Source and Accuracy Statement for the April 2002 CPS Microdata File for Child Support

Table of Contents

Tables

Source and Accuracy Statement for the March and April 2002 CPS Microdata File for Child Support

SOURCE OF DATA

The data for this microdata file came from the March and April 2002 Current Population Survey (CPS). The Census Bureau conducts the CPS every month, although this file has only March and April data. The March and April surveys use two sets of questions, the basic CPS and the supplement.

Basic CPS. The monthly 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.

CPS March and April 2002 Supplements. In addition to the basic CPS questions, interviewers asked supplementary questions in March 2002 about the economic situation of persons and families for the previous year. Of the housing units interviewed in March, about 42,500 were also interviewed in April. All household members 15 years of age and older that are a biological parent of children in the household from an absent parent were asked detailed questions about child support and alimony.

In March 2002, there were 4,934 household members eligible of which 1,455 required imputation of child support data. When matching the March 2002 and April 2002 data sets, there were 284 eligible people on the March file that did not match to people on the April file. Child support data for these 284 people were imputed. The remaining 1,171 imputed cases were due to nonresponse to the child support questions. Table 1 gives the sample sizes and the imputation rates by marital status.

Table 1. Sample Sizes and Imputation Rates

Marital Status	Sample Size	Imputed Cases	Rate
Married	1,202	323	27%
Widowed	45	16	36%
Divorced	1,737	498	29%
Separated	576	168	29%
Never Married	1,374	450	33%
Total	4,934	1,455	29%

Basic CPS Sample Design. The present monthly CPS sample was selected from the 1990 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. 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.

About 60,000 occupied households are eligible for interview every month out of the 754 strata. Interviewers are unable to obtain interviews at about 4,500 of these units. This occurs when the occupants are not found at home after repeated calls or are unavailable for some other reason.

The number of households that are eligible for interview in the basic CPS increased from 50,000 to 60,000 in July of 2001. This increase in the number of eligible households is due to the implementation of the State Children's Health Insurance Program (SCHIP) sample expansion. The SCHIP sample expansion increased the monthly CPS sample in states with high sampling errors for low-income uninsured children. With this increase in eligible households, the number of units where interviewers were unable to obtain an interview increased from 3,200 to 4,500.

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 most recent changes were phased in and implementation was completed in 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[1]/non-Hispanic ancestry, and state of residence. The adjusted estimate is called the post-stratification ratio estimate. The independent estimates are calculated based on information from three primary sources:

- 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 March supplement included a further adjustment so husband and wife of a household received the same weight. The independent population estimates include some, but not all, of unauthorized migrants.

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, however, is unknown.

---

[1]Hispanics may be of any race.

<u>Sampling Error</u>. Since the CPS estimates come from a sample, they may differ from figures from a complete census using the same questionnaires, instructions, and enumerators. This possible variation in the estimates due to sampling error is known as "sampling variability."

<u>Nonsampling Error</u>. All other sources of error in the survey estimates are collectively called nonsampling error. Sources of nonsampling error include the following:

- Inability to obtain information about all sample cases (nonresponse).
- Definitional difficulties.
- Differences in interpretation of questions.
- Respondent inability or unwillingness to provide correct information.
- Respondent 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).

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

<u>Nonresponse</u>. The effect of nonresponse cannot be measured directly, but one indication of its potential effect is the nonresponse rate. For the April 2002 basic CPS, the nonresponse rate was 7.3%. The nonresponse rate for the supplement was an additional 6.3%, for a total supplement nonresponse rate of 13.1%.

<u>Undercoverage</u>. 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 is estimated to be about 8 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.

