

How to Use the Data: Accuracy of the Data

1998 Accuracy of the Data

Introduction

The data contained in these Profiles and Summary Tables are based on the American Community Survey (ACS) sample interviewed in 1998. The ACS is designed to provide accurate estimates for the housing units and household population of the nine sites participating in the 1998 ACS. The ACS, like any other statistical activity, is subject to error. The purpose of this documentation is to provide data users with a basic understanding of the ACS sample design, estimation methodology, and accuracy of the ACS data.

Sample Design

Sites -- Nine sites participated in the 1998 ACS: Multnomah County/Portland, OR, Rockland County, NY, Broward County, FL, Fulton County, PA, Douglas County, NE, Otero County, NM, Franklin County, OH, Harris and Ft. Bend Counties, TX, and Richland and Kershaw Counties, SC. The primary sampling unit was the housing unit, including all occupants. Persons living in group quarters were not included in the sample.

Master Address File -- In the urban sites the Census Bureau developed a Master Address File (MAF) that served as the housing unit sampling frame for the survey. The MAF was constructed by an automated match of the sites' 1990 Census Address Control File and the United States Postal Service 1997 Delivery Sequence File. For the Fulton County site the Census Bureau compiled a sampling frame by canvassing and listing each housing unit in the county, then computerizing the list. For Otero County, the Census Bureau combined listing with using the MAF.

Sampling Rates -- In sites other than Broward, Fulton, Otero, Richland and Kershaw Counties, most of the housing units selected into the survey were sampled at a rate of 3 percent. For functioning governmental units in which there were fewer than 1,000 housing units on the sampling frame, the sampling rate was 9 percent. A variable sampling rate was used for the purpose of providing relatively more reliable estimates for small areas. All of Fulton County and a few governmental units in these sites were sampled at 9 percent. For Otero County, the sampling rate was 15 percent. In Broward, four different sampling rates were employed: 2.5 percent for Ft. Lauderdale and Hollywood; 3 percent for Coral Springs and Pembroke Pines; 15 percent for blocks in the smallest governmental units; and .7 percent for the balance of the county.

Data Collection

Three data collection modes were used to conduct the 1998 ACS: Mail, Computer Assisted Telephone Interviewing (CATI), and Computer Assisted Personal Interviewing (CAPI).

These three modes are described below.

Mail Phase -- The Mail phase began with a prenotice letter mailed to each housing unit on the next to last Wednesday of the month preceding the sample month. The ACS questionnaire was mailed one week later, followed by a reminder card one week after that. A replacement questionnaire was mailed three weeks later if the original questionnaire had not yet been checked in at the processing site. Check-in of mail return questionnaires for a sample panel was cut off at the start of the third month following the sample month.

CATI Phase -- Approximately five weeks after the mailout of the original ACS questionnaire, the CATI staff began contacting non-responding sample households by telephone. Late mail returns were removed from the CATI workload on a daily basis. This phase of nonresponse follow-up lasted for approximately four weeks.

CAPI Phase -- The CAPI universe consisted of all outstanding non-response cases remaining after the completion of the CATI phase. A 1 in 3 subsample was selected from the outstanding cases and forwarded to the field interviewers. Field interviewers visited each assigned housing unit and attempted to conduct an interview. Late mail returns were removed from the CAPI workload on a daily basis. The CAPI phase of nonresponse follow-up lasted approximately one month.

Confidentiality of the Data

Confidentiality Edit -- To maintain the confidentiality required by law (Title 13, United States Code), the Census Bureau applies a confidentiality edit to the ACS data to assure that published data do not disclose information about specific individuals, households, or housing units. As a result, a small amount of uncertainty is introduced into the estimates of ACS characteristics. The sample itself provides adequate protection for most areas for which sample data are published since the resulting data are estimates of the actual characteristics. However, small areas require more protection. The confidentiality edit is implemented by identifying a subset of individual housing units from the sample data files as having a unique combination of specified person and household characteristics within a block group. The confidentiality edit is controlled so that the basic structure of the data is preserved.

Errors in the Data

Sampling Error -- The data in the ACS products are estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology. The estimates from the chosen sample also differ from other samples of housing units and persons within those housing units. Sampling error in data arises due to the use of probability sampling, which is necessary to insure the integrity and representativeness of sample survey results. The implementation of statistical sampling procedures provides the basis for the statistical analysis of sample data.

Nonsampling Error -- In addition to sampling error, data users should realize that other types of errors may be introduced during any of the various complex operations used to collect and process survey data. For example, operations such as editing, reviewing, or keying data from questionnaires may introduce error into the estimates. These and other sources of error contribute to the nonsampling error component of the total error of survey estimates. Nonsampling errors may affect the data in two ways. Errors that are introduced randomly increase the variability of the data. Systematic errors which are consistent in one direction introduce bias into the results of a sample survey. The Census Bureau protects against the effect of systematic errors on survey estimates by conducting extensive research and evaluation programs on sampling techniques, questionnaire design, and

data collection and processing procedures. In addition, an important goal of the ACS is to minimize the amount of nonsampling error introduced through nonresponse for sample housing units. One way of accomplishing this is by following up on mail nonrespondents during the CATI and CAPI phases.

Standard Errors -- The standard error is a measure of the deviation of a sample estimate from the average of all possible samples. Sampling errors and some types of nonsampling errors are estimated by the standard error. The sample estimate and its estimated standard error permit the construction of interval estimates with a prescribed confidence that the interval includes the average result of all possible samples. The next section describes the method of calculating standard errors and confidence intervals for the estimates in this ACS product.

