Source and Accuracy of Estimates for Income, Poverty, and Health Insurance Coverage in the United States: 2003

SOURCES OF DATA

The estimates in the report *Income, Poverty, and Health Insurance Coverage in the United States: 2003* come from the 2004 Annual Social and Economic Supplement (ASEC). The Census Bureau conducts the ASEC over a 3-month period, in February, March, and April, with most data collection occurring in the month of March. The ASEC uses two sets of questions: the basic Current Population Survey (CPS) and a set of supplemental questions. The CPS, sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics, is the country's primary source of labor force statistics for the entire population. The ASEC is also sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics.

Basic CPS. The monthly CPS collects primarily labor force data about the civilian noninstitutional population living in the United States. Interviewers ask questions concerning labor force participation about each member 15 years old and over in sample households.

The CPS uses a multi-stage probability sample design with coverage in all 50 states and the District of Columbia. The sample was selected from 1990 Decennial Census files and is continually updated to account for new residential construction. To obtain the sample, the United States was divided into 2,007 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. These 2,007 geographic areas were then grouped into 754 strata, and one geographic area was selected from each stratum.

From the 754 strata, approximately 73,000 households are in sample in March. Based on eligibility criteria, 11 percent of these households are sent directly to Computer-Assisted Telephone Interviewing (CATI). The remaining units are assigned to interviewers for Computer-Assisted Personal Interviewing (CAPI).[1] Of all housing units in sample, about 60,200 are determined to be eligible for interview. Interviewers obtain interviews at about 55,000 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

The Annual Social and Economic Supplement. In addition to the basic CPS questions, interviewers asked supplementary questions for the ASEC. They ask these questions of the civilian noninstitutional population and also of military personnel who live in households with at least one other civilian adult. The additional questions cover the following topics:

- Household and Family Characteristics
- Marital Status
- Geographic Mobility
- Foreign Born Population
- Income from the previous calendar year
- Poverty
- Work Status/Occupation
- Health Insurance Coverage
- Program Participation
- Educational Attainment

Including the respondents from the basic CPS sample, approximately 99,000 housing units are in sample for the ASEC. About 84,500 are determined to be eligible for interview and about 77,600 interviews are obtained.

The additional sample for the ASEC provides more reliable data for Hispanic households, non-Hispanic minority households, and non-Hispanic White households with children 18 years or younger. These households were identified for sample from previous months and the following April. For more information about the households eligible for the ASEC, please refer to:

> Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002, <www.census.gov/prod/2002pubs/tp63rv.pdf>.

[1] For further information on CATI and CAPI and the eligibility criteria, please see: Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002, <www.census.gov/prod/2002pubs/tp63rv.pdf>.

Sample Redesign. Since the introduction of the CPS, the Census Bureau has redesigned the CPS sample several times. These redesigns have improved the quality and accuracy of the data and have satisfied changing data needs. The Census Bureau completely implemented the most recent changes in July 1995.

Estimation Procedure. This survey's estimation procedure adjusts weighted sample results to agree with independent estimates of the civilian noninstitutional population of the United States by age, sex, race, Hispanic ancestry, and state of residence. The independent estimates are based on:

- The 2000 Decennial Census of Population and Housing.
- Statistics on births, deaths, immigration, and emigration.
- Statistics on the size of the armed forces.

The estimation procedure for the ASEC supplement includes a further adjustment so husband and wife of a household receive the same weight. The independent population estimates include some, but not all, unauthorized migrants.

ACCURACY OF ESTIMATES

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

Sampling Error. Since the CPS estimates come from a sample, they may differ from figures from an enumeration of the entire population using the same questionnaires, instructions, and enumerators. For a given estimator, the difference between an estimate based on a sample and the estimate that would result if the sample were to include the entire population is known as sampling error. Standard errors, as calculated by methods described in "Standard Errors and their Use," are primarily measures of the magnitude of sampling error. However, they may include some nonsampling error.

Nonsampling Error. For a given estimator, the difference between the estimate that would result if the sample were to include the entire population and the true population value being estimated is known as nonsampling error. Sources of nonsampling errors include the following:

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

Answers to questions about money income often depend on the memory or knowledge of one person in a household. Recall problems can cause underestimates of income in survey data because it is easy to forget minor or irregular sources of income. Respondents may also misunderstand what the Census Bureau considers money income or may simply be unwilling to answer these questions correctly because the questions are considered too personal. See Appendix C, Current Population Reports, Series P60-184, *Money Income of Households, Families, and Persons in the United States: 1992* for more details.

