**UNITED STATES DEPARTMENT OF COMMERCE**
**Bureau of the Census**
Washington, DC 20233-0001

2014 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS14-RER-02

DSSD 2014 AMERICAN COMMUNITY SURVEY RESEARCH MEMORANDUM SERIES ACS14-R-01

MEMORANDUM FOR ACS Research and Evaluation Advisory Group

From: Patrick J. Cantwell ***signed 1/9/14***
Division Chief, Decennial Statistical Studies Division

Prepared by: Daniel Sommers
American Community Survey Sample Design Branch
Decennial Statistical Studies Division

Subject: Evaluating the Impact of the 2011 Sample Reallocation for the American Community Survey

Attached is the American Community Survey Research and Evaluation report, "Evaluating the Impact of the 2011 Sample Reallocation for the American Community Survey".

If you have any questions about this report, please contact Daniel Sommers at (301-763-3714) or Steven Hefter at (301-763-4082).

Attachment

cc: ACS Research and Evaluation Workgroup

: T. Hughes (ACSO)
M. Asiala (DSSD)
T. Tersine
K. King
S. Hefter
K. Cyffka
P. Davis
D. Keathley

JANUARY 2, 2014

# Evaluating the Impact of the 2011 Sample Reallocation for the American Community Survey

Daniel Sommers and Steven Hefter
Decennial Statistical Studies Division

United States Census Bureau

## I. Background

This document presents an evaluation performed to assess the impact of the 2011 Sample Reallocation for the American Community Survey (ACS) and Puerto Rico Community Survey (PRCS). In 2010 research was conducted to investigate the need to reallocate (or redistribute) the ACS and PRCS sample[1]. Making this determination was a starting point for this research. If in fact there need to reallocate, we would develop methodology to modify the ACS sample design in a way that would allow for a more equitable distribution of sample. The goal being to minimize differences in estimate reliability across areas (primarily census tracts) while maintaining the robust estimates for the larger geographies.

As a replacement for the Census Long Form questions, the ACS is tasked with providing reliable estimates for all levels of geography. From the sample design perspective, the chief mechanism to control reliability is the amount of sample that is allocated to these areas. The sample design is constrained by an annual sample size that is generally unchanged from year to year, instead of a potentially more optimal design whereby the sampling rate is held constant.

Since the inception of the survey, the sample design has been based largely on the sample design for the Census 2000 Long Form. The ACS design used seven sampling strata based on census tract and functioning governmental unit size categories. The design also further stratifies using housing unit response rates. Five of the sampling strata were based on size categories. Two of these strata were split into high and not high housing unit response rate strata, resulting in the seven strata. The goals of the initial research were to 1) further refine these sampling strata, since they contain a large proportion of all addresses on the sampling frame, and utilize more size categories which would provide more control over the sample allocation for these areas, and 2) provide an overall increase to the sampling rates in the smallest geographic areas.

The reallocation research was completed in 2010, and the results were used to determine how the sample design would be modified[2]. This modification was first implemented with the 2011 sample selection. The research that is presented in this document used the 2010 and 2011 samples to evaluate the impact of the reallocation and to answer the following question:

- *Did the implementation of the reallocation of the sample achieve the intended goals?*

Table 1 shows a mapping of the seven original strata, used from 2005-2010, to the current 16 strata. Using this table we can see how the universe of valid housing unit addresses is distributed across the original strata. The second stratum ($200 < GUMOS \leq 800$), and last four strata (TRACTMOS dependent and High Response strata) have the largest proportions of addresses, and therefore we concentrated efforts to refine the sampling rates within these categories.

[1] Throughout the document, references to the ACS will included the PRCS unless otherwise stated.
[2] The documentation for this research can be found in "ACS10-R-5 Equal Coefficient of Variation Plan for Reallocating the ACS Sample".

Table 1. Mapping between Current Strata (2011 – present) and Original Strata (2005 – 2010)

| Valid 2011 Addresses | New Stratum Thresholds | Original Stratum Thresholds | Valid 2010 Addresses |
|---|---|---|---|
| 135,956,068 | - Total | Total - | 135,417,246 |
| 1,119,838 | 0 < GUMOS[3] ≤ 200 | 0 < GUMOS ≤ 200 | 1,141,727 |
| 1,974,467 | 200 < GUMOS ≤ 400 | 200 < GUMOS ≤ 800 | 6,048,816 |
| 3,903,333 | 400 < GUMOS ≤ 800 | | |
| 3,387,738 | 800 < GUMOS ≤ 1,200 | 800 < GUMOS ≤ 1,200 | 3,376,207 |
| 261,546 | 0 < TRACTMOS[4] ≤ 400 | TRACTMOS ≤ 2,000 | 23,931,290 |
| 4,099,459 | 400 < TRACTMOS ≤ 1,000 | | |
| 17,023,174 | 1,000 < TRACTMOS ≤ 2,000 | | |
| 106,440 | 0 < TRACTMOS ≤ 400 H.R. | TRACTMOS ≤ 2,000 High Response (H.R.) | 32,128,879 |
| 3,299,480 | 400 < TRACTMOS ≤ 1,000 H.R. | | |
| 28,385,704 | 1,000 < TRACTMOS ≤ 2,000 H.R. | | |
| 17,968,445 | 2,000 < TRACTMOS ≤ 4,000 | 2,000 < TRACTMOS | 19,603,128 |
| 2,469,315 | 4,000 < TRACTMOS ≤ 6,000 | | |
| 1,124,979 | 6,000 < TRACTMOS | | |
| 38,892,020 | 2,000 < TRACTMOS ≤ 4,000 H.R. | 2,000 < TRACTMOS High Response (H.R.) | 49,187,199 |
| 7,517,133 | 4,000 < TRACTMOS ≤ 6,000 H.R. | | |
| 4,422,997 | 6,000 < TRACTMOS H.R. | | |

Source: 2010 and 2011 American Community Survey Governmental Unit Measure of Size files

## II. Methodology

The evaluation of the reallocation used many of the same methods as was used in the research conducted in 2010. However, we used selected sample and actual sampling rates from 2010 and 2011 instead of simulated sample sizes. In 2011 there was an increase to a 3.54 million annual sample size beginning with the June 2011 panel. For 2011 we selected sample at both the 2.9 and 3.54 million levels, therefore for this research we simply used data from the 2.9 million selected sample, in order to make appropriate comparisons. We again looked at a theoretical five-year coefficient of variation (CV) for a typical tract by tract size category[5]. As in the previous research, we examined how the distribution of sample changed under the new design, again by tract size category. We also looked at the impact of additional weight variation due to the additional sampling rates on total variance at the county level.

