Puerto Rico Community Survey
Accuracy of the Data (2025)

INTRODUCTION

This document describes the accuracy of the 2025 Puerto Rico Community Survey (PRCS) 1-year estimates.[1] The data contained in these data products are based on the PRCS-sample interviewed from January 1, 2025 through December 31, 2025.

The PRCS sample is selected from all municipios in Puerto Rico (PR). Data for Puerto Rico was first released in 2005. In 2006, the PRCS began collecting data from sampled persons in group quarters (GQs) – for example, military barracks, college dormitories, nursing homes, and correctional facilities. Sampled persons in GQs and in housing units (HUs) are included in all 2025 PRCS estimates that are based on the total population.

The PRCS, like any other sample survey, is subject to error. The purpose of this document is to provide data users with a basic understanding of the PRCS sample design, estimation methodology, and accuracy of the PRCS data. The PRCS is sponsored by the U.S. Census Bureau and is part of the Decennial Census Program.

For additional information on the design and methodology of the PRCS, including data collection and processing, visit: https://www.census.gov/programs-surveys/acs/methodology.html. To access other accuracy of the data documents, including the **2025 American Community Survey (ACS) Accuracy of the Data document** and the **2021-2025 PRCS Accuracy of the Data document**[2], visit:

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

---

[1] The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-001-0000005039, Disclosure Review Board (DRB) approval number: CBDRB-FY27-DSEP-002).

[2] The 2021-2025 Accuracy of the Data document will be available after the release of the 5-year products.

Table of Contents

INTRODUCTION **1**
DATA COLLECTION **3**
Housing Unit Addresses 3
Group Quarters 3
SAMPLING FRAME **4**
Housing Unit Addresses 4
Group Quarters 4
SAMPLE DESIGN **5**
Housing Units 5
Group Quarters 7
2024 Sample Sizes for Housing Unit Addresses and Group Quarters 11
WEIGHTING METHODOLOGY **11**
Group Quarters Person Weighting 11
Housing Unit and Household Person Weighting 13
CONFIDENTIALITY OF THE DATA **17**
Title 13, United States Code 17
Disclosure Avoidance 17
ERRORS IN THE DATA **17**
Sampling Error 17
Nonsampling Error 18
MEASURES OF SAMPLING ERROR **18**
Confidence Intervals and Margins of Error 19
Limitations 20
CALCULATION OF STANDARD ERRORS **20**
Approximating Standard Errors and Margins of Error 21
TESTING FOR SIGNIFICANT DIFFERENCES **22**
CONTROL OF NONSAMPLING ERROR **22**
Coverage Error 22
Nonresponse Error 23
Measurement and Processing Error 24

DATA COLLECTION

Housing Unit Addresses

The PRCS employs two modes of data collection:

1. Mailout/Mailback
2. Computer Assisted Personal Interview (CAPI)

The general timing of data collection is as follows. Note that mail responses are accepted during all three months of data collection.

Month 1: Mailable addresses in sample are sent an initial mailing package, which contains a paper questionnaire. An internet response mode is not currently available in Puerto Rico..

Month 2: Continued collection by mail.

Month 3: A sample of mailable nonresponding addresses and unmailable addresses is selected and sent to CAPI.

Group Quarters

Group Quarters data collection spans six weeks, except for Federal prisons that are managed by the Federal Bureau of Prisons (BOP), where the data collection time period is four months. All Federal prisons that are managed by the BOP are assigned to September, where data collection activities are in effect through December.

Field representatives have several options available to them for data collection. They can complete the questionnaire with the resident either in person or over the telephone, conduct a personal interview with a proxy, such as a relative or guardian, or leave a paper questionnaire for residents to complete.

SAMPLING FRAME

Housing Unit Addresses

The universe for the PRCS consists of all valid, residential housing unit addresses in all municipios in Puerto Rico that are eligible for data collection. Beginning with the 2018 sample, we restricted the universe of eligible addresses further to exclude a small proportion of addresses that do not meet a set of minimum address criteria.

The Master Address File (MAF) is a database maintained by the Census Bureau containing a listing of residential, group quarters, and commercial addresses in the U.S. and Puerto Rico. The MAF is updated with the results from various Census Bureau field operations, Geographic Support System partnership files, and state or local government files. The MAF is also normally updated twice each year with the Delivery Sequence Files (DSF) provided by the U.S. Postal Service. These files identify mail drop points and provide the best available source of changes and updates to the housing unit inventory.

