UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

April 23, 2009

MEMORANDUM FOR: Cheryl R. Landman
Chief, Demographic Surveys Division

From: Ruth Ann Killion
Chief, Demographic Statistical Methods Division

Final DSMD Signed

Subject: Source and Accuracy Statement for the June 2008 CPS Microdata File on Fertility and Birth Expectation

Attached is the statement on the source of the data and accuracy of the estimates for the June 2008 CPS Microdata File on Fertility and Birth Expectation.

If you have any questions or need additional information, please contact JulieWalker of the Demographic Statistical Methods Division via email at dsmd.source.and.accuracy@census.gov.

Attachment

cc:

L. Clement (DSD)
G. Weyland
K. Woods
B. Kominski (HHES)
M. O'Connell (POP)
J. Dye
W. Savino (ACSD)
A. Shields

email:

P. Flanagan (DSMD)
J. Scott
S. Adeshiyan
J. View
C. Loudermilk
X. Liu
B. Tran
T. Moore
HSSB (9)

Source of the Data and Accuracy of the Estimates for the June 2008 CPS Microdata File on Fertility and Birth Expectation

Table of Contents

Tables

Source of the Data and Accuracy of the Estimates for the June 2008 CPS Microdata File on Fertility and Birth Expectation

SOURCE OF THE DATA

The data in this microdata file are from the June 2008 Current Population Survey (CPS). The Census Bureau conducts the CPS every month, although this file has only June data. The June survey uses two sets of questions, the basic CPS and a set of supplemental questions. The CPS, sponsored jointly by the 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 Census Bureau and the U.S. Bureau of Labor Statistics also jointly sponsor the supplemental questions for June.

Basic CPS. The monthly CPS collects primarily labor force data about the civilian noninstitutional population living in the United States. The institutionalized population, which is excluded from the population universe, is composed primarily of the population in correctional institutions and nursing homes (91 percent of the 4.1 million institutionalized people in Census 2000). Interviewers ask questions concerning labor force participation about each member 15 years old and over in sample households. Typically, the week containing the nineteenth of the month is the interview week. The week containing the twelfth is the reference week (i.e., the week about which the labor force questions are asked).

The CPS uses a multistage probability sample based on the results of the decennial census, with coverage in all 50 states and the District of Columbia. The sample is continually updated to account for new residential construction. When files from the most recent decennial census become available, the Census Bureau gradually introduces a new sample design for the CPS.

In April 2004, the Census Bureau began phasing out the 1990 sample[1] and replacing it with the 2000 sample, creating a mixed sampling frame. Two simultaneous changes occurred during this phase-in period. First, primary sampling units (PSUs)[2] selected for only the 2000 design gradually replaced those selected for the 1990 design. This involved 10 percent of the sample. Second, within PSUs selected for both the 1990 and 2000 designs, sample households from the 2000 design gradually replaced sample households from the 1990 design. This involved about 90 percent of the sample. The new sample design was completely implemented by July 2005.

In the first stage of the sampling process, PSUs are selected for sample. The United States is divided into 2,025 PSUs. The PSUs were redefined for this design to correspond to the Office of Management and Budget definitions of Core-Based Statistical Area definitions and to improve efficiency in field operations. These PSUs are grouped into 824 strata. Within each stratum, a single PSU is chosen for the sample, with its probability of selection proportional to its population as of the most recent decennial census. This PSU represents the entire stratum from which it was selected. In the case of strata consisting of only one PSU, the PSU is chosen with certainty.

---

1 For detailed information on the 1990 sample redesign, please see reference [1].

2 The PSUs correspond to substate areas (i.e., counties or groups of counties) that are geographically contiguous.

Approximately 72,000 housing units were selected for sample from the sampling frame in June. Based on eligibility criteria, 11 percent of these housing units were sent directly to computer-assisted telephone interviewing (CATI). The remaining units were assigned to interviewers for computer-assisted personal interviewing (CAPI).[3] Of all housing units in sample, about 59,000 were determined to be eligible for interview. Interviewers obtained interviews at about 54,000 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

June 2008 Supplement. In June 2008, in addition to the basic CPS questions, interviewers asked supplementary questions on fertility to women 15 to 44 years of age.

