American Housing Survey

2017 AHS Metropolitan Sample:

Sample Design, Weighting, and Error Estimation

Updated: December 21, 2018

U.S. Census Bureau, Department of Commerce
Department of Housing and Urban Development

Contents

1. Overview

The purpose of this document is to describe the sample design, weighting, and error estimation for the 2017 American Housing Survey Metropolitan Sample (AHS-MS).

For 2015, the U.S. Department of Housing and Urban Development (HUD) and the U.S. Census Bureau selected an entirely new sample for the AHS. In 2017, AHS interviewers visited the same housing units selected in the 2015 sample. Additionally, AHS staff selected newly constructed housing units from a sample using a design similar to the 2015 sample. The AHS sample is composed of an Integrated National Sample and Independent Metropolitan Area Samples (sometimes referred to as "Metro"). The Independent Metro Samples include—

- Representative samples of each of the 15 largest Metropolitan areas, which are also included in the Integrated National Sample (hereafter referred to as the "Top 15").
- Representative samples of 10 extra Metropolitan areas, which are not included anywhere in the Integrated National Sample (hereafter referred to as the "Next 10").

HUD and the Census Bureau intend to survey the Top 15 Metro samples once every 2 years. As such, these are longitudinal panels with a 2-year survey cycles.

For 2017, the 10 selected Metropolitan areas represent one-half of what HUD and the Census Bureau refer to as the "Next 20" group of Metropolitan areas (the first half was included in the 2015 AHS). The Next 20 group of Metropolitan areas is a subset of Metropolitan areas ranging from the 16th to 50th largest, by population.[1] HUD and the Census Bureau intend to survey each member of the Next 20 group of Metropolitan areas once every 4 years. As such, the Next 20 group of Independent Metropolitan Area Samples is a longitudinal panel with a 4-year survey cycle.

The Independent Metro Samples were interviewed between June 26 and October 30, 2017.

2. Metropolitan Sample Design

2.1 Eligible Universe

The universe of interest for the AHS consists of the residential housing units in each Metro area that exist at the time the survey is conducted. The universe includes both occupied and vacant units, but excludes group quarters, businesses, hotels, and motels. Geographically, the survey covers the 25 Core-Based Statistical Areas (CBSAs), as defined by the Office of Management and Budget (OMB), corresponding to each chosen Metropolitan area. For the purposes of this document, "Metro area" and "CBSA" are synonymous, though for consistency and familiarity

---

[1] For more information about how the Next 20 group of Metropolitan areas was selected, see "Metropolitan Area Selection Strategy: 2015 and Beyond"

the term “Metro area” will be used predominately. CBSAs do not always fall entirely within states, as they are defined to be groups of counties with strong commuting ties to a city center.

2.2 Sample Size

For the 2017 AHS Independent Metropolitan Area Samples, approximately 3,000 housing units were originally selected for interview for each Metro area.[2] Table 2.1 below details the exact sample sizes for each in the Top 15 and Next 10:

[2] In the 2014 document “Sample Sizes Determination and Decisions for the 2015 American Housing Survey and Beyond”, Bucholtz and Ash discussed how the sample sizes were determined.

Table 2.1. Sample Size in the 2017 American Housing Survey-Independent Metropolitan Sample Areas

Sample Status	Metropolitan Area (CBSA*)	Total Sample Size (includes HUD oversample for Top 15)
Top 15	Atlanta, GA	3,091
	Boston, MA	3,061
	Chicago, IL	3,097
	Dallas, TX	3,217
	Detroit, MI	3,029
	Houston, TX	3,189
	Los Angeles, CA	3,219
	Miami, FL	3,107
	New York, NY	3,318
	Philadelphia, PA	3,073
	Phoenix, AZ	3,094
	Riverside, CA	3,041
	San Francisco, CA	3,127
	Seattle, WA	3,143
	Washington, DC	3,147
Next 10	Baltimore, MD	2,945
	Birmingham, AL	2,973
	Las Vegas, NV	3,020
	Minneapolis, MN	2,985
	Oklahoma City, OK	3,024
	Richmond, VA	3,006
	Rochester, NY	2,940
	San Antonio, TX	3,111
	San Jose, CA	3,001
	Tampa, FL	2,976

*CBSA- Core-Based Statistical Areas
Source: U.S. Census Bureau, 2017 American Housing Survey.

Several units across each Metro area included for interview were found to be ineligible because the units either no longer existed or did not meet the AHS definition of a housing unit. Of the eligible sample units (both occupied and vacant housing units), some were classified as noninterviews because (1) no one was at home after repeated visits, (2) the respondent refused to be interviewed, or (3) the interviewer was unable to find the unit. This classification produced both unweighted and weighted overall response rates. All of these measures for each sampled Metro area are detailed in Table 2.2 below:

Table 2.2. Interview Activity for the 2017 American Housing Survey-Independent Metropolitan Sample Areas

Sample Status	Metropolitan area (CBSA*)	Unweighted response rate (percent)	Weighted response rate (percent)	Eligible units			Ineligible
				Total	Interviewed	Not Interviewed	
Top 15	Atlanta, GA	76.0	75.9	3,012	2,288	724	79
	Boston, MA	70.9	70.9	2,996	2,124	872	65
	Chicago, IL	74.8	74.8	3,030	2,265	765	67
	Dallas, TX	89.3	89.3	3,164	2,824	340	53
	Detroit, MI	80.1	79.9	2,967	2,376	591	62
	Houston, TX	85.0	85.0	3,028	2,574	454	161
	Los Angeles, CA	85.8	85.7	3,185	2,734	451	34
	Miami, FL	85.3	85.5	3,044	2,598	446	63
	New York, NY	71.6	70.7	3,240	2,320	920	78
	Philadelphia, PA	73.9	74.0	3,006	2,221	785	67
	Phoenix, AZ	79.7	79.6	3,028	2,412	616	66
	Riverside, CA	88.6	88.6	2,990	2,649	341	51
	San Francisco, CA	73.9	73.8	3,095	2,286	809	32
	Seattle, WA	79.6	79.6	3,090	2,461	629	53
	Washington, DC	84.3	84.3	3,115	2,626	489	32
Next 10	Baltimore, MD	76.0	76.0	2,824	2,146	678	121
	Birmingham, AL	79.6	79.6	2,763	2,200	563	210
	Las Vegas, NV	82.0	82.0	2,919	2,395	524	101
	Minneapolis, MN	79.6	79.6	2,922	2,326	596	63
	Oklahoma City, OK	86.6	86.6	2,874	2,490	384	150
	Richmond, VA	77.0	77.0	2,916	2,244	672	90
	Rochester, NY	78.1	78.1	2,834	2,212	622	106
	San Antonio, TX	84.5	84.5	2,958	2,499	459	153
	San Jose, CA	79.5	79.5	2,949	2,345	604	52
	Tampa, FL	78.2	78.2	2,849	2,227	622	127

*CBSA- Core-Based Statistical Areas
Source: U.S. Census Bureau, 2017 American Housing Survey.