Group Quarters

The universe of group quarters for the PRCS consists of all valid GQs in all municipios in Puerto Rico that are eligible for data collection. Due to operational difficulties associated with data collection, the PRCS excludes certain types of GQs from the sampling universe and data collection operations. The weighting and estimation accounts for this segment of the population as they are included in the population controls. The following GQ types are those that are removed from the GQ universe:

- Soup kitchens
- Domestic violence shelters
- Regularly scheduled mobile food vans
- Targeted non-sheltered outdoor locations
- Maritime/merchant vessels
- Living quarters for victims of natural disasters

SAMPLE DESIGN

Housing Units

The PRCS employs a two-phase, two-stage sample design. The first-phase sample consists of two separate address samples: Period 1 and Period 2. These samples are chosen at different points in time. Both samples are selected in two stages of sampling, a first-stage and a second-stage. Subsequent to second-stage sampling, sample addresses are systematically assigned to one of the twelve months of the sample year. The second phase of sampling occurs when the CAPI sample is selected.

The Period 1 sample is selected during September and October of the year prior to the sample year (the 2025 Period 1 sample was selected in September and October of 2024). Approximately half of the sample is selected at this time. Each address in the Period 1 sample is systematically assigned to one of the first six months of the sample year.

Period 2 sampling occurs early in the sample year (the 2025 Period 2 sample was selected during February and March of 2025). This sample accounts for the remaining half of the overall first-phase sample. Period 2 sample addresses are systematically assigned to one of the last six months of the sample year.

A sample of nonresponding addresses and of any addresses deemed unmailable is selected for the CAPI data collection mode.

The following steps are used to select the first-phase and second-phase samples in both periods.

First-Phase Housing Unit Sample Selection

First-Stage Sampling for Housing Units

First stage sampling defines the universe for the second stage of sampling through three steps. First, all addresses that were in a first-phase sample within the past four years are excluded from eligibility. This ensures that no address is in sample more than once in any five-year period. The second step is to select a 20 percent systematic sample of "new" units, i.e. those units that have never appeared on a previous MAF extract within each Municipio. Each new address is systematically assigned to either the current year or to one of the four back-samples. This procedure maintains five relatively equal partitions of the universe. The third step is to randomly assign all eligible addresses to a period.[3]

[3] Most of the period assignments are made during Period 1 sampling. The only assignments in Period 2 sampling are made for addresses that were not part of the process in Period 1, e.g., new addresses.

Assignment of Blocks to a Second-Stage Sampling Stratum for Housing Units

There are sixteen second-stage strata to which blocks in Puerto Rico (PR) can be assigned; in 2025, they were assigned to only five of the strata. The stratum-level rates used in second-stage sampling account for the first-stage selection probabilities. These rates are applied at a block level to addresses in PR by calculating a measure of size for each of the following geographic entities:

- Counties (municipios)
- Places
- Minor Civil Divisions

The measure of size for each area is an estimate of the number of occupied HUs in the area. This is calculated by multiplying the number of PRCS valid addresses by an estimate of the occupancy rate at the block level derived from the most recent Census. A measure of size for each Census Tract is also calculated in the same manner.

Each block is then assigned the smallest positive measure of size from the set of all entities of which it is a part. The 2025 second-stage sampling strata and the overall first-phase sampling rates by Period are shown in Table 1 below.

Calculation of the Second-Stage Sampling Rates for Housing Units

The overall first-phase sampling rates are calculated using the distribution of PRCS valid, eligible addresses by second-stage sampling stratum in such a way as to yield an overall target sample size for the year of approximately 36,000 (18,000 for each period) in PR. The first-phase rates are adjusted for the first-stage sample to yield the second-stage selection probabilities.