We want to use tracts in the reallocation research since they are the primary unit of geography for the ACS sample design as well as ACS estimates. Also, tracts completely cover the United States (U.S.) and Puerto Rico (P.R.) and have useful size and geographic characteristics. Tracts are, in general, smaller than counties, but not so small that we cannot use them as an effective basis for measuring estimate reliability. All governmental entities within the U.S. and P.R. are made up tracts, or parts of tracts, including small entities such as cities, towns, school districts, and American Indian areas. ACS sampling

[3] This is a governmental unit or tract measure of size (GUMOS) that is assigned to every block in the U.S. and P.R. A block's GUMOS is the minimum of the set of all measures of size that correspond to the governmental entities that contain the block. For more information, please see the ACS Design and Methodology Report, Chapter 4.
[4] This is a tract measure of size (TRACTMOS).
[5] Coefficient of Variation is defined as the standard error of an estimate divided by the estimate.

rates are assigned at the census block level. Every block is fully contained within a tract, and most blocks are also contained in a governmental entity. Small governmental areas were specifically targeted by the Census Long Form sample, which used census tract and governmental unit size as the basis for its design, and since the ACS is the replacement for the Census Long Form, the ACS sample design also uses these measures. Since the ACS sample design is geographic based and heavily dependent upon census tract, we can measure the effectiveness of our reallocation using tracts, and make inferences about how well the design impacts small communities.

In addition to the methods above, we expanded the evaluation further. We included an examination of other geographic areas such as places, unified school districts[6], American Indian areas (AINDN), as well as a separate examination within American Indian areas of Alaska Native Village Statistical Areas (ANVSA). We also examined county level CVs and used this in combination with the impact of weight variation at the county level in order to have a good measure of the reallocation's impact on counties. For places, school districts, and AINDN areas, we calculated a five-year theoretical CV for each individual area and then examined the distribution of CVs within and across size class. This is particularly important for looking at the smaller areas in these geographic entities since their estimates are found only in the ACS five year data release. A limitation of this method is that the largest class size category for each of these geographic entities contains a very large proportion of the total housing unit addresses. We left out the largest size class categories for these areas mainly due to their high size variability. For these larger areas we can use the results from the tract analysis to infer how the sample design is performing. A more detailed description of the methodology used in the research is presented below.

A. Evaluating the sample reallocation using the same methodology and metrics from the research that was conducted in 2010[7]. As in that research we calculated the following:

   1. A five year CV using the following expression:

$$CV = \frac{\sqrt{[(1-f)(pq)] / n}}{p}$$

where

DF = Design Factor = 2.1
f = the five year sampling fraction
p = 0.10
q = 0.90

[6] Unified school districts were examined since they contain roughly 120 million housing units, covering most of the country. We could have also examined elementary and secondary districts which are a small subset of school districts where elementary and secondary are separately defined, but these areas only contain roughly 8 million housing units.

[7] The documentation for this research can be found in draft form "ACS10-R-5 Equal Coefficient of Variation Plan for Reallocating the ACS Sample".

n = an estimate of the five year person sample
= average tract size × f × pph
- *tract size = estimated number of occupied housing units in a tract*
- *pph = persons per household = 2.6*

The five year sampling fraction for the 2011 component of the research was a weighted combination of the production rates from M11 and S11. Along with this rate, we used the estimated number of occupied housing units for every size category (used to calculate the average tract size) using data from the M11 edited Master Address File MAF extract for the five year sample calculation[8]. The design factor (DF) of 2.1 is an average of three ACS poverty design factors[9]. We also produced results for 2010 using a weighted combination of the M10 and S10 rates along with counts from the S10 edited MAF extract to calculate the estimated number of occupied housing units, and in turn the five year sampling rate and CV.

2. The impact of additional weight variation on total variance of the estimates at the county level. This involved calculating the squared weight for every sample case in 2010 and 2011:

$$w^2 = (1/\text{rate})^2$$

where rate is the probability of selection assigned to the sample case for the sample year. We summed the squared weights to the county level, and calculated the ratio of the sum of squared weights from the two sample years 2011 to 2010 (new versus original designs):

$$R_{\text{county}} = \sum_{\text{sample 2011}} w^2 \Big/ \sum_{\text{sample 2010}} w^2$$

We then examined the distribution of the square root of $R_{\text{county}}$ by county size class category. This distribution was compared to the 2010 research results.

B. In addition to using the same methods from the previous research, the distribution of CVs by size class categories for several geographic type groups were evaluated. We examined:

1. A theoretical CV for every tract, place, unified school district, AINDN, ANVSA and county (these are referred to as design areas). We used a similar CV calculation as in Section II.A.1, except that we used the selected sample for each individual area rather than a projected hypothetical five-year sample size. We also used the same block level occupancy rate (multiplied by *pph*) from

[8] The edited Master Address File extract is the sampling frame for the ACS.
[9] This design factor was also used in "2009 ACS09-R-1 Analysis of ASR and SSR to Achieve Specified Levels of Reliability (03-16-10)" to examine effects on the ACS survey estimates resulting from an increased annual sample size. We use it here to maintain consistency across research projects dealing with sample size, reallocation and estimate reliability at the tract level.

Section II.A.1 to calculate an expected five-year person sample size. The base was the estimated number of occupied housing units, also multiplied by *pph*. We looked at the distribution of CVs for the selected design areas by size class category, and compared the two years. We examined the median and mean CV for every size category, as well as the first and third quartiles.

2. We looked at the magnitude of the sample shift between 2010 and 2011 by tract size category. This was done by using the same tract size categories previously defined, looking at the distribution of sample for new versus original design.

## III. Assumptions and Limitations

Many of the assumptions in this research are similar to the research conducted in 2010. Here we also include additional assumptions, since this evaluation examines other geographic areas not originally included in the initial research. The assumptions in our research are based on Census 2000 tracts, since the ACS sample design is heavily dependent upon tracts and the five-year ACS estimates are analogous to the smaller area Census Long Form estimates. The measures of estimate reliability are person level measures, and therefore we make assumptions based on occupied housing units and persons per household. We also make assumptions that are time dependent. The assumptions for this study are as follows:

A. The average tract size will remain relatively stable over time (for this research, tract size is a measure of the number of estimated occupied housing units within the tract).