To minimize these errors, the Census Bureau employs quality control procedures in sample selection, wording of questions, interviewing, coding, data processing, and data analysis.

Nonresponse. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the cases eligible for the 2004 ASEC, the basic CPS nonresponse rate was 8.5 percent. The nonresponse rate for the Annual Social and Economic Supplement was an additional 8.2 percent. These two nonresponse rates lead to a combined supplement nonresponse rate of 16.0 percent.

Coverage. The concept of coverage in the survey sampling process is the extent to which the total population that could be selected for sample "covers" the survey's target population. CPS undercoverage results from missed housing units and missed people within sample households. Overall CPS undercoverage for March 2004 is estimated to be about 11 percent. Undercoverage varies with age, sex, and race. Generally, undercoverage is larger for males than for females and larger for Blacks than for non-Blacks.

The Current Population Survey weighting procedure partially corrects for bias due to undercoverage, but biases may still be present when people who are

Table 1.
March 2004 CPS Coverage Ratios

Age	All People			White Only		Black Only		Residual Race		Hispanic[1]	
	Total	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0 to 15 years	0.91	0.91	0.91	0.93	0.92	0.80	0.83	0.95	1.00	0.95	0.95
16 to 19 years	0.87	0.88	0.86	0.90	0.86	0.75	0.76	0.91	1.02	1.01	0.86
20 to 24 years	0.78	0.75	0.82	0.77	0.83	0.64	0.77	0.76	0.77	0.73	0.85
25 to 34 years	0.82	0.79	0.86	0.81	0.88	0.67	0.79	0.78	0.78	0.74	0.88
35 to 44 years	0.88	0.85	0.91	0.87	0.93	0.74	0.79	0.82	0.87	0.82	0.89
45 to 54 years	0.92	0.90	0.93	0.92	0.94	0.76	0.88	0.92	0.98	0.82	0.92
55 to 64 years	0.93	0.93	0.93	0.93	0.93	0.91	0.91	0.89	0.85	0.91	0.91
65 years and older	0.92	0.93	0.91	0.92	0.90	0.93	1.01	1.08	0.84	0.77	0.85
15 years and older	0.88	0.86	0.90	0.88	0.91	0.76	0.84	0.86	0.87	0.81	0.89
0 years and older	0.89	0.87	0.90	0.89	0.91	0.77	0.84	0.88	0.90	0.85	0.91

Note: The Residual Race group includes cases indicating a single race other than White or Black, and cases indicating two or more races.

[1]Hispanics may be of any race.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

missed by the survey differ from those interviewed in ways other than age, race, sex, Hispanic ancestry, and state of residence. How this weighting procedure affects other variables in the survey is not precisely known. All of these considerations affect comparisons across different surveys or data sources.

A common measure of survey coverage is the coverage ratio, calculated as the estimated population before post-stratification divided by the independent population control. Table 1 shows March 2004 CPS coverage ratios for certain age-sex-race-ancestry groups. The CPS coverage ratios can exhibit some variability from month to month.

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

Data users should also use caution when comparing estimates for 1999 to 2003 in *Income, Poverty, and Health Insurance Coverage in the United States: 2003* (which reflect Census 2000-based population controls) with estimates for 1992 to 1998 (from March 1993 CPS to March 1999 CPS), which reflect 1990 census-based population controls and with estimates for 1991 (from March 1992 CPS) and earlier years, which reflect 1980 census-based population controls. Although this change in population controls had relatively little impact on summary measures, such as averages, medians, and percentage distributions, it did have a significant impact on levels. For example, use of Census 2000-based population controls results in about a one percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 2000 and later years will differ from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

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

Based on the results of each decennial census, the Census Bureau gradually introduces a new sample design for the CPS. During this phase-in period, the Census Bureau collects CPS data from sample designs based on different censuses. While most CPS estimates have been unaffected by this mixed sample, geographic estimates are subject to greater error and variability. Users should exercise caution when comparing metropolitan/nonmetropolitan estimates across years with a design change. For more information, see Appendix C, Current Population Reports, Series P60-193, *Money Income in the United States: 1995 (With Separate Data on Valuation of Noncash Benefits).*

A Nonsampling Error Warning. Since the full extent of the nonsampling error is unknown, one should be particularly careful when interpreting results based on small differences between estimates. Even a small amount of nonsampling error can cause a borderline

difference to appear significant or not, thus distorting a seemingly valid hypothesis test. Caution should also be used when interpreting results based on a relatively small number of cases. Summary measures (such as medians and percentage distributions) probably do not reveal useful information when computed on a subpopulation smaller than 75,000.