Calculation of Standard Errors

Generalized Standard Errors

The information provided in Tables A through C can be used to approximate the standard errors of most sample estimates of *totals* and *proportions* in the Profiles and Summary Tables. Tables A and B give the basic standard error for an estimate of a characteristic that would result under a simple random sampling design. The estimates are for person, family, and housing unit characteristics. Design factors by subject are provided in Table C. **Table C has two columns for Broward County, FL. The second column (2) is used for the following places: Coral Springs, Lauderdale, Hollywood, and Pembroke Pines. The first column (1) is used for all other geographies in Broward County, FL.** The term "subject" refers to a characteristic, such as age for persons and tenure for housing units. The design factors reflect the effects of the actual sample design and estimation procedures used for the 1998 American Community Survey. Details of the sample design and estimation procedures are provided elsewhere in this chapter.

To approximate the standard error of an estimate of a total or a proportion using Tables A through C follow the steps described in the next section. A proportion is defined as a ratio of two estimates where the numerator is a subset of the denominator. For example, the proportion of Black lawyers is the ratio of Black lawyers to all lawyers.

An inspection of the formulas used to calculate the simple random sampling standard errors suggests that when dealing with zero estimates or very small estimates of totals and percentages the standard error estimates approach zero. This is also the case for very large estimates of totals and percentages. Zero or small estimates, like any other sample estimates, are still subject to sampling variability and therefore an estimated standard error of zero or close to zero is not adequate.

For the following sites: **Douglas, NE, Rockland, NY, Franklin, OH, Multnomah, OR, Fulton, PA, and Fort Bend & Harris, TX**, when an estimated total is less than 100 or within 100 of the total size of the tabulation area, use a basic standard error of 57. For estimated percentages that are less than 2.5 or greater than 97.5, use the basic standard errors in Table B that are shown in the "2.5 or 97.5" row or use a value of 2.5 for in the formula below Table B.

For **Richland & Kershaw, SC**, when an estimated total is less than 80 or within 80 of the total size of the tabulation area, use a basic standard error of 51. For estimated percentages which are less than 2 percent or greater than 98 percent, use 2 as the value of in the basic standard error formula below Table B.

For **Otero, NM**, when an estimated total is less than 50 or within 50 of the total size of the tabulation area, use a basic standard error of 40. For estimated percentages which are less than 2 percent or greater than 98 percent, use 2 as the value of in the basic standard error formula below Table B.

For **Broward, FL**, there are different cutoffs for different places. For **Coral Springs and Pembroke Pines**, when an estimated total is less than 100 or within 100 of the total size of the tabulation area, use a basic standard error of 57. For estimated percentages that are less than 2.5 or greater than 97.5, use the basic standard errors in Table B that are shown in the "2.5 or 97.5" row or use a value of 2.5 for in the formula below Table B. For **Fort Lauderdale and Hollywood**, when an estimated total is less than 150 or within 150 of the total size of the tabulation area, use a basic standard error of 70. For estimated percentages which are less than 4 percent or greater than 96 percent, use 4 as the value of in the basic standard error formula below Table B. For any **other estimates in Broward, FL**, when an estimated total is less than 425 or within 425 of the total size of the tabulation area, use a basic standard error of 117. For estimated percentages which are less than 11 percent or greater than 89 percent, use 11 as the value of in the basic standard error formula below Table B.

When the denominator of a percentage is zero, the user is referred to the Direct Standard Error Section on Exceptions and Exception #1.

Confidence intervals can be constructed from generalized standard errors just as they are from direct standard errors. However, for estimates other than totals and proportions, generalized standard errors cannot be calculated from Tables A through C.

To get standard errors for estimates of *means*, *medians*, *aggregates* and *per capita amounts* the user should use direct standard errors since generalized standard errors are not provided for these estimates. If two or more tabulation areas are combined to obtain the median value of a characteristic, the user is referred to "Medians" below for a description of how to approximate the standard error. The method relies on linear interpolation. The standard error approximations that result from this method may not be as adequate for characteristics that deviate from the linearity assumption. Users are recommended to exercise caution when analyzing these characteristics.

Use of Tables to Approximate Standard Errors

Tables A through C are used in the following manner to approximate standard errors:

1. Obtain the basic standard error from either Table A (for an estimate of a total) or Table B (for an estimate of a percentage) or use the formula given below the appropriate table. **Obtain the number of persons or number of housing units from Summary Table P1 or H1, respectively for the geography(s) you are interested in. If the estimate is of families, households or housing units then use the number of housing units, otherwise use the number of persons. This is a change from 1996 and 1997 when the site level count of person, families or housing units was used.** Use these numbers to determine which column to look under in Table A.
2. Use Table C to obtain the appropriate design factor for the characteristic; for example, educational attainment or ancestry. Multiply the basic standard error by this factor to get the ACS standard error estimate.

Medians -- For the standard error of the median of a characteristic, it is necessary to examine the distribution from which the median is derived, as the estimated number of persons, households, families or housing units with the characteristic and the distribution of the characteristic affect the standard error. An approximate method is given here. As the first step, compute one-half of the estimated number having the characteristic on which the median is based (refer to this result as B/2). Treat B/2 as if it were an ordinary estimate and obtain its standard error as instructed above. Compute the desired confidence interval about B/2. Starting with the lowest value of the characteristic, cumulate the frequencies in each category of the characteristic until the sum equals or first exceeds the lower limit of the confidence interval about B/2. By linear interpolation, obtain a value of the

characteristic corresponding to this sum. This is the lower limit of the confidence interval of the median. In a similar manner, continue cumulating frequencies until the sum equals or exceeds the count in excess of the upper limit of the interval about B/2. Interpolate as before to obtain the upper limit of the confidence interval for the estimated median.

