

SEP 03 2015



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



MEMORANDUM FOR: Lisa Clement  
Survey Director for CPS & Time Use, Associate Directorate for  
Demographic Programs

From: James B. Treat *James B. Treat*  
Chief, Demographic Statistical Methods Division

Subject: Source and Accuracy Statement for the 2014 Income Consistent  
Annual Social and Economic Supplement Microdata File

Attached is the statement on the source of the data and accuracy of the estimates for the 2014  
Income Consistent Annual Social and Economic Supplement Microdata File.

If you have any questions or need additional information, please contact David Hornick of the  
Demographic Statistical Methods Division via email at [dsmd.source.and.accuracy@census.gov](mailto:dsmd.source.and.accuracy@census.gov).

Attachment

cc:

N. Tillman (CNMP)  
D. Smallwood  
J. Sweeney  
T. Marshall (ADDP)  
G. Weyland  
L. Cheok  
A. Cantu  
V. Velkoff (SEHSD)  
S. Stern  
E. Welniak Jr.  
Carmen DeNavas Walt  
T. Renwick  
B. Proctor

M. Vornovitsky  
J. Smith  
J. Ortman  
H. Pham  
T. Le  
J. Day  
J. Farber (DSMD)  
J. Scott  
Y. Cheng  
T. Kennel  
D. Hornick  
S. Clark  
R. Hoop

**Source of the Data and Accuracy of the Estimates for the  
2014 Annual Social and Economic Supplement Microdata File**

**Table of Contents**

<b>SOURCE OF THE DATA.....</b>	<b>1</b>
Basic CPS.....	1
The 2014 Annual Social and Economic Supplement .....	2
Estimation Procedure .....	3
<b>ACCURACY OF THE ESTIMATES .....</b>	<b>4</b>
Sampling Error.....	4
Nonsampling Error.....	4
Nonresponse.....	5
Sufficient Partial Interview .....	5
Coverage .....	5
Comparability of Data.....	6
A Nonsampling Error Warning.....	7
Estimation of Median Incomes .....	7
Standard Errors and Their Use.....	8
Estimating Standard Errors .....	8
Technical Assistance.....	9
<b>REFERENCES.....</b>	<b>9</b>

**Tables**

Table 1. CPS Coverage Ratios: March 2014 .....	6
--	---

## Source of the Data and Accuracy of the Estimates for the 2014 Annual Social and Economic Supplement Microdata File

### SOURCE OF THE DATA

The data in this microdata file are from the 2014 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). The U.S. Census Bureau conducts the CPS ASEC over a 3-month period in February, March, and April, with most of the data collection occurring in the month of March. The CPS ASEC 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 Bureau of Labor Statistics also jointly sponsor the CPS ASEC.

**Basic CPS.** The monthly CPS collects primarily labor force data about the civilian noninstitutionalized 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 (98 percent of the 4.0 million institutionalized people in Census 2010). 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<sup>1</sup> 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)<sup>2</sup> 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

---

<sup>1</sup> For detailed information on the 2000 sample redesign, please see reference [1].

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

which it was selected. In the case of strata consisting of only one PSU, the PSU is chosen with certainty.

Approximately 72,700 housing units were selected for sample from the sampling frame for the basic CPS. Based on eligibility criteria, 12 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).<sup>3</sup> Of all housing units in sample, about 59,500 were determined to be eligible for interview. Interviewers obtained interviews at about 52,700 of these units. Noninterviews occur when the occupants are not found at home after repeated calls or are unavailable for some other reason.

**The 2014 Annual Social and Economic Supplement.** In addition to the basic CPS questions, interviewers asked supplementary questions for the CPS ASEC. They asked 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 covered 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

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. Including the basic CPS, approximately 97,900 addresses were eligible to receive the improved set of health insurance coverage questions. The improved income questions were implemented to a subsample of these 97,900 addresses using a probability split panel design. Approximately 67,900 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC (traditional questionnaire sample) and the remaining 30,000 addresses were eligible to receive the redesigned income questions (redesigned income sample). The two samples were combined to create a file that contains all of the individuals in both samples with responses that are consistent with the redesigned questionnaire. About 97,900 housing units were determined to be eligible and we obtained interviews at 74,200. The source of data for this S&A is the combined file – the complete CPS ASEC sample.

