

Are State-Level Estimates for the American Housing Survey Feasible?

Ernest Lawley, Stephen Ash, Brian Shaffer, Kathy Zha

U.S. Census Bureau
4600 Silver Hill Road, Washington, DC 20233

Abstract

The American Housing Survey provides key estimates about the housing stock of the United States at national and metropolitan (major cities) levels. As with other major surveys, there is general interest in estimates at lower levels of geography, particularly state-level estimates. This research will investigate if we can currently produce reliable state-level estimates and what methods can we apply to already existing survey data to produce reasonable estimates.

In states where state-level estimates are not feasible, we will also investigate potential design changes that would lead to reliable estimates. Specifically, we will consider the addition of housing units and including more primary sampling areas.

Key Words: Weighting, Estimation, Variance Estimation, Sub-domain Estimation, Sample Design

1. Introduction

The American Housing Survey (AHS) is sponsored by the Department of Housing and Urban Development (HUD) and conducted by the U.S. Census Bureau¹. The AHS is the most comprehensive national housing survey in the United States. It provides data on a wide range of housing subjects, including single-family homes, apartments, manufactured housing, vacant units, family composition, income, housing and neighborhood quality, housing costs, equipment, fuel type, and recent moves. AHS collects National data every 2 years from a sample of housing units (HUs). The national survey, which began in 1973, has sampled the same units since 1985; it also samples new construction to ensure continuity and timeliness of the data. Beginning with the 2015 enumeration, a redesigned AHS sample will sample a new set of housing units; data from these units in this redesigned sample will be collected every 2 years, to include biennial additions of new construction.

AHS is a household survey conducted using a laptop survey questionnaire. AHS generally interviews sampled units between April and September of each enumeration year. Census enumerators collect data by telephone or personal visit. For unoccupied units, data are collected by landlords, rental agents, or neighbors.

¹ Views expressed in this paper are those of the authors and do not reflect the views or policies of the Department of Housing and Urban Development or the U.S. Census Bureau.

Each biennial AHS enumeration consists of two separate surveys, a national survey (AHS-N) and a metropolitan area survey (AHS-MS). AHS-N publishes results every two years; AHS-MS publishes results from each metro on a rotational basis.

The purpose of this study is to determine the feasibility of creating state-level estimates using data collected for the AHS. The AHS sample was designed to select sample cases from and produce estimates in each of the four United States census-defined regions (census-region). Public users of the AHS have requested estimates at smaller levels, to include the nine United States census-defined divisions (census-division) and the fifty states (plus the District of Columbia).

2. AHS Sample Design

The universe of interest for the AHS is the residential housing units in the United States. These residential housing units must exist at the time the survey is conducted. This excludes group quarters and businesses. The AHS-N utilizes a two-stage sample design. The AHS-MS selects housing units directly in a one-stage sample design.

2.1 First Stage—Sample Area Selection

The first stage involved selection of sample areas, also known as Primary Sampling Units (PSUs). A PSU is defined as a county² or a collection of two or more counties for smaller populated counties. Currently, the Census Bureau divides the United States' 3,143 counties into 1,987 PSUs. If a PSU had over 100,000 housing units at the time of selection, we considered the PSU “self-representing”, and included with certainty into the first stage selection process. For 2013, there were 170 self-representing (SR) PSUs. We then allocated the remaining PSUs (each having less than 100,000 housing units) into one of four census-regions, based on each PSU's state: Northeast, North Central, South, and West. Within each census-region, these PSUs were grouped together based on similarities in economics and demographics; nationwide, 224 such groups were created. Each group contained at least two PSUs. The largest group contained 25 PSUs. There was an average of around 7-8 PSUs per group. From each group one PSU was selected proportional to the number of housing units in the PSU to represent all PSUs in the group. We refer to these 224 PSUs as the non self-representing (NSR) PSUs. We spread the first stage sample over 394 PSUs covering 878 counties with coverage in four census-regions in all 50 states and the District of Columbia.

2.2 Second Stage—Housing Unit Selection

The second stage involved selection of housing units within the SR PSUs and selected NSR PSUs. In addition, the second stage was the only stage of sample selection applied to any metro areas included in each year's AHS-MS sample. The AHS consisted of the following types of housing units in the sampled PSUs:

- Housing units selected from the 1980 Census—Systematic sample so that every unit had approximately a 1 in 2,000 chance of being selected;
- New construction in areas requiring building permits—Before each enumeration, a sample of building permits was selected;

² For the purposes of this paper, “county” is used to define a county, parish, borough, or independent city.