Estimation Procedure. This survey's estimation procedure adjusts weighted sample results to agree with independently derived population estimates of the civilian noninstitutional population of the United States and each state (including the District of Columbia). These population estimates, used as controls for the CPS, are prepared monthly to agree with the most current set of population estimates that are released as part of the Census Bureau's population estimates and projections program.

The population controls for the nation are distributed by demographic characteristics in two ways:

- Age, sex, and race (White alone, Black alone, and all other groups combined).
- Age, sex, and Hispanic origin.

The population controls for the states are distributed by race (Black alone and all other race groups combined), age (0-15, 16-44, and 45 and over), and sex.

The independent estimates by age, sex, race, and Hispanic origin, and for states by selected age groups and broad race categories, are developed using the basic demographic accounting formula whereby the population from the latest decennial data is updated using data on the components of population change (births, deaths, and net international migration) with net internal migration as an additional component in the state population estimates.

The net international migration component in the population estimates includes a combination of the following:

- Legal migration to the United States.
- Emigration of foreign-born and native people from the United States.
- Net movement between the United States and Puerto Rico.
- Estimates of temporary migration.
- Estimates of net residual foreign-born population, which include unauthorized migration.

3 For further information on CATI and CAPI and the eligibility criteria, please see reference [2].

Because the latest available information on these components lags the survey date, it is necessary to make short-term projections of these components to develop the estimate for the survey date.

ACCURACY OF THE 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; 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. There are several sources of nonsampling error, which may occur during the development or execution of the survey. It can occur because of circumstances created by the interviewer, the respondent, the survey instrument, or the way the data are collected and processed. For example, errors could occur because:

- The interviewer records the wrong answer, the respondent provides incorrect information, the respondent estimates the requested information, or an unclear survey question is misunderstood by the respondent (measurement error).
- Some individuals which should have been included in the survey frame were missed (coverage error).
- Responses are not collected from all those in the sample or the respondent is unwilling to provide information (nonresponse error).
- Values are estimated imprecisely for missing data (imputation error).
- Forms may be lost, data may be incorrectly keyed, coded, or recoded, etc. (processing error).

To minimize these errors, the Census Bureau applies quality control procedures during all stages of the production process, including the design of the surveys, the wording of questions, the review of the work of interviewers and coders, and the statistical review of reports.

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

Nonresponse. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the June 2008 basic CPS, the household-level nonresponse rate was 7.95 percent. The person-level nonresponse rate for the fertility supplement was an additional 2.9 percent.

Since the basic CPS nonresponse rate is a household-level rate and the fertility supplement nonresponse rate is a person-level rate, we cannot combine these rates to derive an overall nonresponse rate. Nonresponding households may have fewer persons than interviewed ones, so combining these rates may lead to an overestimate of the true overall nonresponse rate for persons for the fertility supplement.

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. Missed housing units and missed people within sample households create undercoverage in the CPS. Overall CPS undercoverage for June 2008 is estimated to be about 12 percent. CPS coverage varies with age, sex, and race. Generally, coverage is larger for females than for males and larger for non-Blacks than for Blacks. This differential coverage is a general problem for most household-based surveys.

The CPS weighting procedure partially corrects for bias from undercoverage, but biases may still be present when people who are missed by the survey differ from those interviewed in ways other than age, race, sex, Hispanic origin, 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 poststratification divided by the independent population control. Table 1 shows June 2008 CPS coverage ratios by age and sex for certain race and Hispanic groups. The CPS coverage ratios can exhibit some variability from month to month.

Table 1. CPS Coverage Ratios: June 2008

Age group	Total			White only		Black only		Residual race		Hispanic	
	All people	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.89	0.89	0.88	0.89	0.89	0.82	0.80	0.97	0.95	0.90	0.88
16-19	0.89	0.88	0.89	0.88	0.91	0.84	0.82	0.98	0.84	0.91	0.93
20-24	0.78	0.78	0.79	0.79	0.80	0.68	0.77	0.78	0.78	0.88	0.85
25-34	0.82	0.79	0.85	0.81	0.86	0.63	0.78	0.80	0.88	0.77	0.90
35-44	0.88	0.86	0.91	0.88	0.92	0.74	0.84	0.86	0.87	0.82	0.92
45-54	0.90	0.89	0.90	0.90	0.91	0.85	0.88	0.79	0.83	0.87	0.90
55-64	0.90	0.89	0.91	0.90	0.92	0.82	0.87	0.86	0.91	0.85	0.90
65+	0.93	0.93	0.93	0.92	0.94	0.99	0.93	0.88	0.82	0.90	0.90
15+	0.88	0.86	0.89	0.88	0.90	0.78	0.84	0.84	0.85	0.84	0.90
0+	0.88	0.87	0.89	0.88	0.90	0.79	0.83	0.87	0.88	0.86	0.90