The additional sample for the CPS 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 CPS ASEC, please refer to reference [2].

---

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

Full information on the CPS sample is available within the documentation accompanying the annual release of reports and data, released in September 2014.

These research data are contained in three SAS data sets:

1. Household – 97,926 records
2. Family – 86,224 records
3. Person – 199,556 records

The Income-Consistent file was created by modeling income sources most affected by the redesign (interest, dividend and retirement income) for the traditional questionnaire to match the conditional distribution of responses whom were asked the redesigned questionnaire. The imputation model was fit to the responses of the redesigned questionnaire and used to impute the income values of the respondents of the traditional questionnaire. Thus, the Income-Consistent file contains the actual interest, dividend, and retirement income responses for the sample receiving the redesigned questionnaire; but imputed values of interest, dividends, and retirement income for the persons receiving the traditional questionnaire. The imputed income values provide the responses that the person who received the traditional questions may have given had they been asked the redesign questionnaire. This modeling process was done 10 times independently, creating 10 separate implicates, to allow researchers to incorporate the imputation uncertainty into their calculations. A single implicate was chosen for further processing, with modeled interest dividend, and retirement income variables from this implicate in the income consistent file for those receiving the traditional questionnaire. All recodes on the file use this single implicate. The variables within these data sets mirror those of the redesigned ASEC 2014 public use file, with the addition of variables from the other nine implicates for interest, dividend, and retirement income.

Further information regarding creation of the Income Consistent Research file can be found in the following Census Bureau working paper: **Bridging a Survey Redesign Using Multiple Imputation: An Application to the 2014 CPS ASEC**, Jonathan Rothbaum, reference [3] <<http://www.census.gov/content/dam/Census/library/working-papers/2015/demo/SEHSD-WP2015-15.pdf>>.

**Estimation Procedure.** This survey's estimation procedure adjusts weighted sample results to agree with independently derived population estimates of the civilian noninstitutionalized 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 2010 Decennial Census 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.

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 that 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 who 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 survey, 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. The basic CPS household-level nonresponse rate, based on the 97,900 addresses eligible to receive the 2013 CPS ASEC questionnaire, was 10.22 percent. The household-level CPS ASEC nonresponse rate was an additional 11.29 percent. These two nonresponse rates lead to a combined supplement nonresponse rate of 20.36 percent.

**Sufficient Partial Interview.** A sufficient partial interview is an incomplete interview in which the household or person answered enough of the questionnaire for the supplement sponsor to consider the interview complete. The remaining supplement questions may have been edited or imputed to fill in missing values. Insufficient partial interviews are considered to be nonrespondents. Refer to the supplement overview attachment in the technical documentation for the specific questions deemed critical by the sponsor as necessary to be answered in order to be considered a sufficient partial interview.

**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 March 2014 is estimated to be about 15.0 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 March 2014 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: March 2014

Age group	All people	Total		White only		Black only		Residual race		Hispanic	
		Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-15	0.85	0.85	0.86	0.88	0.90	0.78	0.75	0.72	0.75	0.83	0.84
16-19	0.84	0.83	0.84	0.85	0.87	0.75	0.80	0.80	0.70	0.82	0.84
20-24	0.72	0.71	0.73	0.75	0.76	0.62	0.66	0.57	0.67	0.74	0.81
25-34	0.81	0.79	0.83	0.82	0.86	0.66	0.76	0.70	0.72	0.76	0.84
35-44	0.86	0.83	0.88	0.85	0.91	0.76	0.84	0.71	0.76	0.75	0.87
45-54	0.87	0.85	0.88	0.86	0.89	0.81	0.82	0.83	0.79	0.78	0.87
55-64	0.89	0.88	0.90	0.89	0.92	0.84	0.79	0.80	0.86	0.82	0.83
65+	0.92	0.91	0.92	0.92	0.92	0.87	0.90	0.93	0.90	0.88	0.91
15+	0.85	0.84	0.87	0.86	0.89	0.76	0.80	0.76	0.77	0.78	0.85
0+	0.85	0.84	0.86	0.86	0.89	0.76	0.79	0.75	0.77	0.79	0.85