- Housing units missed in the 1980 Census—A special study identified addresses missed or inadequately defined in the 1980 Census, a sample of these units was selected;
- Other housing units added since the 1980 Census—To include extra units added in buildings or manufactured/mobile home parks where AHS already had sample units, a sample of these extra units was selected;
- Housing units selected from the 2000 Census—Housing units captured in the 2000 Census not previously captured, a sample of these housing units was selected.

Note: For the 2015 AHS redesign, we selected a sample from the Census Bureau’s 2015 Master Address File (MAF), updates for new construction in future enumerations will be selected from the MAF.

3. AHS Estimation

Each housing unit in the AHS represented itself and approximately 2,000 other housing units. The exact number it represented is its “weight”. We calculated each sampled unit’s weight in six steps (below) to create a final weight. AHS aggregated final weights to create estimates.

- Basic weight—This weight reflected the housing unit’s probability of selection, with rare exceptions the AHS-N weight was 2,148. The AHS-MS basic weight varied from metro to metro, depending on total housing units in each metro area;
- Sample adjustment—An adjustment was made to the units to account for the introduction of housing units selected from the 2000 Census, the addition of supplemental sample in five metro areas (Chicago, Detroit, New York, Northern New Jersey, and Philadelphia), and for an oversample of subsidized housing units. This adjustment was made to ensure the additional sample would not inflate the national housing unit estimates;
- Noninterview adjustment—Adjustment made for refusals and occupied units where no one was home; did not include units that the Census Bureau could not locate. We assumed that units missed were similar in some ways to units interviewed by the AHS. By grouping similar units into cross-classified “cells”, the earlier weight of each interviewed case in each cross-classified cell was multiplied by the following factor:

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

- PSU adjustment—Sample cases located in NSR PSUs were adjusted so their weights were representative of the United States rather than within-PSU weight by multiplying the earlier weight by the following factor:

$$\frac{\text{Total Housing units in all areas that could have been chosen as NSR PSUs}}{\text{Housing units estimated from the AHS sample of NSR PSUs}}$$

- New construction adjustment—Adjusted each sampled case for known deficiencies in sampling new construction weight. Using an independent estimate, Census Bureau’s Survey of Construction and Survey of Manufactured Home Placements, AHS sampled units similar to each other were grouped into cross-classified “cells” and the following factor was created and multiplied to the earlier weight:

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

- Demographic adjustment—To ensure comparability among Census Bureau surveys, an independent estimate from the Census Bureau’s Population Division was used create the following factor and multiplied to the earlier weight:

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

Demographic adjustment was done for Black/non-Black groups, Hispanic/non-Hispanic groups, and an adjustment for Vacant housing units.

AHS raked adjustments for new construction and demographics until marginal totals were consistent.

4. AHS Variance Estimation

AHS used replication methods for variance estimation; there were two types of replication variance estimation techniques, Balanced Repeated Replication (BRR) and Successive Differences Replication (SDR). BRR was used for NSR PSU cases and SDR was used for SR PSU cases. For each sample case, the unbiased weight (basic weight X sample adjustment) was multiplied by replicate factors to produce unbiased replicate weights. We further adjusted these unbiased weights through the noninterview adjustment, PSU adjustment (for NSR PSU cases), new construction adjustment, and demographic adjustment (to include necessary raking) just as the full sample was weighted. By applying all of the weighting adjustments to each replicate, the final replicate weights reflected the impact of the weighting adjustments on the variance. Replicate factors using a combination of BRR and SDR measured the two-stage variance in NSR PSUs and the one-stage variance in SR PSUs, respectively. Refer to McCarthy (1966), Wolter (1985; chapter 3), and Särndal et al. (1992, section 11.4). In SR PSU strata and for the AHS-MS sample, no PSUs were selected so BRR was not appropriate. Since the variation of SR PSUs comes entirely from selecting units within each PSU, the SDR technique was used. Refer to Wolter (1985), Fay and Train (1995), and Ash (2014). The AHS created 160 replicates for variance estimation.

