the nation, assuming a 6 percent unemployment rate. Within each state, the sample was designed to meet reliability requirements of 8 percent CVs or lower on the annual average unemployment level, given a 6 percent unemployment rate.

Before sampling begins, primary sampling units (PSUs) are defined and grouped into strata. PSUs consist of one or more contiguous counties. The largest PSUs, as of the 1990 census, are designated as self-representing PSUs. The remaining PSUs are grouped to form non-self-representing strata of approximately equal size within the state using a clustering algorithm. During the first stage of sample selection, one PSU is selected from each non-self-representing strata based on population size. There are currently 428 self-representing and 326 non-self-representing PSUs in the CPS sample.

Within a selected PSU, multiple sampling frames are used in the next stage of selection. The determination of which frames should be used is based on the type of enumeration during the last decennial census and the availability of housing permits. If complete addresses are available for a particular block from the decennial address files and the block is in an area requiring building permits, the sample is selected from the decennial files and from a frame of addresses created by sampling building permits issued for post-April 1990 construction. Otherwise, a sample of areas (one or more contiguous blocks) is selected; a field representative of the Census Bureau lists all the housing units in the selected areas; and a sample is selected from the listed units.

At this stage of sampling, ultimate sampling units (USUs) are selected using systematic sampling. USUs are clusters of four housing units. The clusters from the decennial address lists are usually geographically compact. The clusters from the area frame, while in expectation should consist of four units, may contain more or less if there has been major building or demolition within the area since the census. (Group quarters are sampled separately from housing units on the decennial census list.)

Multiple samples are selected all at once, each sample being a 'neighbor' or geographically close to the previous sample. Each sample is divided into 8 sub-samples (referred to as panels or rotation groups.) Each month a new rotation group is introduced for its first month of interviewing; the rotation group which had its eighth interview in the preceding month is retired permanently. Thus, over the decade until the results of the next census are available, the specific sample of units changes, but their general characteristics do not. Each rotation group is interviewed for 4 consecutive months, out of sample for the next 8 months, and then interviewed for 4 months. This 4-8-4 rotation pattern implies a 75 percent overlap month-to-month and a 50 percent overlap year-to-year.

The additional Hispanic sample is from the previous November's basic CPS sample. If a person is identified as being of Hispanic origin from the November interview and is still residing at the same address in March, that housing unit is eligible for the March survey.

This amounts to a near doubling of the Hispanic sample since there is no overlap of housing units between the basic CPS samples in November and March.

C. Data collection and acquisition

As in every other month, Bureau of the Census field representatives attempt to collect data from the sample units during the week of the 19th. For the first month-in-sample interview, the interviewer visits the sample address to determine if the sample unit exists, if it is occupied, and if some responsible adult will provide the necessary information. If someone at the sample unit agrees to the interview, the interviewer uses a laptop computer to administer the interview. A typical field representative's workload consists of about 37 sample addresses, some of which may be nonexistent or vacant.

More than 90 percent of month-in-sample 2 through 4 and 6 through 8 interviews are conducted by telephone. Interviewers continue to visit households without telephones, with poor English-language skills, or which decline a telephone interview. Approximately 15 percent of the sample each month is interviewed by telephone from one of three centralized telephone interviewing centers: Hagerstown, Maryland; Jeffersonville, Indiana; and Tucson, Arizona. Interviewing at these centers begins on Sunday afternoon of interviewing week and ends on Wednesday. Cases which have not been contacted at this point are transferred to the field staff, either for telephone or personal visit interviews. Field interviewers have until Tuesday of the following week to complete all their assigned cases.

One of the most difficult tasks an interviewer faces is obtaining respondent cooperation. CPS is a voluntary survey, and persuading residents to take the time to answer questions completely and accurately as possible is sometimes a difficult challenge. In March of recent years, between 92 and 93 percent of the eligible households have provided basic labor force information, while only between 80 and 82 percent of the eligible households have completed the Annual Demographic Survey supplement.

In households which have not previously been interviewed, the interview begins with questions about the housing unit and the demographic characteristics of the people who consider this address their usual residence. Otherwise, after updating the household roster for any changes, the interview begins with the labor force portion of the interview. Labor force information is collected for each civilian 15 years of age or older, although the data for 15-year olds are not used in official BLS estimates.

After the labor force information has been collected for all eligible household members, the Annual Demographic Survey questions are asked. Typically, a single respondent reports for everyone in the household, although telephone call-backs to obtain particular items of information known only by someone else in the household is fairly common.

The specific questions to be asked appear on the computer screen, and the interviewer has been trained to ask each question exactly as it is worded. Based upon the response

entered by the interviewer, the computerized questionnaire (or data collection instrument) determines the next question to be asked. While some questions require simple "yes" or "no" responses, others require the interviewer to categorize a response into a set of predetermined categories. Occasionally, interviewers type in answers directly, rather than selecting predetermined categories, although, with the exception of the responses to questions about the name of an employer, job duties, and occupation, the interviewer rarely has to type a lengthy response.

In continuing households, information collected in previous interviews is used to shorten portions of the interview and reduce spurious month-to-month changes. For example, employed persons who verify they are still working for the same employer are asked if their job duties have changed and if the description of their job last month is accurate, rather than having this information re-asked each month.

