

Puerto Rico Community Survey Multiyear Accuracy of the Data (5-year 2014-2018)

INTRODUCTION

The data contained in these data products are based on the Puerto Rican Community Survey (PRCS) sample. For the 5-year data products, interviews from January 1, 2014 through December 31, 2018 were used.

In general, PRCS estimates are period estimates that describe the average characteristics of population and housing over a period of data collection. The 2014-2018 5-year PRCS estimates from January 1, 2014 to December 31, 2018, respectively. Multiyear estimates cannot be used to say what is going on in any particular year in the period, only what the average value is over the full period.

The PRCS sample is selected from all municipios in Puerto Rico (PR). In 2006, the PRCS began collection of data from sampled persons in group quarters (GQs) – for example, military barracks, college dormitories, nursing homes, and correctional facilities. Persons in group quarters are included with persons in housing units (HUs) in all 2014-2018 5-year PRCS estimates based on the total population.

The PRCS, like other statistical activity, is subject to error. The purpose of this documentation is to provide data users with a basic understanding of the PRCS sample design, estimation methodology, and accuracy of the 2014-2018 5-year PRCS estimates. The PRCS is sponsored by the U.S. Census Bureau, and is part of the Decennial Census Program.

Additional information on the design and methodology of the PRCS, including data collection and processing, can be found at:

<https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html>

The **Multiyear Accuracy of the Data from the American Community Survey** can be at:

<https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>

Table of Contents

INTRODUCTION.....	1
DATA COLLECTION	3
SAMPLE DESIGN.....	3
WEIGHTING METHODOLOGY.....	4
ESTIMATION METHODOLOGY FOR MULTIYEAR ESTIMATES.....	5
CONFIDENTIALITY OF THE DATA	6
Title 13, United States Code.....	6
Disclosure Avoidance	6
Data Swapping	7
Synthetic Data	7
ERRORS IN THE DATA.....	7
Sampling Error.....	7
Nonsampling Error	7
MEASURES OF SAMPLING ERROR.....	8
Confidence Intervals and Margins of Error	8
Limitations.....	10
CALCULATION OF STANDARD ERRORS	10
Approximating Standard Errors and Margins of Error	12
TESTING FOR SIGNIFICANT DIFFERENCES	12
CONTROL OF NONSAMPLING ERROR.....	12
Coverage Error	13
Nonresponse Error	14
Measurement and Processing Error	15

DATA COLLECTION

The PRCS employs three modes of data collection:

- Mailout/Mailback
- Computer Assisted Telephone Interview (CATI)¹
- Computer Assisted Personal Interview (CAPI)

The general timing of data collection is:

- Month 1: Addresses determined to be mailable are sent a questionnaire via the U.S. Postal Service.
- Month 2: All mail non-responding addresses with an available phone number are sent to CATI.
- Month 3: A sample of mail non-responses without a phone number, CATI non-responses, and unmailable addresses are selected and sent to CAPI.

SAMPLE DESIGN

Sampling rates are assigned independently at the census block level. A measure of size is calculated for each municipio. The measure of size is an estimate of the number of occupied housing units in the municipio. This is calculated by multiplying the number of PRCS addresses by an estimate of the occupancy rate from the 2010 Census and the PRCS at the block level. A measure of size for each Census Tract is also calculated in the same manner.

Each block is then assigned the smallest measure of size from the set of all entities of which it is a part. Sampling rates are shown in Table 1. Beginning in 2011 the PRCS (along with the ACS) implemented a sample reallocation, increasing the number of second-stage sampling strata from seven to 16. Not all of the 16 sampling strata are applicable in Puerto Rico. Table 1 gives only the sampling rates for the PRCS that are in applicable strata.

¹ Note that all CATI operations ended at the end of September, 2017.

Table 1. 2014 Through 2018 Average Sampling Rates for Puerto Rico by Sampling Stratum

Stratum Thresholds	Sampling Rate
$0 < \text{MOS}^1 \leq 200$	15.00%
$200 < \text{MOS} \leq 400$	10.00%
$400 < \text{MOS} \leq 800$	7.00%
$400 < \text{MOS} < 1,200$	3.92%
$1,200 \leq \text{MOS}$ -and- $\text{TRACTMOS}^2 < 400$	4.40%
$1,200 \leq \text{MOS}$ -and- $400 \leq \text{TRACTMOS} < 1,000$	3.17%
$1,200 \leq \text{MOS}$ -and- $1,000 \leq \text{TRACTMOS} < 2,000$	1.75%
$1,200 \leq \text{MOS}$ -and- $2,000 \leq \text{TRACTMOS} < 4,000$	0.94%
$1,200 \leq \text{MOS}$ -and- $4,000 \leq \text{TRACTMOS} < 6,000$	0.52%
$1,200 \leq \text{MOS}$ -and- $6,000 \leq \text{TRACTMOS}$	0.49%

¹MOS = Measure of size of the smallest governmental entity

²TRACTMOS = Census Tract measure of size.