5. State-Level Estimation

We evaluated several methods for producing state-level estimates. The quality of each method was determined based on the coefficient of variation (CV); the larger the standard error relative to the mean, the less reliable the estimate. We considered CVs over 15% as too high. State-level estimates were also compared the Census Bureau’s Population

Division Housing Unit 2013 Totals as a parity check to see if estimates are in the ballpark.

Table 5.0.1: POP Division's 2013 HU Totals
 "True" values used to compare to State-Level Estimates
 (Used for Tables 5.1.1, 5.2.2, 5.3.1)

State	HU Count	State	HU Count
Alabama	2,189,938	Montana	485,771
Alaska	307,399	Nebraska	807,034
Arizona	2,892,325	Nevada	1,186,879
Arkansas	1,329,681	New Hampshire	616,537
California	13,790,495	New Jersey	3,578,141
Colorado	2,247,238	New Mexico	905,135
Connecticut	1,487,982	New York	8,126,026
Delaware	412,009	North Carolina	4,394,261
District of Columbia	302,947	North Dakota	339,313
Florida	9,047,612	Ohio	5,123,997
Georgia	4,109,896	Oklahoma	1,682,256
Hawaii	526,257	Oregon	1,684,035
Idaho	676,192	Pennsylvania	5,565,157
Illinois	5,289,423	Rhode Island	461,640
Indiana	2,809,447	South Carolina	2,158,652
Iowa	1,349,617	South Dakota	370,291
Kansas	1,239,706	Tennessee	2,840,914
Kentucky	1,936,565	Texas	10,255,642
Louisiana	1,990,835	Utah	1,006,106
Maine	723,128	Vermont	323,920
Maryland	2,404,012	Virginia	3,412,460
Massachusetts	2,813,536	Washington	2,928,217
Michigan	4,525,141	West Virginia	879,449
Minnesota	2,368,711	Wisconsin	2,633,330
Mississippi	1,283,165	Wyoming	265,438
Missouri	2,719,001	National	132,802,859

Before exploring state-level methods, we examined each state's SR and NSR housing unit percentages. We looked at 1980 Census percentages because we selected the design in 1985, i.e. these percentages were all that was available to us for the old design. We already knew that SR PSUs were by definition self-representing, thus could be used wholly when included in the state-level estimate. SR PSUs represented only themselves, and not other PSUs. Additionally, SR PSUs did not cross state lines. On the other hand, not only did selected NSR PSUs represent their own state, but they also represented other PSUs within the census-region not necessarily in the same state. Knowing this, we assumed that those states with higher percentages of housing units in SR PSUs would have a more accurate state-level estimate.

Table 5.0.2: 1980 SR vs NSR Percentages by State

State	% SR	% NSR	State	% SR	% NSR	State	% SR	% NSR
Alabama	39.2%	60.8%	Kentucky	42.1%	57.9%	North Dakota	0.0%	100.0%
Alaska	0.0%	100.0%	Louisiana	50.6%	49.4%	Ohio	68.2%	31.8%
Arizona	74.7%	25.3%	Maine	55.0%	45.0%	Oklahoma	50.1%	49.9%
Arkansas	22.0%	78.0%	Maryland	86.6%	13.4%	Oregon	52.1%	47.9%
California	91.4%	8.6%	Massachusetts	93.6%	6.4%	Pennsylvania	75.6%	24.4%
Colorado	64.5%	35.5%	Michigan	63.4%	36.6%	Rhode Island	100.0%	0.0%
Connecticut	100.0%	0.0%	Minnesota	54.9%	45.1%	South Carolina	47.5%	52.5%
Delaware	62.3%	37.7%	Mississippi	16.1%	83.9%	South Dakota	0.0%	100.0%
District of Columbia	100.0%	0.0%	Missouri	53.3%	46.7%	Tennessee	62.0%	38.0%
Florida	79.5%	20.5%	Montana	0.0%	100.0%	Texas	61.4%	38.6%
Georgia	46.0%	54.0%	Nebraska	30.3%	69.7%	Utah	62.6%	37.4%
Hawaii	75.4%	24.6%	Nevada	56.1%	43.9%	Vermont	26.6%	73.4%
Idaho	0.0%	100.0%	New Hampshire	64.6%	35.4%	Virginia	56.0%	44.0%
Illinois	74.4%	25.6%	New Jersey	98.3%	1.7%	Washington	62.4%	37.6%
Indiana	47.0%	53.0%	New Mexico	31.9%	68.1%	West Virginia	22.1%	77.9%
Iowa	20.0%	80.0%	New York	85.5%	14.5%	Wisconsin	42.2%	57.8%
Kansas	36.0%	64.0%	North Carolina	38.2%	61.8%	Wyoming	0.0%	100.0%
						National	65.5%	34.5%

