

# Appendix A.

## Definitions and Explanations

**SYMBOLS.** A dash (–) represents zero or rounds to zero. The symbol “B” means that the base for the derived figure from the Current Population Survey is less than 75,000. For data from the Survey of Income and Program and Participation (SIPP), the “B” is used if the base is less than 200,000. The symbol “C” means the confidence includes zero. An “X” means not applicable, and “NA” means not available.

**Rounding of estimates.** Individual figures are rounded to the nearest thousand without being adjusted to group totals, which are independently rounded; percentages are based on the unrounded numbers.

**Metropolitan-nonmetropolitan residence.** The population residing in metropolitan statistical areas (MSA's) constitutes the metropolitan population. MSA's are defined by the Office of Management and Budget for use in the presentation of statistics by agencies of the Federal Government. An MSA is a geographic area consisting of a large population nucleus, together with adjacent communities which have a high degree of economic and social integration with that nucleus. The definitions specify a boundary around each large city so as to include most or all its suburbs. Entire counties form the MSA building blocks, except in New England where cities and towns are used. The former term SMSA was changed to MSA in 1983.

An area qualifies for recognition as an MSA if (1) it includes a city of at least 50,000 population, or (2) it includes a Census Bureau-defined urbanized area of at least 50,000 population with a total metropolitan population of at least 100,000 (75,000 in New England). In addition to the county containing the main city or

urbanized area, an MSA may include other counties having strong commuting ties to the central county. If specified conditions are met, certain large MSA's are designated as consolidated MSA's (CMSA's) and divided into component primary MSA's (PMSA's).

In July 1985, the CPS began carrying the metropolitan statistical area definitions announced by the Office of Management and Budget on June 30, 1984. Figures published from the CPS in the early 1980's and throughout most of the 1970's referred to metropolitan areas as defined on the basis of the 1970 census. Since there are important differences in the population classified as metropolitan using the 1970 and 1984 definitions, comparisons should be avoided.

The new CPS metropolitan estimates have consistently been higher than independent estimates of the metropolitan population prepared by the Census Bureau; the new CPS nonmetropolitan estimates have been lower than the independent estimates. The apparent overestimation of metropolitan population and underestimation of nonmetropolitan population in the CPS relative to the Census Bureau's independent estimates should be taken into account when using the data.

**Age.** The age classification is based on the age of the person at his or her last birthday. The adult universe (i.e., population of marriageable age) now comprises persons 15 years old and over. Prior to 1980 the adult universe was 14 years old and over.

**Race.** The population is divided into three groups on the basis of race: White, Black, and other races. The last category includes American In-

dians, Japanese, Chinese, and any other race except White and Black.

**Persons of Hispanic origin.** Persons of Hispanic origin in this report were determined on the basis of a question that asked for self-identification of the person's origin or descent. Respondents were asked to select their origin (or the origin of some other household member) from a “flash card” listing ethnic origins. Persons of Hispanic origin, in particular, were those who indicated that their origin was Mexican, Puerto Rican, Cuban, Central or South American, or some other Hispanic origin. Persons of Hispanic origin may be of any race.

**Marital status.** The marital status classification identifies four major categories: never married, married, widowed, and divorced. These terms refer to the marital status at the time of the enumeration.

The category “married” is further divided into “married, spouse present,” “separated,” and “other married, spouse absent.” A person was classified as “married, spouse present” if the husband or wife was reported as a member of the household, even though he or she may have been temporarily absent on business or on vacation, visiting, in a hospital, etc., at the time of the enumeration. Persons reported as separated included those with legal separations, those living apart with intentions of obtaining a divorce, and other persons permanently or temporarily separated because of marital discord. The group “other married, spouse absent” includes married persons living apart because either the husband or wife was employed and living at a considerable distance from home, was serving away from home in the Armed Forces, had moved to another area, or had a different place of residence for any

other reason except separation as defined above.

**Household.** A household consists of all the persons who occupy a housing unit. A house, an apartment or other group of rooms, or a single room is regarded as a housing unit when it is occupied or intended for occupancy as separate living quarters; that is, when the occupants do not live and eat with any other persons in the structure and there is direct access from the outside or through a common hall.

A household includes the related family members and all the unrelated persons, if any, such as lodgers, foster children, wards, or employees who share the housing unit. A person living alone in a housing unit or a group of unrelated persons sharing a housing unit as partners is also counted as a household. The count of households excludes group quarters.

**Group quarters.** As of 1983, group quarters were defined in the Current Population Survey as noninstitutional living arrangements for groups not living in conventional housing units or groups living in housing units containing ten or more unrelated persons or nine or more persons unrelated to the person in charge. (Prior to 1983, group quarters included housing units containing five or more persons unrelated to the person in charge.) Examples of persons in group quarters include a person residing in a rooming house, in staff quarters at a hospital, or in a half-way house. Beginning in 1972, residents of institutions have not been included in the Current Population Survey.

**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 is the "reference person" to whom the relationship of all other household members, if any, is recorded.

Prior to 1980, the husband was always considered the householder in married-couple households. The number of householders is equal to the number of households. Also, the number of family householders is equal to the number of families.