Table 1. First-phase Sampling Rate Categories for Puerto Rico

Sampling Stratum #	Type of Area	Rate Definitions	2025 Sampling Rates Period 1	2025 Sampling Rates Period 2
3	$400 \leq$ MOS[1] $<$ 800	7.00%	7.00%	7.00%
5	$1200 \leq$ MOS *and* $0 <$ TRACTMOS[2] $\leq$ 400	3.50 × BR	4.79%	4.80%
7	$1200 \leq$ MOS *and* $400 <$ TRACTMOS ≤ 1000	2.80 × BR	3.83%	3.84%
9	$1200 \leq$ MOS *and* $1000 <$ TRACTMOS ≤ 2000	1.70 × BR	2.33%	2.33%
11	$1200 \leq$ MOS *and* $2000 <$ TRACTMOS ≤ 4000	BR[3]	1.37%	1.37%

Note: A subset of sampling strata is listed here because not all of the stateside sampling strata contained addresses on the frame for Puerto Rico.
[1]MOS = measure of size (estimated number of occupied housing units) of the smallest governmental entity
[2]TRACTMOS = the measure of size (MOS) at the Census Tract level
[3]BR = base sampling rate

Second-Stage Sample Selection for Housing Units

After each block is assigned to a second-stage sampling stratum in each period, a systematic sample of addresses is selected from the second-stage universe (first-stage sample) within each municipio.

Sample Month Assignment for Housing Units

After the second stage of sampling, all sample addresses are systematically assigned to a sample month. Addresses selected during Period 1 sampling are systematically assigned to the first six months of the sample year; sample addresses selected during Period 2 sampling are systematically assigned to the last six months of the sample year.

Second-Phase Housing Unit Sample Selection – CAPI Sampling

All addresses determined to be unmailable are sampled for the CAPI phase of data collection at a rate of 2-in-3 (unmailable addresses are not eligible for any other mode of data collection). All mailable addresses for which no response has been obtained prior to CAPI are sampled at a rate of 1-in-2. Puerto Rico CAPI rates are summarized in Table 2.

Table 2. Second-Phase (CAPI) Sampling Rates for Puerto Rico

Address Characteristics	CAPI Sampling Rate
Unmailable addresses	66.7%
Mailable addresses	50.0%

Since 2020, resource constraints have prevented the selection of the full number of addresses for CAPI interviewing based on the sampling rates in Table 2. Instead, there are monthly, varying caps on the total CAPI sample. The sampling rates for the strata are adjusted proportionately to meet the monthly cap.

Group Quarters

The 2025 group quarters (GQ) sampling frame was divided into two strata: a small GQ stratum and a large GQ stratum. Small GQs are defined to have expected populations of fifteen or fewer residents, while large GQs have expected populations of more than fifteen residents.

Samples were selected in two phases within each stratum. In general, GQs were selected in the first phase and then persons/residents were selected in the second phase. Both phases differ between the two strata. GQs were assigned to one or more months in 2025 – it was in these months that their person samples were selected, with the exception of BOP GQs. See the Group Quarter Sample Month Assignment Method and Bureau of Prisons Group Quarters sections below.

Small GQ Stratum

First Phase of Sample Selection for Small GQs

There are two stages of selecting small GQs for sample.

1. First stage

   The small GQ universe is divided into five groups that are approximately equal in size, similar to what is done during the HU address sampling. All new small GQs are systematically assigned to one of these five groups on a yearly basis, with about the same probability (20 percent) of being assigned to any given group. Each group represents a second-stage sampling frame, from which GQs are selected once every five years. The 2025 second-stage sampling frame was used in 2020 as well, and is currently to be used in 2030, 2035, etc.

2. Second stage

   GQs were systematically selected from the 2025 second-stage sampling frame. Each GQ had the same second-stage probability of being selected.

Second Phase of Sample Selection for Small GQs

Persons were selected for sample from each GQ that was selected for sample in the first phase of sample selection. If fifteen or fewer persons were residing at a GQ at the time a field representative (interviewer) visited the GQ, then all persons were selected for sample. Otherwise, if more than fifteen persons were residing at the GQ, then the instrument selected a systematic sample of ten persons from the GQ's roster.

Target Sampling Rate (Probability of Selection) for Small GQs

The target sampling rate is the overall probability of selecting any given person in a GQ; it is around this probability that the sample design is based. This probability reflects both phases of sample selection. The target rate for Puerto Rico in 2025 was 2.45 percent.

The sample was designed so that the second-phase sampling rate would be one hundred percent for small GQs, (i.e., select the entire expected population of fifteen or fewer persons for sample in every small sampled GQ). This means the probability of selecting any person in a small GQ was designed to equal the probability of selecting the small GQ itself (2.45 percent in 2025).

Large GQ Stratum

First Phase of Sample Selection for Large GQs

All large GQs were eligible to be sampled in 2025, as has been the case every year since the inception of the GQ sampling in 2006. This means there was only a single stage of sampling in this phase. This stage consists of systematically assigning "hits" to GQs, where each hit represents ten persons to be sampled.