Addresses determined to be unmailable do not go to the CATI phase of data collection and are subsampled for the CAPI phase of data collection at a rate of 2-in-3. Subsequent to CATI, all addresses for which no response has been obtained are subsampled. This subsample is sent to the CAPI data collection phase. Beginning with the CAPI sample for the January 2006 panel (March 2006 data collection), the CAPI subsampling rate was based on the expected rate of completed mail and CATI interviews at the tract level.

Table 2. 2014 Through 2018 CAPI Subsampling Rates for Puerto Rico

Address and Tract Characteristics	2006 through 2018 CAPI Subsampling Rates
Unmailable addresses	66.7%
Mailable addresses (June through December)	50.0%

For a more detailed description of the PRCS sampling methodology, see the PRCS Accuracy of the Data document. This document is available for 2018 as well as prior data years at: <https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>.

WEIGHTING METHODOLOGY

The multiyear estimates should be interpreted as estimates that describe a time period rather than a specific reference year. For example, a 5-year estimate for the poverty rate of a given area describes the total set of people who lived in that area over those five years much the same way as a 1-year estimate for the same characteristic describes the set of people who lived in that area over one year. The only fundamental difference between the estimates is the number of months of collected data, which are considered in forming the estimate. For this reason, the estimation procedure used for the multiyear estimates is an extension of the 2018 1-year estimation procedure. In this document only the procedures that are unique to the multiyear estimates are discussed.

To weight the 5-year estimates, 60 months of collected data are pooled together. The pooled data are then reweighted using the procedures developed for the 2018 1-year estimates with a

few adjustments. These adjustments concern geography, month-specific weighting steps, and population controls. In addition to these adjustments, there is one multiyear specific model-assisted weighting step.

Some of the weighting steps use the month of tabulation in forming the weighting cells within which the weighting adjustments are made. One such example is the variation in monthly response adjustment. In these weighting steps, the month of tabulation is used independently of year. Thus, for the 5-year, sample cases from May 2014, May 2015, May 2016, May 2017, and May 2018 are combined.

Since the multiyear estimates represent estimates for the period, the controls are not a single year's population estimates from the Population Estimates Program, but rather are an average of these estimates over the period. The population controls by age and sex are obtained by taking a simple average of the 1-year population estimates of the municipio or weighting area by age and sex. For example, the 2014-2018 control total used for males age 20-24 in a given municipio would be obtained by averaging the 1-year population estimates for that demographic group for 2014, 2015, 2016, 2017, and 2018. The version or vintage of estimates used is always that of the last year of the period since these are considered to be the most up to date and are created using a consistent methodology.

The GQ weighting methodology imputes GQ person records into the 2014-2018 PRCS 5-year. See the American Community Survey Accuracy of the Data (2018) for details on the GQ imputation.

In addition, a finite population correction (FPC) factor is included in the creation of the replicate weights for the 5-year data at the tract level. It reduces the estimate of the variance and the margin of error by considering the sampling rate. A two-tiered approach was used. One FPC was calculated for mail, internet, and CATI respondents and another for CAPI respondents. The CAPI was given a separate FPC to take into account the fact that CAPI respondents are subsampled. The FPC is not included in the 1-year data because the sampling rates are relatively small and thus the FPC does not have an appreciable impact on the variance.

For more information on the replicate weights and replicate factors, see the Design and Methodology Report at: <https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html>

ESTIMATION METHODOLOGY FOR MULTIYEAR ESTIMATES

For the 1-year estimation, the tabulation geography for the data is based on the boundaries defined on January 1 of the tabulation year, which is consistent with the tabulation geography used to produce the population estimates. All sample addresses are updated with this geography prior to weighting. For the multiyear estimation, the tabulation geography for the data is referenced to the final year in the multiyear period. For example, the 2014-2018 period uses the 2018 reference geography. Thus, all data collected over the period of 2014-2018 in the

blocks that are contained in the 2018 boundaries for a given place are tabulated as though they are a part of that place for the entire period.