Every sample unit of the 2017 Independent Metro sample was asked a core set of questions. Housing units within these 25 Metro areas were also randomly split into two samples, and each of these samples were asked a separate set of additional questions from rotating topical modules. In 2017, one set of the split samples was asked questions on the topical module of commuting, while the other split sample set was asked questions on the topical module of disaster planning. Other modules in other survey cycles include housing counseling, arts and culture, healthy homes, and food security.

2.3 Sample Selection

Each sample within a Metro area forms a representative sample of housing units for that respective Metro area. The sampling process involved selecting housing units systematically from a list of all housing units within each of the Metro areas, which correspond to 25 of the self-representing primary sampling units described in the Integrated National Sample paper. The housing unit list, known as the Master Address File (MAF), is a data set maintained by the Census Bureau based on updates from the prior decennial census and semiannual updates from the United States Postal Service (USPS) Delivery Sequence File, which itself consists of the addresses and mail routes serviced by the USPS. The MAF is updated semiannually in January and July, using information provided by the USPS. The 2017 AHS sample was based on the July 2016 MAF.[3]

To ensure the sample was representative of the target population, the Census Bureau stratified the MAF housing units in each Metro area into one of the following categories (known as strata).

- A HUD-assisted unit (as of 2013).
- Trailer or mobile home.
- Owner-occupied and one unit in structure.
- Owner-occupied and two or more units in structure.
- Renter-occupied and one unit in structure.
- Renter-occupied and two or more units in structure.
- Vacant and one unit in structure.
- Vacant and two or more units in structure.
- Other units, such as houseboats and recreational vehicles.

The information to create the stratification was based on the 2010 decennial census and a 2013 list of HUD-assisted units.[4] The sample rate for each stratum was constant and chosen to achieve as close to a sample of 3,000 housing units in each Metro area as mathematically possible.

---

[3] A small number of housing units (about 130) in remote rural areas, derived from another list known as the Coverage Improvement list, were added to the sample.

[4] In practice, the MAF was merged to both the 2010 decennial census and the 2013 HUD-assisted data, thereby permitting stratification of all housing units using the aforementioned housing characteristics.

3. Weighting

Each housing unit in the AHS sample represents itself and between 450 and 4,000 other units.[5] The exact number it represents is its "weight." The weight was calculated in four steps for two purposes: to minimize sampling errors and errors from incomplete data and to force consistency with published estimates of certain housing and household characteristics that are believed to come from a more reliable data source.

3.1 Step 1: Base Weight Calculation

Every housing unit in the MAF had a positive probability of being selected into the AHS sample. The reciprocal of this probability of selection is referred to as the base weight and accounts for a sample housing unit's probability of selection in the Metro sample selection process.

3.2 Step 2: Noninterview Adjustment Factor

Many eligible housing units selected for the AHS have potential respondents who do not complete an interview. Some are never home, refuse to answer, or had a language barrier, and sometimes, although rarely, the housing unit cannot be accessed by passable roads or the address cannot be found. These sample housing units result in a noninterview, which is also referred to as "unit nonresponse" and is different from "item nonresponse," which covers instances where an interviewee declines to answer a subset of AHS questions.

The noninterview adjustment factor (NAF) deals exclusively with unit nonresponse by expanding the weights of completed interviews to account for similar noninterviews. The calculation of the NAF involves three components—

1. Define NAF cells.
2. Calculate the NAF.
3. Collapse cells, if necessary.

Defining and calculating the NAF cells is a way of reducing the bias due to differential nonresponse. To reduce this nonresponse bias, the Census Bureau formed cells that include sample units that are homogenous to each other within the cells and heterogeneous between cells. Homogeneity and heterogeneity for sample units are measured with respect to the household's propensity to respond to the AHS interview.

Table 3.1 summarizes the variables used in combination to define cells of the noninterview adjustment. Research conducted prior to 2015 determined the variables that best group sample units into cells with homogenous propensity to complete an AHS interview.

---

[5] The mean value of the weights is 1,940. The median value of the weights is 1,309. The lower bound 5th percentile of weights is 464. The upper bound 95th percentile of weights is 3,982.

Table 3.1: Noninterview Cells

Variable	Level Defined	Values
Core-Based Statistical Area (CBSA; 2013)	CBSA	One of 25 values
Type of housing unit	Housing Unit	(1) House, apartment or flat (2) Mobile home (3) Other
Core-Based Statistical Area (CBSA; 2013)	County	(1) Metropolitan area: principal city (2) Metropolitan area: nonprincipal city (3) Micropolitan area
Quartiles of median income	Census block group	Four values for each of the four quartiles
Urban/rural status (2010)[a]	Area	(1) Urban (includes urban cluster or urban area) (2) Rural

[a] See https://www.census.gov/geo/reference/ua/urban-rural-2010.html for more information on the 2010 Urban and Rural classifications.

Note that is possible for a CBSA to have a "rural" component within it because urban/rural status is defined both independently of CBSAs and at a lower level of geography.

With the cells defined, the NAF within each cell is calculated as

$$\text{NAF} = \frac{\text{Interviews} + \text{Noninterviews}}{\text{Interviews}}$$

For both the numerator and the denominator of the NAF, weighted estimates of the number of interviews and noninterviews were used. The estimates were weighted using the product of the base weight (step 1) and housing unit calibration (step 2).

Lastly, cells of the NAF were collapsed if they have fewer than 25 sample housing units or the NAF is greater than 2.0. This avoided two potential problems: (1) unstable NAF estimates due to small cell counts and (2) large variances due to large adjustment factors.

It is important to note that some housing units selected for the AHS have respondents who complete enough questions in an interview for it to be considered a completed interview. However, if the respondents did not answer all the questions in the split sample modules, the housing unit is considered a noninterview for the split sample modules and will not have a value for the split sample weight.

3.3 Step 3: Housing and Demographic Adjustment Factors

The last step of calculating the weights is applying the Ratio Adjustment Factors (RAFs) to the weights to improve the coverage and reduce the variance of estimates. This step involves adjusting AHS weights to be consistent with known estimates of housing units and population from other data sources believed to be of superior quality or accuracy—these are referred to as "control totals." The RAF reduces the variance of an estimate when the control totals are associated with the estimated variable of interest.