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

- Statistical Policy Working Paper 3, *An Error Profile: Employment as Measured by the Current Population Survey*, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1978, <www.fcsm.gov/working-papers/spp.html>.
- Technical Paper 63RV, *The Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002, <www.census.gov/prod/2002pubs/tp63rv.pdf>.

Estimation of Median Incomes. The Census Bureau has changed the methodology for computing median income over time. The Census Bureau has computed medians using either Pareto interpolation or linear interpolation. Currently, we are using linear interpolation to estimate all medians. Pareto interpolation assumes a decreasing density of population within an income interval; whereas, linear interpolation assumes a constant density of population within an income interval. The Census Bureau calculated estimates of median income and associated standard errors for 1979 through 1987 using Pareto interpolation if the estimate was larger than $20,000 for people or $40,000 for families and households. This is because the width of the income interval containing the estimate is greater than $2,500.

We calculated estimates of median income and associated standard errors for 1976, 1977, and 1978 using Pareto interpolation if the estimate was larger than $12,000 for people or $18,000 for families and households. This is because the width of the income interval containing the estimate is greater than $1,000. All other estimates of median income and associated standard errors for 1976 through 2003 and almost all of the estimates of median income and associated standard errors for 1975 and earlier were calculated using linear interpolation.

Thus, use caution when comparing median incomes above $12,000 for people or $18,000 for families and households for different years. Median incomes below those levels are more comparable from year to year since they have always been calculated using linear interpolation. For an indication of the comparability of medians calculated using Pareto interpolation with medians calculated using linear interpolation, see Series P-60, No. 114, *Money Income in 1976 of Families and Persons in the United States.*

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

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

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

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

The tables in *Income, Poverty, and Health Insurance Coverage in the United States: 2003* list estimates followed by a number labeled "90-percent confidence interval (+/-)." This number can be added to or subtracted from the estimates to calculate upper and lower bounds of the 90-percent confidence interval. For example, Table 5 in *Income, Poverty, and Health Insurance Coverage in the United States: 2003* shows the numbers for health insurance. For the statement "the percent of people without health insurance rose to 15.6 percent in 2003," the 90-percent confidence

interval for the estimate, 15.6 percent, is 15.6 (± 0.2) percent, or 15.4 percent to 15.8 percent. The tables also display asterisks in the last columns for significant differences.

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

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

Generalized Variance Parameters. It is possible to compute and present an estimate of the standard error based on the survey data for each estimate in a report, but there are a number of reasons why this is not done. A presentation of the individual standard errors would be of limited use, since one could not possibly predict all of the combinations of results that may be of interest to data users. Additionally, variance estimates are based on sample data and have variances of their own. Therefore, some method of stabilizing these estimates of variance, for example, by generalizing or averaging over time, may be used to improve their reliability.

Experience has shown that certain groups of estimates have a similar relationship between their variance and expected value. Modeling or generalization may provide more stable variance estimates by taking advantage of these similarities. The generalized variance function is a simple model that expresses the variance as a function of the expected value of the survey estimate. The parameters of the generalized variance function are estimated using direct replicate variances. These generalized variance parameters provide a relatively easy method to obtain approximate standard errors for numerous characteristics. In this source and accuracy statement, Table 2 provides generalized variance parameters for characteristics from the ASEC data. Table 3 provides factors to approximate parameters for CPS estimates prior to 2003. Table 4 provides parameters for CPS Hispanic estimates prior to 1984. Table 5 provides ASEC parameters for income and nonincome characteristics for Asian and Pacific Islanders, American Indian and Alaskan Natives, Native Hawaiian and Other Pacific Islanders, and people of two or more races.

Tables 6 and 7 contain the year-to-year correlation coefficients for poverty and income characteristics. Table 8 contains the correlation coefficients for comparing race categories that are subsets of one another.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number shown in *Income, Poverty, and Health Insurance Coverage in the United States: 2003* can be obtained using the formula:

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

Here x is the size of the estimate and a and b are the parameters in Tables 2, 4, and 5 associated with the particular type of characteristic. When calculating standard errors from cross-tabulations involving different characteristics, use the set of parameters for the characteristic that will give the largest standard error.