**Head versus householder.** Beginning with the 1980 CPS, the Bureau of the Census discontinued the use of the terms "head of household" and "head of family." Instead, the terms "householder" and "family householder" are used. Recent social changes have resulted in greater sharing of household responsibilities among the adult members and, therefore, have made the term "head" increasingly inappropriate in the analysis of household and family data. Specifically, the Census Bureau has discontinued its longtime practice of always classifying the husband as the reference person (head) when he and his wife are living together.

In this report, the term "householder" is used in the presentation of data that had previously been presented with the designation "head." The householder is the first adult household member listed on the questionnaire. The instructions call for listing first the person (or one of the persons) in whose name the home is owned or rented. If a home is owned jointly by a married couple,

either the husband or the wife may be listed first, thereby becoming the reference person, or householder, to whom the relationship of other household members is to be recorded.

**Reference person.** The reference person is the person with regard to whom the relationship of other persons in the household is recorded. The household reference person is the person listed as the householder (see definition of "Householder"). The subfamily reference person is either the single parent or the husband/wife in a married-couple situation.

**Family.** A family is a group of two persons or more (one of whom is the householder) related by birth, marriage, or adoption and residing together; 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 any unrelated persons (unrelated subfamily members and/or secondary individuals) who may be residing there are included. 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. See the definition of family.

**Related subfamily.** A related subfamily is a married couple with or

without children, or one parent with one or more own never-married children under 18 years old, living in a household and related to, but not including, the person or couple who maintains the household. One 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 count of families.

**Unrelated subfamily.** An unrelated subfamily (formerly called a secondary family) is a married couple with or without children, or a single parent with one or more own never-married children under 18 years old living in a household. Unrelated subfamily members are not related to the householder. An unrelated subfamily may include persons such as guests, partners, roommates, or resident employees and their spouses and/or children. The number of unrelated subfamily members is included in the total number of household members, but is not included in the count of family members.

Beginning in 1989, any person(s) who is not related to the householder and who is not the husband, wife, parent, or child in an unrelated subfamily is counted as an unrelated individual.

**Family group.** A family group is any two or more persons (not necessarily including a householder) residing together, and related by birth, marriage, or adoption. A household may be composed of one such group, more than one, or none at all. The count of family groups includes family households, related subfamilies, and unrelated subfamilies.

**Married couple.** A married couple, as defined for census purposes, is a

husband and wife enumerated as members of the same household. The married couple may or may not have children living with them. The expression "husband-wife" or "married-couple" before the term "household," "family," or "subfamily" indicates that the household, family, or subfamily is maintained by a husband and wife. The number of married couples equals the count of married-couple families plus related and unrelated married-couple subfamilies.

**Unmarried couple.** An unmarried couple is composed of two unrelated adults of the opposite sex (one of whom is the householder) who share a housing unit with or without the presence of children under 15 years old.

**Unrelated individuals.** Unrelated individuals are persons of any age who are not members of families or subfamilies.

**Nonfamily householder.** A nonfamily householder is a person maintaining a household while living alone or exclusively with persons to whom they are not related.

**Secondary individuals.** Secondary individuals are persons of any age who reside in a household, but are not related to the householder (except unrelated subfamily members). Persons who reside in group quarters are also secondary individuals. Examples of a secondary individual include (1) a guest, partner, roommate, or resident employee; (2) a foster child; or (3) a person residing in a rooming house, a halfway house, staff quarters at a hospital, or other type of group quarters.

**Own children and related children.** "Own" children in a family are sons and daughters, including step-

children and adopted children, of the householder. Similarly, "own" children in a subfamily are sons and daughters of the married couple or parent in the subfamily. (All children shown as members of related subfamilies are own children of the person(s) maintaining the subfamily.) "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 never-married children; however, "own children under 25" and "own children of any age," as the terms are used here, include all children regardless of marital status. The totals include never-married children living away from home in college dormitories.

The count of related children in families was formerly restricted to never-married children. However, beginning with data for 1968 the Bureau of the Census includes ever-married children under the category of related children. This change added approximately 20,000 children to the category of related children in March 1968.

**Tenure.** A housing unit is "owned" if the owner or co-owner lives in the unit, even if it is mortgaged or not fully paid for. A cooperative or condominium unit is "owned" only if the owner or co-owner lives in it. All other occupied units are classified as "rented," including units rented for cash rent and those occupied without payment of cash rent.

**Years of school completed.** Education refers to the years of school completed by the person. Data on years of school completed were derived from the combination of answers to two questions, (a) "What is the highest grade of school

that the person has attended?" and (b) "Did the person finish this grade?"

**Labor force and employment status.** The definitions of labor force and employment status in this report are related to the civilian population 15 years and over. Persons shown here are classified as in the labor force if they were employed as civilians or unemployed during the survey week.

**Employed.** Employed persons comprise (1) all civilians who, during the specified week, did any work at all as paid employees or in their own business or profession, or on their own farm, or who worked 15 hours or more as unpaid workers on a farm or in a business operated by a member of the family, and (2) all those who were not working but who had jobs or businesses from which they were temporarily absent because of illness, bad weather, vacation, or labor management dispute, or because they were taking time off for personal reasons, whether or not they were paid by their employers for time off, and whether or not they were seeking other jobs. Excluded from the employed group are persons whose only activity consisted of work around the house (own home housework, painting or repairing own home, etc.) or volunteer work for religious, charitable, and similar organizations.