The process of applying adjustment factor is called "raking." Ratio adjustments are a method of adjusting sample weights with control totals; their implementation is fairly straightforward.

$$\text{RAF} = \frac{\text{Independent Estimate}}{\text{AHS Sample Estimate}}$$

The calculation of the RAFs for AHS includes five steps:

1. Choose control totals and their adjustment priority order.
2. Define cells.
3. Calculate RAF iteratively, in order of importance (called raking).
4. Collapse cells.
5. Repeat raking until no further change is observed.

Table 3.2 provides information about the RAFs and their order of implementation. It is important to note the adjustment priority order reflects the importance placed on ensuring the AHS estimates, as adjusted, match the control totals. In other words, HUD and the Census Bureau place greater priority on adjusting AHS weights to match new construction control totals than the other control totals.

Additional information about the RAF and raking process is contained in Appendix A. Examples of the RAF process are contained in Appendix B.

Table 3.2 Ratio Adjustment Factor Details

Adjustment Priority Order	Control Total	Cell Definition	Data Source(s)
1	Number of HUD*-assisted housing units	CBSA*/County and three categories of HUD assistance (Public Housing, Private-Project Based, and Vouchers)	HUD*
2	Number of total persons	CBSA*/County	Census Population Division
3	Number of Black persons	CBSA*/County	Census Population Division
4	Number of persons aged 65+	CBSA*/County	Census Population Division
5	Number of Hispanic persons	CBSA*/County	Census Population Division
6	Number of total housing units	CBSA*/County	Census Population Division

*CBSA = Core-Based Statistical Area; HUD = U.S. Department of Housing and Urban Development.

4. Nonsampling Errors

All numbers from the AHS, except for sample size, are estimates. As in other surveys, two types of general errors occur: sampling errors and nonsampling errors. Sampling errors are discussed in Section 5. The definition of nonsampling errors is—

> Nonsampling errors arise mainly due to misleading definitions and concepts, inadequate sampling frames, unsatisfactory questionnaires, defective methods of data collection, tabulation, coding, incomplete coverage of sample units, and so on These errors are unpredictable and not easily controlled. Unlike sampling error, this error may increase with increases in sample size. If not properly controlled, nonsampling error can be more damaging than sampling error for large-scale household surveys.[6]

The various types of nonsampling errors are discussed in the following sections.

4.1 Coverage Errors

Coverage errors arise from the failure to give some units in the target population any chance of selection into the sample (undercoverage), or giving units more than one chance of selection (overcoverage). Because of deficiencies in the sampling lists, the housing units in the survey may not represent all housing units in the country. The Census Bureau attempts to address the deficiencies by adjusting the raw numbers from the survey proportionally so that the numbers published match independent estimates of the total number of housing units. This is part of the

[6] https://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_7.pdf

weighting production process described in Section 3. Table 4.1 lists the sources of coverage errors. AHS users do not have to take any additional steps to account for coverage error.

Table 4.1 Sources of Coverage Errors

Type of Unit	Type of Coverage Error
Housing units with P.O.* Box address or without 911 address[a]	The MAF* includes 911 addresses because they can be located and does not include P.O. Box addresses.
New construction	Eligible units will be added but there is a lag between the time the unit is eligible and when it is added to the MAF*.
Group quarters	Eligible units could be missed because of incorrect answers to questions used to screen out group quarters, which are ineligible units for the AHS*.

*P.O. = Post Office;MAF = Master Address File; AHS = American Housing Survey.

[a] A number assigned to a structure that, in conjunction with a street or road name, identifies the location of the structure in the event of an emergency.

4.2 Nonresponse Error

Some respondents refuse the interview or cannot be located. HUD and the Census Bureau correct for nonresponse by implementing NAFs into the weighting process, as discussed in Section 3. AHS users do not have to take any additional steps to account for nonresponse error.

4.3 Measurement Errors from Missing Responses to Questions

Some respondents participate in an interview but refuse to answer questions or do not know a particular answer. For certain questions, HUD and the Census Bureau impute missing responses using various imputation techniques. The Census Bureau does not know how close the imputed values are to the actual values. For other items, “not reported” is used as an answer category. The items with the most missing data are primarily those that people forget or consider sensitive: mortgages, other housing costs, and income.

Incompleteness can cause large errors. A missing response in even 10 percent of sample units represents about 13.7 million homes (about 137 million homes are in the United States). To give users a sense of the bias caused by missing data, table 4.2 provides estimates for Errors for Incomplete Data Bias.

Table 4.2 Errors for Incomplete Data Bias for 2017 AHS-MS [a] (numbers in thousands) – Top 15

When the AHS estimate of the number of housing units with a characteristic is…	…the chances are 90 percent that the complete value* is within the range of plus or minus						
(Size of Estimate)	Atlanta, GA	Boston, MA	Chicago, IL	Dallas, TX	Detroit, MI	Houston, TX	Los Angeles, CA	Miami, FL
0	4.5	3.9	7.6	5.5	3.8	5.1	9.2	5.0
10	5.1	4.5	8.2	6.1	4.4	5.7	9.8	5.6
100	10.5	9.8	13.6	11.5	9.7	11.1	15.1	11.0
250	19.5	18.8	22.5	20.4	18.7	20.1	24.1	19.9
500	34.4	33.7	37.4	35.4	33.6	35.0	39.0	34.9
750	49.3	48.6	52.4	50.3	48.5	49.9	54.0	49.8
1,000	64.3	60.9	67.3	65.2	57.9	64.8	68.9	64.7
2,000	22.5	1.2	117.6	52.9		40.9	128.6	36.9
3,000			57.9				107.4	
4,000							47.7	
4,600							11.8	

(Size of estimate)	New York, NY	Philadelphia, PA	Phoenix, AZ	Riverside, CA	San Francisco, CA	Seattle, WA	Washington, DC
0	15.8	4.9	3.8	3.1	3.6	3.1	4.6
10	16.4	5.5	4.4	3.7	4.2	3.7	5.2
100	21.7	10.9	9.7	9.0	9.5	9.1	10.6
250	30.7	19.8	18.7	18.0	18.5	18.1	19.6
500	45.6	34.8	33.6	32.9	33.4	33.0	34.5
750	60.5	49.7	48.6	47.9	48.4	47.9	49.4
1,000	75.5	64.6	58.2	36.5	51.9	38.4	64.4
2,000	135.2	34.1					25.8
3,000	194.9						
4,000	253.6						
5,000	193.9						
6,000	134.2						
7,000	74.5						
7,900	20.7						

[a] AHS-MS = American Housing Survey – Metropolitan Sample
Source: U.S. Census Bureau, 2017 American Housing Survey.
* Complete Value means the value derived if there were no missing data.