Notes: (1) The Residual race group includes cases indicating a single race other than White or Black, and cases indicating two or more races.
(2) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.

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 be careful when comparing the data from this microdata file, which reflects Census 2000-based controls, with microdata files from March 1994 through December 2002, which reflect 1990 census-based controls. Ideally, the same population controls should be used when comparing any estimates. In reality, the use of same population controls is not practical when comparing trend data over a period of 10 to 20 years. Thus, when it is necessary to combine or compare data based on different controls or different designs, data users should be aware that changes in weighting controls or weighting procedures can create small differences between estimates. See the discussion following for information on comparing estimates derived from different controls or different sample designs.

Microdata files from previous years reflect the latest available census-based controls. Although the most recent 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 controls results in about a one percent increase from the 1990 census-based controls in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 2003 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 population subgroups than for the total population.

Note that certain microdata files from 2002, namely June, October, November, and the 2002 ASEC, contain both Census 2000-based estimates and 1990 census-based estimates and are subject to the comparability issues discussed previously. All other microdata files from 2002 reflect the 1990 census-based controls.

Users should also exercise caution because of changes caused by the phase-in of the Census 2000 files (see "Basic CPS"). During this time period, CPS data are collected from sample designs based on different censuses. Three features of the new CPS design have the potential of affecting published estimates: (1) the temporary disruption of the rotation pattern from August 2004 through June 2005 for a comparatively small portion of the sample, (2) the change in sample areas, and (3) the introduction of the new Core-Based Statistical Areas (formerly called metropolitan areas). Most of the known effect on estimates during and after the sample redesign will be the result of changing from 1990 to 2000 geographic definitions. Research has shown that the national-level estimates of the metropolitan and nonmetropolitan populations should not change appreciably because of the new sample design. However, users should still exercise caution when comparing metropolitan and nonmetropolitan estimates across years with a design change, especially at the state level.

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

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. The Census Bureau recommends that data users incorporate information about nonsampling errors into their analyses, as nonsampling error could impact the conclusions drawn from the results. 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 references [2] and [3].

Standard Errors and Their Use. The sample estimate and its standard error enable one to construct a confidence interval. A confidence interval is a range about a given estimate that has a specified probability of containing the average result of all possible samples. 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, but one can say with specified confidence that the interval includes the average estimate calculated from all possible samples.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that the population parameters are different. An example of this would be comparing the percentage of men who were part-time workers to the percentage of women who were part-time workers.

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 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 associated with nonsampling error. Bias is the average over all possible samples of the differences between the sample estimates and the true value.

Generalized Variance Parameters. While it is possible to compute and present an estimate of the standard error based on the survey data for each estimate in a report, 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, data users have access to CPS microdata files, and it is impossible to compute in advance the standard error for every estimate one might obtain from those data sets. Moreover, variance estimates are based on sample data and have variances of their own. Therefore, some methods 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 similar relationships between their variances and expected values. Modeling or generalizing 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 3 provides the generalized variance parameters for labor force estimates, and Tables 4 and 5 provides generalized variance parameters for characteristics from the June 2008 supplement. Tables 6 and 7 provide factors and population controls to derive U.S. state and regional parameters.

The basic CPS questionnaire records the race and ethnicity of each respondent. With respect to race, a respondent can be White, Black, Asian, American Indian and Alaskan Native (AIAN), Native Hawaiian and Other Pacific Islander (NHOPI), or combinations of two or more of the preceding. A respondent's ethnicity can be Hispanic or non-Hispanic, regardless of race.

The generalized variance parameters to use in computing standard errors are dependent upon the race/ethnicity group of interest. The following table summarizes the relationship between the race/ethnicity group of interest and the generalized variance parameters to use in standard error calculations.