D. Definition of the survey units

Household - A household consists of all the persons who occupy a house, an apartment, or other group of rooms, or a room, which constitutes a housing unit. A group of rooms or a single room is regarded as a housing unit when it is occupied as separate living quarters; that is, when the occupants do not live and eat with any other person in the structure, and when there is direct access from the outside or through a common hall. The count of households excludes persons living in group quarters, such as rooming houses, military barracks, and institutions. Inmates of institutions (mental hospitals, rest homes, correctional institutions, etc.) are not included in the survey.

Family - A family is a group of two persons or more (one of whom is the householder) residing together and related by birth, marriage, or adoption. All such persons (including related subfamily members) are considered as members of one family. Beginning with the 1980 CPS, unrelated subfamilies (referred to in the past as secondary families) are no longer included in the count of families, nor are the members of unrelated subfamilies included in the count of family members.

Family Household - A family household is a household maintained by a family (as defined above), and may include among the household members any unrelated persons (unrelated subfamily members and/or unrelated individuals) who may be residing there. The number of family households is equal to the number of families. The count of family household members differs from the count of family members, however, in that the family household members include all persons living in the household, whereas family members include only the householder and his/her relatives.

Householder - The householder refers to the person (or one of the persons) in whose name the housing unit is owned or rented (maintained) or, if there is no such person, any adult member, excluding roomers, boarders, or paid employees. If the house is owned or rented jointly by a married couple, the householder may be either the husband or the wife. The person designated as the householder on the file is the "reference person" on

the CPS-260 control card to whom the relationship of all other household members, if any, is recorded.

Householder With No Other Relatives in Household - A householder who has no relatives living in the household. This is the entry for a person living alone. Another example is the designated householder of an apartment shared by two or more unrelated individuals.

Householder With Other Relatives (Including Spouse) in Household - The person designated as householder if he/she has one or more relatives (including spouse) living in the household.

Nonfamily Householder - A nonfamily householder (formerly called a primary individual) is a person maintaining a household while living alone or with nonrelatives only.

Other Relative of Householder - Any relative of the householder other than his spouse, child (including natural, adopted, foster, or step child), sibling, or parent; for example, grandson, daughter-in-law, etc.

Nonrelative of Householder With No Own Relatives in Household - A nonrelative of the householder who has no relative(s) of his own in the household. This category includes such nonrelatives as a ward, a lodger, a servant, or a hired hand, who has no relatives of his own living with him in the household.

Nonrelative of Householder With Own Relatives (Including Spouse) in Household - Any household member who is not related to the householder but has relatives of his own in the household; for example, a lodger, his spouse, and their son.

Own Child - A child related by birth, marriage, or adoption to the family householder.

Related Children - Related children in a family include own children and all other children in the household who are related to the householder by birth, marriage, or adoption. For each type of family unit identified in the CPS, the count of own children under 18 years old is limited to single (never married) children; however, "own children under 25" and "own children of any age," include all children regardless of marital status. The totals include never-married children living away from home in college dormitories.

Related Subfamily - A related subfamily is a married couple with or without children, or one parent with one or more own single (never married) children under 18 years old, living in a household and related to, but not including, the householder or spouse. The most common example of a related subfamily is a young married couple sharing the home of the husband's or wife's parents. The number of related subfamilies is not included in the number of families.

Unrelated Individuals - Unrelated individuals are persons of any age (other than inmates of institutions) who are not living with any relatives. An unrelated individual may be (1)

a nonfamily householder living alone or with nonrelatives only, (2) a roomer, boarder, or resident employee with no relatives in the household, or (3) a group quarters member who has no relatives living with him/her. Thus, a widow who occupies her house alone or with one or more other persons not related to her, a roomer not related to anyone else in the housing unit, a maid living as a member of her employer's household but with no relatives in the household, and a resident staff member in a hospital living apart from any relatives are all examples of unrelated individuals.

Unrelated Subfamily - An unrelated subfamily is a family that does not include among its members the householder and relatives of the householder. Members of unrelated subfamilies may include persons such as guests, roomers, boarders, or resident employees and their relatives living in a household. The number of unrelated subfamily members is included in the number of household members but is not included in the count of family members. Persons living with relatives in group quarters were formerly considered as members of families. However, the number of such unrelated subfamilies is so small that persons in these unrelated subfamilies are included in the count of secondary individuals.

E. Contents

Income

For each person in the sample who is 15 years old and over, questions are asked on the amount of money income received in the preceding calendar year from each of the following sources: (1) money wages or salary; (2) net income from nonfarm self-employment; (3) net income from farm self-employment; (4) Social Security or railroad retirement; (5) Supplemental Security Income; (6) public assistance or welfare payments; (7) interest (on savings or bonds); (8) dividends, income from estates or trusts, or net rental income; (9) veterans' payment or unemployment and workmen's compensation; (10) private pensions or government employee pensions; (11) alimony or child support, regular contributions from persons not living in the household, and other periodic income.

Although income statistics refer to receipts during the preceding year, the characteristics of the person such as age, labor force status, etc., and the composition of households refer to the time of the survey. The income of the household does not include amounts received by persons who are members of the household during all or part of the income year if these persons no longer reside with the household at the time of enumeration. On the other hand, household income includes amounts reported by persons who did not reside with the household during the income year but who were members of the household at the time of enumeration.