When interpolation is required in the upper open-ended interval of a distribution to obtain a confidence bound, use 1.5 times the lower limit of the open-ended confidence interval as the upper limit of the open-ended interval.

Direct Standard Errors

Methodology Used -- Direct estimates of the standard errors were calculated for all estimates reported in this product. They are provided in the Profiles and in the Summary Tables estimates for medians, means, aggregates, and per capita amounts. The standard errors, in most cases, are calculated using standard variance estimation software using a methodology that takes into account the sample design and estimation procedures.

Exceptions -- There are two cases for which the direct standard error estimates are not appropriate.

1. There are no sample observations available to compute an estimate of a proportion or per capita amount or an estimate of its standard error. The estimate is represented in the tables by "-" and the lower and upper bounds by "***".
2. Only a small number of identical values are reported and used to calculate an aggregate, median, mean or per capita amount. In this case, there are too few sample observations to compute a stable estimate of the standard error. The lower and upper bounds are represented in the tables by "**".

Sums and Differences -- The standard errors estimated from these tables are for individual estimates. Additional calculations are required to estimate the standard errors for sums of and differences between two sample estimates. The estimate of the standard error of a sum or difference is approximately the square root of the sum of the two individual standard errors squared; that is, for standard errors and of estimates and :

This method, however, will underestimate (overestimate) the standard error if the two items in a sum are highly positively (negatively) correlated or if the two items in a difference are highly negatively (positively) correlated.

Ratios -- Frequently, the statistic of interest is the ratio of two variables, where the numerator may or may not be a subset of the denominator. The standard error of the ratio between two sample estimates is approximated as follows:

Confidence Intervals

Confidence Intervals -- A sample estimate and its estimated standard error may be used to construct confidence intervals about the estimate. These intervals are ranges that will contain the average value of the estimated characteristic that results over all possible samples, with a known probability.

For example, if all possible samples that could result under the 1998 ACS sample design were independently selected and surveyed under the same conditions, and if the estimate and its estimated standard error were calculated for each of these samples, then:

1. Approximately 68 percent of the intervals from one estimated standard error below the estimate to one estimated standard error above the estimate would contain the average result from all possible samples;
2. Approximately 90 percent of the intervals from 1.65 times the estimated standard error below the estimate to 1.65 times the estimated standard error above the estimate would contain the average result from all possible samples.
3. Approximately 95 percent of the intervals from two estimated standard errors below the estimate to two estimated standard errors above the estimate would contain the average result from all possible samples.

The intervals are referred to as 68 percent, 90 percent, and 95 percent confidence intervals, respectively.

Confidence Intervals of Ratios, Sums, and Differences -- Confidence intervals also may be constructed for the ratio, sum of, or difference between two sample figures. This is done by first computing the ratio, sum, or difference, then obtaining the standard error of the ratio, sum, or difference (using the formulas given earlier), and finally forming a confidence interval for this estimated ratio, sum, or difference as above. One can then say with specified confidence that this interval includes the ratio, sum, or difference that would have been obtained by averaging the results from all possible samples.

Lower and Upper Bounds -- The lower and upper bounds presented in the summary tables and profiles are the bounds based upon a 90% confidence interval.

Limitations -- The user should be careful when computing and interpreting confidence intervals.

1. The estimated standard errors included in this data product do not include all portions of the variability due to nonsampling error that may be present in the data. In particular, the standard errors do not reflect the effect of correlated errors introduced by interviewers, coders, or other field or processing personnel. Thus, the standard errors calculated represent a lower bound of the total error. As a result, confidence intervals formed using these estimated standard errors may not meet the stated levels of confidence (i.e., 68, 90, or 95 percent). Thus, some care must be exercised in the interpretation of the data in this data product based on the estimated standard errors.
2. Zero or small estimates; very large estimates -- The value of almost all ACS characteristics is greater than or equal to zero by definition. For zero or small estimates use of the method given previously for calculating confidence intervals relies on large sample theory, and may result in negative values which for most characteristics are not admissible. In this case the lower limit of the confidence interval is set to zero by default. A similar caution holds for estimates of totals close to a control total or estimated ratios near one, where the upper limit of the confidence interval is set to its largest admissible value. In these situations the level of confidence of the adjusted range of values is less than the prescribed confidence level.
3. Small tabulation areas -- Estimates for small tabulation areas (particularly small areas with high nonresponse rate) are based on small samples. Again, using large sample theory methodology to construct confidence intervals may result in values that are not admissible for the characteristic of interest or in overstating the confidence for a given range. The user should exercise caution in the analysis of data for these areas.

Examples

We will present some examples based on 1998 to demonstrate the use of the formulas.

Example 1 - Total Estimate

The estimated number of males, never married is 457 and females, never married is 426 from summary table P20 in Otero County, NM, Tract 4.01, but we are interested in the total number of people who have never been married. So the estimate of never married people is $457 + 426 = 883$. To determine the basic standard error, we use the formula below Table A. In this formula is our estimate of 883 and N is determined from summary table P1 for Tract 4.01 to be 5,423.

$$\text{BasicSE}(883) = = 155.$$

The design factor (from Table C) for "Marital Status" for Otero County is 0.5. The approximate standard error estimate for the estimated number of never married people is determined by multiplying the basic standard error 155 by the design factor 0.5. This yields an estimated standard error of 78. (The level of precision on each calculation is the same as for the estimates.)