Table 2. Estimation Groups of Interest and Generalized Variance Parameters

Race/ethnicity group of interest	Generalized variance parameters to use in standard error calculations
Total population	Total or White
Total White, White AOIC, or White non-Hispanic population	Total or White
Total Black, Black AOIC, or Black non-Hispanic population	Black
Total Asian, AIAN, NHOPI; Asian, AIAN, NHOPI AOIC; or Asian, AIAN, NHOPI non-Hispanic population	Asian, AIAN, NHOPI
Populations from other race groups	Asian, AIAN, NHOPI
Hispanic population	Hispanic
Two or more races – employment/unemployment and educational attainment characteristics	Black
Two or more races – all other characteristics	Asian, AIAN, NHOPI

Notes: (1) AIAN, NHOPI are American Indian and Alaska Native,

Native Hawaiian and Other Pacific Islander, respectively.

(2) AOIC is an abbreviation for alone or in combination. The AOIC population for a race group of interest includes people reporting only the race group of interest (alone) and people reporting multiple race categories including the race group of interest (in combination).

(3) Hispanics may be any race.

(4) Two or more races refers to the group of cases self-classified as having two or more races.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number from this microdata file can be obtained by 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 Table 3 or 4 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 1

Suppose there were 2,598,000 unemployed women of ages 15 to 44 in the civilian labor force. Use the appropriate parameters from Table 3 and Formula (1) to get

Illustration 1	
Number of unemployed females in the civilian labor force (x)	2,598,000
a parameter (a)	-0.000031
b parameter (b)	2,782
Standard error	84,000
90-percent confidence interval	2,460,000 to 2,736,000

The standard error is calculated as

$$s_x = \sqrt{-0.000031 \times 2{,}598{,}000^2 + 2{,}782 \times 2{,}598{,}000} = 84{,}000$$

The 90-percent confidence interval is calculated as 2,598,000 ± 1.645 × 84,000.

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 the parameter from Table 3 or 4 as indicated by the numerator.

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 ($0 \leq p \leq 100$), and b is the parameter in Table 3 or 4 associated with the characteristic in the numerator of the percentage.

Illustration 2

Suppose that 6.4 percent of the 61,692,000 women 15 to 44 years old had a child in the last year. Use the appropriate parameter from Table 4 and Formula (2) to get

Illustration 2	
Percentage of women aged 15-44 who had a child in the last year (p)	6.4
Base (x)	61,692,000
b parameter (b)	2,016
Standard error	0.14
90-percent confidence interval	6.2 to 6.6

The standard error is calculated as

$$s_{x,p} = \sqrt{\frac{2{,}016}{61{,}692{,}000} \times 6.4 \times (100.0 - 6.4)} = 0.14$$

The 90-percent confidence interval for the estimated percentage of women aged 15 to 44 who had a child in the last year is from 6.2 to 6.6 percent (i.e., $6.4 \pm 1.645 \times 0.14$).

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

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

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 result in accurate estimates of the standard error 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 3

Suppose that of the 3,960,000 women in 2008 between 15-44 years of age who had a child in the previous year, 57.1 percent were in the labor force, and of the 3,974,000 women in 2006 between

15-44 years of age who had a child in the previous year, 55.9 percent were in the labor force. Use the appropriate parameters from Table 3 and Formulas (2) and (3) to get

Illustration 3			
	2008 (*x*)	2006 (*y*)	Difference
Percentage mothers of infant children in labor force (*p*)	57.1	55.9	1.2
Base	3,960,000	3,974,000	-
b parameter (*b*)	2,782	2,782	-
Standard error	1.31	1.31	1.86
90-percent confidence interval	54.9 to 59.3	53.7 to 58.1	-1.9 to 4.3

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{1.31^2 + 1.31^2} = 1.86$$

The 90-percent confidence interval around the difference is calculated as $1.2 \pm 1.645 \times 1.86$. Since this interval does include zero, we cannot conclude with 90 percent confidence that the percentage of women in 2008 between 15-44 years of age who had a child in the previous year who were in the labor force is greater than the percentage of women in 2006 between 15-44 years of age who had a child in the previous year who were in the labor force.