Data on consumer income collected in the CPS by the Bureau of the Census cover money income received (exclusive of certain money receipts such as capital gains) before payments for personal income taxes, Social Security, union dues, Medicare deductions, etc. Also, money income does not reflect the fact that some households receive part of

their income in the form of nonmoney transfers such as food stamps, health benefits, subsidized housing, and energy assistance; that many farm households receive nonmoney income in the form of rent free housing and goods produced and consumed on the farm; or that nonmoney income is received by some nonfarm residents that often takes the form of the use of business transportation and facilities, or full or partial contributions for retirement programs, medical and educational expenses, etc. These elements should be considered when comparing income levels. Moreover, readers should be aware that for many different reasons there is a tendency in household surveys for respondents to under report their income. From an analysis of independently derived income estimates, it has been determined that wages and salaries tend to be much better reported than such income types as public assistance, Social Security, and net income from interest, dividends, rents, etc.

Labor Force

Persons are classified as in the labor force if they are employed, unemployed, or in the Armed Forces during the survey week. The "civilian labor force" includes all civilians classified as employed or unemployed. The file includes labor force data for civilians age 15 and over. However, the official definition of the civilian labor force is age 16 and over.

1. Employed - Employed persons comprise (1) all civilians who, during the survey week did any work at all as paid employees or in their own business or profession, or on their own farm, or who work 15 hours or more as unpaid workers on a farm or a business operated by a member of the family; and (2) all those who have jobs but who are not working because of illness, bad weather, vacation, or labor-management dispute, or because they are taking time off for personal reasons, whether or not they are seeking other jobs. These persons would have an Labor Force Status Recode (LFSR) of 1 or 2 respectively in character 145 of the person record which designates "at work" and "with a job, but not at work." Each employed person is counted only once. Those persons who held more than one job are counted in the job at which they worked the greatest number of hours during the survey week. If they worked an equal number of hours at more than one job, they are counted at the job they held the longest.

2. Unemployed - Unemployed persons are those civilians who, during the survey week, have no employment but are available for work, and (1) have engaged in any specific job seeking activity within the past 4 weeks such as registering at a public or private employment office, meeting with prospective employers, checking with friends or relatives, placing or answering advertisements, writing letters of application, or being on a union or professional register; (2) are waiting to be called back to a job from which they had been laid off; or (3) are waiting to report to a new wage or salary job within 30 days. These persons would have an LFSR code of 3 or 4 in character 145 of the person record. The unemployed includes job leavers (persons who quit or otherwise terminate their employment voluntarily and immediately begin looking for work), job losers (persons whose employment ends involuntarily, who immediately begin looking for work, and those persons who are already /on layoff), new job entrants (persons who never worked at

a full-time job lasting two weeks or longer), and job re-entrants (persons who previously worked at a full-time job lasting two weeks or longer but are out of the labor force prior to beginning to look for work).

3. *Not in Labor Force* - All civilians 15 years old and over who are not classified as employed or unemployed. These persons are further classified as major activity: keeping house, going to school, unable to work because of long-term physical or mental illness, and other. The "other" group includes, for the most part, retired persons. Persons who report doing unpaid work in a family farm or business for less than 15 hours are also classified as not in the labor force. For persons not in the labor force, data on previous work experience, intentions to seek work again, desire for a job at the time of interview, and reasons for not looking for work are asked only in those households that are in the fourth and eighth months of the sample, i.e., the "outgoing" groups, those which had been in the sample for three previous months and would not be in for the subsequent month.

Finally, it should be noted that the unemployment rate represents the number of persons unemployed as a percent of the civilian labor force 16 years old and over. This measure can also be computed for groups within the labor force classified by sex, age, marital status, race, etc. The job loser, job leaver, reentrant, and new entrant rates are each calculated as a percent of the civilian labor force 16 years old and over; the sum of the rates for the four groups thus equals the total unemployment rate.

Layoff - A person who is unemployed but expects to be called back to a specific job. If he/she expects to be called back within 30 days, it is considered a temporary layoff; otherwise, it is an indefinite layoff.

Geographic Concepts

Geographic Division - An area composed of contiguous States, with Alaska and Hawaii also included in one of the divisions. (A State is one of the 51 major political units in the United States.) The nine geographic divisions have been largely unchanged for the presentation of summary statistics since the 1910 census.

Regions - There are four regions: Northeast, Midwest (formerly North Central),¹ West, and South. States and divisions within regions are presented below.

NORTHEAST REGION

New England Division

Connecticut
Maine
Massachusetts
New Hampshire
Rhode Island
Vermont

Middle Atlantic Division

New Jersey
New York
Pennsylvania

MIDWEST REGION

East North Central Division

Illinois
 Indiana
 Michigan
 Ohio
 Wisconsin
 North Dakota
 South Dakota

West North Central Division

Iowa
 Kansas
 Minnesota
 Missouri
 Nebraska

 1. The Midwest Region was designated as the North Central Region until June 1964.

WEST REGION**Mountain Division**

Arizona
 Colorado
 Idaho
 Montana
 Nevada
 Utah
 Wyoming
 New Mexico

Pacific Division

Alaska
 California
 Hawaii
 Oregon
 Washington

SOUTH REGION**East South Central Division**

Alabama
 Kentucky
 Mississippi
 Tennessee

West South Central Division

Arkansas
 Louisiana
 Oklahoma
 Texas

South Atlantic Division

Delaware
 District of Columbia
 Florida
 Georgia
 Maryland
 North Carolina
 South Carolina
 Virginia
 West Virginia