Table 4.3 Errors for Incomplete Data Bias for 2017 AHS-MS [a] (numbers in thousands) – Next 10

When the AHS estimate of the number of housing units with a characteristic is…	…the chances are 90 percent that the complete value* is within the range of plus or minus							
(Size of estimate)	Baltimore, MD	Birmingham, AL	Las Vegas NY	Minneapolis, MN	Oklahoma City, OK	Richmond, VA	Rochester, NY	San Antonio, TX
0	2.3	1.0	1.8	2.9	1.1	1.0	0.9	1.8
10	2.9	1.6	2.4	3.5	1.7	1.6	1.5	2.4
100	8.3	7.0	7.7	8.8	7.1	7.0	6.9	7.7
250	17.2	15.9	16.7	17.8	16.1	16.0	14.7	16.7
500	32.1	2.1	25.6	32.7	5.6	2.8		25.2
750	26.8		10.7	44.5				10.3
1,000	11.9			29.6				
1,400				5.7				

(Size of estimate)	San Jose, CA	Tampa, FL
0	1.4	2.8
10	2.0	3.4
100	7.3	8.8
250	16.3	17.7
500	12.7	32.7
750		42.8
1,000		27.9
1,400		4.0

[a] AHS-MS = American Housing Survey – Metropolitan Sample
Source: U.S. Census Bureau, 2017 American Housing Survey.
* Complete Value means the value derived if there were no missing data.

Tables 4.2 and 4.3 are intended to be used only when a particular survey estimate is based on one or more variables with completeness rates of 50 to 90 percent. The values in table 4.2 are based on a 1990 analysis[7] by the Census Bureau, which estimated the standard error from missing data to be—

> .0012 x U + .0363 x MIN(A, U-A), where A is any count of housing units with a characteristic from the AHS and U is the total number of housing units in the given Metropolitan area (both in thousands, result also in thousands).

Due to the large number of variables in the AHS, HUD and the Census Bureau typically do not publish completeness rates for individual survey estimates. AHS users who are interested in completeness rates should consider using the AHS public use file (PUF) microdata to estimate completeness rates. When using the PUF to estimate completeness rates, users should be aware of the following.

[7] *How Response Error, Missing Data and Undercoverage Bias Survey Data*, P. Burke et. al. 1990.

- PUF variables with a value of “not applicable” should not be considered missing. Not applicable means the question corresponding to the variable was not asked of the AHS respondent because they were not “in scope” for the question. For instance, if a respondent reported living in an apartment building, the respondent will not be asked questions about mobile home features.
- PUF variable with a value of “not reported” should be considered missing. A PUF variable will have a value of “not reported” if HUD and the Census Bureau did not develop an imputation process for the variable.
- For PUF variables for which HUD and the Census Bureau developed an imputation process, the variable will have a corresponding edit/imputation flag variable indicating whether the value of the variable was imputed for the respondent. The edit/imputation flag variables are the same as the variable name but are preceded by the letter “J.” For instance, if a respondent did not report a value for the variable HFUEL (heating fuel), but the respondent’s value was imputed, the variable JHFUEL will equal “2,” indicating an imputation.
- The edit/imputation flag will take a value of “1” if the respondent’s reported value was edited. These edited values should not be considered missing.

  The correct way to calculate a completeness rate in the AHS is the following:

$$\frac{(\text{sum of respondents with reported values} - \text{sum of respondents with imputed values})}{(\text{sum of respondents with reported values} + \text{sum of respondents with “nonreported” values})}$$

4.4 Measurement Error from Inaccurate Responses to Questions

Wrong answers happen because people misunderstand questions, cannot recall the correct answer, or do not want to give the right answer. See *American Housing Survey for the United States: 2005*[8] for more discussion on this topic.

4.5 Question Validity Errors

In order to avoid the failure to design a survey question that accurately measures the construct of interest, HUD and the Census Bureau carefully test each new survey question to ensure it is measuring the construct of interest. Although some respondents possibly misinterpret the question, HUD and the Census Bureau do not have any additional information to estimate validity error rates. AHS users do not have to take any additional steps to account for validity error.

4.6 Processing Errors

After the data are collected, errors that can be introduced include data capture errors, data coding and classification errors, and data editing and imputation errors. HUD and the Census

[8] https://www.census.gov/prod/2006pubs/h150-05.pdf

Bureau carefully test all aspects of the data capture, coding, classification, editing, and imputation procedures. Although mistakes are possible, HUD and the Census Bureau believe they are minimal. If a processing error is discovered, HUD and the Census Bureau will let AHS users know and, in some cases, will publish revised estimates. AHS users do not have to take any additional steps to account for processing error.

4.7 Additional Considerations

The AHS is a longitudinal survey conducted every 2 years. Many AHS users compare current-year AHS estimates with prior-year estimates. Users should be aware that HUD and the Census Bureau often make small changes to the text of various questions between surveys. AHS users comparing estimates with prior-year surveys should consult the document "Changes Between Surveys" that is published with each new AHS.[9]

5. Sampling Errors

Error from sampling reflects how estimates from a sample vary from the actual value if all housing units had been interviewed under the same conditions. A confidence interval is a range that contains the actual value with a specified probability.

Users of the AHS PUFs can use replicate weights to create standard errors for any estimate. For further information, see "Guide to Estimating Variances Using Replicate Weights," which is available on the Census AHS website (https://www.census.gov/programs-surveys/ahs/tech-documentation/help-guides/quickguide_repweight.html).

For users of the AHS Summary Tables, Generalized Variance Functions (GVFs) are a convenient tool for quick and easy estimation of sampling errors. The text below describes how to calculate sampling errors for counts, percents, medians, and differences using GVFs.

5.1 Sampling Errors for Estimated Totals

Most published estimates from the AHS reflect weighted estimated totals of housing units. The error from sampling for a weighted estimated total is approximated using the following GVF for constructing a 90-percent confidence interval.

$$1.645 \times \sqrt{b \times A + a \times A^2},$$

where *A* is the weighted estimated total of housing units in thousands from the AHS and *a* and *b* are GVF parameters that vary depending on the characteristic being estimated.

Tables 5.1 and 5.2 include the values of *a* and *b* for each of the 25 Metro areas included in the Metro sample, broken down by full and split sample estimates, and by occupancy type.