<u>Standard Errors of Ratios</u>. Certain estimates may be calculated as the ratio of two numbers. The standard error of a ratio, *x*/*y*, may be computed using

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

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 persons in those families or households with a certain characteristic. If there is at least one person with the characteristic in every family or household, use 0.7 as an estimate of *r*. An example of this type is the mean number of children per family with children.

For all other types of ratios, *r* is assumed to be zero. Examples are the average number of children per family and the family poverty rate. If *r* is actually positive (negative), then this procedure will provide an overestimate (underestimate) of the standard error of the ratio.

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 4
Suppose there were 33,266,000 ever-married women 15-44 years old and 28,426,000 never-married women 15-44 years old. The ratio of ever-married women, *x*, to never-married women, *y*, is 1.17. Use the appropriate parameters from Table 4 and Formulas (1) and (4) to get

Illustration 4			
	Ever-married (x)	Never-married (y)	Ratio
Women 15-44	33,266,000	28,426,000	1.17
a parameter (a)	-0.000020	-0.000020	-
b parameter (b)	4,687	4,687	-
Standard error	366,000	342,000	0.019
90-percent confidence interval	32,664,000 to 33,868,000	27,863,000 to 28,989,000	1.14 to 1.20

Using Formula (5) with $r = 0$, the estimate of the standard error is

$$s_{x/y} = \frac{33{,}266{,}000}{28{,}4256{,}000}\sqrt{\left[\frac{366{,}000}{33{,}266{,}000}\right]^2 + \left[\frac{342{,}000}{28{,}426{,}000}\right]^2} = 0.019$$

The 90-percent confidence interval is calculated as $1.17 \pm 1.645 \times 0.019$.

Standard Errors of Fertility Ratios. The standard error of a fertility ratio is a function of the number of children ever born per 1,000 women and the number of women in a given category. The formula for the standard error of a fertility ratio is

$$s_{x,y} = x\sqrt{a + \frac{b}{xy} + \frac{c}{1{,}000y}} \qquad (5)$$

where *a*, *b*, and *c* are the parameters from Table 5, *x* is the number of children ever born or expected per 1,000 women and *y* is the number of women in thousands. This formula should be used when calculating standard errors for estimates involving the possibility of more than one event per women, i.e., number of children ever born. For data involving at most one event per woman, convert the ratio to a percentage and use Formula (2) and the parameters in Table 3 or 4 to calculate the standard errors.

Illustration 5
Suppose that 10,748,000 women 40-44 years old had 1,901 children ever born per 1,000 women. Use Formula (5) and the parameters in Table 5 to get

Illustration 5	
Children ever born (x)	1,901
Base (1,000y)	10,748
a parameter (a)	+0.0000013
b parameter (b)	810
c parameter (c)	1,479
Standard error	25
90-percent confidence interval	1,859 to 1,943

The standard error is calculated as

$$s_{x,y} = 1{,}901\sqrt{0.0000013 + \frac{810}{1{,}901 \times 10{,}748} + \frac{1{,}479}{1{,}000 \times 10{,}748}} = 25$$

The 90-percent confidence interval is from 1,859 to 1,943 children ever born per 1,000 women (i.e., $1{,}901 \pm 1.645 \times 25$). 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 Quarterly or Yearly Averages. For information on calculating standard errors for labor force data from the CPS which involve quarterly or yearly averages, please see the "Explanatory Notes and Estimates of Error: Household Data" section in *Employment and Earnings*, a monthly report published by the U.S. Bureau of Labor Statistics.

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

Since the CPS is designed to produce both state and national estimates, the proportion of the total population sampled and the sampling rates differ among the states. In general, the smaller the population of the state the larger the sampling proportion. For example, in Vermont approximately 1 in every 250 households was sampled each month. In New York the sample is about 1 in every 2,000 households. Nevertheless, the size of the sample in New York is four times larger than in Vermont because New York has a larger population.

Standard Errors of State Estimates. The standard error for a state may be obtained by determining new state-level a and b parameters and then using these adjusted parameters in the standard error formulas mentioned previously. To determine a new state-level b parameter (b_{state}), multiply the b parameter from Table 3 or 4 by the state factor from Table 6. To determine a new state-level a parameter (a_{state}), use the following:

(1) If the a parameter from Table 3 or 4 is positive, multiply it by the state factor from Table 6.