B. The average tract size for a particular tract size category represents all tracts in that category. This assumption is used when examining estimate reliability for a typical tract.

C. The number of persons per household will remain stable across all tracts over time. We use this assumption when examining five-year estimate reliability.

D. There is a 100 percent interview rate and the delete rate is zero for the sample selected. This is important since we want to capture general properties of the sample design, and not survey performance.

E. We assume the largest geographic areas, in general, have a tract size distribution that is similar to the national tract size distribution. We can then use the results from the tract analysis for these areas.

## IV. Results

A. Comparison of Tracts and Sampling Strata Rates, 2010 and 2011

Results from the 2010 research were used to determine the changes to the sample design that were implemented for the 2011 sample. The conclusions were based on how well the sample design performed across tract size categories. The measure of

design performance was how well the design minimized the differences in estimate reliability across tract size categories.

Table 2 shows sampling rates for 2010 and 2011, respectively, by sampling stratum and compares these to the rates produced in the reallocation research from 2010. It is important again to note that all rates and sample sizes use a targeted annual sample size of 2.9 million. The final rates used for 2010 and 2011 are a weighted average of rates from the main and supplemental processes[10]. The sampling rates under the research estimates columns are derived from simulations using data only from the M11 sampling frame. As we can see, the rates from the simulations and the production years are approximately the same.

Table 2 also shows the results for the theoretical five year CV that was calculated from the research simulations and compares this to a theoretical five year CV calculated from the weighted sample year rates. Here we used stratum average size (size being the number of estimated occupied housing units) as the typical geography for each stratum to calculate the theoretical five year CV. For strata five through 12 we used the average tract size. Average place size was used for strata one through four. We also included other geographic entities for the 2010 research, but here we see that this makes little difference. As would be expected with similar rates, the two CV calculations (research simulation and production) are about the same. This is because the rates and CVs in the table are dependent upon several components, the two most influential in terms of affecting change (since here we fix the sample size) are the distribution of valid addresses across the sampling strata and the amount of influence supplemental sampling had on the total annual sample for the year.

We can conclude from this that the 2010 research gave us a good basis for representing the sample year 2010 (given the changes in the frame between M10 and S10 due to address canvassing). Additionally, it was a good basis to use for simulating what we expected to see in 2011, since the changes to the sampling frame between the phases of S10, M11, and S11 were relatively stable and predictable. In general, Table 2 shows that the 2010 simulations were very close to what we saw in the 2010 and 2011 sample years. Therefore the reallocation is producing the sampling rates and projected CVs that we expected to see.

[10] Each sample year the ACS produces two sets of sampling rates, one set for main and one for supplemental. Therefore we calculate a single annual sampling rate for each size category as the (proportion of valid addresses in size category from main)×(main sampling rate) + (proportion of valid addresses in size category from supplemental)×(supplemental sampling rate).

Table 2. Comparison of Results from 2010 Reallocation Research and Production Data

| 2011 Production Results | | | Research | | 2010 Production Results | | Research | |
|---|---|---|---|---|---|---|---|---|
| New Stratum[11] | Sampling Rate | CV | Sampling Rate | CV | Sampling Rate | CV | Sampling Rate | CV |
| 1 | 15.00% | 22% | 15.00% | 23% | 10.00% | 40% | 10.00% | 40% |
| 2 | 10.00% | 22% | 10.00% | 23% | 6.67% | 26% | 6.75% | 26% |
| 3 | 7.00% | 23% | 7.00% | 22% | | | | |
| 4 | 3.84% | 26% | 3.83% | 26% | 3.33% | 28% | 3.38% | 28% |
| 5 | 4.77% | 41% | 4.79% | 41% | 2.18% | 65% | 2.23% | 64% |
| 6 | 4.37% | 44% | 4.40% | 44% | 2.00% | 69% | 2.05% | 69% |
| 7 | 3.83% | 29% | 3.83% | 29% | 2.20% | 41% | 2.23% | 40% |
| 8 | 3.53% | 30% | 3.52% | 30% | 2.03% | 41% | 2.05% | 41% |
| 9 | 2.33% | 28% | 2.32% | 28% | 2.20% | 29% | 2.23% | 29% |
| 10 | 2.14% | 29% | 2.14% | 29% | 2.03% | 30% | 2.05% | 30% |
| 11 | 1.36% | 28% | 1.37% | 28% | 1.62% | 26% | 1.67% | 25% |
| 12 | 1.26% | 29% | 1.26% | 28% | 1.49% | 27% | 1.54% | 26% |
| 13 | 0.82% | 28% | 0.82% | 29% | 1.61% | 19% | 1.67% | 19% |
| 14 | 0.76% | 29% | 0.75% | 29% | 1.49% | 20% | 1.54% | 20% |
| 15 | 0.48% | 28% | 0.48% | 28% | 1.60% | 15% | 1.67% | 14% |
| 16 | 0.44% | 28% | 0.44% | 28% | 1.48% | 15% | 1.54% | 15% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

In addition to looking at the sampling rates and theoretical CVs by sampling strata, we also wanted a more generalized summary of how the reallocation was performing. By looking only at tract size categories, along with a theoretical CV, we wanted to capture the overall trend in the sample shift. Tables 3 and 4 give these more generalized summaries. Total tracts and average tract size use data from the M10 and M11 sampling frames (since we use estimated occupied housing units for the size categories). We did not use S10 and S11 data because we do not place tracts into size categories during those phases for assigning sampling rates; instead, we use categories from the main phase. In both tables we again see results similar to those from the 2010 research.

From Table 3 we can also see that the changes in the frame between 2010 and 2011 did not have a large impact on the average tract size. One assumption from the 2010 research is that the typical tract size for each size category will not change over time. Here we see that typical changes in the frame between years should not be a concern when using average tract size to measure sample design performance.

[11] We ordered the new strata by sampling rate and numbered them 1 through 16.