**Unemployed.** Unemployed persons are those civilians who, during the survey week, had no employment but were available for work and (1) had 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) were waiting to be called back to a job from which they had been laid off; or (3) were waiting to report to a new wage or salary job within 30 days.

**Not in the labor force.** All civilians who are not classified as employed or unemployed are defined as "not in the labor force." This group who are neither employed nor seeking work includes persons engaged only in own home housework, attending school, or unable to work because of long-term physical or mental illness; persons who are retired or too old to work, seasonal workers for whom the survey week fell in an off season, and the voluntary idle. Persons doing only unpaid family work (less than 15 hours) are also classified as not in the labor force.

**Income.** Current Population Survey (CPS), data on income cover money income only, prior to deduction for taxes, received from such sources as wages or salaries, net income from self-employment, Social Security, dividends, interest, public assistance and welfare, unemployment compensation, government pensions, and veterans payments. Certain money receipts such as capital gains are not included.

In data from the Survey of Income and Program Participation (SIPP), the cash income concept includes the sum of all income received from any of the sources listed in table A-1. Rebates, refunds, loans and capital gain or loss amounts from the sale of assets, and interhousehold transfers of cash such as allowances are not included.

Accrued interest on Individual Retirement Accounts, KEOGH retirement plans, and U.S. Saving bonds are also excluded. This definition differs somewhat from that used in the annual income reports based on the March CPS income supplement questionnaire. The data in those reports, published in the Current Population Reports, Series P-60, are based only on income received in a regular or periodic manner and, therefore, exclude lump-sum or one-time payments, such as inheritances or insurance settlements which are included as income in SIPP. Educational assistance, which is included in the March CPS income concept, is not included in the SIPP income concept.

The income amounts represent amounts actually received during the month, before deductions for income and payroll taxes, union dues, Part B Medicare premiums, etc.

The SIPP income definition includes three types of earnings: wages and salary, nonfarm self-employment, and farm self-employment. The definition of nonfarm self-employment and farm self-employment is not based on the net difference between gross receipts or sales and operating expenses, depreciation, etc. The monthly amounts for these income types are based on the salary or other income received from the business by the owner of the business or farm during the 4-month period. Earnings from all jobs and self-employment are included.

While the income amounts from most sources are recorded monthly for the 4-month reference period, property income amounts such as interest, dividends, and rental income, were recorded as totals for the 4-month period. These totals were distributed equally between

months of the reference period for purposes of calculating poverty status for SIPP.

**Poverty.** The poverty definition used here is that adopted for official Government use by the Office of Management and Budget and consists of a set of money income

thresholds that vary by family size and composition. Families or individuals with income below a particular threshold are classified as below the poverty level. The poverty thresholds are updated every year to reflect changes in the Consumer Price Index. These thresholds are

based on money income only and do not include the value of noncash benefits such as employer-provided health insurance, food stamps, or Medicaid. For a more detailed explanation, see Bureau of the Census, Current Population Reports, Series P-60.

Table A-1.

**Income Sources Included  
In Monthly Cash Income**

<b>Earnings From Employment</b>	Payments from a sickness, accident, or disability insurance policy purchased on your own
Wages and salary	Aid to Families with Dependent Children (AFDC), (ADC)
Nonfarm self-employment income	General assistance or General relief
Farm self-employment income	Indian, Cuban, or Refugee assistance
<b>Income From Assets (Property Income)</b>	Foster child care payments
Regular/passbook savings accounts in a bank, savings and loan or credit union	Other welfare
Money market deposit accounts	Child support payments
Certificate of deposit	Alimony payments
NOW, Super NOW, or other interest-earning checking accounts	Pensions from a company or union
Money market funds	Federal Civil Service or other Federal civilian employee pensions
U.S. Government securities	U.S. Military retirement
Municipal or corporate bonds	National Guard or Reserve Forces retirement
Other interest-earning assets	State government pensions
Stocks or mutual fund shares	Local government pensions
Rental property	Income from paid-up life insurance policies or annuities
Mortgages	Estates and trusts
Royalties	Other payments for retirement, disability or survivors, G.I. Bill/VEAP education benefits
Other financial investments	Income assistance from a charitable group
<b>Other Income Sources</b>	Other unemployment compensation (Trade Adjustment Act benefits, strike pay, other)
Social Security	Veterans' compensation or pensions
U.S. Government	Money from relatives or friends
Railroad Retirement	Lump sum payments
Federal Supplemental Security Income	Income from roomers or boarders
State Administered	National Guard or Reserve pay
Supplemental Security Income	Incidental or casual earnings
State unemployment compensation	Other cash income not included elsewhere
Supplemental Unemployment Benefits	
Black Lung payments	
Worker's compensation	
State temporary sickness or disability benefits	
Employer or union temporary sickness policy	

# Appendix B.

## Source and Accuracy of Estimates

### Source of Data

Most estimates in this report come from data obtained from the Current Population Survey (CPS) conducted in March of years 1987 through 1990. The Bureau of the Census conducts the survey every month, although this report uses mostly March data for its estimates. Data from November 1988 were used for the voting estimates. Also, some estimates come from 1960 through 1990 decennial census data. The March and November surveys use two sets of questions: the basic CPS and the supplements.

