

Appendix B.

Sample Design and Weighting

SAMPLE SIZE

The 2001 national data are from a sample of housing units interviewed between August and November 2001. The same basic sample of housing units is interviewed every 2 years until a new sample is selected. The U.S. Census Bureau updated the sample by adding newly constructed housing units and units discovered through coverage improvement efforts every enumeration. For the 2001 American Housing Survey-National (AHS-N), approximately 55,700 sample housing units were selected for interview. About 2,100 of these units were found to be ineligible because the unit no longer existed or because the units did not meet the AHS-N definition of a housing unit.

Of the 53,600 eligible sample units, about 5,400 were classified (both occupied and vacant housing units), as “Type A” noninterviews because (a) no one was at home after repeated visits, (b) the respondent refused to be interviewed, or (c) the interviewer was unable to find the unit. This classification produced a 90-percent overall response rate.

SAMPLE SELECTION

The Census Bureau has interviewed the current sample of housing units since 1985. First, the United States was divided into areas made up of counties or groups of counties and independent cities known as primary sampling units (PSUs). A sample of these PSUs was selected. Then a sample of housing units was selected within these PSUs.

Selection of sample areas. The sample for AHS is spread over 394 PSUs. These PSUs cover 878 counties and independent cities with coverage in all 50 states and the District of Columbia.

If there were over 100,000 housing units in a PSU at the time of selection, the PSU is known as a self-representing PSU, because it was removed from the probability sampling operation. It was in sample with certainty. The sample from the PSU represents only that PSU. There are 170 self-representing PSUs.

The Census Bureau grouped the remaining PSUs and selected one PSU per group, proportional to the number of housing units in the PSU, to represent all PSUs in the group. These selected PSUs are referred to as nonself-representing PSUs. The sample nonself-representing PSUs for AHS are a subsample of the Current Population Survey’s (CPS) sample areas based on the 1980 census.

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

- Housing units selected from the 1980 census
- New construction in areas requiring building permits
- Housing units missed in the 1980 census
- Other housing units added since the 1980 census

Housing units selected from the 1980 census. The Census Bureau picked a systematic sample so every unit had a 1 in 2,148 chance of being included in the AHS.

In areas where addresses are complete (at least 96 percent of units having a house number and street name) and permits are required for new construction, housing units receiving 1980 census long-form questionnaires were sorted by the following items:

- PSU
- Central city, urbanized area, urban outside urbanized area, rural
- Owner, renter, vacant for rent, vacant for sale, other types of vacants
- Number of rooms
- Value of home or gross rent
- Mobile home or not a mobile home

In areas where addresses are not complete or permits are not required for new construction, land areas were sorted using a formula incorporating the following items:

- PSU
- Central city, urbanized area, urban outside urbanized area, rural
- Median value of home
- Number of children under 6 years old
- Number of elderly people
- Number of owner-occupied homes
- Number of mobile homes
- Number of homes lacking some plumbing
- Number of owner-occupied homes whose value is below \$45,000
- Number of renter-occupied homes with rent below \$200
- Number of Black and Hispanic people
- Number of 1-room homes

New construction in areas requiring building permits. In areas that require building permits for new construction, the Census Bureau selected a sample of permits. These permits do not cover mobile homes or conversion of older buildings to residential use.

Housing units missed in the 1980 census. The Census Bureau conducted a special study that identified units at addresses missed or inadequately defined in the 1980 census. A sample of these identified units was selected.

Housing units added since the 1980 census. If extra units are added in buildings or mobile home parks where AHS already has sample units, a sample of these extra units was selected. To find when whole buildings are built (in addition to building permits, mentioned above) or are converted from nonresidential to residential use, the Census Bureau listed all residential buildings in a sample of areas around the country, found any additional buildings, and selected a sample of their units.

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 five steps. The purpose of these steps is to minimize both sampling errors and errors from incomplete data. The result of the steps is also to force consistency with some major categories of data in other Census Bureau surveys. Therefore, figures on these categories do not actually depend on the AHS sample, but on the other surveys.

In 2001, the weighting procedures were changed for this publication by switching from 1980 census-based geography to 1990 census-based geography in the various steps of the weighting. Refer to Table H in Appendix C that compares the housing unit totals for 1980-based weights with 1990-based weights. Differences in the 1980 and 1990 geography-based weights are reflected in steps (2.), (3.), and (5.). The 1980 geography-based weights will continue to be included on the microdata file.