Table 3. A Theoretical Five Year CV by Tract Size Category - 2010 and 2011 Production Data

| Tract Size Thresholds (Tract MOS) | Total Tracts 2010 | Total Tracts[12] 2011 | Average 2010 Tract Size | Average 2011 Tract Size | CV 2010 | CV 2011 | Research Simulation (original) | Research Simulation (current) |
|---|---|---|---|---|---|---|---|---|
| Totals | 64,439 | 64,412 | | | | | | |
| 0 – 400 | 983 | 1,071 | 292 | 292 | 66% | 42% | 66% | 41% |
| 401 – 1,000 | 9,729 | 9,943 | 772 | 768 | 41% | 30% | 41% | 30% |
| 1,001 – 2,000 | 31,401 | 30,875 | 1,486 | 1,486 | 30% | 29% | 29% | 29% |
| 2,001 – 4,000 | 20,086 | 19,997 | 2,626 | 2,642 | 27% | 29% | 26% | 29% |
| 4,001 – 6,000 | 1,787 | 1,935 | 4,680 | 4,689 | 20% | 28% | 19% | 29% |
| 6,001 + | 453 | 591 | 8,358 | 8,368 | 15% | 28% | 15% | 28% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

In Table 4 we see the shift in sample going from 2010 (original stratification) to 2011 (current stratification). Table 4 does not include sample cases that are ungeocoded, since only geocoded addresses can be assigned to a tract[13]. There is an increase in sample for the tracts in the smaller size categories and a decrease in the larger areas. This shift is what drives the equalizing of CVs across size categories seen in Table 3. The tracts in the larger areas still have enough sample to produce reliable estimates, while those in the smaller areas show a very large expected improvement in estimate reliability.

Table 4. Distribution of Sample for the 2010 and 2011 Sample Years

| Tract Size Thresholds (Tract MOS) | 2010 Sample | 2011 Sample | Research Simulation (original) | Research Simulation (current) |
|---|---|---|---|---|
| Totals | 2,818,026 | 2,871,664 | 2,837,551 | 2,859,978 |
| 0 – 400 | 11,442 | 21,829 | 11,330 | 21,127 |
| 401 – 1,000 | 253,342 | 390,985 | 253,427 | 391,918 |
| 1,001 – 2,000 | 1,313,477 | 1,422,468 | 1,299,697 | 1,426,162 |
| 2,001 – 4,000 | 1,018,250 | 917,179 | 1,027,744 | 906,799 |
| 4,001 – 6,000 | 153,567 | 90,151 | 161,550 | 87,250 |
| 6,001 + | 67,948 | 29,052 | 83,803 | 26,722 |

Ungeocoded sample cases cannot be included in these totals, since they are not assigned to a tract.

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

B. Examining other Geographic Areas and Impact from Reallocation

When looking at CVs for places, school districts, AINDN, ANVSA and counties, we see a similar trend for the smaller areas as in the tract analysis. On average, CVs for the smaller size categories are lower. Below, we look at the analysis for these areas. We do not have a comparison to the research done in 2010, since for that research we did not examine these areas in as much detail. One of the assumptions made in the

[12] Note that the differences between 2010 and 2011 are larger than normal due to results that include the Census Address Canvassing operation (M11) and without the operation (M10).

[13] Ungeocoded addresses are addresses on the Edited Master Address File extracts that do not have census tract and block codes.

2010 research was that the increased rates that are held constant across years for the smaller geographic entities would translate into improved CVs for small places, school districts, etc. However, for this research we looked at the distribution of CVs for these areas, attempting to capture an overall picture of the changes we expect to see in their estimate reliability.

C. Places

First, we look at the change in CVs moving from 2010 to 2011. In Table 5 we can see that the areas in the smallest two size classes (0 – 200 and 201 – 400) have a drop in the theoretical five year CV. This makes sense, since the largest increases to the sampling rates were in the two corresponding sampling strata. We also see a change in the third (401 – 800) and fourth (801 – 1,200) size categories, which correspond to the third and fourth strata. However, there is an increase in the CV within the fourth (801 – 1,200) and fifth (1,201 – 2,000) size classes, and a decrease within the larger size class. This is likely a result of the change in sampling rates moving from the first three, fixed rate strata to strata nine and 10 (there is a significant drop in rates when moving across these size categories). We left out the largest places where nationally, the majority of estimated occupied housing units are located. However, here we are interested in the smaller areas, and the overall trend in the change of CVs for these size categories is consistent with our expectations.

Table 5. Mean and Median Theoretical Five Year CVs for Places by Size Class (2010 - 2011)

| Place Size Thresholds (Estimated Occupied HU) | Total 2010 Places | Estimated Occupied HUs 2010 | Total 2011 Places | Estimated Occupied HUs 2011 | Median CV | | Mean CV | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2010 | 2011 | 2010 | 2011 |
| Totals | 22,894 | 26,351,604 | 24,930 | 27,149,030 | | | | |
| 0 – 200 | 6,930 | 676,044 | 8,196 | 779,647 | 38% | 23% | 41% | 29% |
| 201 – 400 | 3,636 | 1,053,255 | 3,996 | 1,153,883 | 34% | 24% | 37% | 30% |
| 401 – 800 | 3,769 | 2,173,655 | 3,950 | 2,276,393 | 25% | 24% | 30% | 31% |
| 801 – 1,200 | 2,102 | 2,069,910 | 2,176 | 2,136,360 | 30% | 27% | 32% | 31% |
| 1,201 – 2,000 | 2,216 | 3,445,212 | 2,317 | 3,591,092 | 32% | 33% | 32% | 34% |
| 2,001 – 4,000 | 2,486 | 7,036,904 | 2,500 | 7,103,512 | 24% | 25% | 24% | 26% |
| 4,001 – 6,000 | 1,114 | 5,459,412 | 1,148 | 5,618,589 | 18% | 19% | 18% | 20% |
| 6,001 – 8,000 | 641 | 4,437,212 | 647 | 4,489,554 | 15% | 15% | 15% | 16% |

Totals row only includes Places that have 8,000 or fewer estimated occupied housing units

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

We also looked at the distribution of the CVs for places by size class. Table 6 compares the first and third quartile, as well as the median CV for each size class by sample year.