(2) If the *a* parameter in Table 3 or 4 is negative, calculate the new state-level *a* parameter as follows:

$$a_{state} = \frac{-b_{state}}{POP_{state}} \tag{6}$$

where POP_{state} is the state population found in Table 6.

To determine state-level parameters for the fertility ratio parameters found in Table 5, multiply all parameters by the state factor from Table 6.

Note: The Census Bureau recommends the use of 3-year averages to compare estimates across states and 2-year averages to evaluate changes in state estimates over time.

Standard Errors of Regional Estimates. To compute standard errors for regional estimates, follow the steps for computing standard errors for state estimates found in "Standard Errors of State Estimates" using the regional factors found in Table 7.

Illustration 6

Suppose that of 22,684,000 women 15-44 years old in the South, 43.6 percent remain childless. Use Formula (2) and the appropriate parameter and factor from Tables 4 and 7 to get:

Illustration 6	
Percent of childless women in South (*p*)	43.6
Base (*x*)	22,684,000
b parameter (*b*)	2,016
South regional factor	1.07
Regional b parameter (b_{region})	2,157
Standard error	0.48
90-percent confidence interval	42.8 to 44.4

Obtain the region-level *b* parameter by multiplying the *b* parameter in Table 4 by the regional factor in Table 7. This gives $b_{region} = 2{,}016 \times 1.07 = 2{,}157$. The standard error of the estimate of the percentage of women 15-44 years old in the South who are childless can then be found by using Formula (2) and the new region-level *b* parameter. The standard error is calculated as

$$s_{x,p} = \sqrt{\frac{2{,}157}{22{,}684{,}000} 43.6 \times (100 - 43.6)} = 0.48$$

and the 90-percent confidence interval for the percentage of women 15-44 years old in the South who are childless is calculated as $43.6 \pm 1.645 \times 0.48$.

Standard Errors of Groups of States. The standard error calculation for a group of states is similar to the standard error calculation for a single state. First, calculate a new state group factor for the group of states. Then, determine new state group *a* and *b* parameters. Finally, use these adjusted parameters in the standard error formulas mentioned previously.

Use the following formula to determine a new state group factor:

$$\text{state group factor} = \frac{\sum_{i=1}^{n}(\text{POP}_i \times \text{state factor}_i)}{\sum_{i=1}^{n}\text{POP}_i} \qquad (7)$$

where POP_i and $state\ factor_i$ are the population and factor for state *i* from Table 6. To obtain a new state group *b* parameter ($b_{state\ group}$), multiply the *b* parameter from Table 3 or 4 by the state group factor obtained by Formula (7). To determine a new state group *a* parameter ($a_{state\ group}$), use the following:

(1) If the *a* parameter from Table 3 or 4 is positive, multiply it by the state group factor determined by Formula (7).

(2) If the *a* parameter from Table 3 or 4 is negative, calculate the new state group *a* parameter as follows:

$$a_{\text{state group}} = \frac{-b_{\text{state group}}}{\sum_{i=1}^{n}\text{POP}_i} \qquad (8)$$

To determine state group-level parameters for the fertility ratio parameters found in Table 5, multiply all parameters by the state group factor calculated by Formula (7).

<u>Illustration 7</u>
Suppose the state group factor for the state group Illinois-Indiana-Michigan was required. The appropriate factor would be

$$\textit{state group factor} = \frac{12{,}729{,}855 \times 1.13 + 6{,}285{,}887 \times 1.11 + 9{,}917{,}501 \times 1.13}{12{,}729{,}855 + 6{,}285{,}887 + 9{,}917{,}501} = 1.13$$

<u>Technical Assistance</u>. If you require assistance or additional information, please contact the Demographic Statistical Methods Division via e-mail at <u>dsmd.source.and.accuracy@census.gov</u>.