F. Quality of dataReviewing and revising the data

While some edits are built into the data collection instrument, others are performed after the case has been transferred from the interviewer to the central computer. Imputations are made for missing and inconsistent data items. The textual responses to questions on industry and occupation are transmitted to Jeffersonville, Indiana, where staff perform a computerized coding operation and transmit the resulting codes back to the central computer. In addition, the responses to several questions may be used to create new variables. For example, within the data collection instrument, labor force status

(employed, unemployed, or not in the labor force) is derived from responses to several questions. During post-interview processing, employed persons are classified as being full or part time. The final step in preparing the database is the assignment of a series of weights, both to households and each person. The final CPS weight is used to produce basic demographic and labor force estimates. The weights are a function of the probability of selection and reflect any needed subsampling, nonresponse adjustment, and post-stratified ratio adjustment to population control totals. Further editing, imputation, coding, and weighting are performed if there is a supplement. However, none of this additional processing will affect the information collected or derived from the labor force portion of the CPS interview.

Imputation of Unreported Data Items

When a response is not obtained for a particular data item, or an inconsistency in reported items is detected, an "imputed" response is entered in the field. Imputation is performed using a "hot deck" method, whereby a response from another sample person with similar demographic and economic characteristics is used for the nonresponse. The imputation procedure is performed one item at a time. In a typical month, the imputation rate for demographic items is less than 1 percent. The rates for labor force items and school enrollment are slightly over 1 percent. Over all earnings items, the imputation rate is near 10 percent with some items having much higher and others much lower nonresponse rates.

Weighting

For all CPS data files a single weight is prepared and used to compute the monthly labor force status estimates. An additional weight was 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. However, the difference in content of the March CPS supplement requires the presentation of additional weights: a household weight, a family weight, and a March supplement weight. 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 both with and without supplement questions.

The final weight, which is the product of several adjustments, is used to produce population estimates for the various items covered in the regular 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.

Differences in the questionnaire, sample and data uses for the March CPS supplement result in the need for additional adjustment procedures to produce what is called the March supplement weight. The sample for the March CPS supplement is expanded to include members of the Armed Forces who are living in civilian housing or with the family on a military base, as well as additional Hispanic households which are not included in the monthly labor force estimates.

The expanded sample and the need to have a husband and wife receive the same weight has resulted in a weighting system which produces the March supplement weight. The March supplement weight should be used for producing estimates from March supplement data.

Finally, household and family weights are the weights assigned from the householder or reference person after all adjustments have been made and should be used when tabulating estimates of families-households.

Earnings Data - Beginning in 1982, usual hourly and weekly earnings data appear on the Annual Demographic File (ADF) 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 now 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 a simple ratio-estimation to the person's labor force status by age, race, and sex. 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.

Accuracy of results

A sample survey estimate has two types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error. The nature of the sampling error is known given the survey design. The full extent of the nonsampling error, however, is unknown.

Sampling error - Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and

enumerators. This possible variation in the estimates due to sampling error is known as "sampling variability."

Nonsampling error - All other sources of error in the survey estimates are collectively called nonsampling error. Sources of nonsampling error include the following:

- Inability to get information about all sample cases.
- Definitional difficulties.
- Differences in interpretation of questions.
- Respondent inability or unwillingness to provide correct information.
- Respondent inability to recall information.
- Errors made in data collection, such as recording and coding data.
- Errors made in processing the data.
- Errors made in estimating values for missing data.
- Failure to represent all units with the sample (undercoverage).

Two types of nonsampling error that can be examined to a limited extent are nonresponse and coverage.

Nonresponse - The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the March 2001 basic CPS, the nonresponse rate was 8.03%. The nonresponse rate for the supplement was an additional 8.5%, for a total supplement nonresponse rate of 15.85%.

Coverage - The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for sample "covers" the survey's target population. CPS undercoverage results from missed housing units and missed people within sample households. Overall CPS undercoverage is estimated to be about 8 percent. CPS undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks and other races combined than for Whites. As described previously, ratio estimation to independent age-sex-race-Hispanic population controls partially corrects for the bias due to undercoverage. However, biases exist in the estimates to the extent that missed people in missed households or missed people in interviewed households have different characteristics from those of interviewed people in the same age-sex-race-ancestry-state group.

A common measure of survey coverage is the coverage ratio, the estimated population before post-stratification divided by the independent population control. Table 1 shows CPS coverage ratios for age-sex-race groups for a typical month. The CPS coverage ratios can exhibit some variability from month to month. Other Census Bureau household surveys experience similar coverage.

