

Appendix B.

Sample Design and Weighting

SAMPLE SIZE

The 2003 national data are from a sample of housing units interviewed between late-May and mid-September 2003. 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. For the 2003 American Housing Survey-National (AHS-N), approximately 63,300 sample housing units were selected for interview. About 2,250 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 61,050 eligible sample units, about 5,650 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 an unweighted overall response rate of 91 percent. The weighted overall response rate was 92 percent.

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
- Manufactured/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 manufactured/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 manufactured/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 manufactured/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.

Supplemental Metropolitan Sample

In 2003, the Census Bureau reinstated units in six metropolitan areas. The data for these areas are based on AHS-National sample because the AHS-MS sample in these six areas was dropped to reduce costs. These metropolitan areas are:

- Chicago, IL
- Detroit, MI
- New York-Nassau-Suffolk-Orange, NY
- Northern New Jersey
- Los Angeles-Long Beach, CA
- Philadelphia, PA-NJ

Most of these metropolitan areas are consistent with the 1993 Office of Management and Budget (OMB) definitions of the metropolitan statistical area (MSA), consolidated metropolitan statistical area (CMSA), or primary metropolitan statistical area (PMSA) with the following exceptions:

- Chicago, IL, does not include DeKalb County from the 1993 OMB definition for the Chicago, IL PMSA.
- Detroit, MI, includes Livingston County in addition to the 1993 OMB definition of the Detroit, MI PMSA.
- New York-Nassau-Suffolk-Orange, NY, does not include Pike county, PA, from the 1993 OMB definition for the New York-Nassau-Suffolk-Orange, NY-PA PMSAs.
- Northern New Jersey does not include Warren County, PA, from the 1993 OMB definition for Newark NJ PMSA.
- Philadelphia, PA-NJ, does not include Salem County, NJ, from the 1993 OMB definition of the Philadelphia, PA-NJ-PMSA.

In order to provide more reliable sample estimates for the six metropolitan areas, the Census Bureau used sample cases from the basic sample, along with an extra sample that had been selected for possible sample supplementation. The extra sample is referred to as the supplemental sample. In 1987 and 1991, some of this sample was used for rural supplementation. However, most of the supplemental sample was interviewed for the first time in 1995. Table C provides the size of the supplemental sample added in each of the six metropolitan areas.

Table C. 2003 Supplemental Sample Size for Each of the Six AHS-National-Based Metropolitan Areas

Metropolitan area	Supplemental sample size
Chicago, IL	1,818
Detroit, MI	1,115
Los Angeles-Long Beach, CA	2,041
New York-Nassau-Suffolk-Orange, NY	137
Northern New Jersey.....	112
Philadelphia, PA-NJ.....	1,209

In all of the metropolitan areas except Northern New Jersey and New York, the supplemental sample units included units selected from the 1980 census and any new construction since the 1980 census. In Northern New Jersey and New York very little supplemental sample was needed. Only 1980 census renters in urban areas in a few counties were added to the sample.

The Census Bureau used all of the 2003 AHS-National basic and supplemental sample for the following areas:

- Chicago
- Detroit
- Northern New Jersey
- Philadelphia

In Los Angeles, all of the AHS-National sample from the urbanized areas of this MS, and only the supplemental sample from urban areas outside urbanized areas and from rural areas was used. This was done for confidentiality reasons.

In New York, the Census Bureau used different samples for the user file and the publication. For the publication, the AHS-National basic and supplemental sample in all areas was used. For the user file, the AHS-National basic and supplemental sample, after excluding the urbanized area cases in Orange County, was used. This was done for confidentiality reasons.

Interview activity. Table D summarizes the interview activity for the six AHS-National metropolitan areas. The table provides the response rate, number of eligible units (comprised of completed interviews and noninterviews), and the number of units visited but ineligible for interview.