Table 3. Parameters for Computation of Standard Errors for Labor Force Characteristics: June 2008		
Characteristic	a	b
Total or White		
Civilian labor force, employed	-0.000016	3,068
Not in labor force	-0.000009	1,833
Unemployed	-0.000016	3,096
Civilian labor force, employed, not in labor force, and unemployed		
Men	-0.000032	2,971
Women	-0.000031	2,782
Both sexes, 16 to 19 years	-0.000022	3,096
Black		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000151	3,455
Men	-0.000311	3,357
Women	-0.000252	3,062
Both sexes, 16 to 19 years	-0.001632	3,455
Hispanic		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000141	3,455
Men	-0.000253	3,357
Women	-0.000266	3,062
Both sexes, 16 to 19 years	-0.001528	3,455
Asian, AIAN, NHOPI		
Civilian labor force, employed, not in labor force, and unemployed		
Total	-0.000346	3,198
Men	-0.000729	3,198
Women	-0.000659	3,198
Both sexes, 16 to 19 years	-0.004146	3,198

Notes:
(1) These parameters are to be applied to basic CPS monthly labor force estimates.
(2) AIAN, NHOPI are American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, respectively.
(3) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.
(4) The Total or White, Black, and Asian, AIAN, NHOPI parameters are to be used for both alone and in combination race group estimates.
(5) For nonmetropolitan characteristics, multiply the a and b parameters by 1.5. If the characteristic of interest is total state population, not subtotaled by race or ethnicity, the a and b parameters are zero.
(6) For foreign-born and noncitizen characteristics for Total and White, the a and b parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Hispanic, and Asian, AIAN, NHOPI parameters.
(7) For the groups self-classified as having two or more races, use the Asian, AIAN, NHOPI parameters for all employment characteristics.

Table 4. Parameters for Computation of Standard Errors for Fertility and Birth Expectation Characteristics: June 2008

Characteristic	Persons		Households, etc.	
	a	b	a	b
FERTILITY				
Total or White	-0.000033	2,016	(X)	(X)
Black	-0.000141	2,016	(X)	(X)
Hispanic	-0.000271	3,397	(X)	(X)
Asian, AIAN, NHOPI and two or more races	-0.000381	2,016	(X)	(X)
NUMBER OF BIRTHS				
Total or White	-0.000059	3,676	(X)	(X)
Black	-0.000257	3,670	(X)	(X)
Hispanic	-0.000493	6,186	(X)	(X)
Asian, AIAN, NHOPI and two or more races	-0.000694	3,670	(X)	(X)
MARITAL STATUS, HOUSEHOLD & FAMILY CHARACTERISTICS				
Total or White	-0.000020	4,687	-0.000008	1,860
Black	-0.000148	6,733	-0.000037	1,683
Hispanic	-0.000308	11,347	-0.000077	2,836
Asian, AIAN, NHOPI and two or more races	-0.000395	6,733	-0.000099	1,683
INCOME				
Total or White	-0.000009	2,207	-0.000008	2,016
Black	-0.000055	2,527	-0.000048	2,201
Hispanic	-0.000115	4,259	-0.000101	3,709
Asian, AIAN, NHOPI and two or more races	-0.000148	2,527	-0.000129	2,201
EDUCATIONAL ATTAINMENT				
Total or White	-0.000009	2,131	-0.000008	1,860
Black and two or more races	-0.000053	2,410	-0.000037	1,683
Hispanic	-0.000074	2,745	-0.000077	2,836
Asian, AIAN, NHOPI	-0.000141	2,410	-0.000099	1,683
NATIVITY – Born in:				
Mexico, other N. America, S. America	-0.000034	9,942	(X)	(X)
Europe	-0.000019	5,712	(X)	(X)
Asia, Africa, Oceania	-0.000031	9,310	(X)	(X)
United States	-0.000017	4,997	(X)	(X)

Notes: (1) These parameters are to be applied to the June 2008 Fertility and Birth Expectation Supplement data.

(2) Fertility includes number of women by number of children ever born, percent childless, and number of women who have had a child in the last year.

(3) AIAN, NHOPI are American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, respectively.

(4) Hispanics may be any race. For a more detailed discussion on the use of parameters for race and ethnicity, please see the "Generalized Variance Parameters" section.

(5) The Total or White, Black, and Asian, AIAN, NHOPI parameters are to be used for both alone and in combination race group estimates.

(6) For nonmetropolitan characteristics, multiply the *a* and *b* parameters by 1.5. If the characteristic of interest is total state population, not subtotaled by race or ethnicity, the *a* and *b* parameters are zero.