In general, a GQ has either Z or Z+1 hits assigned to it. The value for Z is dependent on both the GQ's expected population size and its target sampling rate. When this rate is multiplied by a GQ's expected population, the result is a GQ's expected person sample size. If a GQ's expected person sample size is less than ten, then Z = 0; if it is at least ten but less than twenty, then Z = 1; if it is at least twenty but less than thirty, then Z = 2; and so on. See below for a detailed example.

If a GQ has an expected person sample size that is less than ten, then this method effectively gives the GQ a probability of selection that is proportional to its size; this probability is the expected person sample size divided by ten. If a GQ has an expected person sample size of ten or more, then it is in sample with certainty and is assigned one or more hits.

Second Phase of Sample Selection for Large GQs

Persons were selected within each GQ to which one or more hits were assigned in the first phase of selection. There were ten persons selected at a GQ for every hit assigned to the GQ. The persons were systematically sampled from a roster of persons residing at the GQ at the time of an interviewer's visit. The exception was if there were far fewer persons residing in a GQ than expected – in these situations, the number of persons to sample at the GQ would be reduced to reflect the GQ's actual population. In cases where fewer than ten persons resided in a GQ at the time of a visit, the interviewer would select all of the persons for sample.

Target Sampling Rate (Probability of Selection) for Large GQs

As for small GQs, the target sampling rate is the probability of selecting any given person in a GQ in Puerto Rico. This probability reflects both phases of sample selection. The target sampling rate for Puerto Rico in 2025 was 2.45 percent. Note that this is the same rate as for persons in small GQs.

As an example, suppose a GQ had an expected population of 250, and the target sampling rate was 2.29%, meaning any given person in a GQ had about a 1-in-43⅔ chance of being selected. This rate, combined with the GQs expected population of 250, means that the expected number of persons selected for sample in this GQ would be 5.725 (2.29% × 250). Since this is less than ten, this GQ would have either 0 or 1 hits assigned to it (Z =

0). The probability of it being assigned a hit would be the GQ's expected person sample size of 5.725 divided by 10, or 57.25%.

As a second example, suppose a GQ had an expected population of 1,000 and the target sampling rate was 4.30%; this means any given person in a GQ had about a 1-in-23.26 chance of being selected. This rate, combined with the GQ's expected population of 1,000, means that the expected number of persons selected for sample in the GQ would be 43 (4.30% × 1,000); this GQ would be assigned either four or five hits (Z = 4).

Group Quarters Sample Month Assignment

All small sample GQs and large sample GQ hits were assigned to a month in which to be interviewed (interview months) – these were the months in which interviewers would visit a GQ to select a person sample and conduct interviews. All small GQs, all large GQs that were assigned only one hit, all sampled military facilities, and all sampled correctional facilities (regardless of how many hits a military or correctional facility was assigned) were assigned to a single interview month. Federal prisons managed by the BOP were assigned to September; all of the others were randomly assigned to an interview month. Most small GQs and large GQ hits, that were not federal prisons managed by the BOP, could be assigned to any of the twelve months of the sample year. The exceptions were for college dormitories, whose hits were randomly assigned to non-summer months only, i.e., January through April and September through December; and for military ships, whose hits were randomly assigned to only the last ten months of the year, i.e., March through December.

Large sample GQs with multiple hits, but that were not in any of the categories above, had their hits randomly assigned to an interview month. Hits in each GQ were assigned to different interview months, e.g., a GQ with four hits might have had its hits assigned to January, April, June, and December. If either a college dormitory had more than eight assigned hits, a military ship had more than ten assigned hits, or any other large GQ had more than twelve assigned hits, then the randomization process of assigning hits to interview months would repeat itself for the excess hits. For example, if a GQ had fifteen hits assigned to it, and it was neither a college dormitory nor a military ship, then there would be three interview months in which two hits were assigned and nine interview months in which one hit was assigned.

Bureau of Prison Group Quarters

Since 2006, the Census Bureau and the BOP have had an inter-agency agreement, and the BOP has delivered the most recent list of federal prisons (managed by the BOP) as well as the most recent roster of federal inmates in those prisons to the Census Bureau. From 2006 to 2015, the BOP sampling identified the number of people to sample during the same procedure as the non-BOP sample. Beginning in 2016, a separate sampling frame was created using the facilities file delivered from the BOP and the BOP sample was selected from this frame.