Table 6. Comparison of 2010 - 2011 Median, First, and Third Quartile CVs (Places)

| Place Size Thresholds (Estimated Occupied HU) | Year | Q1 | Median | Q3 |
|---|---|---|---|---|
| 0 – 200 | 2010 | 32% | 38% | 46% |
| | 2011 | 19% | 23% | 29% |
| 201 – 400 | 2010 | 31% | 34% | 38% |
| | 2011 | 22% | 24% | 27% |
| 401 – 800 | 2010 | 23% | 25% | 30% |
| | 2011 | 21% | 24% | 33% |
| 801 - 1,200 | 2010 | 27% | 30% | 34% |
| | 2011 | 25% | 27% | 34% |
| 1,201 - 2,000 | 2010 | 29% | 32% | 35% |
| | 2011 | 29% | 33% | 38% |
| 2,001 - 4,000 | 2010 | 22% | 24% | 27% |
| | 2011 | 22% | 25% | 30% |
| 4,001 - 6,000 | 2010 | 17% | 18% | 19% |
| | 2011 | 17% | 19% | 21% |
| 6,001 - 8,000 | 2010 | 14% | 15% | 16% |
| | 2011 | 14% | 15% | 18% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

D. American Indian/Alaska Native Areas[14]

Sample design enhancements between 2010 and 2011 also had an impact on American Indian areas. Many of these areas are small, so as a result of the reallocation, they receive, in general, higher initial sampling rates. Tables 7 and 8 provide the same summary presented for places. As expected, there is significant improvement in the theoretical five year CV for areas in the smaller size categories.

Table 7. Mean and Median Theoretical Five Year CVs for AINDN by Size Class (2010 - 2011)

| AINDN Size Thresholds (Estimated Occupied HU) | Total 2010 AINDN | Estimated Occupied HUs 2010 | Total 2011 AINDN | Estimated Occupied HUs 2011 | Median CV | | Mean CV | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2010 | 2011 | 2010 | 2011 |
| Totals | 536 | 289,136 | 580 | 312,786 | | | | |
| 0 – 200 | 351 | 23,736 | 368 | 24,448 | 40% | 23% | 40% | 23% |
| 201 – 400 | 54 | 14,784 | 64 | 16,776 | 27% | 16% | 28% | 17% |
| 401 – 800 | 34 | 19,371 | 40 | 22,880 | 22% | 17% | 21% | 16% |
| 801 – 1,200 | 24 | 22,524 | 27 | 26,405 | 19% | 17% | 20% | 16% |
| 1,201 – 2,000 | 24 | 37,304 | 30 | 47,961 | 15% | 12% | 16% | 13% |
| 2,001 – 4,000 | 36 | 99,127 | 39 | 108,774 | 12% | 11% | 13% | 11% |
| 4,001 – 6,000 | 8 | 38,229 | 8 | 37,815 | 8% | 8% | 9% | 8% |
| 6,001 – 8,000 | 5 | 34,061 | 4 | 27,727 | 8% | 8% | 8% | 8% |

Totals row only includes AINDN that have 8,000 or fewer estimated occupied housing units

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

[14] These theoretical five year CVs do not take into account the personal visit full follow-up procedure that began in August, 2011 for most AINDN, including all ANVSA. All mail and telephone non-responding cases now go directly to the personal visit mode of data collection.

Table 8. Comparison of 2010 - 2011 Median, First, and Third Quartile CVs (AINDN)

| AINDN Size Thresholds (Estimated Occupied HU) | Year | Q1 | Median | Q3 |
|---|---|---|---|---|
| 0 – 200 | 2010 | 33% | 40% | 46% |
| | 2011 | 19% | 23% | 26% |
| 201 – 400 | 2010 | 25% | 27% | 31% |
| | 2011 | 14% | 16% | 21% |
| 401 – 800 | 2010 | 18% | 22% | 25% |
| | 2011 | 10% | 17% | 21% |
| 801 - 1,200 | 2010 | 17% | 19% | 26% |
| | 2011 | 8% | 17% | 23% |
| 1,201 - 2,000 | 2010 | 13% | 15% | 20% |
| | 2011 | 9% | 12% | 15% |
| 2,001 - 4,000 | 2010 | 10% | 12% | 15% |
| | 2011 | 7% | 11% | 14% |
| 4,001 - 6,000 | 2010 | 7% | 8% | 11% |
| | 2011 | 6% | 8% | 10% |
| 6,001 - 8,000 | 2010 | 8% | 8% | 10% |
| | 2011 | 6% | 8% | 9% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

We do not see the same increase in CVs across the 1,200 and 4,000 size classes as we saw with places. This is likely due to how sampling rates are determined for AINDN. For these areas, the block level proportion of Census 2010 respondents that identified themselves as American Indian is used as a second factor in determining a block's stratum. This second component allows AINDN areas to be sampled, in general, at higher rates[15]. Examining the quartiles, we see overlap in the range of CVs across these size classes and comparable CVs across sample year.

E. Alaska Native Village Statistical Areas (ANVSA)[16]

The ACS has focused on improving estimate reliability in ANVSAs[17]. The reasoning is that these areas traditionally have had high proportions of non-mailable addresses). This means that large portions of these addresses are only eligible for the Computer Assisted Person Interviewing (CAPI) mode of data collection. We want to examine the impact the reallocation has had on these areas. Tables 9 and 10 show theoretical CVs for ANVSAs for 2010 and 2011. The CVs across size categories follow a similar pattern as the overall group of AINDN areas. The increased CV for the 2,000 to 4,000 group is most likely attributed to the lower sampling rate for stratum 11 in 2011. However, the overall trend is what was expected: we see large improvements in estimate reliability for ANVSAs in the small size categories.

[15] The final measure of size (MOS) for these areas is a summation of this block level estimate: (estimated number of occupied housing units) × (proportion of Census 2010 respondents that identified themselves as Native American alone or in combination with other census defined race groups). This tends to lower the MOS, resulting in higher sampling rates.

[16] There is some overlap of ANVSA and areas identified as remote Alaska (these are hard to reach areas of Alaska which the ACS refers to as remote Alaska). Here we include all ANVSA.

[17] The 116 ANVSAs is a subset of the 536 American Indian areas.