(7) For foreign-born and noncitizen characteristics for Total and White, the *a* and *b* parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Asian, AIAN, NHOPI, and Hispanic parameters.

(8) For the group self-classified as having two or more races, use the Asian, AIAN, NHOPI parameters for all characteristics except employment, unemployment, and educational attainment, in which case use Black parameters.

Table 5. Parameters for Computation of Standard Errors for Fertility Ratios: June 2008

a	b	c
0.0000013	810	1,479

NOTE: Multiply the parameters by 1.3 to get foreign-born parameters.

Table 6. Populations and Factors for State Parameters and Standard Errors: June 2008					
State	Factor	Population	State	Factor	Population
Alabama	1.09	4,583,016	Montana	0.25	951,307
Alaska	0.18	664,321	Nebraska	0.47	1,752,284
Arizona	1.13	6,376,229	Nevada	0.65	2,595,975
Arkansas	0.70	2,803,168	New Hampshire	0.37	1,304,197
California	1.14	36,245,901	New Jersey	1.14	8,594,920
Colorado	1.14	4,858,854	New Mexico	0.51	1,966,261
Connecticut	0.91	3,448,533	New York	1.16	19,043,727
Delaware	0.23	859,705	North Carolina	1.13	9,022,815
District of Columbia	0.18	579,025	North Dakota	0.17	625,159
Florida	1.10	18,074,293	Ohio	1.13	11,300,599
Georgia	1.11	9,507,425	Oklahoma	0.94	3,563,651
Hawaii	0.31	1,253,645	Oregon	1.00	3,743,011
Idaho	0.35	1,504,451	Pennsylvania	1.13	12,232,097
Illinois	1.13	12,729,855	Rhode Island	0.30	1,037,036
Indiana	1.11	6,285,887	South Carolina	1.11	4,366,704
Iowa	0.79	2,952,957	South Dakota	0.18	785,973
Kansas	0.74	2,736,072	Tennessee	1.12	6,121,849
Kentucky	1.11	4,183,658	Texas	1.14	23,870,728
Louisiana	1.09	4,229,294	Utah	0.54	2,681,607
Maine	0.42	1,303,324	Vermont	0.19	615,810
Maryland	1.16	5,543,802	Virginia	1.12	7,555,752
Massachusetts	1.11	6,373,535	Washington	1.15	6,452,500
Michigan	1.13	9,917,501	West Virginia	0.41	1,787,833
Minnesota	1.11	5,168,314	Wisconsin	1.13	5,546,998
Mississippi	0.73	2,869,017	Wyoming	0.15	523,029
Missouri	1.15	5,803,806			

NOTES: (1) The state population counts in this table are for the 0+ population.
(2) For foreign-born and noncitizen characteristics for Total and White, the *a* and *b* parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Asian AIAN, NHOPI, and Hispanic.

Table 7. Populations and Factors for Regional Parameters and Standard Errors: June 2008		
Region	Factor	Population
Midwest	1.06	65,605,405
Northeast	1.06	53,953,179
South	1.07	109,521,735
West	1.02	69,817,091

NOTES: (1) The state population counts in this table are for the 0+ population.
(2) For foreign-born and noncitizen characteristics for Total and White, the *a* and *b* parameters should be multiplied by 1.3. No adjustment is necessary for foreign-born and noncitizen characteristics for Black, Asian, AIAN, NHOPI, and Hispanic.

References

[1] Bureau of Labor Statistics. 1994. *Employment and Earnings*. Volume 41 Number 5, May 1994. Washington, DC: Government Printing Office.

[2] U.S. Census Bureau. 2006. *Current Population Survey: Design and Methodology*. Technical Paper 66. Washington, DC: Government Printing Office. (http://www.census.gov/prod/2006pubs/tp66.pdf)

[3] Brooks, C.A. and Bailar, B.A. 1978. *Statistical Policy Working Paper 3 - An Error Profile: Employment as Measured by the Current Population Survey*. Subcommittee on Nonsampling Errors, Federal Committee on Statistical Methodology, U.S. Department of Commerce, Washington, DC. (http://www.fcsm.gov/working-papers/spp.html)