Appendix
Source and Reliability of Estimates

Source of Data

Most estimates in this report come from data obtained in March or June of years 1960 through 1988 in the Current Population Survey (CPS). The Bureau of the Census conducts the survey every month, although this report uses mostly March and June data for its estimates. Also, some estimates come from 1960 and 1970 decennial census data. The March and June surveys use two sets of questions, the basic CPS and the supplements.

For other sources see the reference lists associated with each section of the report.

Basic CPS. The basic CPS collects primarily labor force data about the civilian noninstitutional population. Interviewers ask questions concerning labor force participation about each member 14 years old and over in every sample household.

The present CPS sample was selected from the 1980 decennial census files with coverage in all 50 States and the District of Columbia. The sample is continually updated to account for new residential construction. It is located in 729 areas comprising 1,973 counties, independent cities, and minor civil divisions. About 59,500 occupied households are eligible for interview every month. Interviewers are unable to obtain interviews at about 2,500 of these units because the occupants are not home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has redesigned the CPS sample several times to improve the quality and reliability of the data and to satisfy changing data needs. The most recent changes were completely implemented in July 1985.

March and June supplements. In addition to the basic CPS questions, interviewers asked supplementary questions, in March, about relationship, marital status and living arrangements, and, in June, about marital history, fertility and birth expectations.

In order to obtain more reliable data for the Hispanic population, the March CPS sample was increased by about 2,500 eligible housing units, interviewed the previous November, that contained at least one sample person of Hispanic origin. In addition, the sample included persons in the Armed Forces living off post or with their families on post.

Estimation procedure. This survey's estimation procedure inflates weighted sample results to independent estimates of the civilian noninstitutional population of the United States by age, sex, race and Hispanic/non-Hispanic categories. The independent estimates were based on statistics from decennial censuses of population; statistics on births, deaths, immigration and emigration; and statistics on the size of the Armed Forces. The independent population estimates used for 1980 to present were based on updates to controls established by the 1980 decennial census. Data previous to 1980 were based on independent population estimates from the most recent decennial census. For more details on the change in independent estimates, see the section entitled "Introduction of 1980 Census Population Controls" in an earlier report (Series P-60, No. 133). The estimation procedure for the March supplement included a further adjustment so the husband and wife in a household received the same weight.

The estimates in this report for 1984 through 1988 also employ a revised survey weighting procedure for persons of Hispanic origin. In previous years, weighted sample results were inflated to independent estimates of the noninstitutional population by age, sex, and race. There was no specific control of the survey estimates for the Hispanic population. Since then, the Bureau of the Census developed independent population controls for the Hispanic population by sex and detailed age groups. Revised weighting procedures incorporate these new controls. The independent population estimates include some, but not all, undocumented immigrants.

Divorce rates. Data on divorce rates were obtained from the Vital Statistics Reports published by the National Center for Health Statistics (NCHS).

Reliability of Estimates

Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. A sample survey estimate has two possible types of error: sampling and nonsampling. The accuracy of an estimate depends on both types of error, but the full extent of the nonsampling error is unknown. Consequently, one should be particularly careful when interpreting results based on a relatively small number of cases or on small differences between estimates. The standard errors for CPS estimates primarily indicate the magnitude of sampling error. They also partially measure the effect of some nonsampling errors in responses and enumeration, but do not measure systematic biases in the data. (Bias is the average over all possible samples of the differences between the sample estimates and the desired value.)

Nonsampling variability. Nonsampling errors can be attributed to many sources. These sources include the inability to obtain information about all cases in the sample, definitional difficulties, differences in the interpretation of questions, respondents' inability or unwillingness to provide correct information or to recall information, errors made in data collection such as in recording or coding the data, errors made in processing the data, errors made in estimating values for missing data, and failure to represent all units with the sample (undercoverage).

CPS undercoverage results from missed housing units and missed persons within sample households. Compared to the level of the 1980 decennial census, overall CPS undercoverage is about 7 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 persons in missed households or missed persons in interviewed households have different characteristics from those of interviewed persons in the same age–sex–race–Hispanic group. Furthermore, the independent population controls have not been adjusted for undercoverage in the 1980 census.

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 and Technical Paper 40, *The Current Population Survey: Design and Methodology*, Bureau of the Census, U.S. Department of Commerce.

Sampling variability. Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors are primarily measures of sampling variability, although they may include some nonsampling error. The sample estimate and its standard error enable one to construct a confidence interval, 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 using 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 1.6 standard errors below the estimate to 1.6 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 also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis appearing in this report is that the population parameters are different. An example of this would be comparing one person households to married-couple households.

Tests may be performed at various levels of significance, where a significance level is the probability of concluding that the characteristics are different when, in fact, they are the same. Unless otherwise qualified, all statements of comparison in the text have passed a hypothesis test at the 0.10 level of significance or better. This means that the absolute value of the estimated difference between characteristics is greater than or equal to 1.6 times the standard error of the difference.

For a discussion of the computation of standard errors and for examples of their application, please see the individual reports referenced in each section of this report.

Reliability of divorce rates. Data on the number of divorces are from administrative records supplied by the individual states to NCHS and are not subject to sampling error.

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. Use caution when comparing results from different sources.

Caution should also be used when comparing estimates in this report, which reflect 1980 census-based population controls, with estimates for 1980 and earlier years, which reflect 1970 census-based population controls. This change in population controls had relatively little impact on summary measures such as means, medians, and percentage distributions, but did have a significant impact on levels. For example, use of 1980 based population controls results in about a 2-percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1981 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.

Since no independent population control totals for Hispanics were used before 1984, compare Hispanic estimates over time cautiously.

Note when using small estimates. Summary measures (such as medians and percentage distributions) are shown only when the base is 75,000 or greater. Because of the large standard errors involved, summary measures would probably not reveal useful information when computed on a smaller base. However, estimated numbers are shown even though the relative standard errors of these numbers are larger than those for corresponding percentages. These smaller estimates permit combinations of the categories to suit data users' needs. Take care in the interpretation of small differences. For instance, even a small amount of nonsampling error can cause a borderline difference to appear significant or not, thus distorting a seemingly valid hypothesis test.

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Official Business
Penalty for Private Use, $300

FIRST-CLASS MAIL
POSTAGE & FEES PAID
CENSUS
PERMIT No. G–58