To calculate the lower and upper bounds of the 90 percent confidence interval around 883 using the final standard error, simply multiply 78 by 1.65, then add and subtract the product from 883. Thus the 90 percent confidence interval for this estimate is

$$[883 - 1.65(78)] \text{ to } [883 + 1.65(78)] \text{ or } 754 \text{ to } 1012.$$

Example 2 - "Small" Total Estimate

The estimated number of children 5 years old from summary table P9 in Otero County, NM, Tract 4.01 is 34. Since this estimate is below the small estimate cutoff for Otero of 50, we use a basic standard error of 40.

The design factor (from Table C) for "Age and Sex" for Otero County is 0.7. The approximate standard error estimate for the estimated number of never married people is determined by multiplying the basic standard error 40 by the design factor 0.7. This yields an estimated standard error of 28. (The level of precision on each calculation is the same as for the estimates.)

To calculate the lower and upper bounds of the 90 percent confidence interval around 34 using the final standard error, simply multiply 28 by 1.65, then add and subtract the product from 34. Thus the 90 percent confidence interval for this estimate is found to be $[34 - 1.65(28)]$ to $[34 + 1.65(28)]$ or -12 to 80, but since the lower bound cannot be negative as described in the Limitations section, the lower bound is given a value of 0. So thus the confidence interval is 0 to 80.

Example 3 - Proportion or Percentage Estimate

The estimated number of males in Otero County, Tract 4.01 is 2,567 from summary table P5 and the population for the tract is 5,423. So the percentage of males in Tract 4.01 is 47.3 percent. The base of the estimated percentage is 5,423. Using the formula below Table B to obtain the basic standard error, we get:

$$\text{BasicSE}(47.3) = = 3.9$$

The design factor for "Age and Sex" for Otero County, NM is 0.7. Multiply 3.9 by the design factor 0.7 to approximate the ACS standard error estimate. The standard error estimate is found to be 2.7.

To calculate the lower and upper bounds of the 90 percent confidence interval around 47.3 percent using the final standard error, simply multiply 2.7 by 1.65, then add and subtract the product from 47.3. Thus the 90 percent confidence interval for this estimated percentage is $[47.3 - 1.65(2.7)]$ to $[47.3 + 1.65(2.7)]$ or 42.8 to 51.8.

Example 4 - Combining geographies

We are interested in estimating the number of females in Otero County, NM in tracts 4.01 and 4.02. The number of females in Otero County, NM, Tract 4.01 is 2,856 and the number of females in Otero County, NM, Tract 4.02 is 1,896 from summary table P5. The total population from summary table P1 is 5,423 for Tract 4.01 and 4,089 for Tract 4.02. The total number of females in the two tracts is $2,856 + 1,896 = 4,752$, and the total population is $5,423 + 4,089 = 9,512$. To determine the basic standard error, we use the formula below Table A. In this formula is our estimate of the number of females, 4,752, and N is population for the two tracts, 9,512.

$$\text{BasicSE}(4752) = = 277.$$

The design factor (from Table C) for "Age and Sex" for Otero County is 0.7. The approximate standard error estimate for the estimated number of females is determined by multiplying the basic standard error 277 by the design factor 0.7. This yields an estimated standard error of 194. (The level of precision on each calculation is the same as for the estimates.)

To calculate the lower and upper bounds of the 90 percent confidence interval around 4,752 using the final standard error, simply multiply 194 by 1.65, then add and subtract the product from 4,752. Thus the 90 percent confidence interval for this estimate is $[4,752 - 1.65(194)]$ to $[4,752 + 1.65(194)]$ or 4432 to 5072.

Estimation Procedure

The estimates that appear in this product were obtained from a raking ratio estimation procedure that resulted in the assignment of two sets of weights: a weight to each sample person record and a weight to each sample housing unit record. For any given tabulation area, a characteristic total was estimated by summing the weights assigned to the persons, households, families or housing units possessing the characteristic in the tabulation area. Estimates of person characteristics were based on the person weight. Estimates of family, household or housing unit characteristics were based on the housing unit weight.

Each sample person or housing unit record was assigned exactly one weight to be used to produce estimates of all characteristics. For example, if the weight given to a sample person or housing unit had the value 6, all characteristics of that person or housing unit would be tabulated with the weight of 6. The estimation procedure, however, did assign weights varying from person to person or housing unit to housing unit.

The estimation procedure used to assign the weights was performed independently within each of the 1998 ACS sites.

Initial Housing Unit Weighting Factors - This process produced the following factors:

- **Base Weight (BW)** - This factor was assigned to every housing unit based on its sampling stratum and was the inverse of the housing unit's sampling rate. The 1998 sampling rates are shown in the table below.