Illustration No. 1

In *Income, Poverty, and Health Insurance Coverage in the United States: 2003*, Table 3 shows that there were 35,861,000 people below poverty in the United States in 2003. Use the appropriate parameters from Table 2 and Formula (1) to get

Number, x	35,861,000
a parameter	-0.000018
b parameter	5,282
Standard error	408,000
90% conf. int.	35,190,000 to 36,532,000

The standard error is calculated as

$$s_x = \sqrt{-0.000018 \times 35{,}861{,}000^2 + 5{,}282 \times 35{,}861{,}000} = 408{,}000$$

The 90-percent confidence interval is calculated as 35,861,000 ± 1.645 × 408,000.

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 both 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

Table 2.
CPS Standard Error Parameters for Income and Nonincome Characteristics: 2003

Characteristic	Total or White		Black		Hispanic	
	a	b	a	b	a	b
BELOW POVERTY LEVEL						
People						
Total	-0.000018	5,282	-0.000065	5,282	-0.000131	5,282
Male	-0.000038	5,282	-0.000133	5,282	-0.000256	5,282
Female	-0.000036	5,282	-0.000127	5,282	-0.000268	5,282
Age						
Under 15	-0.000063	4,072	-0.000167	4,072	-0.000325	4,072
Under 18	-0.000049	4,072	-0.000134	4,072	-0.000269	4,072
15 and over	-0.000021	5,282	-0.000080	5,282	-0.000164	5,282
15 to 24	-0.000032	1,998	-0.000094	1,998	-0.000187	1,998
25 to 44	-0.000024	1,998	-0.000080	1,998	-0.000149	1,998
45 to 64	-0.000029	1,998	-0.000138	1,998	-0.000328	1,998
65 and over	-0.000058	1,998	-0.000380	1,998	-0.000961	1,998
Households, Families, and Unrelated Individuals						
Total	0.000052	1,243	0.000052	1,243	0.000052	1,243
ALL INCOME LEVELS						
People						
Total	-0.000005	1,249	-0.000022	1,430	-0.000044	1,430
Male	-0.000010	1,249	-0.000045	1,430	-0.000087	1,430
Female	-0.000010	1,249	-0.000042	1,430	-0.000091	1,430
Age						
15 to 24	-0.000020	1,249	-0.000067	1,430	-0.000134	1,430
25 to 44	-0.000015	1,249	-0.000058	1,430	-0.000107	1,430
45 to 64	-0.000018	1,249	-0.000099	1,430	-0.000235	1,430
65 and over	-0.000036	1,249	-0.000272	1,430	-0.000688	1,430
Households, Families, and Unrelated Individuals						
Total	-0.000005	1,140	-0.000019	1,245	-0.000039	1,245
NONINCOME CHARACTERISTICS						
People						
Employment Status	-0.000008	1,586	-0.000154	3,296	-0.000187	3,296
Educational Attainment	-0.000005	1,206	-0.000021	1,364	-0.000029	922
Health Insurance	-0.000009	2,652	-0.000047	3,809	-0.000094	3,809
Total, Marital Status, Other						
Some household members	-0.000009	2,652	-0.000047	3,809	-0.000094	3,809
All household members	-0.000011	3,222	-0.000069	5,617	-0.000139	5,617
Households, Families, and Unrelated Individuals						
Total	-0.000004	1,052	-0.000014	952	-0.000030	952

Notes: To obtain parameters prior to 2003, multiply by the appropriate factor in Table 3. For nonmetropolitan residence categories, multiply the a and b parameters by 1.5. For foreign-born and noncitizen characteristics for Total and White, multiply the a and b parameters by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Blacks and Hispanics. For regional estimates, multiply the a and b parameters by 0.89, 0.91, 1.14, and 1.23 for Northeast, Midwest, South, and West, respectively. The Total or White and Black parameters are to be used for both alone and in combination race group estimates. Hispanics may be of any race.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table 3.
CPS Year Factors for Income and Poverty Estimates (1946-2002)

Year of Estimate	Total/White	Black[1]		Hispanic
	a and b	a	b	a and b
2002	1.00	1.00	1.00	1.00
2000(expanded)-2001	1.00	2.22	1.00	1.00
1995-2000(basic)	1.97	4.36	1.97	1.97
1989-1994	1.82	4.03	1.82	1.82
1988	2.02	4.48	2.02	2.12
1984-1987	1.70	3.77	1.70	1.70
1981-1983	1.70	3.77	1.70	2.38
1972-1980	1.52	3.37	1.52	2.13
1966-1971	1.52	3.37	1.52	3.58
1956-1965	2.28	5.06	2.28	5.38
1946-1955	3.42	7.58	3.42	8.07

Note: For races not listed, use the factors for total.