States with very high estimates of percent SR may yield good state-level estimates, particularly California, Connecticut, District of Columbia, Maryland, Massachusetts, New Jersey, New York and Rhode Island. This is because SR PSU weights were restricted to only represent housing units in-state, thus we concluded that states with high percentages of SR housing units as robust. Selected NSR PSU weights tend to represent housing units in-state as well as other housing units out-of-state. Because (not selected) NSR PSUs “borrow” weights from other (selected) NSR PSUs (most likely in other states), it would be safe to assume that states with higher percentages of NSR housing units may yield less reliable estimates due to the dynamic nature of shuffling these weights around. As a result of these effects, we chose to focus on states at or near 75%. In addition to the states already listed, we also considered other candidates—Arizona, Florida, Hawaii, Illinois, and Pennsylvania.

5.1 Method 1—Summing Weighted Sample Cases by State

The first method simply used the current weights without any changes. The caveat here was that since the NSR portion of the sample was allocated by census-region (and not by state), this method would produce estimates with large variances because the first stage was not stratified by state.

Table 5.1.1: AHS Estimates of Total Housing Units (HUs)

State	# HUs	% diff	CV	State	# HUs	% diff	CV
Alabama	2,289,524.2	4.5%	26.7%	Montana	555,458.2	14.3%	53.4%
Alaska	304,288.3	-1.0%	115.4%	Nebraska	914,924.1	13.4%	25.6%
Arizona	3,058,825.2	5.8%	24.4%	Nevada	793,971.4	-33.1%	4.4%
Arkansas	1,022,535.9	-23.1%	32.8%	New Hampshire	383,230.1	-37.8%	10.3%
California	13,759,175.2	-0.2%	3.1%	New Jersey	3,305,923.2	-7.6%	4.7%
Colorado	2,643,196.6	17.6%	20.0%	New Mexico	1,036,050.5	14.5%	24.2%
Connecticut	1,514,348.7	1.8%	4.2%	New York	7,495,204.7	-7.8%	4.9%
Delaware	529,108.4	28.4%	52.8%	North Carolina	4,106,719.6	-6.5%	16.5%
District of Columbia	277,385.6	-8.4%	6.9%	North Dakota	275,035.9	-18.9%	67.2%
Florida	9,275,990.6	2.5%	7.3%	Ohio	5,611,276.8	9.5%	11.7%
Georgia	3,616,034.7	-12.0%	15.2%	Oklahoma	1,730,955.8	2.9%	24.3%
Hawaii	320,938.0	-39.0%	7.6%	Oregon	1,908,518.3	13.3%	25.0%
Idaho	542,317.2	-19.8%	62.0%	Pennsylvania	6,286,118.2	13.0%	8.8%
Illinois	5,119,199.6	-3.2%	7.1%	Rhode Island	533,717.5	15.6%	7.9%
Indiana	2,464,530.1	-12.3%	16.7%	South Carolina	2,383,305.7	10.4%	28.1%
Iowa	1,430,151.6	6.0%	27.1%	South Dakota	551,856.6	49.0%	52.3%
Kansas	966,352.6	-22.0%	28.3%	Tennessee	2,624,422.9	-7.6%	14.9%
Kentucky	2,469,631.8	27.5%	20.9%	Texas	10,141,169.0	-1.1%	7.7%
Louisiana	2,026,449.3	1.8%	22.5%	Utah	723,442.5	-28.1%	28.5%
Maine	946,023.1	30.8%	38.5%	Vermont	341,105.5	5.3%	78.7%
Maryland	2,567,502.7	6.8%	13.2%	Virginia	3,129,995.1	-8.3%	10.8%
Massachusetts	2,913,724.9	3.6%	8.2%	Washington	2,917,768.0	-0.4%	20.9%
Michigan	4,850,039.8	7.2%	8.3%	West Virginia	1,157,608.2	31.6%	41.3%
Minnesota	2,240,630.7	-5.4%	16.6%	Wisconsin	2,657,070.6	0.9%	15.0%
Mississippi	1,330,491.3	3.7%	37.9%	Wyoming	264,113.3	-0.5%	71.4%
Missouri	2,524,778.1	-7.1%	14.6%	National	132,832,136	0.0%	