**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 15 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 and includes 1,973 counties, independent cities, and minor civil divisions. About 60,000 occupied housing units are eligible for interview every month. Interviewers are unable to obtain interviews at about 2,600 of these units because the occupants are not found at home after repeated calls or are unavailable for some other reason.

Since the introduction of the CPS, the Bureau of the Census has

Description of Current Population Survey

Time period	Number of sample areas	Housing units eligible <sup>1</sup>	
		Interviewed	Not interviewed
1990 .....	729	57,400	2,600
1989 .....	729	53,600	2,500
1986 to 1988.....	729	57,000	2,500
1985 .....	<sup>2</sup> 629/729	57,000	2,500
1982 to 1984.....	629	59,000	2,500
1980 to 1981.....	629	65,500	3,000
1977 to 1979.....	614	55,000	3,000
1973 to 1976.....	461	48,500	2,500
1972 .....	449	45,000	2,000
1967 to 1971.....	449	48,000	2,000
1963 to 1966.....	357	33,500	1,500
1960 to 1962.....	333	33,500	1,500

<sup>1</sup>Excludes about 2,500 Hispanic households added in March from the previous November sample. (See "March Supplement.")

<sup>2</sup>The CPS was redesigned following the 1980 Decennial Census of Population and Housing. During phase-in of the new design, housing units from the new and old designs were in the sample.

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.

The table summarizes changes in the CPS designs for the years for which data appear in this report.

**March Supplement.** In addition to the basic CPS questions, interviewers asked supplementary questions in March about marital status, educational attainment, and geographical mobility.

To obtain more reliable data for the Hispanic-origin population, the March CPS sample was increased by about 2,500 eligible housing units. These housing units were interviewed the previous November and 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.

**November Supplement.** In addition to the basic CPS questions, interviewers asked supplementary questions in November 1988 about voting in the presidential election.

**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 1981 to present were based on updates to controls established by the 1980 Decennial Census. Data before 1981 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 of a household received the same weight.

The estimates in this report for 1985 and later 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 non-institutional 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.

### Accuracy 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 errors: sampling and nonsampling. The accuracy of an estimate depends on both types of errors, 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.** There are several sources of nonsampling error including the following:

- Inability to get information about all sample cases.
- Definitional difficulties.
- Differences in interpretation of questions.
- Respondents' inability or unwillingness to provide correct information.
- Respondents' 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).

CPS undercoverage results from missed housing units and missed persons within sample households. Compared with 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 caused by 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.

**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.

CPS estimates in this report (which reflect 1980 Census-based population controls) may differ from 1990 Census results. Population controls incorporating 1990 Census results will be used for CPS estimates beginning in 1993.

Caution should also be used when comparing estimates in this report 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 percent distributions. It 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 sub-population groups than for the total population.

Since no independent population control totals for persons of Hispanic origin were used before 1985, 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. These estimates may not be reliable for 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.

**Sampling Variability.** Sampling variability is variation that occurred by chance because a sample was surveyed rather than the entire population. Standard errors, as calculated by methods described next, are primarily measures of sampling variability, although they may include some nonsampling errors.

#### **Standard Errors and Their Use.**

A number of approximations are required to derive, at a moderate cost, standard errors applicable to all the estimates in this report. Instead of providing an individual standard error for each estimate, parameters are provided to calculate standard errors for various types of characteristics. These parameters are listed in table B-1. For information on how to calculate standard errors for Census data see the census reports.

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 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 1.645 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.

Some statements in the report may contain estimates followed by a number in parentheses. This number can be added to and subtracted from the estimate to calculate upper and lower bounds of the 90-percent confidence interval. For example, if a statement contains the phrase "grew by 1.7 percent ( $\pm 1.0$ )," the 90 percent confidence interval for the estimate, 1.7 percent, is 0.7 percent to 2.7 percent.

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 appearing in this report is that the population parameters are different. An example of this would be comparing White voters to Black voters.

Tests may be performed at various levels of significance. The significance level of a test is the probability of concluding that the characteristics are different when, in fact, they are the same. 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 1.645 times the standard error of the difference.

**Standard Errors of Estimated Numbers.** Use the following formula to compute the approximate standard error,  $s_x$ , of an estimated number shown in this report.

$$s_x = \sqrt{ax^2 + bx} \quad (1)$$

Here,  $x$  is the size of the estimate and  $a$  and  $b$  are the parameters in table B-1 associated with the particular type of characteristic. When calculating standard errors for numbers from cross-tabulations involving different characteristics, use the set of parameters for the characteristic that will give the largest standard error.