Type of Area	Broward FL	Otero NM	Fulton PA	Kershaw & Richland SC	All other counties
All blocks other than those specified below	0.7%	15.0%	n/a	5.0%	3.0%
Blocks in smallest governmental units (fewer than 800 HUs)	15.0%	15.0%	9.0%	15.0%	9.0%
Blocks in Fort Lauderdale and Hollywood, FL	2.5%				
Blocks in Coral Springs and Pembroke Pines, FL	3.0%				

- Sample Reduction Factor (SRF)** -- This factor was assigned only in the two sites (Broward FL and Douglas NE) where there was a substantial reduction in sampling rates between 1997 and 1998. Since the survey is based on the month a response was received (not when the questionnaire was mailed), January and February 1998, in these counties, had an unusually large number of responses compared to the other ten months of 1998. To reduce the possible bias to the editing and weighting, these "carried-over" records were subsampled down to the standard 1998 sampling rates. The sample reduction factor reflects the increased weight assigned to the addresses that were retained.
- CAPI Subsampling Factor (SSF)** -- The weights of the CAPI cases were adjusted to reflect the results of CAPI subsampling. This factor was assigned to each record as follows:

Selected in CAPI subsampling: SSF = 3.0
 Not selected in CAPI subsampling: SSF = 0.0
 Not a CAPI case: SSF = 1.0

For Otero county, NM, some addresses were unmailable. A two-thirds sample of these were sent directly to CAPI and for these cases SSF = 1.5.
- Variation in Monthly Response by Mode (VMS)** -- This factor made the total weight of the Mail, Delivery, CATI, and CAPI records to be tabulated in a month equal to the total base weight of all cases originally mailed for that month. The value of VMS for Mail and Delivery cases was 1.0. For CATI and CAPI cases, VMS was computed and assigned based on the following groups.

Site x Month
- Noninterview Factor (NIF)** -- This factor adjusted the weight of all responding occupied housing units to account for both responding and nonresponding housing units. This factor was computed in two stages: NIF1 and NIF2. NIF1 was computed and assigned to occupied housing units based on the following groups.

Site x County x Building Type (single or multi-unit) x Tract

After having adjusted each occupied housing unit for NIF1, a ratio adjustment NIF2 was computed and assigned to occupied housing units based on the following groups.

Site x Building Type x Month

NIF was then computed for each occupied housing unit as the product of NIF1 and NIF2. Vacancies were assigned a value of NIF = 1.0. Nonresponding housing units were now assigned a weight of 0.0.
- Noninterview Factor - Mode (NIFM)** -- This factor adjusted the weight of just the responding CAPI occupied housing units to account for both CAPI respondents and all nonrespondents. This factor was computed as if NIF had not already been assigned to every occupied housing unit record. This factor was not used directly but rather as part of computing the next factor: MBF. NIFM was computed and assigned to occupied CAPI housing units based on the following groups.

Site x Building Type x Month

Mail, Delivery and CATI cases received a value of NIFM = 1.0. Vacancies received a value of NIFM = 1.0.
- Mode Bias Factor (MBF)** -- This factor made the total weight of the housing units in the groups below the same as if NIFM had been used instead of NIF. MBF was computed and assigned to occupied housing units based on the following groups.

Site x Tenure (Owner or renter) x Month x Marital Status (married/widowed or other)

Vacancies received a value of MBF = 1.0.
- First Housing Unit Post-Stratification Factor (HPF1)** -- This factor made the number of housing units in a tract equal to the tract counts from the January 1999 Master Address File (MAF), after all factors through HPF1 had been applied. HPF1 was computed and assigned to all housing units based on the following groups.

Site x County x Tract

Person Weighting Factors -- Initially the person weight of each person in an occupied housing unit was the product of the weighting factors of their associated housing unit (BW x . . . x HPF1). At this point everyone in the household would have the same weight. These person weights were then individually adjusted based on each person's age, race, sex, and Hispanic origin as described below:

- Person Post-Stratification Factor (PPSF)** - This factor was applied to individuals based on their age, race, sex and Hispanic origin. It adjusted the person weights so that the weighted sample counts matched county population control counts by age, race, sex, and Hispanic origin. These population control counts were independently derived by the Census Bureau's Population Division using demographic analysis.

This was an iterative procedure that first computes PPSF using the following groups:

County x Race (White, Black, Other) x Sex x Age Groups used for Race adjustment

After applying the value of PPSF computed above, a second stage value of PPSF was computed using the following groups:

County x Hispanic Origin (Hispanic, Non-Hispanic) x Sex x Age Groups used for Hispanic adjustment

The above two steps were repeated up to six times, or until the change to PPSF from one iteration to the next became small.

- **Rounding** -- The final product of all person weights (BW x . . . x HPF1 x PPSF) was rounded to an integer. Rounding was performed so that the sum of the rounded weights was within one person of the sum of the unrounded weights for any of the groups listed below:

County
County x Race
County x Race x Hispanic Origin
County x Race x Hispanic Origin x Sex
County x Race x Hispanic Origin x Sex x Age
County x Race x Hispanic Origin x Sex x Age x Tract
County x Race x Hispanic Origin x Sex x Age x Tract x Block

For example, the number of White, Hispanic, Males, Age 30 estimated for a county using the rounded weights was within one of the number produced using the unrounded weights.

Final Housing Unit Weighting Factors -- This process produced the following factors:

- **Principal Person Factor (PPF)** -- This factor adjusted for differential response depending on the race, Hispanic origin, sex, and age of the principal person in the household. The principal person was defined as the female spouse of the responding householder. If there was no such person, then the responding householder was the principal person. The value of PPF for a housing unit was the PPSF of the principal person.
- **Second Housing Unit Post-stratification Factor (HPF2)** -- This factor made the number of housing units in a tract again equal to the January 1999 MAF control count totals after all factors (BW x . . . x HPF1 x PPF) had been applied. HPF2 was computed and assigned to all housing units based on the following groups.