Table 5.1.1 shows estimated housing unit counts for each state (# HUs) by aggregating all (SR and NSR) HU weights within each state. As predicted, these estimates produced large variances (many CV values exceeding the threshold of 15%). The % diff column shows how the estimated total (# HUs) differed from Pop Divisions 2013 HU Totals (each state's "control total", refer to Table 5.0.1). Positive % diff values show that each state's estimate was greater than its control total; conversely, negative % diff values show that each state's estimate was less than its control total. Estimates in many states appeared significantly different—notice large values of % diff (positive and negative) in many states (South Dakota, Hawaii, New Hampshire, Nevada, etc.). Highlighted states indicated those states where SR percentages were at or near 75% and higher (refer to Table 5.0.2). Notice Hawaii's % diff value. AHS estimation undercounted Hawaii's housing unit total by 39%. Further investigation found that for the 1985-2013 design of the AHS, Hawaii had no NSR PSUs selected; essentially Hawaii's entire estimate was based on Hawaii's SR PSU. Since there existed no NSR sample in Hawaii, we concluded that Hawaii's state-level estimate was inaccurate, and always will be for this particular design; that is, Hawaii did not yield a feasible state-level AHS estimate.

5.2 Method 2—Synthetic Estimation Using Proportions

SR PSUs were selected with certainty; HU weights from SR PSUs contributed only to the state that the SR PSU resided. NSR PSUs were stratified into one of four census-regions. Though NSR PSUs were stratified together by census-region with other NSR PSUs, oftentimes PSUs in these groups were located in several different states. Due to old records, we only knew the PSUs selected into the sample. The strata that these PSUs represented, as well as the other PSUs in those strata were not recorded; thus, for this study, we synthetically distributed each selected PSU by census-region into each state. We used the 1980 Census Housing Unit counts to come up with state proportions within each census-region.

For example, using the 1980 Census, California had an SR housing unit count of 8.5 million and an NSR housing unit count of 799,000. The West Region had an SR housing unit count of 12.6 million and an NSR housing count of 4.5 million, respectively. The respective by census-region percentages of California's housing unit counts were 67.2% (SR) and 17.9% (NSR). In a nutshell, California comprised of 67.2% of housing units in the entire West region's SR areas (think urban areas like Los Angeles and San Francisco) and 17.9% of housing units in the West region's NSR areas (think of rural areas). Being that counties belonging to SR areas were selected with certainty, these counties represented only themselves; their entire weighted values were factored into California's estimate. However, due to ambiguity in how NSR PSUs were selected (remember a PSU is defined as a combination of two or more counties), and what other census-region PSUs were represented with each selected PSU, we proportionally distributed each NSR PSU. That is, the total weighted housing estimate for each NSR PSU within the West Region was distributed among the 13 states in the census-region (California would have received 17.9% of each NSR PSU's weighted estimate).

Table 5.2.1 provides an example of a hypothetical NSR PSU in the West Region with a weighted count of 10,000 housing units would have been distributed.

Note: Each NSR sample case was distributed proportionally among all the states within census-region, regardless of the state where the sample case was located.

Table 5.2.1 Proportional Distribution of a Hypothetical NSR PSU with a Weighted Count of 10,000 HUs

State	Region NSR Proportion	Slice= 10,000 X Proportion	State	Region NSR Proportion	Slice= 10,000 X Proportion
Alaska	0.0364	364	Nevada	0.0334	334
Arizona	0.0629	629	New Mexico	0.0772	772
California	0.1786	1,786	Oregon	0.1159	1,159
Colorado	0.0949	949	Utah	0.0410	410
Hawaii	0.0184	184	Washington	0.1419	1,419
Idaho	0.0839	839	Wyoming	0.0421	421
Montana	0.0734	734	REGION	1.0000	10,000

This method was also applied to replicate weights. Sample cases selected in SR PSUs had each of their entire replicate weights applied to its respective state. Sample cases selected in NSR PSUs had each of their replicate weights distributed in the same fashion as their sample weight. CVs were then calculated using replicate weights.