*Illustration*

Suppose that 19,818,000 persons 65 years old and over reported voting in the 1988 presidential

election. Use the appropriate parameters from table B-1 and formula (1) to get

Number, $x$	19,818,000
a parameter	-0.000028
b parameter	3,346
Standard error	235,000
90% conf. int.	19,431,000 to 20,205,000

The standard error is calculated as

$$s_x = \sqrt{-0.000028 \times 19,818,000^2 + 3,346 \times 19,818,000} = 235,000$$

The 90-percent confidence interval is calculated as 19,818,000  $\pm$  1.645x235,000.

A conclusion that the average estimate derived from all possible samples lies within a range computed in this way would be correct for roughly 90 percent of all possible samples.

**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 parameter from table B-1 indicated by the numerator.

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

$$s_{x,p} = \sqrt{(b/x) p (100-p)} \quad (2)$$

Here,  $x$  is the total number of persons, 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 B-1 associated with the characteristic in the numerator of the percentage.

*Illustration*

Suppose that of the 17,232,000 females 65 years old and over, 39.7 percent were living with their spouses. Use the appropriate parameter from table B-1 and formula (2) to get

Percentage, $p$	39.7
Base, $x$	17,232,000
b parameter	4,785
Standard error	0.8
90% conf. int.	38.4 to 41.0

The standard error is calculated as

$$s_{x,p} = \sqrt{\frac{4,785}{17,232,000} \times 39.7 \times (100.0 - 39.7)} = 0.8$$

The 90-percent confidence interval for the percentage of females 65 years old and over living with their spouses is calculated as 39.7  $\pm$  1.645x0.8.

**Standard Error of a Difference.**

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

$$s_{x-y} = \sqrt{s_x^2 + s_y^2} \quad (3)$$

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.

#### *Illustration*

Suppose that 2,747,000 persons 70-74 years old,  $x$ , and 3,051,000 persons 75 years old and over,  $y$ , completed high school. Use the appropriate parameters from table B-1 and formulas (1) and (3) to get

	$x$	$y$	difference
Number	2,747,000	3,051,000	304,000
a parameter	-0.000021	-0.000021	-
b parameter	2,743	2,743	-
Standard error	88,000	90,000	124,000
90% conf. int.	2,606,000 to 2,888,000	2,903,000 to 3,199,000	100,000 to 508,000

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{88,000^2 + 90,000^2} = 124,000$$

The 90-percent confidence interval around the difference is calculated as  $304,000 \pm 1.645 \times 124,000$ . Since this interval does not contain zero, we can conclude, at the 10-percent significance level, that the number of persons 75 years old and over who completed high school is greater than the number of persons 70-74 years old who did.

## Types and Quality of Data Available on the Elderly in the 1990 Census

A decennial census provides rich subject-matter and geographic detail for local areas and the nation that is not possible in a sample survey. Census counts by age, sex, and race are used as the denominator of epidemiological measures. Thus, the quality of census data are critical. First, we will discuss the quality of data available on the elderly population, particularly as it affects denominators in epidemiological measures. Second, we will discuss some of the types of data available from the 1990 census and evaluation studies.

### Data Quality

In 1712, the Chinese Emperor Tsing Shen Tsu complained, "I have examined the census report of the viceroys of the provinces and have found them inaccurate." Data were not perfect then and are not now although there have been vast improvements. Data users should carefully consider the quality of the information they are using from censuses, surveys, and vital statistics. All data, whether from a complete enumeration of the population or from a sample, are subject to coverage and content errors. Data based on a sample are also subject to sampling error. Data on the older population have some particular problems with respect to these sources of error.

Errors in the data are of two types: sampling errors and nonsampling errors. Sampling error

affects those items collected from a sample of the population (see part II of table B-3). Sampling error occurs when a portion of the population is surveyed to represent the entire population. Data based on a sample are estimates that would differ somewhat from data based on a complete enumeration of all households or persons in group quarters. Sampling error can be measured based on the actual sample observed. In the census, about 1 in 6 households and 1 in 6 persons in group quarters received the sample form.

The deviation of the sample estimate from the average of all possible samples (which approximates a complete enumeration) is called "sampling error." The sampling error is a function of the observed sampling size; as the sample size becomes smaller, sampling error increases. Thus, for local areas with a small population, or when the group of interest is small, such as the population 85 years and over, sampling error may be quite large and should be accounted for in analysis. Each census report with sample data contains an appendix explaining the calculation of sampling error and its interpretation.

Nonsampling errors occur in the collection and processing of data. They are often difficult to measure and identify. Nonsampling errors may be random or in a consistent direction which biases the data. Nonsampling errors are of two basic types: coverage and content errors. Coverage errors result in persons being missed or counted erroneously (for example, counted

more than once). Content errors include errors by respondents and interviewers, processing errors, and those occurring when the data item is not completed (that is, nonresponse). Errors in age data include misstatement of age, a preference for giving an age or year of birth that ends in "0" or "5," and ages that are not known or not given.

Coverage errors occur when whole households are missed and when persons within households are missed or counted more than once. For example, an older couple may be traveling in their trailer and not receive their census form in the mail. In another type of coverage error, the same household may be counted twice. This might occur, for example, if a retired couple from the Northeast goes to their second home in Florida for the winter ("snowbirds"). There are census procedures to catch persons who may be travelling and to avoid counting in both places, but such errors do occur.