Site x County x Tract

- **Rounding** -- The final product of all housing unit weights (BW x . . . x HPF1 x PPF x HPF2) was rounded to an integer. Rounding was performed so that total rounded weight was within one housing unit of the total unrounded weight for any of the groups listed below:

Site
Site x County
Site x County x Tract
Site x County x Tract x Block

Control of Nonsampling Error

As mentioned earlier, sample data are subject to nonsampling error. This component of error could introduce serious bias into the data, and the total error could increase dramatically over that which would result purely from sampling. While it is impossible to completely eliminate nonsampling error from a survey operation, the Census Bureau attempts to control the sources of such error during the collection and processing operations. Described below are the primary sources of nonsampling error and the programs instituted for control of this error. The success of these programs, however, is contingent upon how well the instructions actually were carried out during the survey.

Undercoverage -- It is possible for some sample housing units or persons to be missed entirely by the survey. The undercoverage of persons and housing units can introduce biases into the data. A major way to avoid undercoverage in a survey is to ensure that its sampling frame, for ACS an address list in each site, is as complete and accurate as possible.

The source of addresses in the three urban sites was a new product, the Master Address File (MAF), currently being developed by the Census Bureau. The MAF is created by combining the 1990 Census Address Control File and the Delivery Sequence File of the United States Postal Service. An attempt is made to assign all appropriate geographic codes to each MAF address via an automated procedure using the Census Bureau TIGER files. A manual coding operation based in the appropriate regional offices is attempted for addresses which could not be automatically coded. The MAF was used as the source of addresses for selecting sample housing units and mailing questionnaires. TIGER produced the location maps for personal visit CAPI assignments.

In Fulton County, PA the Census Bureau conducted a manual listing and map-spotting operation in the summer of 1995. Interviewers hand delivered ACS questionnaires to the addresses from this listing.

In the CATI and CAPI nonresponse follow-up phases, efforts were made to minimize the chances that housing units that were not part of the sample were interviewed in place of units in sample by mistake. If a CATI interviewer called a mail nonresponse case and was not able to reach the exact address, no interview was conducted and the case was eligible for CAPI. During CAPI follow-up, the interviewer had to locate the exact address for each sample housing unit. In some multi-unit structures the interviewer could not locate the exact sample unit or found a different number of units than expected. In these cases the interviewers were instructed to list the units in the building and follow a specific procedure to select a replacement sample unit.

Respondent and Interviewer Error -- The person answering the questionnaire or responding to the questions posed by an interviewer could serve as a source of error, although the questions were phrased as clearly as possible based on testing, and detailed instructions for completing the questionnaire were provided to each household. In addition, respondents' answers were edited for completeness, and problems were followed up as necessary.

Interviewer monitoring -- The interviewer may misinterpret or otherwise incorrectly enter information given by a respondent; may fail to collect some of the information for a person or household; or may collect data for households that were not designated as part of the sample. To control these problems, the work of interviewers was monitored carefully. Field staff were prepared for their tasks by using specially developed training packages that included hands-on experience in using survey materials. A sample of the households interviewed by CAPI interviewers was reinterviewed to control for the possibility that interviewers may have fabricated data.

Item Nonresponse -- Nonresponse to particular questions on the survey questionnaire and instrument allows for the introduction of bias into the data, since the characteristics of the nonrespondents have not been observed and may differ from those reported by respondents. As a result, any imputation procedure using respondent data may not completely reflect this difference either at the elemental level (individual person or housing unit) or on the average.

Some protection against the introduction of large biases is afforded by minimizing nonresponse. In the ACS, nonresponse for the CATI and CAPI operations was reduced substantially by the requirement that the automated instrument receive a response to each question before the next one could be asked. For mail responses, the clerical edit and follow-up operations were aimed at obtaining a response for every question on selected questionnaires. Values for any items that remain unanswered were imputed by computer using reported data for a person or housing unit with similar characteristics.

Clerical Review -- Questionnaires returned by mail were edited for completeness and acceptability. They were reviewed by clerks for content omissions and population coverage. If necessary, a telephone follow-up was made to obtain missing information. Potential coverage errors were included in this follow-up, as well as questionnaires with too many omissions to be accepted as returned.

Processing Error -- The many phases involved in processing the survey data represent potential sources for the introduction of nonsampling error. The processing of the survey questionnaires includes the clerical editing, follow-up by telephone, and keying of data from completed questionnaires; the manual coding of write-in responses; and the electronic data processing. The various field, coding and computer operations undergo a number of quality control checks to insure their accurate application.

Automated Editing -- After data collection was completed, any remaining incomplete or inconsistent information was imputed during the final automated edit of the collected data. Imputations, or computer assignments of acceptable codes in place of unacceptable entries or blanks, were needed most often when an entry for a given item was lacking or when the information reported for a person or housing unit on that item was inconsistent with other information for that same person or housing unit. As in other surveys and previous censuses, the general procedure for changing unacceptable entries was to assign an entry for a person or housing unit that was consistent with entries for persons or housing units with similar characteristics. Assigning acceptable values in place of blanks or unacceptable entries enhances the usefulness of the data.