1. **Basic weight.** The Census Bureau assigned each unit a weight to reflect its probability of selection. With rare exceptions, this weight is 2,148.
2. **Noninterview adjustment.** An adjustment was made for refusals and occupied units where no one was home. The calculations for this adjustment do not include units the Census Bureau could not locate. The earlier weight was multiplied by the following factor:

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

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

This adjustment is done separately for groups defined by cross-classifying the following data items if prior year data for the indicated items is available:

- Four census regions
- 1990 Central city, suburb, or nonmetropolitan
- 1990 Urban or rural
- Mobile home or not a mobile home
- Owner/for sale or renter/for rent
- Number of units in structure*
- Number of rooms*
- Occupied, vacant year round, or seasonal/migratory vacant*

(*If known from a previous survey; otherwise, the Census Bureau substituted whether or not units were drawn from building permits for these items.)

For seasonal/migratory vacants and year-round vacants other than those for rent or for sale, units were cross-classified only by census region and 1990 central city/suburb/ nonmetropolitan.

3. **PSU adjustment.** The Census Bureau adjusted for differences that existed in 1990 between the number of 1990 census housing units estimated from the AHS sample of nonself-representing (NSR) PSUs and the 1990 census counts outside the self-representing PSUs. The earlier weight was multiplied by the following factor:

$$\frac{\text{1990 census housing units in all areas that could have been chosen as nonself-representing PSUs}}{\text{1990 census housing units estimated from the AHS sample of nonself-representing PSUs}}$$

This adjustment is done separately for groups defined by cross-classifying:

- Owner, renter, or vacant (four census regions)
- 1990 Central city, suburb, or nonmetropolitan
- 1990 Urban or rural
- Hispanic or non-Hispanic householder (only in South and West regions)
- Black or non-Black householder (only in South region)

4. **New construction adjustment.** The Census Bureau adjusted 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 is done separately for groups defined by cross-classifying:

- Four census regions
- Mobile home or not a mobile home
- Number of units in structure
- Year built (pre-1980 and 5-year categories after 1980 as shown in the publication)

Independent estimates are based on the Census Bureau's Survey of Construction and Manufactured Homes Survey. Note that final AHS figures for the categories above are not really based on the AHS sample findings, but on the independent sources.

5. **Demographic adjustment.** Comparability among the surveys was ensured by multiplying the earlier weight by the following factor:

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

This adjustment is done in two steps for occupied units. First, the factors were computed and applied for the Hispanic or non-Hispanic groups defined by cross-classifying:

- Four census regions
- Owner or renter
- Hispanic or non-Hispanic householder
- Husband-wife, other male householder, or other female householder
- Age of householder

Next, the demographic adjustment is repeated with the same cells, except classified by the Black or non-Black groups, rather than the Hispanic or non-Hispanic groups.

Vacant for sale, vacant for rent, other year-round vacant and seasonal/migratory vacant units were cross-classified only by the four census regions and 1990 central city, suburb, or nonmetropolitan.

The percentage of occupied and vacant units was based on the AHS itself. The distribution within occupied and vacant units is from the Census Bureau's Current Population Survey for occupied units, and from the Housing Vacancy Survey for vacant units. The grand total number of all housing units in the United States is based on the 1990 census adjusted to account for new and lost units. Note that final AHS figures for the categories above are not really based on the AHS sample findings, but on the independent sources.

Repetitions. The new construction and demographic adjustments were repeated to help match both sets of independent estimates simultaneously. These adjustments were repeated until every cell's factor is between 0.98 and 1.02 or the change in each factor from one repetition to the next is fewer than 0.015.

Small cells. In each step of weighting, many items were cross-classified; so some cells may have few cases. When a cell is too small (fewer than 30 cases for the noninterview adjustment or fewer than 50 cases for the demographic adjustment) or the adjustment factor is too extreme (greater than 1.5 for the noninterview adjustment or outside a range of 0.5 to 2.0 for the demographic adjustment) the Census Bureau combined the cell with one or more other cells that are similar in most respects. Cells for the PSU adjustment or the new construction adjustment were not combined.