Monetary values for the PRCS multiyear estimates are inflation-adjusted to the final year of the period. For example, the 2014-2018 PRCS 5-year estimates are tabulated using 2018-adjusted dollars. These adjustments use the national Consumer Price Index (CPI) since a regional-based CPI is not available for the entire country.

For a more detailed description of the PRCS estimation methodology, see the Accuracy of the Data document. This document is available for 2018 and prior data years at:

<https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>

CONFIDENTIALITY OF THE DATA

The Census Bureau has modified or suppressed some data on this site to protect confidentiality. Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's data can be identified.

The Census Bureau's internal Disclosure Review Board sets the confidentiality rules for all data releases. A checklist approach is used to ensure that all potential risks to the confidentiality of the data are considered and addressed.

Title 13, United States Code

Title 13 of the United States Code authorizes the Census Bureau to conduct censuses and surveys. Section 9 of the same Title requires that any information collected from the public under the authority of Title 13 be maintained as confidential. Section 214 of Title 13 and Sections 3559 and 3571 of Title 18 of the United States Code provide for the imposition of penalties of up to five years in prison and up to \$250,000 in fines for wrongful disclosure of confidential census information.

Disclosure Avoidance

Disclosure avoidance is the process for protecting the confidentiality of data. A disclosure of data occurs when someone can use published statistical information to identify an individual who has provided information under a pledge of confidentiality. For data tabulations, the Census Bureau uses disclosure avoidance procedures to modify or remove the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual, the Census Bureau has taken steps to disguise or suppress the original data while making sure the results are still useful. The techniques used by the Census Bureau to protect confidentiality in tabulations vary, depending on the type of data. All disclosure avoidance procedures are done prior to the whole person imputation into not-in-sample GQ facilities.

Data Swapping

Data swapping is a method of disclosure avoidance designed to protect confidentiality in tables of frequency data (the number or percent of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases when creating a table. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas that have similar characteristics (such as the same number of adults and same number of children). Because the swap often occurs within a neighboring area, there is no effect on the marginal totals for the area or for totals that include data from multiple areas. Because of data swapping, users should not assume that tables with cells having a value of one or two reveal information about specific individuals. Data swapping procedures were first used in the 1990 Census, and were used again in Census 2000 and the 2010 Census.

Synthetic Data

The goals of using synthetic data are the same as the goals of data swapping, namely to protect the confidentiality in tables of frequency data. Persons are identified as being at risk for disclosure based on certain characteristics. The synthetic data technique then models the values for another collection of characteristics to protect the confidentiality of that individual.

Note: The data use the same disclosure avoidance methodology as the original 1-year data. The confidentiality edit was previously applied to the raw data files when they were created to produce the 1-year estimates and these same data files with the original confidentiality edit were used to produce the 5-year estimates.

ERRORS IN THE DATA

Sampling Error

The data in PRCS products are estimates of the actual figures that would be obtained by interviewing the entire population. The estimates are a result of the chosen sample, and are subject to sample-to-sample variation. Sampling error in data arises due to the use of probability sampling, which is necessary to ensure the integrity and representativeness of sample survey results. The implementation of statistical sampling procedures provides the basis for the statistical analysis of sample data. Measures used to estimate the sampling error are provided in the next section.

Nonsampling Error

Other types of errors might be introduced during any of the various complex operations used to collect and process survey data. For example, data entry from questionnaires and editing may introduce error into the estimates. Another potential source of error is the use of controls in the weighting. These controls are based on Population Estimates and are designed to reduce variance and mitigate the effects of systematic undercoverage of groups who are difficult to enumerate. However, if the extrapolation methods used in generating the Population Estimates

do not properly reflect the population, error can be introduced into the data. This potential risk is offset by the many benefits the controls provide to the PRCS estimates, which include the reduction of issues with survey coverage and the reduction of standard errors of PRCS 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, or errors that consistently skew the data 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.

An important goal of the PRCS 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. For more information, please see the section entitled “Control of Nonsampling Error”.

MEASURES OF SAMPLING ERROR

Sampling error is the difference between an estimate based on a sample and the corresponding value that would be obtained if the entire population were surveyed (as for a census). Note that sample-based estimates will vary depending on the particular sample selected from the population. Measures of the magnitude of sampling error reflect the variation in the estimates over all possible samples that could have been selected from the population using the same sampling methodology.