The Current Population Survey weighting procedure uses ratio estimation whereby sample estimates are adjusted to independent estimates of the national population by age, race, sex and Hispanic ancestry. This weighting partially corrects for bias due to 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, and Hispanic ancestry. 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, the estimated population before post-stratification divided by the independent population control. Table 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 2. CPS Coverage Ratios

Age	Non-Black		Black		All Races		
	Male	Female	Male	Female	Male	Female	Total
0 to 14 years.	0.942	0.951	0.880	0.904	0.932	0.943	0.937
15 to 19 years.	0.864	0.910	0.885	0.751	0.867	0.884	0.876
20 to 24 years.	0.823	0.877	0.707	0.757	0.808	0.859	0.834
25 to 29 years.	0.863	0.919	0.755	0.810	0.850	0.903	0.877
30 to 34 years.	0.880	0.950	0.671	0.833	0.855	0.934	0.895
35 to 44 years.	0.899	0.940	0.684	0.863	0.875	0.930	0.903
45 to 54 years.	0.938	0.961	0.778	0.953	0.923	0.960	0.942
55 to 64 years.	0.932	0.953	0.834	0.929	0.923	0.951	0.938
65 to 74 years.	0.932	0.977	0.939	0.958	0.932	0.975	0.956
75 years and older	1.019	1.008	0.910	0.961	1.011	1.004	1.007
15 years and older	0.902	0.945	0.767	0.858	0.887	0.934	0.912
0 years and older	0.911	0.946	0.802	0.871	0.898	0.936	0.917

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.

A number of changes were made in data collection and estimation procedures beginning with the January 1994 CPS. The major change was the use of a new questionnaire. The questionnaire was redesigned to measure the official labor force concepts more precisely, to expand the amount of data available, to implement several definitional changes, and to adapt to a computer-assisted interviewing environment. The April child support supplement changes included refining the screening of potential respondents, restructuring the questionnaire to accommodate computerizing the survey, revising terminology that refers to types of child support agreements or awards, increasing the detail in questions about the amount of child support due, including overdue child support in the amount of child support due, and adding new questions on pass-through payments (child support collected for public assistance recipients by a state enforcement office, some of which passes through to recipients). Changes to the April CPS supplement in 1994 do not allow for comparisons with CPS data collected before that year.

Caution should also be used when comparing data from this microdata file, which reflect 2000 census-based population controls, with microdata files from April 1994- 2001, which reflect 1990 census-based population controls. Microdata files from previous years reflect the latest available census-based population controls. Although this change in population controls had relatively little impact on summary measures, such as means, medians, and percentage distributions, it did have a significant impact on levels. For example, use of 2000 based population controls results in about a one percent increase from the 1990 based population controls in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 2001 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, CPS data are collected from sample designs based on different censuses. While most CPS estimates were unaffected by this mixed sample, geographic estimates are subject to greater error and variability. Users should exercise caution when comparing estimates across years for metropolitan/nonmetropolitan categories.

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 probably do not reveal useful information when computed on a base[2] 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.

- Technical Paper 63RV, *Current Population Survey: Design and Methodology*, U.S. Census Bureau, U.S. Department of Commerce, 2002.

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 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 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

[2]subpopulation

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 parameters using sample estimates. The most common type of hypothesis is that two population parameters are different. An example of this would be comparing the percentages 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 parameters 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. To estimate the standard error of a CPS estimate, the Census Bureau uses replicated variance estimation methods. 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 of the differences, over all possible samples, between the sample estimates and the true value.

Generalized Variance Parameters. Consider all the possible estimates of characteristics of the population that are of interest to data users. Now consider all the subpopulations such as racial groups, age ranges, etc. Finally, consider every possible comparison or ratio combination. The list would be completely unmanageable. Similarly, a list of standard errors to go with every estimate would be unmanageable.

Through experimentation, we have found that certain groups of estimates have similar relationships between their variances and expected values. We provide a generalized method for calculating standard errors for any of the characteristics of the population of interest. The generalized method uses generalized variance parameters for groups of estimates. These parameters are in Table 3 for basic CPS monthly labor force estimates, and Table 4 for April supplement data.

Standard Errors of Estimated Numbers. The approximate standard error, s_x, of an estimated number from this microdata file can be obtained using this 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 for numbers from cross-tabulations involving different characteristics, use the factor or set of parameters for the characteristic which will give the largest standard error.

