Appendix B.
Source and Accuracy of Estimates

Source of Data

Estimates in this report primarily come from data obtained from the Current Population Survey (CPS) conducted in March of 1980 through 1993. Some estimates come from 1960 through 1990 decennial census data. The Bureau of the Census conducts the CPS survey every month, although this report uses mostly the March survey data. Data from November 1992 were used for the voting estimates. The March and November CPS 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. The United States was divided into 1,973 geographic areas. In most states, a geographic area consisted of a county or several contiguous counties. In some areas of New England and Hawaii, minor civil divisions are used instead of counties. A total of 729 geographic areas was selected for sample. 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.

Table B-1.
Description of Current Population Survey

Time period	Number of sample areas	Housing units eligible[1]	
		Interviewed	Not interviewed
1990 to 1993	729	57,400	2,600
1989	729	53,600	2,500
1986 to 1988	729	57,000	2,500
1985	[2]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	46,500	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

[1]Excludes about 2,500 Hispanic households added in March from the previous November sample. (See "March Supplement.")
[2]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.

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.

Table B-1 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 1992 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 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.

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 1990 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.

A common measure of survey coverage is the coverage ratio, the estimated population before the post-stratification ratio estimate divided by the independent population control. Table B-2 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 B-2.
CPS Coverage Ratios

Age	Non-Black		Black		All Persons		
	Male	Female	Male	Female	Male	Female	Total
0-14	.929	.964	.850	.838	.916	.943	.929
15	.933	.895	.763	.824	.905	.883	.895
16-19	.881	.891	.711	.802	.855	.877	.866
20-29	.847	.897	.660	.811	.823	.884	.854
30-39	.904	.931	.680	.845	.877	.920	.899
40-49	.928	.966	.816	.911	.917	.959	.938
50-59	.953	.974	.896	.927	.948	.969	.959
60-64	.961	.941	.954	.953	.960	.942	.950
65-69	.919	.972	.982	.984	.924	.973	.951
70+	.993	1.004	.996	.979	.993	1.002	.998
15+	.914	.945	.767	.874	.898	.927	.918
0+	.918	.949	.793	.864	.902	.931	.921

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 began to be used for CPS estimates beginning with the 1994 surveys.

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 subpopulation 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-3.

Table B-3.
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
Educational Attainment—March 1992 and 1993				
Total or White	−0.000021	2,532	−0.000011	1,899
Black	−0.000247	3,425	−0.000071	1,716
Hispanic	−0.000371	3,425	−0.000142	1,716
Geographical Mobility—March 1993				
Total or White	−0.000025	7,130	-0.000011	1,899
Black	−0.000025	7,130	-0.000071	1,716
Hispanic	−0.000589	7,130	-0.000142	1,716
Marital Status—March 1993				
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
Voting—November 1992				
Total or White	−0.000017	3,011	-0.000011	1,899
Black	−0.000216	4,408	-0.000084	1,716
Hispanic	−0.000540	7,428	-0.000210	2,892
Poverty—1992				
Total or White	−0.000040	9,502	−0.000093	2,243
Black	−0.000322	9,502	−0.000093	2,243
Hispanic	−0.000470	9,502	−0.000093	2,243
65 and over	-0.000113	3,607	(X)	(X)
Income—1992				
Total or White	−0.000012	2,254	−0.000012	2,058
Black	−0.000122	2,577	−0.000109	2,243
Hispanic	−0.000182	2,577	−0.000175	2,243

Note: Multiply the above parameters by 0.83, 0.93, 0.98, and 1.37 for the Northeast, Midwest, South, and West, respectively. Multiply the above parameters by 1.5 for outside metropolitan.

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 (±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.

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

Here, x is the size of the estimate and a and b are the parameters in table B-3 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 1992 presidential election. Use the appropriate parameters from table B-3 and formula (1) to get

Number, x	19,818,000
a parameter	−0.000017
b parameter	3,011
Standard error	230,000
90% conf. int.	19,440,000 to 20,196,000

The standard error is calculated as

$$s_x = \sqrt{-0.000017 \times 19{,}818{,}000^2 + 3{,}011 \times 19{,}818{,}000}$$

$$= 230{,}000$$

The 90-percent confidence interval is calculated as 19,818,000 ± 1.645x230,000.

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-3 indicated by the numerator.

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

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

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-3 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-3 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 $39.7 \pm 1.645 \times 0.8$.

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

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

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-3 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	2,532	2,532	-
Standard error	82,000	87,000	120,000
90% conf. int.	2,612,000 to 2,882,000	2,908,000 to 3,194,000	107,000 to 501,000

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{82{,}000^2 + 87{,}000^2} = 120{,}000$$

The 90-percent confidence interval around the difference is calculated as $304{,}000 \pm 1.645 \times 120{,}000$. 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.

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

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

Data Quality

Data users always 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 in a census or survey. 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. Sampling error can be measured based on the actual sample observed. In the census, about one in six households and one in six 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. 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 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.[1]

For 1990, results from demographic analysis show different coverage patterns for males and females. For females, 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 results indicate a relatively large undercount of persons 85 years and over. For males, 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). These results, especially for the group 85 years and over are subject to change based on further research. There are problems with the files used for comparison (for example, Medicare files do not purge all deaths).[2]

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.

[1]U.S. Bureau of the Census, ''The Coverage of Population in the 1980 Census,'' Evaluation and Research Reports, PHC80-E4, Washington, DC, U.S. Government Printing Office, 1988, table 3.3.

[2]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.

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.

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 1990 census, if there had been no imputation for missing data, 14.7 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 decreased to 12.6.

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.'' 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.[3]

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 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 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

[3]Greville developed an adjustment technique described in Mortimer Speigelman, Introduction to Demography, rev. ed., Cambridge, MA: Harvard University Press, 1968. Speigelman 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 essentially correct if both population and deaths were biased in the same direction and in about the same proportion.

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.[4] 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 include 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 contributes to more meaningful understanding of the census results.

[4]J. Gregory Robinson, Prithwis Das Gupta, and Bashir Ahmed, U.S. Bureau of the Census, ''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.

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 are available for single years to the end category, ''105 years and over'' for the United States, and sub-state statistical and administrative 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 is available (and is known as ''PUMSO'').

Finally, reports have been issued that evaluate the quality of 1990 census data. These reports focus on coverage and content evaluation and provide additional insight into the uses and limitations of data on America's 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); Outreach Survey (respondent attitudes towards and the census); and ethnographic studies on response and coverage problems.

Table B-4
Items in the 1990 Census

I. Information collected from households:[1]

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

II. Information collected from a sample of households:[1]

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

[1]Persons in group quarters, including institutions, are asked population items only.