Table 9. Mean and Median Theoretical Five Year CVs for ANVSA by Size Class (2010 - 2011)

| ANVSA Size Thresholds (Estimated Occupied HU) | Total 2010 ANVSA | Estimated Occupied HUs 2010 | Total 2011 ANVSA | Estimated Occupied HUs 2011 | Median CV | | Mean CV | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2010 | 2011 | 2010 | 2011 |
| Totals | 116 | 28,119 | 129 | 36,049 | | | | |
| 0 – 200 | 95 | 9,695 | 103 | 10,981 | 41% | 23% | 41% | 23% |
| 201 – 400 | 13 | 3,104 | 13 | 3,252 | 26% | 14% | 27% | 16% |
| 401 – 800 | 1 | 626 | 2 | 1,072 | 16% | 10% | 16% | 10% |
| 801 – 1,200 | 3 | 2,556 | 4 | 3,750 | 18% | 12% | 16% | 12% |
| 1,201 – 2,000 | 2 | 3,029 | 4 | 5,883 | 22% | 13% | 22% | 15% |
| 2,001 – 4,000 | 1 | 2,345 | 2 | 4,446 | 11% | 21% | 11% | 21% |
| 4,001 – 6,000 | 0 | 0 | 0 | 0 | NA | NA | NA | NA |
| 6,001 – 8,000 | 1 | 6,764 | 1 | 6,665 | 6% | 6% | 6% | 6% |

Totals row only includes ANVSA that have 8,000 or fewer estimated occupied housing units

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

Table 10. Comparison of 2010 - 2011 Median, First, and Third Quartile CVs (ANVSA)

| ANVSA Size Thresholds (Estimated Occupied HU) | Year | Q1 | Median | Q3 |
|---|---|---|---|---|
| 0 – 200 | 2010 | 34% | 41% | 46% |
| | 2011 | 19% | 23% | 26% |
| 201 – 400 | 2010 | 23% | 26% | 28% |
| | 2011 | 13% | 14% | 17% |
| 401 – 800 | 2010 | 16% | 16% | 16% |
| | 2011 | 9% | 10% | 11% |
| 801 - 1,200 | 2010 | 13% | 18% | 19% |
| | 2011 | 7% | 12% | 17% |
| 1,201 - 2,000 | 2010 | 20% | 22% | 24% |
| | 2011 | 11% | 13% | 19% |
| 2,001 - 4,000 | 2010 | 11% | 11% | 11% |
| | 2011 | 11% | 21% | 32% |
| 4,001 - 6,000 | 2010 | NA | NA | NA |
| | 2011 | NA | NA | NA |
| 6,001 - 8,000 | 2010 | 6% | 6% | 6% |
| | 2011 | 6% | 6% | 6% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

## F. School Districts[18]

We examined unified school districts by looking at the theoretical CVs across district size categories using the same methods used for places and American Indian areas.

[18] We looked only at unified school districts, since these hold the largest number of housing units compared to other school district types.

Table 11. Mean and Median Theoretical Five Year CVs for School Districts by Size Class (2010 - 2011)

| School District -Unified - Size Thresholds (Estimated Occupied HU) | Total 2010 Unified School Districts | Estimated Occupied HUs 2010 | Total 2011 Unified School Districts | Estimated Occupied HUs 2011 | Median CV | | Mean CV | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2010 | 2011 | 2010 | 2011 |
| Totals | 8,144 | 21,811,385 | 8,040 | 21,767,367 | | | | |
| 0 – 200 | 252 | 27,442 | 238 | 26,464 | 32% | 20% | 34% | 21% |
| 201 – 400 | 462 | 140,136 | 425 | 129,972 | 29% | 20% | 29% | 20% |
| 401 – 800 | 1,032 | 613,756 | 986 | 585,911 | 22% | 19% | 22% | 19% |
| 801 – 1,200 | 847 | 842,458 | 837 | 830,762 | 21% | 19% | 22% | 19% |
| 1,201 – 2,000 | 1,357 | 2,138,748 | 1,374 | 2,162,228 | 21% | 19% | 22% | 20% |
| 2,001 – 4,000 | 2,136 | 6,209,550 | 2,138 | 6,244,799 | 19% | 17% | 19% | 18% |
| 4,001 – 6,000 | 1,203 | 5,900,331 | 1,182 | 5,802,520 | 16% | 16% | 16% | 16% |
| 6,001 – 8,000 | 855 | 5,938,964 | 860 | 5,984,711 | 14% | 14% | 14% | 14% |

Totals row only includes Unified School Districts that have 8,000 or fewer estimated occupied housing units

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

Table 12. Comparison of 2010 - 2011 Median, First, and Third Quartile CVs (School Districts)

| School District -Unified- Size Thresholds (Estimated Occupied HU) | Year | Q1 | Median | Q3 |
|---|---|---|---|---|
| 0 – 200 | 2010 | 28% | 32% | 38% |
| | 2011 | 18% | 20% | 24% |
| 201 – 400 | 2010 | 26% | 29% | 32% |
| | 2011 | 18% | 20% | 22% |
| 401 – 800 | 2010 | 20% | 22% | 24% |
| | 2011 | 17% | 19% | 20% |
| 801 - 1,200 | 2010 | 19% | 21% | 24% |
| | 2011 | 16% | 19% | 21% |
| 1,201 - 2,000 | 2010 | 18% | 21% | 25% |
| | 2011 | 15% | 19% | 23% |
| 2,001 - 4,000 | 2010 | 16% | 19% | 22% |
| | 2011 | 14% | 17% | 21% |
| 4,001 - 6,000 | 2010 | 14% | 16% | 18% |
| | 2011 | 13% | 16% | 18% |
| 6,001 - 8,000 | 2010 | 13% | 14% | 15% |
| | 2011 | 12% | 14% | 16% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

A majority of housing unit addresses in school districts fall into the larger size categories. We assume these large areas show a similar distribution as we saw in Table 3 for tracts.

The smaller size categories show improvement in estimate reliability between 2010 and 2011. In general the reallocation was meant to improve estimate reliability for the smaller geographic entities, and these results show this improvement.

G. Counties

We examined counties by size class categories, focusing on the smallest counties. These counties were expected to benefit the most from the reallocation. Additionally, many counties also are expected to receive an additional reduction in variance due to a decrease in weight variation. This aspect of the reallocation is addressed in the next section. The results for counties are shown in Table 13. We see a general improvement for the smallest areas and a slight improvement across the remaining areas.