[9] An example of this document, which shows the changes between 2015 and 2017, is at https://www2.census.gov/programs-surveys/ahs/2017/2017%20AHS%20Historical%20Changes.pdf

Table 5.1. Generalized Variance Function Parameters for Metro Area Estimates - Top 15

		Full Sample		Split Samples	
Metropolitan Area	Domain	*a*	*b*	*a*	*b*
Atlanta, GA	Total Units	-0.000393	0.90	-0.000817	1.88
	Owner Occupied	-0.000560	0.98	-0.001137	2.05
	Renter Occupied	-0.000676	1.06	-0.001235	2.08
Boston, MA	Total Units	-0.000431	0.84	-0.001069	2.09
	Owner Occupied	-0.000580	0.91	-0.000914	2.03
	Renter Occupied	-0.000904	1.00	-0.000717	2.10
Chicago, IL	Total Units	-0.000390	1.50	-0.000906	3.48
	Owner Occupied	-0.000474	1.66	-0.001048	3.72
	Renter Occupied	-0.000561	1.76	-0.001335	3.83
Dallas, TX	Total Units	-0.000287	0.80	-0.000686	1.92
	Owner Occupied	-0.000396	0.94	-0.000795	2.05
	Renter Occupied	-0.000435	0.94	-0.000852	2.03
Detroit, MI	Total Units	-0.000382	0.73	-0.000815	1.55
	Owner Occupied	-0.000511	0.80	-0.001044	1.68
	Renter Occupied	-0.000846	0.80	-0.001852	1.77
Houston, TX	Total Units	-0.000335	0.87	-0.000736	1.91
	Owner Occupied	-0.000480	0.99	-0.000924	2.04
	Renter Occupied	-0.000602	1.00	-0.001113	2.13
Los Angeles, CA	Total Units	-0.000311	1.45	-0.001161	5.39
	Owner Occupied	-0.000621	1.74	0.001229	3.73
	Renter Occupied	-0.000574	1.68	0.000384	4.24
Miami, FL	Total Units	-0.000322	0.82	-0.001144	2.90
	Owner Occupied	-0.000563	0.99	-0.000237	2.31
	Renter Occupied	-0.000523	0.91	0.000047	2.23
New York, NY	Total Units	-0.000403	3.22	-0.001095	8.74
	Owner Occupied	-0.000625	3.74	-0.000989	8.65
	Renter Occupied	-0.000707	3.82	-0.001049	8.48
Philadelphia, PA	Total Units	-0.000402	1.00	-0.000882	2.19
	Owner Occupied	-0.000504	1.09	-0.000957	2.29
	Renter Occupied	-0.000573	1.11	-0.001188	2.36
Phoenix, AZ	Total Units	-0.000349	0.67	-0.000787	1.50
	Owner Occupied	-0.000548	0.81	-0.001116	1.71
	Renter Occupied	-0.000580	0.75	-0.001092	1.57
Riverside, CA	Total Units	-0.000326	0.51	-0.000682	1.06
	Owner Occupied	-0.000537	0.61	-0.001057	1.24
	Renter Occupied	-0.000626	0.57	-0.001460	1.21
San Francisco, CA	Total Units	-0.000368	0.66	-0.000981	1.77
	Owner Occupied	-0.000553	0.77	-0.001194	1.75
	Renter Occupied	-0.000578	0.76	-0.001337	1.82
Seattle, WA	Total Units	-0.000354	0.56	-0.000831	1.32
	Owner Occupied	-0.000568	0.65	-0.000748	1.39
	Renter Occupied	-0.000666	0.64	-0.000586	1.35
Washington, DC	Total Units	-0.000306	0.72	-0.000840	1.98
	Owner Occupied	-0.000465	0.84	-0.000690	1.84
	Renter Occupied	-0.000673	0.86	-0.000842	2.03

Source: U.S. Census Bureau, 2017 American Housing Survey.

Note that "Total Units" means all eligible housing units, including vacant units

Table 5.2 Generalized Variance Function Parameters for Metro Area Estimates - Next 10

		Full Sample		Split Samples	
Metropolitan Area	Domain	*a*	*b*	*a*	*b*
Baltimore, MD	Total Units	-0.000389	0.45	-0.000846	0.98
	Owner Occupied	-0.000536	0.52	-0.001071	1.07
	Renter Occupied	-0.000786	0.52	-0.001539	1.14
Birmingham, AL	Total Units	-0.000394	0.20	-0.000846	0.44
	Owner Occupied	-0.000486	0.22	-0.001017	0.46
	Renter Occupied	-0.000845	0.25	-0.001717	0.51
Las Vegas, NV	Total Units	-0.000352	0.32	-0.000770	0.69
	Owner Occupied	-0.000613	0.37	-0.001134	0.74
	Renter Occupied	-0.000545	0.35	-0.001218	0.78
Minneapolis, MN	Total Units	-0.000382	0.55	-0.000809	1.17
	Owner Occupied	-0.000511	0.61	-0.001009	1.28
	Renter Occupied	-0.000776	0.60	-0.001406	1.27
Oklahoma City, OK	Total Units	-0.000348	0.20	-0.000743	0.43
	Owner Occupied	-0.000473	0.22	-0.000824	0.44
	Renter Occupied	-0.000586	0.22	-0.001234	0.48
Richmond, VA	Total Units	-0.000388	0.21	-0.000858	0.45
	Owner Occupied	-0.000544	0.23	-0.001115	0.49
	Renter Occupied	-0.000776	0.23	-0.001659	0.52
Rochester, NY	Total Units	-0.000377	0.18	-0.000834	0.40
	Owner Occupied	-0.000525	0.20	-0.001061	0.43
	Renter Occupied	-0.000964	0.22	-0.001640	0.46
San Antonio, TX	Total Units	-0.000329	0.29	-0.000746	0.67
	Owner Occupied	-0.000449	0.33	-0.000748	0.67
	Renter Occupied	-0.000517	0.33	-0.000891	0.69
San Jose, CA	Total Units	-0.000366	0.25	-0.000816	0.56
	Owner Occupied	-0.000597	0.28	-0.001053	0.59
	Renter Occupied	-0.000724	0.29	-0.001249	0.61
Tampa, FL	Total Units	-0.000392	0.56	-0.000841	1.19
	Owner Occupied	-0.000563	0.64	-0.001067	1.30
	Renter Occupied	-0.000699	0.67	-0.001178	1.39

Source: U.S. Census Bureau, 2017 American Housing Survey.