Illustration. Suppose you want to calculate the standard error and a 90 percent confidence interval of the number of unemployed females in the civilian labor force when the number of unemployed females in the civilian labor force is about 3,737,000. Use Formula (1) and the appropriate parameters from Table 3 to get:

Number, x	3,737,000
a parameter	-0.000033
b parameter	2,693
Standard error	98,000
90% conf. int.	3,576,000 to 3,898,000

where the standard error is calculated as

$$s_x = \sqrt{-0.000033 \times 3{,}737{,}000^2 + 2{,}693 \times 3{,}737{,}000} = 98{,}000$$

The 90-percent confidence interval is calculated as 3,737,000 ± 1.645×98,000.

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 from 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 indicated by the numerator.

The approximate standard error, $s_{x,p}$, of an estimated percentage can be obtained by using the following 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. In 2002, of the 11,291,000 custodial mothers in the United States, 31.2% were never married. Using the appropriate parameter from Table 4 and Formula (2) gives

Percentage, p	31.2
Base, x	11,291,000
b parameter	6,249
Standard error	1.09
90% conf. int.	29.4 to 33.0

where the standard error is calculated as

$$s_{x,p} = \sqrt{\frac{6,249}{11,291,000} \times 31.2 \times (100 - 31.2)} = 1.09$$

The 90-percent confidence interval is calculated as $31.2 \pm 1.645 \times 1.09$.

<u>Standard Error of a Difference</u>. 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.

For information on calculating standard errors for labor force data from the CPS which involve differences in consecutive quarterly or yearly averages, consecutive month-to-month differences in estimates, and consecutive year-to-year differences in monthly estimates see “Explanatory Notes and Estimates of Error: Household Data” in *Employment and Earnings*, a monthly report published by the Bureau of Labor Statistics.

Illustration. In 2001, of the 6,212,000 custodial mothers that were due child support, 2,821,000 or 45.4% received the full amount of child support due. Of the 712,000 custodial fathers that were due child support, 278,000 or 39.0% received the full amount of child support due.

	x	y	difference
Percentage, p	45.4	39.0	6.4
Base, x	6,212,000	712,000	-
b parameter	2,943	2,943	-
Standard error	1.08	3.14	3.32
90% conf. int.	43.6 to 47.2	33.8 to 44.2	0.9 to 11.9

The standard error of the difference is calculated as

$$s_{x-y} = \sqrt{1.08^2 + 3.14^2} = 3.32$$

The 90-percent confidence interval for the estimated difference between the households is calculated as $6.4 \pm 1.645 \times 3.32$.

Because this interval does not include zero, we can conclude with 90-percent confidence that the percentage of custodial mothers due child support who received the full amount due is greater than the percentage of custodial fathers due child support who received the full amount due.

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 400 households is 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.

Computation of Standard Errors for State Estimates. Standard errors for a state may be obtained by computing national standard errors, using formulas described earlier, and multiplying these by the appropriate f factor from Table 5. An alternative method for computing standard errors for a state is to multiply the a and b parameters in Table 3 or 4 by f^2 and then use these adjusted parameters in the standard error formulas.

Illustration. Suppose you want to calculate the standard error for the percentage of people 18 years old and over living in the state of Florida who had completed a bachelor's degree or more. Suppose about 2,863,000 (24.2 percent) people had completed at least a bachelor's degree when there were about 11,846,000 people aged 18 and over living in Florida. Following the first method mentioned above, use the appropriate parameter from Table 4 and Formula (2) to get:

Percentage, p	24.2
Base, x	11,846,000
b parameter	2,841
State f factor	1.07
Standard error	0.66

Table 5 shows the f factor for Florida to be 1.07. Thus, the standard error on the estimate of the percentage of people 18 and over in Florida state who had completed college is approximately 0.66 x 1.07 = 0.71.

Following the alternative method mentioned above, obtain the needed state parameter by multiplying the parameter in Table 4 by the f^2 factor in Table 5 for the state of interest. For example, for educational attainment for total or white in Florida this gives b = 2,841 x 1.14 = 3,239. The standard error of the estimate of the percentage of people 18 and older in Florida state who had completed college can then be found by using formula (2), the base of 11,846,000 and the new b parameter, 3,239. This gives a standard error of 0.71.