Table 13. Mean and Median Theoretical Five Year CV for Counties by Size Class (2010 - 2011)

| County Size Thresholds (Estimated Occupied HU) | Total 2010 Counties | Estimated Occupied HUs 2010 | Total 2011 Counties | Estimated Occupied HUs 2011 | Median CV | | Mean CV | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2010 | 2011 | 2010 | 2011 |
| Totals: | 1,296 | 5,105,123 | 1,285 | 5,076,710 | | | | |
| 0 – 200 | 5 | 543 | 5 | 544 | 39% | 20% | 40% | 19% |
| 201 – 400 | 26 | 8,354 | 23 | 7,399 | 27% | 20% | 28% | 20% |
| 401 – 800 | 47 | 28,784 | 46 | 27,546 | 21% | 18% | 21% | 17% |
| 801 - 1,200 | 75 | 75,248 | 77 | 77,396 | 19% | 16% | 19% | 16% |
| 1,201 - 2,000 | 137 | 219,506 | 135 | 217,265 | 16% | 14% | 17% | 15% |
| 2,001 - 4,000 | 393 | 1,178,605 | 390 | 1,174,531 | 15% | 13% | 15% | 13% |
| 4,001 - 6,000 | 332 | 1,633,209 | 333 | 1,641,044 | 13% | 12% | 13% | 12% |
| 6,001 - 8,000 | 281 | 1,960,874 | 276 | 1,930,985 | 12% | 11% | 11% | 11% |

Totals row only includes Counties that have 8,000 or fewer estimated occupied housing units

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

Table 14. Comparison of 2010 - 2011 Median, First, and Third Quartile CVs (Counties)

| County Size Thresholds (Estimated Occupied HU) | Year | Q1 | CV Median | Q3 |
|---|---|---|---|---|
| 0 – 200 | 2010 | 34% | 39% | 46% |
| | 2011 | 18% | 20% | 20% |
| 201 – 400 | 2010 | 24% | 27% | 30% |
| | 2011 | 17% | 20% | 22% |
| 401 – 800 | 2010 | 18% | 21% | 23% |
| | 2011 | 14% | 18% | 20% |
| 801 - 1,200 | 2010 | 17% | 19% | 21% |
| | 2011 | 13% | 16% | 18% |
| 1,201 - 2,000 | 2010 | 14% | 16% | 20% |
| | 2011 | 12% | 14% | 17% |
| 2,001 - 4,000 | 2010 | 12% | 15% | 18% |
| | 2011 | 10% | 13% | 16% |
| 4,001 - 6,000 | 2010 | 11% | 13% | 15% |
| | 2011 | 10% | 12% | 14% |
| 6,001 - 8,000 | 2010 | 10% | 12% | 13% |
| | 2011 | 9% | 11% | 12% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

H. Impact of Weight Variation on Total Variance

We wanted to examine the potential effect the new strata may have on total variance at the county level. By increasing the number of strata, we increased the number of sampling rates and also introduced the possibility of introducing more county level variation due to additional weights. In order to assess this, we looked at the distribution of the square root of the ratio (R) of the sum of the squared weights from the 2011 sample to the sum of the squared weights from 2010 sample. Similar work was done in the 2010 research using simulated samples. When comparing Table 15 to the results in 2010 a similar pattern of increasing weight variation from smaller counties to larger counties is observed. The majority of counties fall into the range between 0.8 and 1.2. Ideally, counties would have a ratio of 1, which would translate to no additional weight variation for the county due to the reallocation. The fact that most counties fall into this range is a positive result. Counties with $\sqrt{R} \geq 1.2$ show that there is expected to be, roughly, a greater than 20 percent increase on total variance due to additional weight variation. Large counties tend to have more tracts, and also a larger range of tract and geographic entity sizes. With increased strata, we would expect these results for the larger counties. We also see 1,589 counties with $\sqrt{R} < 1$. We expect these counties to have less weight variation under the new allocation. Many of these counties are also small, so these are areas that are generally receiving more sample as a result of the reallocation. This translates into another component of improved estimate reliability for small areas over the previous design.

Table 15. Distribution of County Level Ratios, 2011 to 2010 (2.9 million sample size)

| County Size (2011 Valid Addresses) | $\sqrt{R} < 0.8$ | $0.8 \leq \sqrt{R} < 1$ | $1 \leq \sqrt{R} < 1.2$ | $1.2 \leq \sqrt{R} < 1.4$ | $1.4 \leq \sqrt{R} < 1.6$ | $1.6 \leq \sqrt{R}$ |
|---|---|---|---|---|---|---|
| Totals: | 102 | 1,487 | 1,328 | 168 | 51 | 7 |
| 0 - 1K | 33 | 32 | 0 | 0 | 0 | 0 |
| 1K - 2K | 11 | 123 | 9 | 0 | 0 | 0 |
| 2K - 3K | 19 | 126 | 13 | 0 | 0 | 0 |
| 3K - 4K | 13 | 122 | 24 | 2 | 0 | 0 |
| 4K - 5K | 10 | 122 | 33 | 0 | 0 | 0 |
| 5K - 6K | 7 | 122 | 29 | 2 | 0 | 0 |
| 6K - 7K | 2 | 78 | 38 | 2 | 0 | 0 |
| 7K - 8K | 0 | 86 | 44 | 1 | 1 | 0 |
| 8K - 9K | 1 | 82 | 31 | 1 | 0 | 0 |
| 9K - 10K | 2 | 68 | 40 | 1 | 0 | 0 |
| 10K - 11K | 0 | 56 | 34 | 2 | 0 | 0 |
| 11K - 12K | 1 | 50 | 40 | 6 | 0 | 0 |
| 12K - 13K | 0 | 36 | 36 | 1 | 0 | 1 |
| 13K - 14K | 1 | 37 | 35 | 4 | 1 | 0 |
| 14K - 15K | 1 | 26 | 26 | 1 | 1 | 0 |
| 15K - 20K | 1 | 115 | 151 | 11 | 3 | 2 |
| 20K - 25K | 0 | 74 | 107 | 10 | 3 | 0 |
| 25K - 100K | 0 | 114 | 441 | 70 | 27 | 4 |
| 100K - 500K | 0 | 14 | 175 | 46 | 15 | 0 |
| 500K + | 0 | 4 | 22 | 8 | 0 | 0 |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files
County Level Ratio = $[2011\ \sum_{sample} (w^2)] \div [2010\ \sum_{sample} (w^2)]$, w = (1/bwgt)

Results from the 2010 research gave confirmation of our assumptions about county level CVs; the largest counties already have low CVs, and therefore any relatively small increase in total variance for these areas due to increased weight variation may be acceptable. In order to get a better understanding of this impact, we looked at the five year theoretical CV by size classes. We examined the county level median CV for each size class based on data from both 2010 and 2011 (same CVs used in Table 15). We again wanted an overall summary of the impact from weight variation for all the counties, but we also wanted to look only at the counties where the square root of the ratio, R, was ≥ 1.2 (226 counties). Table 16 shows the results of this analysis. We again see that counties are distributed as in Table 15, but here we have collapsed the columns containing counties with the largest ratios. Next to this column we show the median theoretical CV for counties in each size class, followed by the median CV for only the group of counties in the size class that have $\sqrt{R} \geq 1.2$. Once we calculated the CV for each of these 226 counties where we expect to see an increase in weight variation and the general assumptions in this research, we calculated a new median, using the new CV for each size category.