Table 5.2.2: AHS Synthetic State-Level Estimates

State	# HUs	% diff	CV	State	# HUs	% diff	CV
Alabama	2,159,107.9	-1.4%	2.5%	Montana	581,645.5	19.7%	4.4%
Alaska	288,330.4	-6.2%	4.4%	Nebraska	838,476.1	3.9%	2.8%
Arizona	2,501,887.7	-13.5%	3.6%	Nevada	1,058,426.0	-10.8%	3.0%
Arkansas	1,455,573.4	9.5%	3.1%	New Hampshire	577,437.8	-6.3%	6.6%
California	14,150,520.0	2.6%	1.3%	New Jersey	3,257,597.3	-9.0%	2.5%
Colorado	2,202,867.2	-2.0%	2.0%	New Mexico	887,249.0	-2.0%	3.8%
Connecticut	1,514,348.7	1.8%	4.2%	New York	7,804,213.3	-4.0%	1.1%
Delaware	357,667.8	-13.2%	5.2%	North Carolina	4,213,085.6	-4.1%	1.4%
District of Columbia	277,385.6	-8.4%	6.9%	North Dakota	328,838.3	-3.1%	1.6%
Florida	8,480,598.7	-6.3%	2.0%	Ohio	5,070,091.6	-1.1%	1.2%
Georgia	3,808,649.4	-7.3%	1.6%	Oklahoma	1,875,559.4	11.5%	2.2%
Hawaii	466,492.4	-11.4%	5.0%	Oregon	1,909,375.5	13.4%	2.7%
Idaho	664,274.5	-1.8%	4.4%	Pennsylvania	6,093,937.9	9.5%	1.4%
Illinois	5,375,692.8	1.6%	1.0%	Rhode Island	533,717.5	15.6%	7.9%
Indiana	2,962,585.0	5.5%	1.5%	South Carolina	2,099,272.1	-2.8%	3.0%
Iowa	1,491,281.7	10.5%	1.8%	South Dakota	351,998.0	-4.9%	1.6%
Kansas	1,329,720.6	7.3%	2.5%	Tennessee	2,824,242.3	-0.6%	2.3%
Kentucky	2,225,034.3	14.9%	2.4%	Texas	10,361,033.1	1.0%	2.0%
Louisiana	2,195,523.5	10.3%	2.4%	Utah	846,568.6	-15.9%	3.4%
Maine	717,631.1	-0.8%	5.4%	Vermont	318,876.8	-1.6%	7.1%
Maryland	2,293,341.4	-4.6%	2.5%	Virginia	3,506,350.2	2.8%	1.5%
Massachusetts	2,901,635.5	3.1%	3.1%	Washington	2,937,131.4	0.3%	1.9%
Michigan	4,506,534.3	-0.4%	0.9%	West Virginia	1,136,918.6	29.3%	4.0%
Minnesota	2,315,898.9	-2.2%	1.8%	Wisconsin	2,565,224.3	-2.6%	1.5%
Mississippi	1,409,487.5	9.8%	2.4%	Wyoming	333,294.5	25.6%	4.4%
Missouri	2,469,504.7	-9.2%	1.9%	National	132,832,136	0.0%	

In Table 5.2.2, all CVs looked good (<15%). However, many states' actual estimates were off target when comparing to POP Division's 2013 HU Totals, particularly West Virginia, Wyoming, and Montana. There were 14 states where the synthetic estimate differs from the POP Division's 2013 HU Totals estimate by more than 10%. This suggested that distributing NSR PSUs might yield inconsistent resulting estimates, particularly in those states where there are larger percentages of NSR PSUs. The highlighted states represented those states where around 75% or more housing units were in SR areas (excepting Hawaii, refer to Table 5.0.2).

5.3 Method 3—Creating State-Level Estimates Adjusting to Individual State Control Totals for Housing Units and Population