Evaluation studies performed after the 1980 census showed there was actually a net overcount of persons in the age groups 65 to 69 and 70 to 74, for both Blacks and Whites. Some of this was likely due to errors in reporting age as well as coverage error. At ages 75 and over, the studies concluded there was a net undercount of 0.6 percent for Black

males, 6.4 percent for Black females, 0.9 percent for White males and 2.6 percent for White females.<sup>1</sup>

For 1990, preliminary results from demographic analysis show different coverage patterns for males and females. For females, the initial estimates indicate a net overcount for age groups 65 to 69, 70 to 74, and 75 to 79, for both Blacks and races other than Black (Nonblack). Net undercounts occur at ages over 79 and the preliminary results indicate a relatively large undercount of persons 85 years and over. For males, the preliminary results indicate a net undercount for most age groups (with the exception of a net overcount for ages 75 to 79 for Black and Nonblack males and at ages 80 to 84 for Nonblack males only). We emphasize the preliminary nature of these results, especially for the group 85 years and over. There are problems with the files used for comparison (for example, the Medicare files do not purge all deaths). These, and other data difficulties will be accounted for in the final results which will not be available until after the release of this report.<sup>2</sup>

Some nonsampling errors occur during data collection and processing. The Census Bureau

mailed forms to most households. In most households, one household member fills out the questionnaire even though they may not know accurate information (such as age) for every household member. Sometimes, census takers visited respondents door-to-door. If a census taker does not understand a question, he or she may give seemingly authoritative but incorrect advice to respondents on how to answer. This can affect the data. In institutions such as nursing homes, the questionnaires are often filled out by staff using administrative records and their own knowledge and guesses. In larger institutions, the extra work can be a tedious, burdensome process and nonresponse to particular questions is often quite high. Clerical processing of forms in census offices can also lead to errors if workers make clerical errors or do not follow procedures. For the 1990 census, much of the processing has been automated to reduce the extent of clerical error.

Questionnaires may be returned with incomplete or inconsistent information. Nonresponse may be total, in which a respondent does not complete any items on the questionnaire, or partial, in which only some questions that should have been answered actually are answered. In institutions, such as nursing homes, the information may not be available in the administrative records and nonresponse rates, especially for social and economic characteristics, may be unusually high. For example, neither a patient nor the institution staff may be aware of an income source that goes directly to the patient's family.

<sup>1</sup>U.S. Bureau of the Census, "The Coverage of Population in the 1980 Census," *Evaluation and Research Reports*, PHC80-E4 US Government Printing Office, 1988, Washington, DC, table 3.3.

<sup>2</sup>U.S. Bureau of the Census, J. Gregory Robinson, Bashir Ahmed, Prithwis Das Gupta, and Karen A. Woodrow, "Estimating Coverage of the 1990 United States Census: Demographic Analysis," *Proceedings of the Social Statistics Section of the American Statistical Association*, 1992.

If efforts to obtain missing information fail, the computer "imputed", or filled in, the missing or inconsistent information. This imputation for missing data is based on the observed responses of a household with similar characteristics such as household size and race. In group quarters, it is based on the responses of others in the group quarters. In the 1980 census, if there had been no imputation for missing data, 11.1 percent of the population for which age was observed would have been shown as aged 65 or older; after imputation, however, the proportion of the population aged 65 or older increased to 11.3 percent.

Nonresponse can introduce bias into the data, as the characteristics of the nonrespondents have not been observed directly and may be different from those imputed. Each census report contains an appendix with a table showing the percentage of responses to particular items that were imputed. Data users should consult these appendices, especially when using information subject to nonresponse or misreporting, such as income. A high percentage of allocation indicates that particular caution is warranted in using the information.

Additional errors occur that affect the quality of census data. A respondent may misreport information, either intentionally or by misunderstanding the intent of the question. For example, respondents may misreport income intentionally. Or, they may simply not have understood that they should have included income amounts from a particular source such as self-employment.

Errors in the statement of age may affect total error in data for the elderly more than coverage errors. This is especially true in data before the 1990 census around age 65 and among the oldest old (especially centenarians) because of the misreporting of age. In modern censuses, "year of birth" is asked in addition to "age" which has reduced this error considerably. Nevertheless, reporting error remains. Age reporting error found in the 1990 census data is described in appendix C. Sometimes people misreport their age because they do not know or remember their age. Some give a "rounded-off" age and numbers ending in "0" or "5" occur more frequently than they should, a phenomenon known as "age heaping."<sup>3</sup> These errors are especially important when data are for single years of age and less important when grouped in 5- or 10-year age groups. Historical data may need to be adjusted as the errors are often sufficient to affect death rates.<sup>4</sup>

<sup>3</sup>In some cultures, certain numbers are particularly avoided (such as 13 in the West and 4 in the Orient). For a discussion of the various indexes of age preference and methods of adjusting reported single-year-of-age data, see Henry S. Shryock and Jacob S. Siegel, and Associates, *The Methods and Materials of Demography*, U.S. Government Printing Office, Washington, DC, 1971, pp. 205-211.