Table A. Unadjusted Standard Error for Estimated Totals

[Based on a 3 percent simple random sample. For estimates from the PUMS multiply standard errors from the table or the formula by 1.75 for all sites except Broward County, FL, and Kershaw and Richland Counties, SC. For Broward multiply standard errors by 1.68, and for Richland and Kershaw multiply standard errors by 1.43]

Estimate ¹	Size of the publication Area ²											
	500	1,000	2,500	5,000	7,500	10,000	25,000	50,000	100,000	250,000	500,000	3,500,000
100	57	57	57	57	57	57	57	57	57	57	57	57
250	64	78	85	88	88	89	89	90	90	90	90	90
500	.	90	114	121	123	124	126	127	127	127	127	127
750	.	78	130	144	148	150	153	155	155	155	156	156
1,000	.	.	139	161	167	171	176	178	179	179	180	180
2,500	.	.	.	201	232	246	270	277	281	283	284	284
5,000	232	284	360	381	392	398	400	402
7,500	246	412	454	474	485	489	492
10,000	440	509	539	557	563	568
15,000	440	583	642	675	686	695
25,000	636	779	853	876	896
50,000	899	1137	1206	1262
75,000	779	1303	1436	1540
100,000	1393	1608	1772
250,000	2010	2740
500,000	3723

1 To get a better approximation of the standard error of an estimated total use the formula below.

N = Size of geographic area
 = Estimate of characteristic total

2 The population estimate of the tabulation area if the estimate is a person characteristic, or the estimate of housing units in the tabulation area if the estimate is a housing or family characteristic.

Table B. Unadjusted Standard Error in Percentage Points for Estimated Percentages

[Based on a 3 percent simple random sample. For estimates from the PUMS multiply standard errors from the table or the formula by 1.75 for all sites except Broward County, FL, and Kershaw and Richland Counties, SC. For Broward multiply standard errors by 1.68, and for Richland and Kershaw multiply standard errors by 1.43]

Estimated Percentage	Base of the Estimated Percentage ¹												
	500	750	1,000	2,500	5,000	7,500	10,000	25,000	50,000	100,000	250,000	500,000	
2.5 or 97.5	4.0	3.2	2.8	1.8	1.3	1.0	0.9	0.6	0.4	0.3	0.2	0.1	
5 or 95	5.5	4.5	3.9	2.5	1.8	1.4	1.2	0.8	0.6	0.4	0.2	0.2	
10 or 90	7.6	6.2	5.4	3.4	2.4	2.0	1.7	1.1	0.8	0.5	0.3	0.2	
15 or 85	9.1	7.4	6.4	4.1	2.9	2.3	2.0	1.3	0.9	0.6	0.4	0.3	

Estimated Percentage	Base of the Estimated Percentage ¹											
	500	750	1,000	2,500	5,000	7,500	10,000	25,000	50,000	100,000	250,000	500,000
20 or 80	10.2	8.3	7.2	4.5	3.2	2.6	2.3	1.4	1.0	0.7	0.5	0.3
25 or 75	11.0	9.0	7.8	4.9	3.5	2.8	2.5	1.6	1.1	0.8	0.5	0.3
30 or 70	11.7	9.5	8.2	5.2	3.7	3.0	2.6	1.6	1.2	0.8	0.5	0.4
35 or 65	12.1	9.9	8.6	5.4	3.8	3.1	2.7	1.7	1.2	0.9	0.5	0.4
40 or 60	12.5	10.2	8.8	5.6	3.9	3.2	2.8	1.8	1.2	0.9	0.6	0.4
45 or 55	12.7	10.3	8.9	5.7	4.0	3.3	2.8	1.8	1.3	0.9	0.6	0.4
50	12.7	10.4	9.0	5.7	4.0	3.3	2.8	1.8	1.3	0.9	0.6	0.4

1 For an estimated percentage not shown in the table, use the formula below to get standard error approximations. Use this table only for proportions, that is, where the numerator is a subset of the denominator.

B= Base of estimated percentage
 = Estimated percentage

Table C. Standard Error Design Factors - 1998 ACS Test

[Design factors are site specific, use the appropriate column.]

Characteristics	Rockland, NY	Multnomah, OR	Fulton, PA	Douglas, NE	Otero, NM	Franklin, OH	Houston, TX	Broward, FL		Columbia, SC
								1	2	
POPULATION										
Families	2.0	1.5	1.3	1.7	0.8	1.6	2.0	4.0	2.0	1.4
Households	3.0	3.2	1.7	2.3	1.3	2.8	2.7	5.2	2.8	2.8
Age and Sex	1.5	1.4	0.9	1.4	0.7	1.4	1.6	3.2	1.8	1.2
Race	2.6	3.1	2.0	2.3	1.2	2.8	3.1	5.9	3.2	2.1
Hispanic Origin	2.3	2.5	2.0	2.3	1.4	2.9	2.8	5.5	2.1	1.5
Marital Status	1.2	1.2	0.8	1.2	0.5	1.2	1.4	2.7	1.5	1.0
Ancestry	2.0	2.0	1.7	2.1	1.0	1.8	2.1	4.0	2.2	1.5
Household Size	1.3	1.3	0.9	1.3	0.6	1.3	1.4	2.8	1.6	1.1
Household Type and Relationship	1.3	1.4	0.7	1.2	0.6	1.3	1.7	3.0	1.7	1.1
Presence and Age of Children	1.4	1.4	0.9	1.3	0.7	1.3	1.5	2.8	1.6	1.1
Functional Limitation and Ability to Go Outside Alone	1.5	1.6	1.0	1.5	0.7	1.8	2.1	3.4	2.1	1.3
Work Disability	1.5	1.5	0.8	1.4	0.6	1.4	1.7	3.3	1.9	1.2
Place of Birth, Year of Entry, and Residence 5 years ago	2.7	2.0	1.6	2.2	0.9	2.2	2.5	4.2	2.3	1.8
Language Spoken at Home and English Ability	2.2	2.4	1.7	2.7	1.3	3.6	2.6	5.1	2.8	3.6
Educational Attainment, and School Enrollment	1.9	1.7	1.0	1.7	0.8	1.8	1.9	3.8	2.2	1.5
Family Type	1.3	1.2	0.9	1.2	0.6	1.2	1.4	2.7	1.5	1.0
Employment Status	1.3	1.3	0.8	1.3	0.6	1.3	1.5	2.8	1.7	1.1
Industry, Occupation, and Class of Worker	1.4	1.4	0.9	1.4	0.6	1.4	1.6	3.0	1.7	1.1
Hours Per Week and Weeks Worked in past 12 months	1.3	1.3	0.8	1.2	0.6	1.3	1.4	2.8	1.6	1.0
Number of Workers in Family	1.4	1.3	1.0	1.3	0.6	1.3	1.4	2.8	1.6	1.1
Place of Work, Means of Transportation to Work, Travel Time to Work, Private Vehicle Occupancy, and Time Leaving Home to Go to Work	1.6	1.5	1.0	1.6	0.7	1.5	1.7	3.4	1.9	1.3
Type of Income in the Past 12 Months	1.8	2.0	1.2	1.8	0.8	2.0	2.4	3.5	2.2	1.7
Household Income in the Past 12 Months	1.2	1.2	0.8	1.2	0.6	1.2	1.3	2.5	1.4	1.0
Family Income in the Past 12 Months	1.2	1.2	0.8	1.1	0.6	1.1	1.2	2.5	1.4	0.9