Table 1. CPS Coverage Ratios			
	Non-Black	Black	All People

Age	M	F	M	F	M	F	Total
0-14	0.929	0.964	0.850	0.838	0.916	0.943	0.929
15	0.933	0.895	0.763	0.824	0.905	0.883	0.895
16-19	0.881	0.891	0.711	0.802	0.855	0.877	0.866
20-29	0.847	0.897	0.660	0.811	0.823	0.884	0.854
30-39	0.904	0.931	0.680	0.845	0.877	0.920	0.899
40-49	0.928	0.966	0.816	0.911	0.917	0.959	0.938
50-59	0.953	0.974	0.896	0.927	0.948	0.969	0.959
60-64	0.961	0.941	0.954	0.953	0.960	0.942	0.950
65-69	0.919	0.972	0.982	0.984	0.924	0.973	0.951
70+	0.993	1.004	0.996	0.979	0.993	1.002	0.998
15+	0.914	0.945	0.767	0.874	0.898	0.927	0.918
0+	0.918	0.949	0.793	0.864	0.902	0.931	0.921

Comparability of data - Data obtained from the CPS and other sources are not entirely comparable. This results from differences in interviewer training and experience and in differing survey processes. This is an example of nonsampling variability not reflected in the standard errors. Therefore, caution should be used when comparing results from different sources.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The questionnaire was redesigned to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The March supplemental income questions were also modified for adaptation to computer-assisted interviewing, although there were no changes in definitions and concepts. See Appendix C of Report P - 60 No. 188 on "Conversion to a Computer Assisted Questionnaire" for a description of these changes and the effect they had on the data. Due to these and other changes, one should use caution when comparing estimates from data collected before 1994 with estimates from data collected in 1994 and later.

Caution should also be used when comparing data from this microdata file, which reflects 1990 census-based population controls, with microdata files from March 1993 and earlier

years, which reflect 1980 census-based population controls. Although this change in population controls had relatively little impact on summary measures such as averages, medians, and percentage distributions, it did have a significant impact on levels. For example, use of 1990 based population controls results in about a one percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1994 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

Caution should also be used when comparing Hispanic estimates over time. No independent population control totals for people of Hispanic ancestry were used before 1985.

Based on the results of each decennial census, the Census Bureau gradually introduces a new sample design for the CPS (For detailed information on the 1990 sample redesign, see the Department of Labor, Bureau of Labor Statistics report, *Employment and Earnings*, Volume 41 Number 5, May 1994). During this phase-in period, CPS data are collected from sample designs based on different censuses. While most CPS estimates were unaffected by this mixed sample, geographic estimates are subject to greater error and variability. Users should exercise caution when comparing estimates across years for metropolitan/ nonmetropolitan categories.

A Nonsampling error warning - Since the full extent of the nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between estimates. Even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures probably do not reveal useful information when computed on a base (subpopulation) smaller than 75,000.

For additional information on nonsampling error including the possible impact on CPS data when known, refer to:

- Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey*, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978
- Technical Paper 63, *The Current Population Survey: Design and Methodology*, Census Bureau, U.S. Department of Commerce, 2000.

Standard errors and their use - The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range that would include the average result of all possible samples with a known probability. For example, if all possible samples were surveyed under essentially the same general conditions and the same sample design, and if an estimate and its standard error were calculated from each sample, then approximately 90 percent of the intervals from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate would include the average result of all possible samples.

A particular confidence interval may or may not contain the average estimate derived from all possible samples. However, one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Standard errors may be used to perform hypothesis testing. This is a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that the population parameters are different. An example of this would be comparing the percentage of Whites with a college education to the percentage of Blacks with a college education.

Tests may be performed at various levels of significance. A significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. For example, to conclude that two parameters are different at the 0.10 level of significance, the absolute value of the estimated difference between characteristics must be greater than or equal to 1.645 times the standard error of the difference.

The Census Bureau uses 90 percent confidence intervals and 0.10 levels of significance to determine statistical validity. Consult standard statistical texts for alternative criteria.

Estimating standard errors - To estimate the standard error of a CPS estimate, the Census Bureau uses replicated variance estimation methods. These methods primarily measure the magnitude of sampling error. However, they do measure some effects of nonsampling error as well. They do not measure systematic biases in the data due to nonsampling error. (Bias is the average of the differences, over all possible samples, between the sample estimates and the true value.)

Generalized Variance Parameters - Consider all of the possible estimates of characteristics of the population that are of interest to data users. Now consider all of the subpopulations such as racial groups, age ranges, etc. Finally, consider every possible comparison or ratio combination. The list would be completely unmanageable. Similarly, a list of standard errors to go with every estimate would be unmanageable.

Through experimentation, we have found that certain groups of estimates have similar relationships between their variances and expected values. We provide a generalized method for calculating standard errors for any of the characteristics of the population of interest. The generalized method uses parameters for groups of estimates. These parameters are in Table 2, for basic CPS monthly labor force estimates, and Table 3, for March supplement data, including the Hispanic supplement.

Standard errors of estimated numbers - The approximate standard error, s_x , of an estimated number from this microdata file can be obtained using this formula:

$$s_x = \sqrt{ax^2 + bx}$$

Formula 1

Here x is the size of the estimate and a and b are the parameters in Table 2 or 3 associated with the particular type of characteristic. When calculating standard errors for numbers from cross-tabulations involving different characteristics, use the factor or set of parameters for the characteristic which will give the largest standard error.

For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages see "Explanatory Notes and Estimates of Error: Household Data" in *Employment and Earnings*, a monthly report published by the Bureau of Labor statistics.

Standard errors of estimated percentages - The reliability of an estimated percentage, computed using sample data for both numerator and denominator, depends on the size of the percentage and its base. Estimated percentages are relatively more reliable than the corresponding estimates of the numerators of the percentages, particularly if the percentages are 50 percent or more. When the numerator and denominator of the percentage are in different categories, use the factor or parameter from Table 2 or 3 indicated by the numerator.