The third method we examined was to take each individual state's selected sample and apply state-level control totals for housing unit totals and population with raking ratio adjustments. For Method 1, we simply calculated estimates with the original national weight. For each sample case, we formulated a base weight (i.e. some sort of "take-every" value), and applied various weighting factors to the base weight to come up with a final weight (refer to Section 3 "AHS Estimation" of this paper). Once we obtained a

final weight for each interviewed sample case (respondent), we then controlled (at the census-region level) our estimates to a more reliable source, particularly the Census Bureau's Population Division totals for Housing Units (HUs) and Population (POP). Using current methodology, we raked our estimates to these two control totals. For state-level estimation, we performed the controlling and raking process at the state level (rather than the census-region level). We limited the scope of the study and focus only on the 12 states containing SR percentages near 75% and above (refer to Table Table 5.0.2-- Arizona, California, Connecticut, District of Columbia, Florida, Illinois, Maryland, Massachusetts, New Jersey, New York, Pennsylvania, and Rhode Island). Due to large NSR populations in the other states, as well as the unknowns involved in NSR sample selection, we deemed the other 39 states non-feasible for state-level estimation (optimistically we hope to include more states in a future study). Also, (the 160) replicate weights were provided for these 12 states; they went through the same process of state-level controlling and raking. We calculated CVs for overall Housing Unit totals using replicate weights. After adjusting and raking each of the 12 states using state-level control totals for housing units and population, along with doing the same for each of the 160 replicates, Table 5.3.1 (below) illustrated total housing unit values.

Table 5.3.1: AHS State-Level Estimates Using State-Level HU and POP Controls
Total Housing Units

State	Wtd Sum	Control	%Diff	CV
Arizona	2,771,040.8	2,892,325	-4.2%	0.0%
California	13,393,213.4	13,790,495	-2.9%	0.1%
Connecticut	1,488,831.7	1,487,982	0.1%	0.1%
District of Columbia	310,185.4	302,947	2.4%	3.7%
Florida	9,051,617.7	9,047,612	0.0%	0.0%
Illinois	5,293,559.3	5,289,423	0.1%	0.1%
Maryland	2,325,205.3	2,404,012	-3.3%	0.0%
Massachusetts	2,801,902.1	2,813,536	-0.4%	0.0%
New Jersey	3,578,854.5	3,578,141	0.0%	0.0%
New York	8,127,836.6	8,126,026	0.0%	0.0%
Pennsylvania	5,565,365.6	5,565,157	0.0%	0.0%
Rhode Island	462,446.2	461,640	0.2%	0.4%

CVs looked excellent, all CV percentages were lower than we considered good (CV < 15%). Percent differences looked good.

Now that we have utilized state-level raked control totals to our estimates, we can next check the feasibility of estimates of a few subdomains of AHS. For this study, we checked the subdomains for Total Occupied, Total Vacant, Seasonal, New Construction, and Mobile Homes ("general" subdomains taken from Table 1-1 of the 2013 AHS Publication). Table 5.3.2 reviews the final estimates for each of these subdomains.

Table 5.3.2: AHS State-Level Estimates Using State-Level HU and POP Controls
 Total Occupied, Total Vacant, Seasonal, New Construction, Mobile Homes

	ESTIMATES					
	Total HUs	Total Occupied	Total Vacant	Seasonal	New Const	Mobile Homes
Arizona	2,771,040.8	2,543,883.4	227,157.5	57,680.9	129,277.8	193,389.9
California	13,393,213.4	12,972,389.0	420,824.4	72,000.8	334,019.6	346,110.5
Connecticut	1,488,831.7	1,389,837.2	98,994.5	9,588.1	16,776.0	10,579.9
District of Columbia	310,185.4	308,350.4	1,835.0	152.9	0.0	0.0
Florida	9,051,617.7	7,975,209.6	1,076,408.1	286,948.7	7,013.4	694,284.2
Illinois	5,293,559.3	5,012,406.9	281,152.5	16,704.5	9,356.8	87,185.7
Maryland	2,325,205.3	2,152,810.2	172,395.1	51,200.7	2,837.3	22,008.2
Massachusetts	2,801,902.1	2,601,670.2	200,232.0	37,100.6	42,377.5	18,805.0
New Jersey	3,578,854.5	3,326,556.6	252,297.8	76,017.6	38,507.2	29,664.8
New York	8,127,836.6	7,542,079.2	585,757.4	128,406.4	121,929.4	172,782.0
Pennsylvania	5,565,365.6	5,013,765.0	551,600.6	72,146.8	55,510.3	188,579.7
Rhode Island	462,446.2	445,548.4	16,897.8	989.3	298.1	9,964.8

What constituted a reasonable sample size within a subdomain to yield a good estimate? Determination of appropriate sample size for subdomains can be answered by observing CVs associated (using replicate weights) with each state's subdomain estimate.