Note that "Total Units" means all eligible housing units, including vacant units

For example, if a user wants to calculate the 90-percent confidence interval of the number of owner occupied homes in the Richmond, VA Metro area (*A* = 315.0), then they would look at the Owner Occupied row within Richmond, VA for the full sample column in Table 5.2 and apply the following formula —

$$1.645 \times \sqrt{0.23 \times 315.0 - 0.000544 \times 315.0^2} = 7.1$$

The 90-percent confidence interval can then be formed by adding to and subtracting from this error to the survey estimate of 315.0 (that is, 315.0 plus or minus 7.1). There is a 90 percent chance that this interval – 315.0 plus or minus 7.1 (307.9 to 322.1) – contains the actual total and a 10 percent chance that it does not.[10]

Numbers in the published estimates are shown in thousands, so 315.0 means 315,000. The formulas are designed to use numbers directly from the published estimates; do not add zeros. The result is also in thousands, so 7.1 means 7,100.

5.2 Sampling Error for Percentages

Any subgroup can be shown as a percentage of a larger group. The error from sampling for a 90-percent confidence interval for this percentage is—

$$1.645 \times \sqrt{\frac{b \times p \times (100 - p)}{A}},$$

where *p* is the percentage; *A* is the weighted denominator, or base of the percentage in thousands; and *b* is the GVF parameter from Tables 5.1 or 5.2.

For example, if a user wants to calculate the 90-percent confidence interval for the percentage of owner occupied homes in the Richmond CBSA that have a garage or carport (60.7%, *A* = 315.0), the formula is—

$$60.7 \pm 1.645 \times \sqrt{\frac{0.23 \times 60.7 \times (100 - 60.7)}{315.0}} = 60.7 \pm 2.2$$

There is a 90 percent chance that the interval of 58.5 percent to 62.9 percent contains the true percentage.

Note that when a ratio *C*/*D* is computed where *C* is not a subgroup of *D* (for example, the number of owner occupied housing units as a ratio of the number of renter occupied), the error from sampling is different. The error from sampling for a 90-percent confidence interval for a ratio *C*/*D* is—

$$1.645 \times \left[\left(\frac{C}{D}\right)\sqrt{\left(\frac{\text{SE for } C}{C}\right)^2 + \left(\frac{\text{SE for } D}{D}\right)^2}\right].$$

The standard error for *C* is equivalent to this equation below; use an analogous formula for *D*.

$$SE\ for\ C = \sqrt{b \times C + a \times C^2}$$

Where *a* and *b* are the GVF parameters from Tables 5.1 or 5.2.

[10] This formula gives 90-percent confidence interval errors. For 95-percent confidence interval errors, multiply by 1.96 instead of 1.645; for 99-percent confidence, multiply by 2.576 instead of 1.645.

5.3 Sampling Error for Differences

Two estimates from the AHS, like 21 and 34 or 34 percent and 55 percent, have a statistically significant difference if their 90-percent confidence intervals do not overlap.

When 90-percent confidence intervals do overlap, numbers are still statistically different if the result of subtracting one from the other is more than—

$$\sqrt{(\text{error for first number})^2 + (\text{error for second number})^2}.$$

The error for the first and second numbers should be interpreted as the error for a 90-percent confidence interval for the first and second numbers, respectively.

5.4 Sampling Error for Medians

Table 5.3 shows how to calculate the error from sampling for a 90-percent confidence interval for medians. This is an approximation of the error. The steps in table 5.3 should only be used when the cumulative number of housing units for which the median applies is larger than 10 percent of the total number of housing units.

When cumulative number of housing units for which the median applies is smaller than 10 percent of the total number of housing units, the confidence interval on medians cannot be estimated reliably. To estimate a median’s sampling error more accurately, use the steps in table 5.4 to find the sampling error on 50 percent and apply it to compute the 90-percent confidence interval for the median.

The steps in Tables 5.3 and 5.4 are based on the 2017 estimates for Total Annual Household Income in the Richmond, VA CBSA, reflected in Table 5.5. AHS estimated that there were 174,498 renter occupied housing units in the Richmond CBSA; so, A = 174.5. The estimated median income is $39,822.00.

Table 5.3. Steps to Compute the 90-Percent Confidence Interval for a Median for Large Bases

Steps for Calculations	The Formula	An Example
How many total units is the median based on (in thousands, exclude "not reported" and "don't know")?	A	174.5
What is the estimated standard error of a 50-percent characteristic with a base equaling the total units?	$\sigma = \sqrt{\frac{b \times 0.5 \times (1-0.5)}{\mathrm{A}}}$	$\sqrt{\frac{0.23(0.5)(1-0.5)}{174.5}} = 0.0182$
What are the end points of the category the median is in?	X – Y	$30,000 – 39,999
What is the width of this category (in dollars, rooms, or whatever the item measures)?	W	$10,000
How many housing units are in this median category (in thousands)?	B	19.9
What is the estimated proportion of the total units falling in the category containing the sample median?	$\mathrm{P} = \frac{\mathrm{B}}{\mathrm{A}}$	$\frac{19.9}{174.5} = 0.114$
Then the standard error from sampling for the median is approximately:	$se_{median} = \frac{\sigma \times \mathrm{W}}{\mathrm{P}}$	$\frac{0.0182 \times \$10{,}000}{0.114} = \$1{,}596.49$
The 90-percent confidence interval for the median is:	$Median \pm 1.645 \times se_{median}$	$\$39{,}822.00 \pm \$2{,}626.23$

Source: U.S. Census Bureau, 2017 American Housing Survey.

Table 5.4. Steps to Compute the Error from Sampling for a 90-Percent Confidence Interval for a Median for Small Bases

Item	Formula	Bottom Limit Example	Top Limit Example
How many total units is the median based on (in thousands, exclude “not reported”)?	A	174.5	
Half the total, for the median (in thousands)	A/2	87.25	
Error from sampling for 50 percent of the base of this median (first line)	$1.645 \times \sqrt{\frac{b \times 0.5 \times (1-0.5)}{\text{A}}}$	$\frac{1.645 \times \sqrt{.21 \times .25}}{\sqrt{174.5}} = 0.029$	
Multiply this percentage by total units to give the error in housing units.	$1.645 \times \sqrt{A \times b \times 0.25}$	$0.030 \times 174.5 = 5.061$	
Bottom of error range (second line minus fourth line, in thousands)	B_{bottom}	82.2*	
Top of error range (second line plus fourth line, in thousands)	B_{top}		92.3*
*Start adding up the housing units in this table, category by category, cumulatively from the beginning of the table, until you exceed the starred number above. What interval does the starred number fall in?		$30,000 – 39,999	$40,000 – 49,999
How many housing units are in all the categories before this one (in thousands)?	C	67.1	87.0
How many housing units are in this category (in thousands)?	D	19.9	18.9
What is the bottom limit of this category (in dollars, rooms, or whatever the item measures)?	E	$30,000	$40,000
What is the bottom limit of the next category (in dollars, rooms, etc.)?	F	$40,000	$50,000
Formula to calculate limits of confidence interval	$\frac{\text{B}-\text{C}}{\text{D}}(\text{F}-\text{E})+\text{E}$	$\frac{82.2-67.1}{19.9}(10{,}000)+30{,}000$	$\frac{92.3-87.0}{18.9}(10{,}000)+40{,}000$

Limits of confidence interval (in dollars, rooms, etc.)		$37,587.94	$42,804.23

Source: U.S. Census Bureau, 2017 American Housing Survey.