The approximate standard error $s_{x,p}$ of an estimated percentage can be obtained by using the following formula:

$$s_{x,p} = \sqrt{(b/x) p(100-p)}$$

Formula 2

Here x is the total number of people, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \leq p \leq 100$) and b is the parameter in Table 2 or 3 associated with the characteristic in the numerator of the percentage.

Standard error of a difference - The standard error of the difference between two sample estimates is approximately equal to

$$s_{x-y} = \sqrt{s_x^2 + s_y^2}$$

Formula 3

where s_x and s_y are the standard errors of the estimates, x and y . The estimates can be numbers, percentages, ratios, etc. This will represent the actual standard error quite accurately for the difference between estimates of the same characteristic in two different areas, or for the difference between separate and uncorrelated characteristics in the same area. However, if there is a high positive (negative) correlation between the two characteristics, the formula will overestimate (underestimate) the true standard error.

For information on calculating standard errors for labor force data from the CPS which involve differences in consecutive quarterly or yearly averages, consecutive month-to-month differences in estimates, and consecutive year-to-year differences in monthly estimates see "Explanatory Notes and Estimates of Error: Household Data" in *Employment and Earnings*, a monthly report published by the Bureau of Labor Statistics.

Standard error of an average for grouped data - The formula used to estimate the standard error of an average for grouped data is

$$s_{\bar{x}} = \sqrt{\frac{b}{y} S^2}$$

Formula 4

In this formula, y is the size of the base of the distribution and b is a parameter from Table 2 or 3. The variance, S^2 , is given by the following formula:

$$S^2 = \sum_{i=1}^c p_i \bar{x}_i^2 - \bar{x}^2$$

Formula 5

where \bar{x} , the average of the distribution, is estimated by

$$\bar{x} = \sum_{i=1}^c p_i \bar{x}_i$$

Formula 6

where

c = the number of groups; i indicates a specific group, thus taking on values 1 through c .

p_i = estimated proportion of households, families or people whose values, for the characteristic (x -values) being considered, fall in group i .

$\bar{x}_i = (Z_{i-1} + Z_i)/2$ where Z_{i-1} and Z_i are the lower and upper interval boundaries, respectively, for group i . \bar{x}_i is assumed to be the most representative value for the characteristic for households, families, and unrelated individuals or people in group i . Group c is open-ended, i.e., no upper interval boundary exists. For this group the approximate average value is

$$\bar{x}_c = \frac{3}{2} Z_{c-1}$$

Formula 7

Standard error of a ratio - Certain estimates may be calculated as the ratio of two numbers. The standard error of a ratio, x/y , may be computed using

$$s_{x/y} = \frac{x}{y} \sqrt{\left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 - 2r \frac{s_x s_y}{xy}}$$

Formula 8

The standard error of the numerator, s_x , and that of the denominator, s_y , may be calculated using formulas described earlier. In Formula (8), r represents the correlation between the numerator and the denominator of the estimate.

For one type of ratio, the denominator is a count of families or households and the numerator is a count of people in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of r . An example of this type is the average number of children per family with children.

For all other types of ratios, r is assumed to be zero. If r is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of

the ratio. Examples of this type are: average number of children per family and the poverty rate.

Standard error of a median - The sampling variability of an estimated median depends on the form of the distribution and the size of the base. One can approximate the reliability of an estimated median by determining a confidence interval about it. (See *Standard errors and their use* for a general discussion of confidence intervals.)

Estimate the 68 percent confidence limits of a median based on sample data using the following procedure.

1. Determine, using Formula (2), the standard error of the estimate of 50 percent from the distribution.
2. Add to and subtract from 50 percent the standard error determined in step 1. These two numbers are the percentage limits corresponding to the 68 percent confidence about the estimated median.
3. Using the distribution of the characteristic, determine upper and lower limits of the 68 percent confidence interval by calculating values corresponding to the two points established in step 2.

Use the following formula to calculate the upper and lower limits.

$$X_{pN} = \frac{pN - N_1}{N_2 - N_1} (A_2 - A_1) + A_1$$

Formula 9

where

X_{pN} = estimated upper and lower bounds for the confidence interval ($0 \leq p \leq 1$). For purposes of calculating the confidence interval, p takes on the values determined in step 2. Note that X_{pN} estimates the median when $p = 0.50$.

N = for distribution of numbers: the total number of units (people, households, etc.) for the characteristic in the distribution.

= for distribution of percentages: the value 1.0.

p = the values obtained in step 2.

A_1, A_2 = the lower and upper bounds, respectively, of the interval containing X_{pN} .

$N_1, N_2 =$ for distribution of numbers: the estimated number of units (people, households, etc.) with values of the characteristic greater than or equal to A_1 and A_2 , respectively.

$=$ for distribution of percentages: the estimated percentage of units (people, households, etc.) having values of the characteristic greater than or equal to A_1 and A_2 , respectively.

4. Divide the difference between the two points determined in step 3 by two to obtain the standard error of the median.

Standard error of estimated per capita deficit - Certain average values in this report represent the per capita deficit for households of a certain class. The average per capita deficit is approximately equal to

$$x = \frac{hm}{p}$$

Formula 10

where

h = number of households in the class,

m = average deficit for households in the class,

p = number of people in households in the class, and

x = average per capita deficit of people in households in the class.

