2017 American Housing Survey Sample Design and Weighting

The 2017 American Housing Survey (AHS) is a longitudinal survey and has two parts: a national and a metropolitan sample design that we refer to as AHS-N and AHS-MS, respectively. The estimates from both AHS-N and AHS-MS will be derived from a sample of housing units interviewed between May and September 2017. In the AHS-N sample design, sample housing units will be interviewed every two years until a new sample is selected. For the sample of AHS-MS, sample housing units follow a longer interviewing cycle. The U.S. Census Bureau updates the sample by adding newly constructed housing units and units discovered through coverage improvement efforts.

For the 2017 AHS, approximately 89,824 sample housing units will be interviewed for the AHS-N sample and approximately 41,887 sample housing units will be interviewed for the AHS-MS.

AHS-N SAMPLE DESIGN

The AHS-N employs a two-stage sample design. First, the U.S. was divided into areas consisting of counties or groups of counties and independent cities known as primary sampling units (PSUs). The sample for AHS includes just over 500 PSUs. These PSUs cover approximately 900 counties and independent cities with coverage in all 50 states and the District of Columbia.

The counties of the top 85 Core Base Statistical Areas (CBSAs), according to the 2010 projected OMB definitions, were included in the first-stage sample design with certainty. These PSUs are known as a self-representing PSUs because they represent themselves and no other PSUs.

The Census Bureau grouped the remaining PSUs into first-stage strata and selected one PSU per strata, proportional to the number of housing units in the PSU, to represent all PSUs in the strata. The selected PSUs are referred to as nonself-representing PSUs because they represent themselves and all of the other PSUs in the same strata.

Allocation of sample housing units. The current national sample size of 34,392 HUs was originally allocated to the sample PSUs proportional to the first-stage strata HU counts from the Census 2010. For each of the top 15 CBSAs in the U.S., we replaced the allocated national sample with approximately 3,000 sample HUs. The long-range plan is to have the top 15 CBSAs as a permanent feature of the national sample design. It will mean that the national sample design will support reasonable estimates for national estimates and for each of the top 15 CBSAs individually.

The AHS-N sample includes an oversample of approximately 5,202 HUs in HUD programs and a bridge sample of approximately 6,000 HUs that will be used to measure any possible break-in-series between the old 1985-2013 and new 2015-2017 sample design. Additionally for 2017, the AHS-N includes an oversample of approximately 4,945 HUs in high poverty tracts.

Selection of sample housing units. The AHS sample within each PSU consists of the following types of units in the sampled PSUs:

☐ All housing units on the Census Bureau Master Address File as of 2014
☐ Housing units in a sample of Coverage Improvement blocks

The second-stage sample design is a systematic random sample from an ordered list. A combination of geographic variables, tenure, and the identification of housing units participating in HUD programs was used to sort the list of HUs prior to selecting the sample.

AHS-N ESTIMATION

Each housing unit in the AHS sample represents itself and over 2,000 other units. The exact number it represents is its "weight." The weight was calculated in six steps and the purpose of these steps is to minimize both sampling errors and errors from incomplete data. The description that follows is consistent with the weighting methodology for the prior cohort interviewed from 1985 to 2013. For 2015, we completed several research projects to improve the weighting; we made improvements in defining geographic cells, research is continuing to improve weighting methodology.

The final weight is the product of the weights and factors described below and was calculated in the order listed below. The calculation of each factor used all of the previously calculated factors.

1. *Base weight.* Each unit is assigned a weight to reflect its probability of selection. The basic weight accounts for the sample selection of PSUs in the first stage and housing units in the second stage. The base weight also accounts for any reductions or expansions of the sample.

2. *First-stage ratio adjustment*. The weights are adjusted for differences between the number of housing units for the complete set of nonself-representing (NSR) PSUs and the estimated number of housing units nonself-representing (NSR) PSUs where only the sample PSUs contribute to the estimate. This can be represented as:

$$\frac{\text{Number of housing units in all PSUs}}{\text{of the nonself-representing PSUs}} \bigg/ \frac{\text{Estimated number housing units in all PSUs}}{\text{of the nonself-representing PSUs}}$$

This adjustment is done separately for groups that have yet to be determined.

3. *Noninterview adjustment*. An adjustment will be made for noninterviews of occupied units. The calculations for this adjustment will not include units the Census Bureau could not locate. The earlier weight will be multiplied by the following factor:

$$\frac{\text{Interviewed units + Units not interviewed}}{\text{Interviewed units}}$$

It is assumed the units missed are similar in some ways to the units interviewed for AHS.

This adjustment is done separately within cells that are related to the propensity to respond and the variable of interest. In 2015 research was conducted and completed to determine the best set of variables to use in the noninterview adjustment.

4.   *Adjustment to Independent Housing Unit totals.*   Independent totals of housing units will be used to adjust for known deficiencies in sampling new construction by multiplying the earlier weight by the following factor:

$$\frac{\text{Independent estimate}}{\text{AHS sample estimate}}$$

This adjustment will be completed separately for groups that have yet to be determined by planned research.

The known totals will be based on the Census Bureau's Survey of Construction and Survey of Manufactured Home Placements. Note that final AHS figures for the categories above are not really based on the AHS sample findings, but on the independent sources.

5.   *Adjustment to Independent Population Totals.* Comparability of population estimates among the surveys will be ensured by multiplying the earlier weight by the following factor:

$$\frac{\text{Independent estimate}}{\text{AHS sample estimate}}$$

This adjustment will be completed separately for groups that have yet to be determined by planned research.

Repetitions. The new construction and demographic adjustments will be repeated to help match both sets of independent estimates simultaneously.

Small groups. In each step of weighting, many items were cross-classified; so some groups may have few cases. When a group is too small (less than 30 cases) or the adjustment factor is too extreme (greater than 2.0 for the noninterview adjustment or outside a range of 0.5 to 2.0 for the demographic adjustment) the Census Bureau combined the group with one or more other groups that are similar in most respects.

AHS-MS SAMPLE DESIGN

The 2017 AHS-MS sample provides information on 14 metropolitan areas interviewed as part of the AHS. The metropolitan areas are all defined by the 2013 Office of Management and Budget (OMB) projected definitions for CBSAs. The supplemental sample of 14 metropolitan areas includes about 3,000 housing units each for a total of 41,887 units.

AHS-MS ESTIMATION

Sample weights for the AHS-MS will be calculated in the same way as the national weights, but without a first-stage ratio adjustment and without an adjustment to independent HU totals.

AHS VARIANCE ESTIMATION

Replicate weights will be calculated for both the AHS-N and AHS-MS and will be used to calculate both the estimates provided in publications and be made available to the public. The AHS-N replicate weights will use Balanced Repeated Replication estimate the variation due to the selection of PSUs in the first-stage and Successive Difference Replication (SDR) to estimate the variation due to the selection of HUs within each of the self-representing PSUs. The AHS-MS replicate weights will use SDR to estimate the variation due to the selection of HUs within each CBSA.