*Starting with this step, this worksheet is equivalent to interpolation

Table 5.5. Annual Household Income Median Estimate for Use with Tables 5.3 and 5.4 (numbers in thousands, except median)

	Number of Housing Units	Cumulative Number of Housing Units
Total annual household income	174.5	
Less than $5,000	13.8	13.8
$5,000 to $9,999	11.2	25.0
$10,000 to $14,999	9.7	34.7
$15,000 to $19,999	7.6	42.3
$20,000 to $24,999	13.9	56.2
$25,000 to $29,999	10.9	67.1
$30,000 to $39,999	19.9	87.0
$40,000 to $49,999	18.9	105.9
$50,000 to $59,999	15.4	121.3
$60,000 to $79,999	22.4	143.7
$80,000 to $99,999	14.9	158.6
$100,000 to $119,999	4.0	162.6
$120,000 or more	11.8	174.4
Median Income (dollars)	$39,822	

Source: U.S. Census Bureau, 2017 American Housing Survey. (Cumulative total is slightly off due to rounding.)

5.5 Additional Considerations

It should be noted that the minimum error from sampling is plus or minus 10 (meaning plus or minus 10,000).[11] If a formula gives an error smaller than 10, use 10.

[11] This minimum error formula is based on the following binomial 90-percent confidence interval on 0: $U \times \left(1 - .1^{4.33/\mathrm{U}}\right) = 10$, (where *U* is the total number of housing units from the AHS). For a 95-percent confidence interval, substitute .05 for .1 in the above formula. For a 99-percent confidence interval, substitute .01 for .1.

Appendix A. Ratio Adjustment Process Details

In the last step of calculating the weights, the Census Bureau applied the Ratio Adjustment Factor (RAF) to the weights to improve the coverage and reduce the variance of estimates. These goals were achieved by adjusting American Housing Survey (AHS) weights to be consistent with control totals of housing units and population. The RAF also reduces the variance of an estimate when the control totals are associated with the estimated variable of interest.

Generally speaking, ratio adjustments are a method of adjusting sample weights with control totals and their implementation is fairly straightforward. Take a control total, *X*, and its corresponding estimate, $\hat{X}$, and multiply sample weights by a factor of $X / \hat{X}$. This calculation results in adjusted sample weights that produce estimates that are much closer to the control total.

The calculation of the RAFs for AHS can be broken down into five steps.

1. Choose Control totals.
2. Define Cells.
3. Calculate the RAF.
4. Collapsing RAF Cells.
5. Repeat Raking.

Step 1. Choose Control Totals

As mentioned previously, the Census Bureau wants control totals *X* that are associated with the variable of interest. Control totals also require a reasonable corresponding estimate $\hat{X}$ from AHS. Both the control total *X* and the AHS estimate $\hat{X}$ should define the same total. For example, a ratio adjustment for the total number of HUD housing units requires that both *X* and $\hat{X}$ represent the same geographic area, apply to the same type of HUD program, and have the same reference period.

A second requirement for the control totals is that they should be a better estimate than the estimate produced from AHS. Again, these control totals are assumed to be more accurate than the AHS estimates and also have no variance.

Based on these two requirements, the following three data sources for control totals, described in table A1, were considered to be suitable for ratios adjustments in 2017.

Table A1. Sources of Control Totals

Control Total Candidates	Data Source
Number of housing units in HUD* programs	HUD*, based on 2017 HUD program data.
Number of new construction housing units	HUD* and the Census Bureau, based on estimates from the 2013–2016 Survey of Construction and Manufactured Housing Survey, which were combined to define the new construction control totals.
Total population and housing unit counts by various characteristics	2017 household population and housing unit demographic analysis projections derived from the 2010 census and estimated for July 1, 2017, by the Census Bureau Population Division.

*HUD = U.S. Department of Housing and Urban Development.

Given all possible control totals available in the sources listed in table A1, HUD and the Census Bureau chose eight sets of totals within these three data sets to use for its RAF, as well as the priority order for which they are applied, which are presented in table A2.

Table A2. Ordered List of Control Totals

Order	Control Total and Source	Cell Definition
1	Number housing units in HUD* programs	CBSA*/County and HUD* program (Public Housing, Private-Project Based, and Vouchers)
2	Number of total persons	CBSA*/County
3	Number of Black persons	CBSA*/County
4	Number of persons aged 65+	CBSA*/County
5	Number of Hispanic persons	CBSA*/County
6	Number of total housing units	CBSA*/County

*HUD = U.S. Department of Housing and Urban Development; CBSA = Core-Based Statistical Area.

Step 2. Define Cells

Control totals within specifically defined groups of housing units, which are referred to as “cells,” were acquired for each of the chosen totals above. Estimates from the AHS were also calculated within these cells, and both of these were used to calculate RAFs.

Table A2 summarizes the cells for each set of ratio adjustments.

All of the ratio adjustments were applied at the CBSA/county level.

Cells defined by HUD programs. The cells for the ratio adjustments of HUD housing units included three types of HUD programs: public housing, private-project based, and vouchers.

Step 3. Calculate the RAF

With the cells defined, AHS started with the first chosen control total—number of HUD-assisted housing units—and calculated ratio adjustment within each cell as—

$$\frac{\text{Control Total}}{\text{AHS Estimated Total}}$$

This factor was then multiplied by the AHS weights to adjust AHS estimated counts within each cell. Ratio adjustments were applied iteratively using each of the remaining chosen control totals and their respectively defined cells in a process that is called raking. Each cell of each rake of table A2 was adjusted with equation 1.

Step 4. Collapsing RAF Cells

RAF cells were collapsed for the same reasons noninterview adjustment factor cells were collapsed: (1) because a small number of sample housing units may produce an unstable estimate of the RAF and (2) to avoid large sample weights. To address both issues, cells are required to have at least 25 housing units, and the RAF must be less than or equal to 2.0. Cells were only collapsed after the first iteration of the raking through all of the chosen control totals in table A2.

Step 5. Repeat Raking

After completing the first iteration of rakes and checking to see which cells need collapsing, raking was repeated using the ratios of chosen control totals over the modified AHS estimates until the AHS estimated totals stopped changing significantly between each raking step.