[1]Blacks have separate factors for the a and b parameter factors due to the new race definitions and how they affected the population control totals.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

the parameter from Table 2, 4, or 5 as indicated by the numerator. However, for calculating standard errors for different characteristics of families in poverty, use the standard error of a ratio equation (see formula (4) in "Standard Errors of Ratios").

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

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

Here x is the total number of people, families, households, or unrelated individuals in the base of the percentage, p is the percentage ($0 \le p \le 100$), and b is the parameter in Table 2, 4, or 5 associated with the characteristic in the numerator of the percentage.

Illustration No. 2

In 2003, there were 44,961,000 out of 288,280,000 people, or 15.6 percent, who did not have health insurance. Use the appropriate parameter from Table 2 and Formula 2 to get

Percentage, p	15.6
Base, x	288,280,000
b parameter	2,652
Standard error	0.1
90% conf. int.	15.4 to 15.8

The standard error is calculated as

$$s_{x,p} = \sqrt{\frac{2{,}652}{288{,}280{,}000} \times 15.6 \times (100-15.6)} = 0.1$$

The 90-percent confidence interval of the percentage of people without health insurance is calculated as $15.6 \pm 1.645 \times 0.1$.

Standard Errors of Differences. The standard error of the difference between two sample estimates is approximately equal to

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

where s_x and s_y are the standard errors of the estimates, x and y. The estimates can be numbers, percentages, ratios, etc. Tables 6 and 7 contain the correlation coefficient, r, for year-to-year comparisons for CPS poverty, income, and health insurance estimates of numbers and proportions. Table 8 contains the correlation coefficient, r, for making comparisons between race categories that are subsets of one another. For example, to compare the number of people in poverty who listed White as their only race to the number of people in poverty who are White in combination with another race, a correlation coefficient is needed to account for the large overlap between the two groups. For making other comparisons (including

Table 4.
CPS Standard Error Parameters for Poverty, Income, and Nonincome Characteristics of Hispanics: 1972 to 1983

Characteristics	1972-1980		1981-1983	
	a	b	a	b
BELOW POVERTY LEVEL				
People				
Total	–0.000063	11,528	–0.001131	12,901
Male	–0.000130	11,528	–0.002307	12,901
Female	–0.000123	11,528	–0.002219	12,901
Age				
Under 15	–0.000052	6,057	–0.001399	6,778
Under 18	–0.000044	6,057	–0.001184	6,778
15 and over	–0.000032	11,528	–0.000421	12,901
15 to 24	–0.000122	4,520	–0.001414	5,058
25 to 44	–0.000097	4,520	–0.000962	5,058
45 to 64	–0.000117	4,520	–0.002147	5,058
65 and over	–0.000153	4,520	–0.006068	5,058
Households, Families, and Unrelated Individuals				
Total	–0.000014	2,420	–0.000237	2,708
ALL INCOME LEVELS				
People				
Total	–0.000020	3,000	–0.000301	3,357
Male	–0.000043	3,000	–0.000615	3,357
Female	–0.000038	3,000	–0.000591	3,357
Age				
15 to 24	–0.000080	3,000	–0.000961	3,357
25 to 44	–0.000065	3,000	–0.000668	3,357
45 to 64	–0.000077	3,000	–0.001459	3,357
65 and over	–0.000147	3,000	–0.004124	3,357
Households, Families, and Unrelated Individuals				
Total	–0.000014	2,420	–0.000237	2,708
Households with children under age 18	–0.000014	2,420	–0.000237	2,708
NONINCOME CHARACTERISTICS				
People				
Employment status	(X)	(X)	(X)	(X)
Educational attainment	–0.000015	2,344	–0.000152	2,623
Total, Marital Status, Other				
Some household members	–0.000026	5,069	–0.000294	5,673
All household members	–0.000044	10,199	–0.000592	11,414
Households, Families, and Unrelated Individuals				
Total	–0.000020	1,626	–0.000022	1,820