Characteristics	Rockland, NY	Multnomah, OR	Fulton, PA	Douglas, NE	Otero, NM	Franklin, OH	Houston, TX	Broward, FL		Columbia, SC
								1	2	
Poverty Status in the Past 12 Months by Age (persons)	1.6	1.6	1.0	1.5	0.7	1.5	1.7	3.3	1.9	1.2
Poverty Status in the Past 12 Months by Age and Household Type & Relationship (persons)	2.0	1.9	1.6	2.1	1.0	2.1	2.8	3.8	1.8	1.6
Poverty Status in the Past 12 Months (families)	1.3	1.2	0.9	1.2	0.6	1.2	1.3	2.7	1.4	1.0
Armed Forces and Veteran Status	1.4	1.4	0.8	1.3	0.6	1.4	1.6	3.0	1.7	1.1
HOUSING										
Age of Householder	1.3	1.2	0.9	1.2	0.6	1.2	1.3	2.6	1.5	1.0
Race of Householder	1.5	1.4	1.1	1.4	0.6	1.3	1.4	3.0	1.7	1.1
Hispanic Origin of Householder	1.3	1.3	1.0	1.4	0.6	1.3	1.4	2.6	1.4	1.1
Condominium Status	1.8	1.4	1.2	1.5	0.6	1.3	1.6	2.8	1.7	1.2
Units in Structure	1.6	1.4	1.2	1.6	0.7	1.4	1.6	2.9	1.6	1.3
Occupied by tenure	1.6	1.4	1.1	1.4	0.6	1.4	1.5	3.0	1.7	1.1
Vacant	1.7	1.8	1.0	1.8	0.7	1.8	1.8	3.8	2.0	1.4
Gross Rent	1.3	1.2	0.8	1.2	0.6	1.2	1.4	2.8	1.5	1.1
Year Structure Built	1.4	1.4	0.9	1.4	0.6	1.3	1.5	3.2	1.8	1.1
Rooms, Bedrooms	1.3	1.3	0.9	1.3	0.6	1.3	1.4	3.0	1.6	1.1
House Heating Fuel	2.0	1.4	1.0	2.1	0.7	1.9	1.7	4.8	2.5	1.2
Heating System	2.5	1.9	1.1	1.6	1.1	1.9	2.0	4.5	2.5	2.2
Air Conditioning	1.4	1.8	1.0	1.7	0.7	1.9	2.1	4.2	2.3	1.5
Source of Water, Plumbing Facilities, and Kitchen Facilities	2.2	1.9	1.3	2.7	0.7	2.0	2.0	2.5	1.7	1.2
Sewage Disposal	1.6	2.4	1.5	1.4	0.9	1.5	1.6	3.5	3.2	1.2
Telephone	1.9	1.5	1.1	1.5	0.7	1.5	1.7	3.4	1.9	1.2
Vehicles Available	1.3	1.3	0.9	1.2	0.6	1.2	1.4	2.8	1.5	1.0
Length at Residence	1.2	1.2	0.9	1.2	0.6	1.2	1.3	2.6	1.5	1.0
Mortgage Status and SMOC	1.2	1.1	0.8	1.1	0.5	1.1	1.2	2.4	1.3	0.9
Gross Rent as a Percent of Household Income	1.3	1.3	0.9	1.2	0.6	1.3	1.4	2.8	1.6	1.1
Value	1.2	1.1	0.9	1.1	0.5	1.1	1.2	2.5	1.3	0.9
Aggregate Persons by Tenure by Units in Structure	2.6	2.4	1.9	2.6	1.1	2.4	2.9	4.0	2.2	2.0

Top of page

[PDF] or denotes a file in Adobe's Portable Document Format. To view the file, you will need the Adobe® Acrobat® Reader available **free** from Adobe. This symbol indicates a link to a non-government web site. Our linking to these sites does not constitute an endorsement of any products, services or the information found on them. Once you link to another site you are subject to the policies of the new site.

Source: U.S. Census Bureau | American Community Survey Office | Page Last Modified: August 25, 2010