Table D. Interview Activity for Each of the Six 2003 AHS-National-Based Metropolitan Areas

Metropolitan area	Unweighted response rate ¹ (percent)	Weighted response rate ² (percent)	Eligible units			Ineligible units ⁴
			Total	Interviewed	Not interviewed ³	
2003 AHS-National total for the six listed MSAs	88	90	14,471	12,803	1,668	485
Chicago, IL	88	90	3,227	2,854	373	114
Detroit, MI	88	89	1,957	1,725	232	44
Los Angeles-Long Beach, CA	90	91	3,489	3,142	347	83
New York-Nassau-Suffolk-Orange, NY	90	91	2,369	2,143	226	112
Northern New Jersey	89	90	1,326	1,174	152	46
Philadelphia, PA-NJ	84	86	2,103	1,765	338	86

¹The unweighted response rate is computed by dividing the unweighted number of interviews by the unweighted total number of cases eligible for interview and multiplying by 100.
²The weighted response rate is computed by dividing the weighted number of interviews by the weighted total number of cases eligible for interview and multiplying by 100.
³Sample units were visited, but occupants were not at home after repeated visits or were unavailable for some other reasons.
⁴Sample units were visited but did not provide information relevant to the housing inventory. This category includes sample units that were found not to be in the sampling frame.

Estimation for AHS-National

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 2003, the weighting procedures were changed for this publication by switching independent estimates from 1990 census-based to 2000 census-based in the various steps of the weighting. In addition, the Census Bureau switched the definition of race from single race-alone categories to multi-race categories grouped together with race-alone categories. Refer to topic Race in Appendix C for more details on race. This change affects steps (3) and (5).

- Basic weight.** The Census Bureau assigned each unit a weight to reflect its probability of selection. With rare exceptions, this weight is 2,148.
- 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
- Manufactured/mobile home or not a manufactured/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.

- 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:

- Four census regions
- Owner, renter, or vacant
- 1990 Central city, suburb, or nonmetropolitan
- 1990 Urban or rural
- Hispanic or non-Hispanic householder (only in South and West regions)
- Black alone or in combination with other races, or non-Black householder (only in South region)

- 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 alone or in combination with other races, 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 2000 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.

Estimation for AHS-National metropolitan areas. The sample housing units were weighted according to a one-stage ratio estimation procedure.

In 2003, the weighting procedures were changed for this publication by switching from 1980 census-based geography to 1990 census-based geography, affecting only step (2.).

In addition, the independent estimates used in the weighting switched from 1990 census-based to 2000 census-based in only step (3.) of the weighting.

1. **Basic weight.** The basic weight is the inverse of the probability of selection. The basic weight varies for each metropolitan area depending on the size of the supplemental sample.
2. **Type A noninterview adjustment.** Before implementation of the ratio estimation procedure, the basic weight for each interviewed sample housing unit was adjusted to account for Type A noninterviews. Type A noninterviews are sample units for which
 - a. Occupants were not home or
 - b. Occupants refused to be interviewed or
 - c. Occupants were unavailable for some other reason

When prior year AHS-National or 1980 census data were available, the Census Bureau used this information to determine the noninterview adjustment cell. The cells include the following characteristics:

- Tenure
- 1990 geography
- Units in structure
- Number of rooms
- Value

When previous data are not available, the Census Bureau computed adjustment factors using geography and tenure.

Within a given cell, the Type A noninterview adjustment factor was equal to the following ratio:

Weighted count of interviewed housing units	+	Weighted count of Type A noninter- viewed housing units
---	---	---

Weighted count of interviewed housing units

Independent estimate of the total housing
inventory for the corresponding geographic
subdivision of the metropolitan area

Sample estimate of the total housing
inventory for the corresponding geographic
subdivision of the metropolitan area

3. Independent total housing unit ratio estimation.

For the ratio estimation procedure described below, each metropolitan area was subdivided into geographic areas consisting of individual counties or a combination of counties.

The ratio estimation procedure reduced the sampling error for most statistics below what would have been obtained by simply weighting the results of the sample by the inverse of the probability of selection. Since the housing population of the sample differed somewhat by chance from the metropolitan area as a whole, one can expect that the sample housing population, or different portions of it, is brought into agreement with known good estimates of the metropolitan area housing population.

The Census Bureau applied the following ratio estimation procedure in all the areas:

The numerator of this ratio was determined by making adjustments to the 2000 census data to account for residential new construction as well as losses to the housing inventory. These estimates were generated at the county level and combined to form geographic subdivisions. For a more detailed description of the determination of these numbers, refer to a description of a similar process at the state level in the *Current Population Report*, Series P25-1123. The denominator was obtained using the existing weight of AHS sample units (that is, the product of the basic weight and the weighting factors).

The computed ratio estimation factor was then applied to all appropriate housing units in the corresponding geographic area of each metropolitan area, and the resulting product was used as the final weight for tabulation purposes.