Note: Data users should multiply the a and b parameters by 1.5 for nonmetropolitan residence categories. The Census Bureau did not publish income data for Hispanics before 1972.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table 5.
CPS Standard Error Parameters for Income and Nonincome Characteristics of Asians and Pacific Islanders, American Indians and Alaskan Natives, Native Hawaiian & Other Pacific Islanders, and People of Two or More Races: 2003

Characteristic	API, AIAN, NH & OPI		Two or More Races	
	a	b	a	b
BELOW POVERTY LEVEL				
People				
Total	-0.000090	5,282	-0.000113	5,282
Male	-0.000179	5,282	-0.000222	5,282
Female	-0.000182	5,282	-0.000231	5,282
Age				
Under 15	-0.000236	4,072	-0.000275	4,072
Under 18	-0.000183	4,072	-0.000221	4,072
15 and over	0.000111	5,282	-0.000143	5,282
15 to 24	-0.000134	1,998	-0.000160	1,998
25 to 44	-0.000104	1,998	-0.000134	1,998
45 to 64	-0.000202	1,998	-0.000279	1,998
65 and over	-0.000578	1,998	-0.000816	1,998
Households, Families, and Unrelated Individuals				
Total	0.000052	1,243	0.000052	1,243
ALL INCOME LEVELS				
People				
Total	-0.000030	1,430	-0.000039	1,430
Male	-0.000060	1,430	-0.000076	1,430
Female	-0.000061	1,430	-0.000079	1,430
Age				
15 to 24	-0.000096	1,430	-0.000115	1,430
25 to 44	-0.000074	1,430	-0.000096	1,430
45 to 64	-0.000145	1,430	-0.000199	1,430
65 and over	-0.000413	1,430	-0.000584	1,430
Households, Families, and Unrelated Individuals				
Total	-0.000026	1,245	-0.000034	1,245
NONINCOME CHARACTERISTICS				
People				
Employment Status	-0.000272	2,749	-0.000154	3,296
Educational Attainment	-0.000029	1,364	-0.000037	1,364
Health Insurance	-0.000065	3,809	-0.000082	3,809
Total, Marital Status, Other				
Some household members	-0.000065	3,809	-0.000082	3,809
All household members	-0.000096	5,617	-0.000120	5,617
Households, Families, and Unrelated Individuals				
Total	-0.000020	952	-0.000026	952

Notes: To obtain parameters prior to 2003, multiply by the appropriate factor in Table 3. For nonmetropolitan residence categories, multiply the a and b parameters by 1.5. For regional estimates, multiply the a and b parameters by 0.89, 0.91, 1.14, and 1.23 for Northeast, Midwest, South, and West respectively. Income data for Asians and Pacific Islanders and American Indians and Alaskan Natives were not collected prior to 1988. Data on people of two or more races and Native Hawaiian Islanders & Other Pacific Islanders were not collected prior to 2003 (2002 estimates). The API, AIAN, HN & OPI parameters are to be used for both alone and in combination race group estimates.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table 6.
CPS Year-to-Year Correlation Coefficients for Poverty Estimates: 1970 to 2003

Characteristics	1972-83, 1984-2000 (basic) or 2000 (expanded)-2003		1999 (basic)-2000 (expanded)		1983-1984		1971-1972		1970-1971	
	People	Families	People	Families	People	Families	People	Families	People	Families
Total	**0.45**	**0.35**	**0.29**	**0.22**	**0.39**	**0.30**	**0.15**	**0.14**	**0.31**	**0.28**
White	0.35	0.30	0.23	0.20	0.30	0.26	0.14	0.13	0.28	0.25
Black	0.45	0.35	0.23	0.18	0.39	0.30	0.17	0.16	0.35	0.32
Other	0.45	0.35	0.22	0.17	0.30	0.30	0.17	0.16	0.35	0.32
Hispanic[1]	0.65	0.55	0.52	0.40	0.56	0.47	0.17	0.16	0.35	0.32

[1]Hispanics may be of any race.

Note: These correlations are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlations are zero.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

Table 7.
CPS Year-to-Year Correlation Coefficients for Income and Health Insurance Estimates: 1960 to 2003

Characteristics	1960-2000 (basic) 2000 (expanded)-2003		1999 (basic)-2000 (expanded)	
	People	Families, households, and unrelated individuals	People	Families, households, and unrelated individuals
Total	**0.30**	**0.35**	**0.19**	**0.22**
White	0.30	0.35	0.20	0.23
Black	0.30	0.35	0.15	0.18
Other	0.30	0.35	0.15	0.17
Hispanic[1]	0.45	0.55	0.36	0.28

[1]Hispanics may be of any race.