To approximate standard errors for these averages, use the formula

$$s_x = \frac{hm}{p} \sqrt{\left(\frac{s_m}{m}\right)^2 + \left(\frac{s_p}{p}\right)^2 + \left(\frac{s_h}{h}\right)^2 - 2r \left(\frac{s_p}{p}\right)\left(\frac{s_h}{h}\right)}$$

Formula 11

In Formula (11), r represents the correlation between p and h .

For one type of average, the class represents households containing a fixed number of people. For example, h could be the number of three-person households. In this case, there is an exact correlation between the number of people in households and the number of households. Therefore, $r = 1$ for such households.

For other types of averages, the class represents households of other demographic types, for example, households in distinct regions, households in which the householder is of a certain age group, and owner-occupied and tenant-occupied households. In this and other cases in which the correlation between p and h is not perfect, use 0.7 as an estimate of r .

Accuracy of state estimates - The redesign of the CPS following the 1980 census provided an opportunity to increase efficiency and accuracy of state data. All strata are now defined within state boundaries. The sample is allocated among the states to produce state and national estimates with the required accuracy while keeping total sample size to a minimum. Improved accuracy of state data was achieved with about the same sample size as in the 1970 design.

Since the CPS is designed to produce both state and national estimates, the proportion of the total population sampled and the sampling rates differ among the states. In general, the smaller the population of the state the larger the sampling proportion. For example, in Vermont approximately 1 in every 400 households was sampled each month. In New York the sample was about 1 in every 2,000 households. Nevertheless, the size of the sample in New York is four times larger than in Vermont because New York has a larger population.

Computation of standard errors for state estimates - Standard errors for a state may be obtained by computing national standard errors, using formulas described earlier, and multiplying these by the appropriate f factor from Table 4. An alternative method for computing standard errors for a state is to multiply the a and b parameters in Table 2 or 3 by f^2 and then use these adjusted parameters in the standard error formulas.

Computation of a factor for groups of states - The factor adjusting standard errors for a group of states may be obtained by computing a weighted sum of the squared factors for the individual states in the group and taking the square root of the result. Depending on the combination of states, the resulting figure can be an overestimate.

The squared factor for a group of n states is given by

$$f^2 = \frac{\sum_{i=1}^n \text{POP}_i \times f_i^2}{\sum_{i=1}^n \text{POP}_i}$$

Formula 12

where POP_i is the state population and f_i^2 is obtained from Table D. The 2001 civilian noninstitutionalized population from the CPS for each state is also given in Table D.

Computation of standard errors for data for combined years - Sometimes estimates for multiple years are combined to improve precision. For example, suppose \bar{x} is an average derived from n consecutive years' data,

$$\bar{x} = \sum_{i=1}^n x_i / n$$

i.e., where the x_i are the estimates for the individual years.

Use the formulas described previously to estimate the standard error, s_x , of each year's estimate. Then the standard error of \bar{x} , $s_{\bar{x}}$, is

$$s_{\bar{x}} = \frac{s_x}{n}$$

Formula 13

where

$$S_x = \sqrt{\sum_{i=1}^n S_{x_i}^2 + 2r \sum_{i=1}^{n-1} S_{x_i} S_{x_{i+1}}}$$

Formula 14

The correlation between consecutive years, r, is 0.35 for non-Hispanic households and 0.55 for Hispanic households. Correlation between nonconsecutive years is zero. The correlations were derived for income estimates but they can be used for other types of estimates where the year-to-year correlation between identical households is high.

Table 2. Parameters for Computation of Standard Errors for Labor Force Characteristics: March 2001		
Characteristic	a	b
Labor Force and Not In Labor Force Data Other than Agricultural Employment and Unemployment		
Total or White	-0.000008	1,586
Men	-0.000035	2,927
Women	-0.000033	2,693
Both sexes, 16 to 19 years	-0.000244	3,005
Black	-0.000154	3,296
Men	-0.000336	3,332
Women	-0.000282	2,944
Both sexes, 16 to 19 years	-0.001531	3,296
Hispanic Ancestry	-0.000187	3,296
Men	-0.000363	3,332
Women	-0.000380	2,944
Both sexes, 16 to 19 years	-0.001822	3,296

Unemployment		
Total or White	-0.000017	3,005
Men	-0.000035	2,927
Women	-0.000033	2,693
Both sexes, 16 to 19 years	-0.000244	3,005
Black	-0.000154	3,296
Men	-0.000336	3,332
Women	-0.000282	2,944
Both sexes, 16 to 19 years	-0.001531	3,296
Hispanic Ancestry	-0.000187	3,296
Men	-0.000363	3,332
Women	-0.000380	2,944
Both sexes, 16 to 19 years	-0.001822	3,296
Agricultural Employment	0.001345	2,989

NOTE: These parameters are to be applied to basic CPS monthly labor force estimates. For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.