<sup>4</sup>Greville developed an adjustment technique described in Mortimer Spiegelman, *Introduction to Demography*, rev. ed., Cambridge, MA: Harvard University Press, 1968. Spiegelman discusses an adjustment technique developed by Greville for historical age data (p. 67) and a blending method for age heaping (pp. 71-75). For death rates, Spiegelman recommends choosing an age grouping for which the death rates would be

Age seems to be exaggerated the most at the oldest ages and among those with lower levels of education. This affects both census and mortality data on the extreme aged. Traditionally, death rates have been unreliable for persons 85 years and older. There have, however, been improvements in these data and we can expect vast improvements as more people reach these ages with higher education and with birth certificates that document year of birth. There also remains plenty of room for additional improvement. Census processing errors have produced unreliable data for counts of persons 90 years and over since at least 1960.<sup>5</sup> For 1990, improvements in the edits greatly reduced processing errors. For example, the reported year of birth and age are tested for consistency which corrects for reporting the wrong century. Additionally, no "child" of a householder can be over age 90 and no "grandchild" over age 75. Relationship and marital status are also compared with age for consistency.

Census error is measured by reinterviews, record matching studies, and demographic analysis. In addition, reinterviews and matching studies are one way to partially measure the effect of imputations for missing data. Another

essentially correct if both population and deaths were biased in the same direction and in about the same proportion.

<sup>5</sup>Ira Rosenwaike, *The Extreme Aged in America*, Westport, CT: Greenwood Press, 1985, Ch. 2; Gregory Spencer, Cynthia M. Taeuber, and Arnold Goldstein, *America's Centenarians*, Current Population Reports, Series P-23, No. 153. U.S. Government Printing Office, Washington, DC, September 1987.

way is to compare the reported census age with death certificate information for those who die close to the time of the census. Neither method is a perfect check as age may be misstated in both a reinterview and on death certificates. Demographic analysis develops estimates of population largely from administrative records such as vital statistics, Medicare data, and immigration statistics.<sup>6</sup> For example, census age distributions can be compared with those from demographic analysis to determine if systematic errors have skewed the distribution.

In summary, data users should be aware of the errors to which the data are subject. Users should review the data to make sure they make sense historically. Census estimates can often be compared with survey estimates to see if the reported trends differ significantly. While census operations included

<sup>6</sup>U.S. Bureau of the Census, J. Gregory Robinson, Prithwis Das Gupta, and Bashir Ahmed, *Evaluating the Quality of Estimates of Coverage Based on Demographic Analysis*, paper presented at the 1990 annual meeting of the Population Association of America, May 3-5, 1990. See also Shryock and Siegel, *Methods and Materials*, pp. 212-228.

procedures to minimize errors, it is impossible to avoid some data problems, such as adamant refusal to respond to the census form. Some census procedures themselves, such as clerical checking and computer editing and imputation, introduce error into the data. Knowledge of the types and extent of errors that may be present will contribute to more meaningful understanding of the census results.

### Types of Data Available

The census asks everyone basic demographic questions on household relationship, sex, race, age, marital status, and Hispanic origin and social and economic questions of a sample of households and persons in group quarters. For the 1990 census, counts of persons, by sex, race, and Hispanic origin will appear for single years to the end category, "105 years and over" for the United States, and sub-state statistical and political divisions.

There are nine main report series from the census as well as summary tape files and public-use microdata files. Public-use microdata samples (PUMS) are computer data files that contain the

edited responses from a sample of individual households. The records contain no identifying information and only large geographic areas are identified to protect the confidentiality of respondents. In addition to the PUMS for the entire population, a file that focuses specifically on the population 60 years and over will be produced (referred to as "PUMSO").

And finally, reports will be issued that evaluate the quality of 1990 census data. Initial reports will focus on coverage. Some content evaluation reports will provide tabulations by age and will provide additional insight into the uses and limitations of data on America's older population. These reports include a Content Reinterview Study (response bias and variance); the Integrated Evaluation of Error Study (evaluates the magnitude of all sources of error, including item nonresponse); Coverage Sampling Research (alternative coverage questions to improve coverage within households); the Master Trace Sample (impact of processing on data); Outreach Survey (respondent attitudes towards the census); and ethnographic studies on response and coverage problems.

**Table B-1.**  
**a and b Parameters and Factors for Calculating Approximate Standard Errors for Persons, Families, Households, Householders, and Unrelated Individuals 65+ in the USA**