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

Table 3. Parameters for Computation of Standard Errors for Labor Force Characteristics: April 2002

Characteristics	a	b
Civilian Labor Force, Employed, and Not in Labor Force		
Total or White	-0.000008	1,586
Men	-0.000035	2,927
Women	-0.000033	2,693
Both Sexes, 16-19 years	-0.000244	3,005
Black	-0.000154	3,296
Men	-0.000336	3,332
Women	-0.000282	2,944
Both Sexes, 16-19 years	-0.001531	3,296
Hispanic ancestry	-0.000187	3,296
Men	-0.000363	3,332
Women	-0.000380	2,944
Both Sexes, 16-19 years	-0.001822	3,296
Unemployment		
Total or White	-0.000017	3,005
Men	-0.000035	2,927
Women	-0.000033	2,693
Both Sexes, 16-19 years	-0.000244	3,005
Black	-0.000154	3,296
Men	-0.000336	3,332
Women	-0.000282	2,944
Both Sexes, 16-19 years	-0.001531	3,296
Hispanic ancestry	-0.000187	3,296
Men	-0.000363	3,332
Women	-0.000380	2,944
Both Sexes, 16-19 years	-0.001822	3,296
Agricultural Employment	0.001345	2,989

NOTE: These parameters are to be applied to basic CPS monthly labor force estimates.

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 Blacks and Hispanics.

Multiply the a and b parameters by 1.5 when tabulating nonmetropolitan estimates.

Table 4. Standard Error Parameters for Child Support Characteristics

Characteristic	Total or White		Black		Hispanic	
	a	b	a	b	a	b
INCOME						
Persons	-0.000014	2,943	-0.000127	3,370	-0.000233	5,679
Families	-0.000012	2,687	-0.000111	2,934	-0.000203	4,945
POVERTY						
Persons Below the Poverty Level	-0.000045	12,448	-0.000343	12,448	-0.000599	20,978
NONINCOME						
Women/Men with Dependent Children Whose Fathers/Mothers are Absent						
Marital Status	-0.000022	6,249	-0.000248	8,977	-0.000432	15,129
SELECTED CHARACTERISTICS OF MEN AND WOMEN						
Education	-0.000013	2,841	-0.000121	3,214	-0.000150	3,660

Multiply the a and b parameters by 1.5 when tabulating nonmetropolitan estimates.

Table 5. State Factors

State	f	f^2	State	f	f^2
Alabama	0.97	0.94	Montana	0.48	0.23
Alaska	0.35	0.12	Nebraska	0.58	0.34
Arizona	1.07	1.15	Nevada	0.59	0.35
Arkansas	0.80	0.64	New Hampshire	0.46	0.22
California	1.22	1.49	New Jersey	0.96	0.92
Colorado	0.82	0.67	New Mexico	0.68	0.46
Connecticut	0.74	0.55	New York	1.00	1.00
Delaware	0.42	0.18	North Carolina	1.05	1.09
Dist. of Col.	0.38	0.14	North Dakota	0.36	0.13
Florida	1.07	1.14	Ohio	1.06	1.13
Georgia	1.30	1.70	Oklahoma	0.85	0.72
Hawaii	0.51	0.26	Oregon	0.82	0.68
Idaho	0.54	0.30	Pennsylvania	1.02	1.04
Illinois	1.04	1.08	Rhode Island	0.41	0.16
Indiana	0.96	0.92	South Carolina	0.91	0.83
Iowa	0.72	0.51	South Dakota	0.37	0.13
Kansas	0.69	0.48	Tennessee	1.16	1.35
Kentucky	0.91	0.83	Texas	1.17	1.37
Louisiana	1.02	1.05	Utah	0.68	0.46
Maine	0.46	0.21	Vermont	0.33	0.11
Maryland	0.96	0.93	Virginia	1.15	1.32
Massachusetts	0.96	0.93	Washington	1.05	1.11
Michigan	1.02	1.05	West Virginia	0.58	0.34
Minnesota	0.90	0.81	Wisconsin	0.91	0.82
Mississippi	0.85	0.73	Wyoming	0.32	0.10
Missouri	1.00	1.00			

Table 6. Region Factors

Characteristic	f	f^2
Northeast	0.95	0.90
Midwest	0.96	0.93
South	1.07	1.14
West	1.07	1.14