- 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 estimates for 2013 in *Income and Poverty in the United States: 2013* and *Health Insurance Coverage in the United States: 2013* (which reflect Census 2010-based controls) with estimates for 1999 to 2010 (from March 2000 CPS to March 2011 CPS), which reflect 2000 census-based controls, and to 1992 to 1998 (from March 1993 CPS to March 1999 CPS), which reflect 1990 census-based controls. Ideally, the same population controls should be used when comparing any estimates. In reality, the use of the 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 could create small differences between estimates. See the following discussion for information on comparing estimates derived from different controls or different sample designs.

Data users should be aware that the redesign of CPS ASEC health insurance constitutes a “break-in-series” to the historical health insurance estimates, so data users should exercise caution when comparing estimates from the redesigned CPS ASEC for data year 2013 to estimates from previous years.

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 2010-based controls results in about a 0.2 percent increase from the 2000 census-based controls in the civilian noninstitutionalized population and in the



number of families and households. Thus, estimates of levels for data collected in 2012 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.

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 were 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 was 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 [4].

**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--depending on the size of the income interval--Pareto for intervals larger than \$2,500 in width, linear otherwise. 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.

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 2013 (2014 CPS ASEC) 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 reference [5].

**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 ASEC 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. While replicate methods provide the most accurate variance estimates, this approach requires more computing resources and more expertise on the part of the user.

Standard errors for the 2014 CPS ASEC Income Consistent microdata file should be calculated using direct variance estimates created from replicate weights. For more information on calculating direct variance estimates, see reference [5]. A discussion about calculating direct variances that include the implicates can be found in the Census Bureau working paper:

**Bridging a Survey Redesign Using Multiple Imputation: An Application to the 2014 CPS ASEC**, Jonathan Rothbaum <<http://www.census.gov/content/dam/Census/library/working-papers/2015/demo/SEHSD-WP2015-15.pdf>>, reference [3].

**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](mailto:dsmd.source.and.accuracy@census.gov).

## REFERENCES

- [1] Bureau of Labor Statistics. 2004. *Employment and Earnings*. “Redesign of the Sample for the Current Population Survey.” Volume 51 Number 11, May 2004. Washington, DC: Government Printing Office. pp 4-6.
- [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/tp-66.pdf>)
- [3] Rothbaum J. 2015. U.S. Census Bureau working paper, *Bridging a Survey Redesign Using Multiple Imputation: An Application to the 2014 CPS ASEC*,. (<http://www.census.gov/content/dam/Census/library/working-papers/2015/demo/SEHSD-WP2015-15.pdf>)
- [4] 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://fcsml.sites.usa.gov/files/2014/04/spwp3.pdf>)
- [5] U.S. Census Bureau. 1978. *Money Income in 1976 of Families and Persons in the United States*. Current Population Reports, P60-114. Washington, DC: Government Printing Office. (<http://www2.census.gov/prod2/popscan/p60-114.pdf>)
- [6] U.S. Census Bureau, July 15, 2009, “Estimating ASEC Variances with Replicate Weights Part I: Instructions for Using the ASEC Public Use Replicate Weight File to Create ASEC Variance Estimates.” ([http://www.bls.census.gov/pub/cps/march/Use\\_of\\_the\\_Public\\_Use\\_Replicate\\_Weight\\_File\\_final\\_PR.doc](http://www.bls.census.gov/pub/cps/march/Use_of_the_Public_Use_Replicate_Weight_File_final_PR.doc))