The person sample for BOP GQs in sample is selected during the BOP sampling operation and is drawn directly from the list of federal inmates on the roster file delivered by the BOP.

2025 Sample Sizes for Housing Unit Addresses and Group Quarters

Counts of sample addresses and GQ persons can be found in two locations on the US Census Bureau website. On data.census.gov, base table B98001 provides housing unit address sample size counts for Puerto Rico, its municipios, and its places; base table B98002 provides group quarters person sample size counts for Puerto Rico. Sample size definitions and sample size counts for Puerto Rico are also available in the Sample Size and Data Quality Section of the ACS website, at https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/.

WEIGHTING METHODOLOGY

The estimates that appear in this product are obtained from a raking ratio estimation procedure that results in the assignment of two sets of weights: a weight to each sample person record and a weight to each sample housing unit record. Estimates of person characteristics are based on the person weight. Estimates of family, household, and housing unit characteristics are based on the housing unit weight. For any given tabulation area, a characteristic total is estimated by summing the weights assigned to the persons, households, families or housing units possessing the characteristic in the tabulation area. Each sample person or housing unit record is 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 has a value 40, all characteristics of that person or housing unit are tabulated with the weight of 40.

The weighting is conducted in two main operations: a group quarters person weighting operation which assigns weights to persons in group quarters, and a household person weighting operation which assigns weights both to housing units and to persons within housing units. The group quarters person weighting is conducted first and the household person weighting second. The household person weighting is dependent on the group quarters person weighting because estimates for total population, which include both group quarters and household population, are controlled to the Census Bureau's official 2025 total resident population estimates.

Group Quarters Person Weighting

Starting with the weighting for the 2011 1-year ACS estimates, the group quarters (GQ) person weighting changed in important ways from previous years' weighting. The GQ population sample was supplemented by a large-scale whole person imputation into not-in-sample GQ facilities. For the 2025 ACS GQ data, roughly 1.5 as many GQ persons were imputed as interviewed. The goal of the imputation methodology was two-fold.

1. The primary objective was to establish representation of municipio by major GQ type group in the tabulations for each combination that exists on the PRCS GQ sample frame.

The seven major GQ type groups are defined by the Population Estimates Program and are given in Table 3.

2. A secondary objective was to establish representation of tract by major GQ type group for each combination that exists on the PRCS GQ sample frame.

Table 3: Population Estimates Program Major GQ Type Groups

Major GQ Type Group	Definition	Institutional / Non-Institutional
1	Correctional Institutions	Institutional
2	Juvenile Detention Facilities	Institutional
3	Nursing Homes	Institutional
4	Other Long-Term Care Facilities	Institutional
5	College Dormitories	Non-Institutional
6	Military Facilities	Non-Institutional
7	Other Non-Institutional Facilities	Non-Institutional

The GQ sampling frame was modified to create an imputation frame from which all imputed GQs were selected from. The frame was updated with the actual populations and GQ type changes from ACS interviews, as well as any subsequent information gathered in other processes since the sampling frame was initially created. The change in populations for ACS GQ interviews was used to calculate a not-in-sample adjustment factor that was used to update the population for all GQs on the frame not selected for sample. This adjustment factor was calculated at the following level:

GQ Major Type × GQ Size Stratum

There were three size strata used for this process: GQs in sample with certainty, GQs with 16 or more persons, and GQs with less than 16 persons.

For all not-in-sample GQ facilities with an expected population of 16 or more persons (large facilities), we imputed a number of GQ persons equal to 2.5% of the expected population. For those GQ facilities with an expected population of fewer than 16 persons (small facilities), we selected a random sample of GQ facilities as needed to accomplish the two objectives given above. For those selected small GQ facilities, we imputed a number of GQ persons equal to 20% of the facility's expected population.

Interviewed GQ person records were then sampled at random to be donors for the imputed persons of the selected not-in-sample GQ facilities. An expanding search algorithm searched for donors within the same specific type of GQ facility and the same municipio. If that failed, the search included all GQ facilities of the same major GQ type group. If that still failed, the search expanded to a specific type within state, then a major GQ type group within state. If no donor was found at that point, then the search was stopped and the imputed person dropped for lack of a donor.

The weighting procedure made no distinction between sampled and imputed GQ person records. The initial weights of person records in the large GQ facilities equaled the observed or

expected population of the GQ facility divided by the number of person records. The initial weights of person records in small GQ facilities equaled the observed or expected population of the GQ facility divided by the number of records, multiplied by the inverse of the fraction of small GQ facilities represented in the weighting to the number on the frame of that tract by major GQ type group combination.

The population totals on the imputation frame are used to ensure that the sub-state distribution of GQ weighting preserves the distribution from the frame. This is accomplished through a series of three constraints:

1. *Tract Constraint (TRCON)* – This factor makes the total weight within each tract by major type group equal the total population from the imputation frame.

2. *County Constraint (CYCON)* – This factor makes the total weight within each municipio by major type group equal the total population from the imputation frame.

3. *State Constraint (STCON)* – This factor makes the total weight within each state by major type group equal the total population from the imputation frame.

As was done in previous years' weighting, we controlled the final weights to an independent set of GQ population estimates produced by the Population Estimates Program for each of the seven major GQ type groups.

Lastly, the final GQ person weight was rounded to an integer. Rounding was performed so that the sum of the rounded weights were within one person of the sum of the unrounded weights for any of the groups listed below:

Major GQ Type Group
Major GQ Type Group × Municipio

Housing Unit and Household Person Weighting

The housing unit and household person weighting use weighting areas built from collections of whole municipios. The 2010 Census data and 2007-2011 ACS 5-year data were used to group municipios of similar demographic and social characteristics. The characteristics considered in the formation included:

- Percent in poverty (the only characteristic using ACS 5-year data)
- Percent renting
- Percent areas with low density of housing units (a proxy for rural areas)
- Race/ethnicity, age, and sex distribution
- Distance between the centroids of the municipios
- Core-based Statistical Area status

Each weighting area was also required to meet a threshold of 400 expected person interviews in the 2011 PRCS. The process also tried to preserve as many municipios that met the

threshold to form their own weighting areas. In total, there are 45 weighting areas formed from the 78 municipios in Puerto Rico.

The estimation procedure used to assign the weights is then performed independently within each of the PRCS weighting areas.

Initial Housing Unit Weighting Factors

This process produced the following factors:

Base Weight (BW)

This initial weight is assigned to every housing unit as the inverse of its block's sampling rate.

CAPI Subsampling Factor (SSF)

The weights of the CAPI cases are adjusted to reflect the results of CAPI subsampling. This factor is assigned to each record as follows:

Selected in CAPI subsampling: SSF = 2.0
Not selected in CAPI subsampling: SSF = 0.0
Not a CAPI case: SSF = 1.0

Some sample addresses are unmailable. A two-thirds sample of these is sent directly to CAPI and for these cases SSF = 1.5.

Variation in Monthly Response by Mode (VMS)

This factor makes the total weight of the Mail, CAPI records to be tabulated in a month equal to the total base weight of all cases originally mailed for that month. For all cases, VMS is computed and assigned based on the following groups:

Weighting Area × Month

Noninterview Factor (NIF)

This factor adjusts the weight of all responding occupied housing units to account for nonresponding housing units. The factor is a ratio adjustment that is computed and assigned to occupied housings units based on the following groups:

Weighting Area × Building Type (single or multi-unit) × Tract

Vacant housing units are assigned a value of NIF = 1.0. Nonresponding housing units are now assigned a weight of 0.0.

Person Weighting Factors

Initially the person weight of each person in an occupied housing unit is the product of the weighting factors of their associated housing unit (BW × … × NIF). At this point, everyone in the household has the same weight. The person weighting is done in a series of three steps, which are repeated until a stopping criterion is met. These three steps form a raking ratio or raking process. These person weights are individually adjusted for each person as described below.

The three steps are as follows:

Municipio Controls Raking Factor (SUBEQRF)

This factor is applied to individuals based on their geography. It adjusts the person weights so that the weighted sample counts equal independent population estimates of total population for the municipio. For those municipios that are their own weighting area, this adjustment factor will be 1.0. Because of later adjustments to the person weights, total population is not assured of agreeing exactly with the official 2025 population estimates for municipios which are not their own weighting area.

Spouse Equalization/Householder Equalization Raking Factor (SPHHEQRF)

This factor is applied to individuals based on the combination of their status of being in a married-couple or unmarried-partner household and whether they are the householder. All persons are assigned to one of four groups:

1. Householder in a married-couple or unmarried-partner household
2. Spouse or unmarried partner in a married-couple or unmarried-partner household (non-householder)
3. Other householder
4. Other non-householder

The weights of persons in the first two groups are adjusted so that their sums are each equal to the total estimate of married-couple or unmarried-partner households using the housing unit weight (BW × … × NIF). At the same time, the weights of persons in the first and third groups are adjusted so that their sum is equal to the total estimate of occupied housing units using the housing unit weight (BW × … × NIF). The goal of this step is to produce more consistent estimates of spouses or unmarried partners and married-couple and unmarried-partner households while simultaneously producing more consistent estimates of householders, occupied housing units, and households.

Demographic Raking Factor (DEMORF)

This factor is applied to individuals based on their age and sex in Puerto Rico (note that there are 13 Age groupings). It adjusts the person weights so that the weighted sample counts equal the independent population estimates by age and sex at the weighting area

level. Because of collapsing of groups in applying this factor, only the total population is assured of agreeing with the official 2025 population estimates at the weighting area level.

These three steps are repeated several times until the estimates for Puerto Rico achieve their optimal consistency with regard to the spouse and householder equalization. The Person Post-Stratification Factor (PPSF) is then equal to the product (SUBEQRF × SPHHEQRF × DEMORF) from all of iterations of these three adjustments.

The unrounded person weight is then the equal to the product of PPSF times the housing unit weight (BW × … × NIF × PPSF).

Rounding

The final product of all person weights (BW × … × NIF × PPSF) is rounded to an integer.

Rounding is performed so that the sum of the rounded weights is within one person of the sum of the unrounded weights for any of the groups listed below:

Municipio
Municipio × Sex
Municipio × Sex × Age
Municipio × Sex × Age × Tract
Municipio × Sex × Age × Tract × Block

For example, the number of Males, Age 30 estimated for a municipio using the rounded weights is within one of the number produced using the unrounded weights.

Final Housing Unit Weighting Factors

This process produces the following factors:

Householder Factor (HHF)

This factor adjusts for differential response depending on the sex and age of the householder. The value of HHF for an occupied housing unit is the PPSF of the householder. Since there is no householder for vacant units, the value of HHF = 1.0 for all vacant units.

Rounding

The final product of all housing unit weights (BW × … × HHF) is rounded to an integer. For occupied units, the rounded housing unit weight is the same as the rounded person weight of the householder. This ensures that both the rounded and unrounded householder weights are equal to the occupied housing unit weight. The rounding for

vacant housing units is then performed so that total rounded weight is within one housing unit of the total unrounded weight for any of the groups listed below:

Municipio
Municipio × Tract
Municipio × Tract × Block

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.

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 that the controls provide to the PRCS estimates, including 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 CAPI phase. 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. 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. All PRCS published margins of error are based on a 90 percent confidence level.

Standard Error = Margin of Error / 1.645
Lower Confidence Bound = Estimate - Margin of Error
Upper Confidence Bound = Estimate + Margin of Error

Note that for 2005, 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, and for every year after 2006, 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, 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, a negative number of people does not make sense, so 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 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.[4]

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).

Approximating Standard Errors and Margins of Error

Previously, this document included formulas for approximating the standard error (SE) and margin of error (MOE) 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 Error".

[4] 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.

TESTING FOR SIGNIFICANT DIFFERENCES

Users may conduct a statistical test to see if the difference between a 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.[5]

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 in each municipio. The source of addresses for the PRCS is the Master Address File (MAF), which was created using the address list for Census 2010. An attempt is made to assign each MAF address to the appropriate geographic codes to 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, which could not be automatically coded.

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

In 2025, 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 Calle 1, Bayamon versus URB Hermosillo, 123 Calle 1, Bayamon, and can introduce overcoverage.

In the 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. 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, B98013, and B98014). Coverage rate definitions and coverage rates for total population for Puerto Rico are also available in the Sample Size and Data Quality Section of the ACS website, 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 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, a subsample of the mail nonrespondents was contacted for by personal visit to attempt an interview. Combined, these efforts resulted in a very high overall response rate for the ACS.

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 for Puerto Rico are also available in the Sample Size and Data Quality Section of the ACS website, 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 the 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 CAPI operation 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 for Puerto Rico are also available in the Sample Size and Data Quality Section of the ACS website, 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 ensure 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 allocated entries for a person or housing unit that was consistent with 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.