Table 5.3.3: State-Level CVs for Each AHS Subdomain Estimate
Total Occupied, Total Vacant, Seasonal, New Construction, Mobile Homes

	CVs					
	Total HUs	Total Occupied	Total Vacant	Seasonal	New Const	Mobile Homes
Arizona	0.0%	1.9%	20.9%	17.6%	21.7%	26.8%
California	0.1%	0.5%	14.3%	17.7%	8.7%	8.8%
Connecticut	0.1%	1.8%	24.1%	71.0%	43.1%	51.8%
District of Columbia	3.7%	5.0%	285.0%			
Florida	0.0%	0.9%	6.7%	11.5%	51.5%	7.0%
Illinois	0.1%	1.0%	17.5%	30.9%	30.6%	20.4%
Maryland	0.0%	1.9%	23.7%	31.7%	105.2%	32.5%
Massachusetts	0.0%	2.0%	26.0%	36.5%	27.5%	48.6%
New Jersey	0.0%	0.8%	9.8%	13.9%	13.9%	28.3%
New York	0.0%	0.8%	9.9%	14.3%	10.7%	15.6%
Pennsylvania	0.0%	1.1%	10.2%	30.7%	16.2%	11.6%
Rhode Island	0.4%	2.8%	62.9%	104.8%		74.8%

On Table 5.3.3, I highlighted CVs greater than 15%. CVs for the subdomain Total Occupied looked good for all states. However, other large subdomain CVs eliminated Connecticut, District of Columbia, Illinois, Maryland, Massachusetts, and Rhode Island (all CVs for other subdomains exceed 15% or were missing due to no sample cases). We were left with the following states:

Table 5.3.4: “Eligible” States and Subdomain CVs

	CVs					
	Total HUs	Total Occupied	Total Vacant	Seasonal	New Const	Mobile Homes
California	0.1%	0.5%	14.3%	17.7%	8.7%	8.8%
Florida	0.0%	0.9%	6.7%	11.5%	51.5%	7.0%
New Jersey	0.0%	0.8%	9.8%	13.9%	13.9%	28.3%
New York	0.0%	0.8%	9.9%	14.3%	10.7%	15.6%
Pennsylvania	0.0%	1.1%	10.2%	30.7%	16.2%	11.6%

Due to insufficient sample size and the effect on the accuracy of estimates, we plan to suppress “grayed-out” subdomains.

6. Concluding Remarks

With limited information on a 30-year old region-based (four census-regions) sample design, we were able to determine that five states, using limited subdomains, would yield feasible state-level estimates. The new 2015 sample design will be division-based using the nine Census-defined divisions. This new design will allow NSR cases to be more closely aligned to states. Additionally, we will have more information, specifically information on what NSR PSUs that selected NSR PSUs are representing. From this information, we can better allocate weights within those NSR PSUs for the synthetic estimation method. With the new 2015 sample redesign, we are currently revisiting how we control our estimates to other surveys. Improvements to these methods will also yield better results with state-level estimation. We hope that as our methodology for state estimates improve, we will be able to provide more state-level weights.

References

- Alexih, L., and Corea, J. (1998). *Three National Surveys: A Statistical Assessment and State Tabulations*. Department of Health and Human Services/ASPE, Prepared Under Contract #HHS-100-012, Delivery Order #22.
- Ash, S. (2014). *Using Successive Difference Replication for Estimating Variances*, Proceedings of Survey Research Methods Section, Miami, FL: American Statistical Association, pp. 3534-3548.
- Cochran, W. (1977). *Sampling Techniques*, 3rd ed. John Wiley and Sons, New York, NY.
- Fay, R., and Train, G. (1995). *Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties*. Proceedings of the Government Statistics Section, Alexandria, VA: American Statistical Association, pp. 154-159.
- Hogg, R., and Craig, A. (1995). *Introduction to Mathematical Statistics*, 5th ed. Prentice-Hall, Englewood Cliffs, NJ.
- McCarthy, P. (1966). *Replication: An Approach to the Analysis of Data from Complex Surveys. Vital Health and Statistics*, U.S. Department of Health, Education and Welfare, Public Health Service, Washington, DC. DHEW Publication No. (PHS) 79-1269, Series 2 – Number 14.
- Särndal C., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York, NY.
- Wolter, K. (1985). *Introduction to Variance Estimation*. Springer-Verlag, New York, NY.