Appendix B provides a detailed example of how AHS uses raking within cells and across chosen totals.

Appendix B. Examples of Ratio Adjustments

This appendix provides two hypothetical examples that demonstrate how the sample weights were adjusted so that they were consistent with a set of control totals. The first example is a ratio adjustment, and it is provided as context because it is a special case of raking—one rake. The second example demonstrates how to complete a more complicated raking adjustment.

For the two examples, assume weights were calculated for a sample and the weights included all weighting adjustments up to a nonresponse adjustment. With these weights, hypothetical totals by two categories of tenure status (owner or renter) and two categories of type of construction (old or new) are assumed. Table B1 summarizes the estimated totals resulting from this hypothetical sample and weights.

Table B1: Estimated Totals

	Owners	Renters	Total
New	110	91	201
Old	97	107	204
Total	207	198	405

Example 1: Ratio Adjustment

Suppose the control totals were as shown in Table B2.

Table B2: Example 1 Control Totals

	Owners	Renters	Total
New	115	105	220
Old	95	105	200
Total	210	210	420

The control totals of Table B2 are used to improve the weights by making the estimates from the weights consistent with the control totals. Table B3 shows the Ratio Adjustment Factor (RAF) that will make the estimated totals consistent with the control totals.

Table B3: Example 1 Ratio Adjustment Factors

	Owners	Renters
New	115/110 = 1.0455	105/91 = 1.1583
Old	95/97 = 0.9794	105/107 = 0.9813

If the factors from Table B3 are applied to the weights of the sample units, then the estimates from the revised weights will be consistent with the totals of Table B2.

Note that ratio-adjusted weights for the combination of owners and new construction is the product of the weight before raking with the RAF, that is,

$$\text{Ratio-adjusted weight} = \text{original weight} \; \square 1.0455 .$$

The ratio-adjusted weights for the other three cells are defined similarly.

Example 2: Raking Adjustment

Table B4 shows different control totals than those of Table B2.

Table B4: Example 2 Control Totals

	Owners	Renters	Total
New	?	?	220
Old	?	?	200
Total	210	210	420

Table B4 does not have the totals for the specific combinations of tenure status and old or new construction; however, totals can be used with raking to improve the weights.

Raking is the repeated application of ratio adjustments to the marginal totals. Ratio adjustments are repeated for each set of marginal totals—the row totals and the column totals in this example. It can be shown that raking will converge to a unique solution.

First, raking the categories of old or new construction is done. This involves adjusting the cells for the totals of old or new construction. Table B5 shows the calculated adjustment factors for the first rake.

Table B5: Factors for First Rake—Old or New Construction

	Ratio	Factor
New	220/201	1.0945
Old	200/204	0.9804

For new construction, the value of 220 came from the marginal control total of new construction (first row) in Table B4, and the value of 201 came from the marginal estimated total of new construction (first row) in Table B1.

The ratios of Table B5 are then applied to the totals, or, equivalently, the weights of the sample units that are used to calculate the total. Table B6 shows the application of the factors from table B5 to the totals of Table B1.

Table B6: New Total for First Rake—Old or New Construction

	Owners	Renters
New	110 x 1.0945 = 120.40	91 x 1.0945 = 99.60
Old	97 x 0.9804 = 95.10	107 x 0.9804 = 104.90

Table B7 shows the result of the first rake—the application of the factors from Table B6 to the totals of Table B1.

Table B7: Revised Totals for First Rake—Old or New Construction

	Owners	Renters	Total
New	120.40	99.60	220.00
Old	95.10	104.90	200.00
Total	215.50	204.50	420.00

After the first rake, the revised estimates are now consistent with the old or new construction column totals, but the estimated row totals are not consistent with the tenure control totals.

The tenure totals are then raked using the revised totals in Table B7. The ratio adjustments are calculated with the revised tenure totals from table B7 and the control totals from Table B4. Table B8 shows the factors needed to adjust the owner or renter columns.

Table B8: Factors for Second Rake—Tenure

	Ratio	Factor
New	210/215.51	0.9745
Old	210/204.505	1.0269

The ratios of table B8 are then applied to the weights of the sample units within owners and renters in Table B9.

Table B9: New Totals for Second Rake—Tenure

	Owners	Renters
New	120.45 x 0.9745 = 117.33	99.645 x 1.0269 = 102.28
Old	95.06 x 0.9745 = 92.67	104.86 x 1.0269 = 107.75

Table B10 shows the complete result of the second rake—the application of the factors from Table B9 to the totals of Table B7.

Table B10: Revised Totals for Second Rake—Tenure

	Owners	Renters	Total
New	117.33	102.28	219.61
Old	92.67	107.72	200.39
Total	210.00	210.00	420.00

With the second rake, the revised estimates are now consistent with the tenure row totals, but the estimated row totals are not consistent with the tenure control totals. However, both the row and the column totals are closer to the control totals. A third rake is done to adjust for the old or new construction totals again. Table B11 shows the factors of the third rake, and Table B12 shows the resultant totals.

Table B11: Factors for Third Rake—Old or New Construction

	Ratio	Factor
New	220/219.61	1.0018
Old	200/200.39	0.9980

Table B12: Revised Totals for Third Rake—Old or New Construction

	Owners	Renters	Total
New	117.54	102.46	220.00
Old	92.49	107.51	200.00
Total	210.03	209.97	420.00

The fourth rake repeats the adjustment for the tenure totals. Table B13 shows the factors of the third rake, and Table B14 shows the resultant totals.

Table B13: Factors for Fourth Rake—Tenure

	Ratio	Factor
Owner	220/210.03	0.9999
Renter	200/2009.97	1.0001

Table B14: Revised Totals for Fourth Rake—Tenure

	Owners	Renters	Total
New	117.52	102.48	220.00
Old	92.48	107.52	200.00
Total	210.00	210.00	420.00

Table B14 shows the final result of the raking. The original estimated totals are now revised so that both the row totals and column totals are consistent with the control totals of old or new construction and tenure.

To clarify how this applies to the weights, note that raking-adjusted weights for the combination of owners and new construction is the product of the weight before raking with the factors of the four rakes, that is,

$$\text{Raking-adjusted weight} = \text{original weight} \times 1.0945 \times 0.9745 \times 1.0018 \times 0.9999$$
$$= \text{original weight} \times 1.0684 \,.$$

The raking-adjusted weights for the other three cells of example 2 were done similarly.

The adjustment factors in the tables were displayed with rounding to four decimal points. No rounding is done in the actual calculation of the raking prior to their application because the raking would not converge if the factors were rounded.