Estimates of the magnitude of sampling errors – in the form of margins of error – are provided with all published PRCS data. The Census Bureau recommends that data users incorporate margins of error into their analyses, as sampling error in survey estimates could impact the conclusions drawn from the results.

Confidence Intervals and Margins of Error

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 PRCS 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 1.645 times the estimated standard error below the estimate to 1.645 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.

Margins of Error

In lieu of providing upper and lower confidence bounds in published PRCS tables, the margin of error is listed. All PRCS published margins of error are based on a 90 percent confidence level. The margin of error is the difference between an estimate and its upper or lower confidence bound. Both the confidence bounds and the standard error can easily be computed from the margin of error:

$$\text{Standard Error} = \text{Margin of Error} / 1.645$$

$$\text{Lower Confidence Bound} = \text{Estimate} - \text{Margin of Error}$$

$$\text{Upper Confidence Bound} = \text{Estimate} + \text{Margin of Error}$$

Note that for 2005 and earlier estimates, PRCS margins of error and confidence bounds were calculated using a 90 percent confidence level multiplier of 1.65. Starting with the 2006 data release, the more accurate multiplier of 1.645 is used. Margins of error and confidence bounds from previously published products will not be updated with the new multiplier. When calculating standard errors from margins of error or confidence bounds using published data for 2005 and earlier, use the 1.65 multiplier.

When constructing confidence bounds from the margin of error, users should be aware of any “natural” limits on the bounds. For example, if a characteristic estimate for the population is near zero, the calculated value of the lower confidence bound may be negative. However, as a negative number of people does not make sense, the lower confidence bound should be reported as zero. For other estimates such as income, negative values can make sense; in these cases, the lower bound should not be adjusted. The context and meaning of the estimate must therefore be kept in mind when creating these bounds. Another example of a natural limit is 100 percent as the upper bound of a percent estimate.

If the margin of error is displayed as ‘*****’ (five asterisks), the estimate has been controlled to be equal to a fixed value and so it has no sampling error. A standard error of zero should be used for these controlled estimates when completing calculations, such as those in the following section.

Limitations

Users should be careful when computing and interpreting confidence intervals.

Nonsampling Error

The estimated standard errors (and thus margins of error) included in these data products do not account for 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 or the effect of imputed values due to missing responses. The standard errors calculated are only lower bounds 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). Some care must be exercised in the interpretation of the data based on the estimated standard errors.

Very Small (Zero) or Very Large Estimates

By definition, the value of almost all PRCS characteristics is greater than or equal to zero. The method provided above for calculating confidence intervals relies on large sample theory, and may result in negative values for zero or small estimates for which negative values are not admissible. In this case, the lower limit of the confidence interval should be set to zero by default. A similar caution holds for estimates of totals close to a control total or estimated proportion 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.

CALCULATION OF STANDARD ERRORS

Direct estimates of margin of error were calculated for all estimates reported. The margin of error is derived from the variance. In most cases, the variance is calculated using a replicate-based methodology known as successive difference replication (SDR) that takes into account the sample design and estimation procedures.

The SDR formula as well as additional information on the formation of the replicate weights, see Chapter 12 of the Design and Methodology documentation at:

<https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html>.

Beginning with the PRCS 2011 1-year estimates, a new imputation-based methodology was incorporated into processing (see the description in the Group Quarters Person Weighting Section). An adjustment was made to the production replicate weight variance methodology to account for the non-negligible amount of additional variation being introduced by the new technique.²

Excluding the base weights, replicate weights were allowed to be negative in order to avoid underestimating the standard error. Exceptions include:

1. *The estimate of the number or proportion of people, households, families, or housing units in a geographic area with a specific characteristic is zero.* A special procedure is used to estimate the standard error.
2. *There are either no sample observations available to compute an estimate or standard error of a median, an aggregate, a proportion, or some other ratio, or there are too few sample observations to compute a stable estimate of the standard error.* The estimate is represented in the tables by “-” and the margin of error by “**” (two asterisks).
3. *The estimate of a median falls in the lower open-ended interval or upper open-ended interval of a distribution.* If the median occurs in the lowest interval, then a “-” follows the estimate, and if the median occurs in the upper interval, then a “+” follows the estimate. In both cases, the margin of error is represented in the tables by “***” (three asterisks).