Note: These correlations are for comparisons of consecutive years. For comparisons of nonconsecutive years, assume the correlatoins are zero.

Source: U.S. Census Bureau, Demographic Statistical Methods Division.

race overlapping where one group is not a complete subset of the other), assume that r equals zero. Making this assumption will result in accurate estimates of standard errors for the difference between two 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 No. 3

In *Income, Poverty, and Health Insurance Coverage in the United States: 2003*, Table 5 shows that the number of people without health insurance in 2003 was 44,961,000 and in 2002 was 43,574,000. The apparent difference is 1,387,000. Use the appropriate parameters, correlation coefficients, and factors from tables 2, 3, and 7 and formulas (1) and (3) to get

	x	y	difference
estimate, x	44,961,000	43,574,000	1,387,000
a parameter	-0.000009	-0.000009	-
b parameter	2,652	2,652	-
correlation, r	-	-	.30
Standard error	318,000	314,000	374,000
90% conf. int.	44,438,000	43,057,000	772,000
	to	to	to
	45,484,000	44,091,000	2,002,000

Table 8.
CPS Correlation Coefficients for a Subsetted Race: 2003

Race 1	Race 2	r
White alone	White alone or in combination	0.99
White alone, not Hispanic	White alone or in combination	0.80
Black alone	Black alone or in combination	0.98
Asian alone	Asian alone or in combination	0.96

Source: US Census Bureau, Demographic Statistical Methods Division.

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{318{,}000^2 + 314{,}000^2 - 2 \times 0.30 \times 318{,}000 \times 314{,}000} = 374{,}000$$

The 90-percent confidence interval around the difference is calculated as 1,387,000 ± 1.645 × 374,000. Since this interval does not include zero, we can conclude with 90 percent confidence that the number of people without health insurance in 2003 was higher than the number of people without health insurance in 2002.

Standard Errors of Ratios. Certain estimates may be calculated as the ratio of two numbers. Compute the standard error of a ratio, x/y, using

$$s_{x/y} = \frac{x}{y}\sqrt{\left(\frac{s_x}{x}\right)^2 + \left(\frac{s_y}{y}\right)^2 - 2r\frac{s_x s_y}{xy}} \qquad (4)$$

The standard error of the numerator, s_x, and that of the denominator, s_y, may be calculated using formulas described earlier. In formula (4), r represents the correlation between the numerator and the denominator of the estimate.

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

For year-to-year and subsetted race correlation coefficients see "Standard Errors of Differences" and Tables 6, 7, and 8. For all other types of ratios, r is assumed to be zero. If r is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of the ratio. Examples of this type are the average number of children per family and the family poverty rate.

Note: For estimates expressed as the ratio of x per 100 y or x per 1,000 y, multiply formula (4) by 100 or 1,000, respectively, to obtain the standard error.

Illustration No. 4

Suppose the number of families below the poverty level, x, was 7,607,000 and the total number of families, y, was 76,232,000. The ratio of families below the poverty level to the total number of families would be 0.100 or 10.0 percent. Use the appropriate parameters from Table 2 and formulas (1) and (4) with r = 0 to get

	x	y	ratio
estimate, x	7,607,000	76,232,000	0.100
a parameter	+0.000052	-0.000004	-
b parameter	1,243	1,052	-
Standard error	112,000	239,000	0.002
90% conf. int.	7,423,000	75,839,000	0.097
	to	to	to
	7,791,000	76,625,000	0.103

The standard error is calculated as

$$s_{x/y} = \frac{7{,}607{,}000}{76{,}232{,}000}\sqrt{\left(\frac{112{,}000}{7{,}607{,}000}\right)^2 + \left(\frac{239{,}000}{76{,}232{,}000}\right)^2} = 0.002$$

The 90-percent confidence interval is calculated as 0.100 ± 1.645 × 0.002.

Other Standard Errors. In the report *Income, Poverty, and Health Insurance Coverage in the United States: 2003*, nine tables provide confidence intervals for most of the estimates discussed in the text. For other estimates, the standard errors can be calculated using the formulas in this Source and Accuracy Statement. For more information or questions on calculating standard errors, e-mail Jana Shepherd at <dsmd.source.and.accuracy@census.gov>.