Characteristic	Persons		Families, households, householders, and unrelated individuals	
	a	b	a	b
<b>Labor Force, Employment, and Occupation - 1990 Annual Averages</b>				
<b>Both Sexes:</b>				
Total or White .....	-0.000008	1,118	-0.000013	967
Black.....	-0.000051	795	-0.000085	688
Hispanic.....	-0.000067	795	-0.000058	688
<b>Male:</b>				
Total or White .....	-0.000013	967	-0.000013	967
Black.....	-0.000065	688	-0.000085	688
Hispanic.....	-0.000118	688	-0.000118	688
<b>Female:</b>				
Total or White .....	-0.000011	829	-0.000011	829
Black.....	-0.000059	590	-0.000058	590
Hispanic.....	-0.000099	590	-0.000099	590
<b>Educational Attainment - March 1989</b>				
Total or White .....	-0.000021	2,743	-0.000013	1,846
Black.....	-0.000247	3,711	-0.000116	1,688
Hispanic.....	-0.000371	3,887	-0.000167	1,747
<b>Geographical Mobility - March 1987</b>				
Total or White .....	-0.000024	6,675	(X)	(X)
Black.....	-0.000024	6,675	(X)	(X)
Hispanic.....	-0.000551	6,675	(X)	(X)
<b>Marital Status - March 1990</b>				
Total or White .....	-0.000026	4,785	-0.000011	1,899
Black.....	-0.000283	6,864	-0.000071	1,716
Hispanic.....	-0.000567	6,864	-0.000142	1,716
<b>Voting - November 1988</b>				
Total or White .....	-0.000028	3,346	(X)	(X)
Black.....	-0.000384	4,899	(X)	(X)
Hispanic.....	-0.000698	8,649	(X)	(X)

Note: Multiply the above parameters by 0.83, 0.93, 0.96, and 1.37 for the Northeast, Midwest, South, and West, respectively. Multiply the above parameters by 1.5 for nonmetropolitan.

Table B-2.  
Items in the 1990 Census

I. Information collected from all households: <sup>1</sup>	II. Information collected from a sample of households: <sup>1</sup>
Population Household relationship Sex Race Age Marital status Spanish/Hispanic origin Housing Number of units in structure Number of rooms in unit Own or rent housing Business at residence Value of owned unit or rent paid Congregate housing (meals included in rent) Vacancy characteristics	Population Social Characteristics Place of birth, citizenship, year of entry Education—enrollment and attainment Ancestry Migration, residence 5 years ago Language spoken at home, ability to speak English Military status Disability limiting work, ability to go outside, or care for personal needs Fertility Economic Characteristics Employment and unemployment, year last worked Place of work and commuting to work Occupation, employer, and type of work Work experience, income in 1989, and sources of income Housing Year moved into residence Number of bedrooms Plumbing and kitchen facilities Telephone Autos, light trucks and vans Fuel use Source of water and method of sewage disposal Age of building Condominium or mobile home status Farm residence Shelter costs, including utilities Real estate taxes and insurance Mortgages and loans

<sup>1</sup> Persons in group quarters, including institutions, are asked population items only.

## Appendix C.

# Age-Race Modifications to the 1990 Census (CPH-L-74 Series)

C-1

Where possible, 1990 census data in this report are from a special file, the CPH-L-74 series. Age and race data have been modified in this series to meet the needs of many users of census data. Essentially, the race statistics were modified to be consistent with the classification used in data sets other than the census, while the age data were adjusted to correspond with the April 1, 1990 census data. These modified data are consistent with the counts of the 1990 census as enumerated. Information about modified data for States are available from Data Users Services Division (301-763-4100). Further information about the modifications in the CPH-L-74 series are available from David Word (301-763-1739) or Gregory Spencer (301-763-1902).

### Race Modification.

There were 9.8 million "Other race" persons included in the 1990 census. Over 98 percent were of Hispanic origin. Such non-specified race persons are not found in data sources other than the census and the category is inconsistent with the Office of Management and Budget Directive 15. The CPH-L-74 series assigns each "other race" person to a specified race. The race assignment rule was: assign each "other

race" person to the specified race reported by a nearby persons with an identical response to the Hispanic-origin question.

The assignment of a specified race was made on an individual basis. That is, no effort was made to minimize racial heterogeneity within households.

### Age Modification.

The following is a portion of the text of a user note which is incorporated in 1990 census products:

Review of detailed 1990 information indicated that respondents tended to provide their age as of the date of completion of the questionnaire, not their age as of April 1, 1990. In addition, there may have been a tendency for respondents to round up their age if they were close to having a birthday. It is likely that approximately 10 percent of persons in most age groups are actually one year younger. For most single years of age, the misstatements are largely offsetting. The problem is most pronounced at age 0 because persons lost to age 1 may not have been fully offset by the inclusion of babies born after April 1, 1990 and because there

may have been more rounding up to age 1 to avoid reporting age as 0 years. Age in completed months was not collected for infants under age 1. The reporting of age one year older than age on April 1, 1990 is likely to have been greater in areas where the census data were collected later in 1990.

About 95 percent of the population provided acceptable birth year responses where were adjusted with the following procedures. The age data for individuals in households were modified by adjusting the reported birth year data by race and sex to correspond with the national level quarterly distribution of births available from the National Center for Health Statistics. Approximately 100 million persons have an age in this modified file which is one year different from what they marked in the 1990 census.

The modification procedure was done separately for each birth year, by sex, for the White, Black, Asian or Pacific Islander, and American Indian, Eskimo, and Aleut populations.

The modification procedure was done separately for each birth year, by sex, for the White, Black, Asian or Pacific Islander, and American Indian, Eskimo, and Aleut populations.

U.S. Department of Commerce  
BUREAU OF THE CENSUS  
Washington, D.C. 20233

**Official Business**  
**Penalty for Private Use, \$300**

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