Calculating Measures of Error Using Variance Replicate Tables

Advanced users may be interested in the Variance Replicate Tables. These augmented PRCS Detailed Tables include sets of 80 replicate estimates, which allow users to calculate measures of error for derived estimates using the same methods that are used to produce the published MOEs on data.census.gov. These methods incorporate the covariance between estimates that the approximation formulas in this document leave out.

The Variance Replicate Tables are available for a subset of the 5-year Detailed Tables for eleven summary levels. These will be released on an annual basis, shortly after the release of the regular 5-year data products.

² For more information regarding this issue, see Asiala, M. and Castro, E. 2012. Developing Replicate Weight-Based Methods to Account for Imputation Variance in a Mass Imputation Application. In JSM proceedings, Section on Survey Research Methods, Alexandria, VA: American Statistical Association.

The Variance Replicate Tables can be found at: <https://census.gov/programs-surveys/acs/data/variance-tables.html>

The Variance Replicate Documentation (including table list and summary level list) can be found at: <https://census.gov/programs-surveys/acs/technical-documentation/variance-tables.html>

Approximating Standard Errors and Margins of Error

Previously, this document included formulas for approximating the standard error (SE) for various types of estimates. For example, summing estimates or calculating a ratio of two or more estimates. These formulas are also found in the Instruction for Statistical Testing documents, which is located at <https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>. In addition, the worked examples have also been placed in the same location in the document called “Worked Examples for Approximating Margins of

TESTING FOR SIGNIFICANT DIFFERENCES

Users may conduct a statistical test to see if the difference between an PRCS estimate and any other chosen estimate is statistically significant at a given confidence level. “Statistically significant” means that it is not likely that the difference between estimates is due to random chance alone.

To perform statistical significance testing, data users will need to calculate a Z statistic. The equation is available in the Instructions for Statistical Testing, which is located at <https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.html>.

Users completing statistical testing may be interested in using the ACS Statistical Testing Tool. This automated tool allows users to input pairs and groups of estimates for comparison. For more information on the Statistical Testing Tool, visit <https://www.census.gov/programs-surveys/acs/guidance/statistical-testing-tool.html>.

CONTROL OF NONSAMPLING ERROR

As mentioned earlier, sample data are subject to nonsampling error. Nonsampling error can introduce serious bias into the data, increasing the total error 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 to control for this error.³

Coverage Error

It is possible for some sample housing units or persons to be missed entirely by the survey (undercoverage). It is also possible for some sample housing units and persons to be counted more than once (overcoverage). Both undercoverage and overcoverage of persons and housing units can introduce bias into the data. Coverage error can also increase both respondent burden and survey costs.

To avoid coverage error in a survey, the frame must be as complete and accurate as possible. For the PRCS, the frame is an address list, the source of which is the Master Address File (MAF). An attempt is made to assign each MAF address to the appropriate geographic codes via an automated procedure using the Census Bureau TIGER (Topologically Integrated Geographic Encoding and Referencing) files. A manual coding operation based in the appropriate regional offices is attempted for addresses that 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 CAPI assignments. Sometimes the MAF contains duplicates of addresses. This could occur when there is a slight difference in the address such as 123 Main Street versus 123 Maine Street, and can introduce overcoverage.

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 mistakenly interviewed instead of units in sample. 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 became eligible for CAPI. During the CAPI follow-up, the interviewer had to locate the exact address for each sample housing unit. If the interviewer could not locate the exact sample unit in a multi-unit structure, or found a different number of units than expected, the interviewers were instructed to list the units in the building and follow a specific procedure to select a replacement sample unit. Person overcoverage can occur when an individual is included as a member of a housing unit but does not meet PRCS residency rules.

Coverage rates give a measure of undercoverage or overcoverage of persons or housing units in a given geographic area. Rates below 100 percent indicate undercoverage, while rates above 100 percent indicate overcoverage. Coverage rates are released concurrent with the release of estimates on data.census.gov in the B98 series of detailed tables (Table IDs B98011, B98012,

³ The success of these programs is contingent upon how well the instructions were carried out during the survey.

B98013, and B980014). Coverage rate definitions and coverage rates are also available at: <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/>.