Table 3. a and b Parameters for Standard Error Estimates for People and Families: March 2001						
Characteristics	Total or White		Black		Hispanic	
	a	b	a	b	a	b
PEOPLE						
Educational Attainment	- 0.000011	2,369	- 0.000103	2,680	- 0.000077	1,811

Employment Characteristics	- 0.000008	1,586	- 0.000154	3,296	- 0.000187	3,296
People by Family Income	- 0.000023	4,901	- 0.000215	5,611	- 0.000239	5,611
Income	- 0.000011	2,454	- 0.000108	2,810	- 0.000120	2,810
Health Insurance	- 0.000008	2,191	- 0.000074	2,661	- 0.000058	1,959
Marital Status, Household and Family Characteristics,						
- Some household members	- 0.000019	5,211	- 0.000209	7,486	- 0.000221	7,486
- All household members	- 0.000023	6,332	- 0.000309	11,039	- 0.000327	11,039
Mobility Characteristics (Movers)						
- Educational Attainment, Labor Force, Marital Status, Household, Family, and Income	- 0.000011	2,869	- 0.000080	2,869	- 0.000085	2,869
US, County, State, Region or MSA	- 0.000029	7,791	- 0.000218	7,791	- 0.000230	7,791
Below Poverty						
- Total	- 0.000038	10,380	- 0.000290	10,380	- 0.000307	10,380
-- Male	- 0.000077	10,380	- 0.000623	10,380	- 0.000596	10,380
-- Female	- 0.000074	10,380	- 0.000543	10,380	- 0.000592	10,380
- Age						
-- Under 15	- 0.000132	8,002	- 0.000826	8,002	- 0.000774	8,002
-- Under 18		8,002		8,002		8,002

	0.000110		0.000691		0.000602	
-- 15 and over	- 0.000048	10,380	- 0.000398	10,380	- 0.000442	10,380
-- 15 to 24	- 0.000101	3,927	- 0.000672	3,927	- 0.000459	3,927
-- 25 to 44	- 0.000048	3,927	- 0.000364	3,927	- 0.000257	3,927
-- 45 to 64	- 0.000063	3,927	- 0.000589	3,927	- 0.000263	3,927
-- 65 and over	- 0.000120	3,927	- 0.001407	3,927	- 0.000749	3,927
Unemployment	- 0.000017	3,005	- 0.000154	3,296	- 0.000187	3,296
FAMILIES, HOUSEHOLDS, OR UNRELATED INDIVIDUALS						
Income	- 0.000010	2,241	- 0.000094	2,447	- 0.000104	2,447
Marital Status, Household and Family Characteristics, Educational Attainment, Population by Age and/or Sex	- 0.000010	2,068	- 0.000072	1,871	- 0.000080	1,871
Poverty	0.000102	2,442	0.000102	2,442	0.000102	2,442

NOTES: These parameters are to be applied to March supplemental data including the Hispanic supplement.

For nonmetropolitan characteristics multiply a and b parameters by 1.5.. If the characteristic of interest is total state population, not subtotaed by race or ancestry, the a and b parameters are zero.

For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.

Table 4. Factors for State Standard Errors and Parameters and State Populations: 2001

State	f	f ²	Population
Alabama	1.01	1.01	3,412,000
Alaska	0.39	0.15	441,000
Arizona	0.98	0.97	3,671,000
Arkansas	0.77	0.59	1,992,000
California	1.14	1.29	25,801,000
Colorado	0.96	0.93	3,184,000
Connecticut	1.00	1.00	2,542,000
Delaware	0.47	0.22	593,000
District of Columbia	0.40	0.16	412,000
Florida	0.98	0.97	12,087,000
Georgia	1.19	1.40	6,043,000
Hawaii	0.59	0.35	891,000
Idaho	0.52	0.27	964,000
Illinois	1.00	1.00	9,228,000
Indiana	1.17	1.38	4,548,000
Iowa	0.84	0.71	2,197,000
Kansas	0.81	0.65	2,009,000
Kentucky	0.96	0.92	3,101,000
Louisiana	0.97	0.95	3,295,000
Maine	0.61	0.37	1,007,000
Maryland	1.17	1.38	4,044,000
Massachusetts	0.90	0.81	4,820,000
Michigan	0.96	0.93	7,579,000
Minnesota	1.05	1.11	3,681,000
Mississippi	0.80	0.64	2,096,000
Missouri	1.17	1.37	4,189,000
Montana	0.45	0.20	696,000
Nebraska	0.65	0.42	1,259,000
Nevada	0.66	0.44	1,440,000
New Hampshire	0.62	0.38	949,000
New Jersey	0.91	0.82	6,312,000

New Mexico	0.63	0.40	1,327,000
New York	0.94	0.89	14,192,000
North Carolina	0.97	0.94	5,846,000
North Dakota	0.40	0.16	477,000
Ohio	1.01	1.02	8,637,000
Oklahoma	0.85	0.73	2,570,000
Oregon	0.93	0.86	2,625,000
Pennsylvania	0.98	0.96	9,295,000
Rhode Island	0.55	0.30	756,000
South Carolina	1.00	1.01	3,059,000
South Dakota	0.41	0.17	556,000
Tennessee	1.16	1.34	4,309,000
Texas	1.10	1.21	15,327,000
Utah	0.66	0.43	1,544,000
Vermont	0.42	0.18	475,000
Virginia	1.22	1.48	5,351,000
Washington	1.21	1.47	4,470,000
West Virginia	0.62	0.39	1,444,000
Wisconsin	1.11	1.23	4,057,000
Wyoming	0.35	0.12	374,000

NOTE: For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics.