

Department of Commerce
United States Census Bureau
OMB Information Collection Request
2020 Census Post-Enumeration Survey Initial and Final Housing Unit Follow-up Operations
OMB Control Number 0607-XXXX

B. Collections of Information Employing Statistical Methods

1. Universe and Respondent Selection

The 2020 Post-Enumeration Survey (PES) sample design has been developed to support the various objectives of the program, which include estimating correct enumerations, erroneous enumerations, and omissions in addition to net coverage for the 2020 Census. The PES is designed to measure the coverage of housing units (HUs) and people, excluding group quarters and people residing in group quarters. The PES will be conducted in the U.S. and in Puerto Rico in selected PES sampled areas. Group quarters (such as college dormitories and correctional facilities) are out of scope because populations can change significantly between census enumeration and PES enumeration operations. Remote areas of Alaska are also out-of-scope for PES because the seasonal nature of addresses and the population throughout the year make it infeasible to accurately conduct the matching and follow-up operations necessary for dual-system estimation. For this reason, the Census Bureau's past post-enumeration surveys have never included remote Alaska.

The PES sample consists of two parts. The Population Sample, P sample, and the Enumeration Sample, E sample, have traditionally defined the samples for dual-system estimation. Both the P sample and the E sample measure the same HU and household population. However, the P sample operations are conducted independent of the census. The E sample consists of census enumerations in the same sample areas as the P sample. For net coverage error estimation, after matching with the census lists and reconciliation, the P sample provides information about the population missed in the census whereas the E sample provides information about erroneous census inclusions. The correct enumeration rate and match rate provide an estimate of the true population size using dual-system estimation. Comparing this number to the census counts yields an estimate of net coverage.

The PES is a multiphase operation designed to measure the net coverage and components of coverage for the household population and HUs in the 2020 Census. The PES sample design comprises a number of distinct processes from forming Basic Collection Units (BCUs), creating the sampling frame, selecting sample BCUs, to eventually selecting addresses for the P sample and E sample. The primary sampling unit for the first stage sample is the BCU, the smallest collection geography for the census. After a sample of BCUs is selected, an address list is created independent of the census in each sample BCU. The approximate PES listing workload is 10,100 BCUs for the United States and 400 for Puerto Rico. Overall, approximately 565,000 HUs

are listed (541,000 in the nation and 24,000 in Puerto Rico). Finally, after subsampling the PES HUs listed during Independent Listing, the final expected P-sample size is approximately 171,500 HUs for the nation and 8,000 for Puerto Rico. The national sample is distributed among the 50 states and the District of Columbia roughly proportional to population size, although there are slight increases in the sample for small states and for American Indian Reservations.

Table 1 summarizes, for the United States and Puerto Rico, the expected 2020 PES listing workloads and P-sample sizes. The E-sample size is expected to be about the same as that for the P sample.

Table 1: 2020 PES Universe and Sample Housing Unit Summary

| Geography | Expected Listing Sample Size | Expected P-sample Size |
|-------------|------------------------------|------------------------|
| U.S. | 541,000 | 171,500 |
| Puerto Rico | 24,000 | 8,000 |
| Total | 565,000 | 179,500 |

The PES sample has three phases of sampling. In the first phase of the PES sampling, BCUs in each state are classified into mutually exclusive and relatively homogeneous groups known as sampling strata. These strata are based on the BCU size and whether the BCU is located on an American Indian Reservation. The four major strata are:

- (1) BCUs with 0 to 2 HUs (small stratum).
- (2) BCUs with 3 to 57 HUs (medium stratum).
- (3) BCUs with 58 or more HUs (large stratum).
- (4) BCUs on American Indian Reservations with three or more HUs (American Indian Reservation stratum).

Using 2010 Census data, the medium and large strata are further split into renter and owner BCUs, resulting in up to six sampling strata being formed in each state and Puerto Rico. The definition of the large and medium groups in 2020 PES is different from the 2010 Census Coverage Measurement (CCM). For 2010 CCM, the medium group contained block clusters with 3 to 79 HUs and the large group contained block clusters with 80 or more HUs. Since the size of the BCUs is smaller than the block clusters on average, this change preserves the same proportion of frame HUs in the medium and large strata between the two designs (2010 and 2020). This yields a similar HU Independent Listing (IL) workload.

BCUs in the larger stratum are selected with higher probability of selection than BCUs in the medium stratum in this first phase because HUs in large BCUs are expected to be subsampled in the third phase. This allows more BCUs to be selected into sample. The non-owners stratum is selected at a higher rate (1.5 times higher) than the BCUs in the owner stratum. This is the

same differential sampling factor used in the 2010 CCM. The differential sampling factor of 1.5 provides a balance between improving the reliability of estimates for the smaller non-owner domain and improving the precision of other domain estimates and the total for a fixed sample size. Within each of the six sampling stratum for each state, the BCUs are sorted and a systematic sample is selected with equal probability.

The second phase of the 2020 PES selects a subsample of BCUs from the first-phase small sampling stratum using a similar method as in the 2010 CCM design. The Census Bureau uses a double-sampling technique by selecting a slightly larger sample of small BCUs in the first phase then selecting a subsample of small BCUs using an updated measure of size. This is done to reduce a BCU's influence on the estimates when more HUs are found than expected. Additionally, small BCU subsampling reduces costs, since conducting interview and follow-up operations in small BCUs is more expensive per HU than in medium or large BCUs.

Using HU counts from both the Independent Listing and the updated census address list, the small BCUs selected in the first phase are restratified by these counts within each state. A systematic sample of BCUs is selected within each stratum with equal probability. All BCUs from the small sampling stratum with ten or more HUs based on the updated counts are retained in sample. All BCUs from the small sampling stratum that are on American Indian Country are also retained in sample. (American Indian Country includes American Indian Reservations and associated trust lands, as well as the American Indian statistical areas.)

The first and second phases of the 2020 PES select the BCU sample. In the third phase of PES sampling, we select a subsample of HUs within large BCUs. For a BCU with 57 or fewer HUs observed, all the HUs are included in the sample. For a BCU with 58 or more HUs observed, a subsample of segments of contiguous HUs is selected to facilitate data collection in the field and to reduce the impact of intraclass correlation on the variance. This phase of sampling results in more similar overall selection probabilities for HUs because the large BCUs will have a higher probability of selection during the first phase.

The sampling frame for the P-sample HUs is created after the PES initial HU matching and follow-up operations. The intent of these HU operations is to identify matches between the independent HU list and an early census HU list. In addition to sending the P sample to the Person Interview, a sample of census units which were missed during the Independent Listing operation will be sent to the Person Interview. While not part of the P sample, these census units are likely to be in the E sample. The P-sample people result from the person interviewing in the P-sample HUs.

The sampling frame for the E-sample HUs consists of the HUs in PES sample areas from the list of 2020 Census enumerations that is available after the P sample is selected. While these two samples are selected at different points in time, we attempt to geographically overlap them to

the extent possible. The E-sample people are the census enumerations in the E-sample HUs with enough information collected.

2. Procedures for Collecting Information

The PES Initial Housing Unit Follow-up questionnaire, D-1303, and the Final Housing Unit Follow-up questionnaires, D-1340, will be used by listers to resolve discrepancies from matching the PES Independent Listing and the census address lists. These questionnaires are translated into Spanish for use in Puerto Rico. The Spanish questionnaires are D-1303 (PR) for Initial Housing Unit Follow-up and D-1340(PR) for Final Housing Unit Follow-up. Listers will contact a member of each HU (or proxy, as a last resort) being followed up to ensure the accuracy of the address lists. Listers will be given the Housing Unit Reference list to aid in following up HUs. The lister will give each respondent in the 50 states and the District of Columbia a copy of the Form D-31(PES-HUFU), U.S. Confidentiality Notice; in Puerto Rico, the lister will give each respondent a copy of the Form D-31(PES-HUFU) PR, Puerto Rico Confidentiality Notice. The Confidentiality Notice explains that any information given to the Census Bureau will be held in strict confidence. It also informs each respondent that participation is mandatory.

For the Initial and Final Housing Unit Follow-up Quality Control (QC), approximately 15 percent of all HUs followed up will be identified for a QC check. The QC listers will locate the HUs identified for the QC check and compare the information collected by the production lister to what they see on the ground. A single attempt will be made to contact a respondent for the QC check. If no one is available, the QC check will be completed by observation as a last resort. BCUs not passing the QC check will be 100 percent verified to ensure the data quality of the Initial and Final Housing Unit Follow-up. QC listers will use the Quality Control Form D-1380 for Initial Housing Unit Follow-up and Quality Control Form D-1325 for Final Housing Follow-up for BCUs in the United States. QC listers will use the Quality Control Form D-1380(PR) for Initial Housing Unit Follow-up and Quality Control Form D-1325(PR) for Final Housing Unit Follow-up for BCUs in Puerto Rico.

3. Methods to Maximize Response

The Initial and Final Housing Unit Follow-up contains the minimum number of questions necessary to obtain the data required for the 2020 PES. The lister will make up to three attempts for each operation to obtain an interview. The lister will explain the reason the Census Bureau is conducting the operation, and respondents will be informed of their legal responsibility to answer the questions. In addition, respondents will be assured that their answers are confidential.

4. Testing of Procedures or Methods

Nothing will be tested during the Initial and Final Housing Unit Follow-up operations.

5. Contacts for Statistical Aspects and Data Collection

Andreana Able
Coverage Measurement Operations Branch
Decennial Statistical Studies Division
U.S. Census Bureau
301-763-0153

Definition of Terms

Components of Census Coverage – The components of census coverage include correct enumerations, erroneous enumerations, whole-person imputations, and omissions. Correct enumerations are people or HUs that were correctly enumerated in the census. Erroneous enumerations are people or HUs that were enumerated in the census but should not have been. Examples of erroneous enumerations are duplicates, nonexistent HUs or people, and people or HUs that were enumerated in the wrong place. Omissions are people and HUs that were not correctly enumerated in the census but should have been. Lastly, whole-person imputations are census records for which all of the demographic characteristics were imputed. Many of these imputations represent people in HUs where we knew the household count but did not obtain sufficient information about the people residing at the HU.

Net Coverage– Reflects the difference between the true population (which is estimated by the Post-Enumeration Survey) and the census count. If the census count was less than the actual number of people or HUs in the population, then we say there was an undercount. If the census count was more than the actual number of people or HUs in the population, then we say there was an overcount.

For more information about the Post-Enumeration Survey Program, please visit the following page of the Census Bureau's website:

https://www.census.gov/coverage_measurement/post-enumeration_surveys/