Nonresponse Error

Survey nonresponse is a well-known source of nonsampling error. There are two types of nonresponse error – unit nonresponse and item nonresponse. Nonresponse errors affect survey estimates to varying levels depending on amount of nonresponse and the extent to which the characteristics of nonrespondents differ from those of respondents. The exact amount of nonresponse error or bias on an estimate is almost never known. Therefore, survey researchers generally rely on proxy measures, such as the nonresponse rate, to indicate the potential for nonresponse error.

Unit Nonresponse

Unit nonresponse is the failure to obtain data from housing units in the sample. Unit nonresponse may occur because households are unwilling or unable to participate, or because an interviewer is unable to make contact with a housing unit. Unit nonresponse is problematic when there are systematic or variable differences in the characteristics of interviewed and non-interviewed housing units. Nonresponse bias is introduced into an estimate when differences are systematic; the nonresponse error of an estimate evolves from variable differences between interviewed and non-interviewed households.

The PRCS made every effort to minimize unit nonresponse, and thus, the potential for nonresponse error. First, the PRCS used a combination of mail, CATI, and CAPI data collection modes to maximize response. The mail phase included a series of three to four mailings to encourage housing units to return the questionnaire. Subsequently, mail nonrespondents (for which phone numbers are available) were contacted by CATI for an interview. Finally, a subsample of the mail and telephone nonrespondents were contacted by personal visit to attempt an interview. Combined, these three efforts resulted in a very high overall response rate for the PRCS.

PRCS response rates measure the percent of units with a completed interview. The higher the response rate (and, consequently, the lower the nonresponse rate), the lower the chance that estimates are affected by nonresponse bias. Response and nonresponse rates, as well as rates for specific types of nonresponse, are released concurrent with the release of estimates on data.census.gov in the B98 series of detailed tables (Table IDs B98021 and B98022). Unit response rate definitions and unit response rates by type are also available at: <https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/>.

Item Nonresponse

Nonresponse to particular questions on the survey can introduce error or bias into the data, as the unknown characteristics of nonrespondents may differ from those of respondents. As a result, any imputation procedure using respondent data may not completely reflect difference either at the elemental level (individual person or housing unit) or on average.

Some protection against the introduction of large errors or biases is afforded by minimizing nonresponse. In the PRCS, item nonresponse for the CATI and CAPI operations was minimized by requiring that the automated instrument receive a response to each question before the next question could be asked. Questionnaires returned by mail were reviewed by computer for content omissions and population coverage and edited for completeness and acceptability. If necessary, a telephone follow-up was made to obtain missing information. Potential coverage errors were included in this follow-up.

Allocation tables provide the weighted estimate of persons or housing units for which a value was imputed, as well as the total estimate of persons or housing units that were eligible to answer the question. The smaller the number of imputed responses, the lower the chance that the item nonresponse is contributing a bias to the estimates. Allocation tables are released concurrent with the release of estimates on data.census.gov in the B99 series of detailed tables with the overall allocation rates across all person and housing unit characteristics in the B98 series of detailed tables (Table IDs B98031 and B98032). Allocation rate definitions and allocation rates by characteristic are also available at:

<https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/>.

Measurement and Processing Error

Measurement error can arise if the person completing the questionnaire or responding an interviewer's questions responds incorrectly. However, to mitigate this risk, the phrasing survey questions underwent cognitive testing and households were provided detailed instructions on how to complete the questionnaire.

Processing error can be introduced in numerous areas during data collection and capture, including during interviews, during data processing and during content editing.

Interviewer monitoring

An interviewer could introduce error by:

1. Misinterpreting or otherwise incorrectly entering information given by a respondent.
2. Failing to collect some of the information for a person or household.
3. Collecting data for households that were not designated as part of the sample.

To control for these problems, the work of interviewers was monitored carefully. Field staff was 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 also reinterviewed to control for the possibility that interviewers may have fabricated data.

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 keying of data from completed questionnaires, automated clerical review, follow-up by telephone, manual coding of write-in responses, and automated data processing. The various field, coding and computer operations undergo a number of quality control checks to insure their accurate application.

Content Editing

After data collection was completed, any remaining incomplete or inconsistent information was imputed during the final content edit of the collected data. Imputations, or computer assignments of acceptable codes in place of unacceptable entries or blanks, were most often needed either when an entry for a given item was missing or when information reported for a person or housing unit was inconsistent with other information for the same person or housing unit. As in other surveys and previous censuses, unacceptable entries were to allocated entries for persons or housing units with similar characteristics. Imputing acceptable values in place of blanks or unacceptable entries enhances the usefulness of the data.