Table 16. Distribution of County Level Ratios with Impact on Theoretical Five Year CV

| | | | | | All Counties | | Restricted to $1.2 \leq \sqrt{R}$ (226 Counties) | | |
|---|---|---|---|---|---|---|---|---|---|
| County Size (Valid Addresses) | $\sqrt{R} < 0.8$ | $0.8 \leq \sqrt{R} < 1$ | $1 \leq \sqrt{R} < 1.2$ | $1.2 \leq \sqrt{R}$ | Median CV | | Median CV | | |
| | | | | | 2010 | 2011 | 2010 | 2011 | $\sqrt{R} \times CV_{2011}$ |
| Totals: | 102 | 1,487 | 1,328 | 226 | | | | | |
| 0 - 1K | 33 | 32 | 0 | 0 | 23.9% | 18.1% | NA | NA | NA |
| 1K - 2K | 11 | 123 | 9 | 0 | 17.9% | 15.1% | NA | NA | NA |
| 2K - 3K | 19 | 126 | 13 | 0 | 15.5% | 13.1% | NA | NA | NA |
| 3K - 4K | 13 | 122 | 24 | 2 | 15.4% | 13.4% | 19.5% | 19.9% | 24.2% |
| 4K - 5K | 10 | 122 | 33 | 0 | 14.2% | 12.3% | NA | NA | NA |
| 5K - 6K | 7 | 122 | 29 | 2 | 13.6% | 12.1% | 10.8% | 9.6% | 12.5% |
| 6K - 7K | 2 | 78 | 38 | 2 | 12.4% | 11.5% | 15.2% | 16.5% | 20.7% |
| 7K - 8K | 0 | 86 | 44 | 2 | 12.1% | 11.1% | 15.1% | 18.7% | 25.0% |
| 8K - 9K | 1 | 82 | 31 | 1 | 11.6% | 10.6% | 15.8% | 18.2% | 23.6% |
| 9K - 10K | 2 | 68 | 40 | 1 | 10.9% | 10.1% | 12.3% | 11.3% | 14.1% |
| 10K - 11K | 0 | 56 | 34 | 2 | 10.8% | 9.8% | 13.4% | 15.4% | 20.0% |
| 11K - 12K | 1 | 50 | 40 | 6 | 10.7% | 10.1% | 12.3% | 13.2% | 17.1% |
| 12K - 13K | 0 | 36 | 36 | 2 | 10.2% | 9.7% | 11.7% | 14.3% | 20.6% |
| 13K - 14K | 1 | 37 | 35 | 5 | 10.1% | 9.2% | 12.8% | 15.2% | 19.4% |
| 14K - 15K | 1 | 26 | 26 | 2 | 9.9% | 9.2% | 10.2% | 11.6% | 16.3% |
| 15K - 20K | 1 | 115 | 151 | 16 | 8.6% | 8.3% | 9.3% | 10.5% | 13.7% |
| 20K - 25K | 0 | 74 | 107 | 13 | 7.7% | 7.3% | 8.7% | 9.5% | 12.2% |
| 25K - 100K | 0 | 114 | 441 | 101 | 5.7% | 5.7% | 5.8% | 6.8% | 9.6% |
| 100K - 500K | 0 | 14 | 175 | 61 | 3.0% | 3.1% | 3.4% | 3.8% | 4.9% |
| 500K + | 0 | 4 | 22 | 8 | 1.5% | 1.6% | 1.4% | 1.5% | 1.9% |

Source: 2010 and 2011 Edited Master Address File extracts, and American Community Survey housing unit sample files

Of note in Tables 15 and 16 is that the 226 counties restricted to $\sqrt{R} \geq 1.2$, we see a slight increase in the 2011 median CV for many of the size classes before impact from additional weight variation. Also, there are several counties in the lower size classes that have large $\sqrt{R}$. These smaller counties are shown in Table 17. For these smaller counties, there appears to be a large proportion of the estimated occupied

housing units in the higher strata (the sampling rates for the county under the new stratification). These counties from Table 16 that have a large $\sqrt{R}$ and also have the largest increase in CV without the weight variation adjustment can be seen in Table 18 (which means lower sampling rates). Also, Custer County Colorado and Bethel Census Area both have their occupied housing units split between the base rate and strata one through three. Counties with a large $\sqrt{R}$ that also have the largest increase in CVs before the adjustment tend to also have a large proportion of their housing units in the higher sampling strata. For example, some counties will have several large tracts or entities and very few smaller entities. If this distribution is extreme enough, there will be a general decrease in the sampling rates for the county under the new stratification. Table 18 shows the counties from Table 16 that have both large $\sqrt{R}$ and the largest increase in CV without the weight variation adjustment.

Table 17. Smallest Counties with $\sqrt{R} \geq 1.2$ (12 of the 226)

| | | | | | | Percent of County Estimated Occupied HU by Substrata Groups | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| State | County | Occ. HU (2011) | CV 2010 | CV 2011 | $\sqrt{R}$ | 01-03 | 04-07 | 08-10 | 11 (BR) | 12-14 | 14-16 |
| CO | Custer County | 2,141 | 19% | 19% | 1.22 | 23% | 0% | 0% | 77% | 0% | 0% |
| VA | Rappahannock County | 3,335 | 20% | 21% | 1.21 | 2% | 0% | 36% | 0% | 62% | 0% |
| ID | Teton County | 4,423 | 13% | 13% | 1.31 | 35% | 0% | 0% | 0% | 65% | 0% |
| AK | Bethel Census Area | 4,957 | 8% | 6% | 1.28 | 58% | 0% | 0% | 42% | 0% | 0% |
| TN | Sequatchie County | 5,664 | 17% | 20% | 1.26 | 0% | 0% | 29% | 0% | 71% | 0% |
| GA | Oglethorpe County | 5,960 | 14% | 13% | 1.26 | 11% | 0% | 25% | 0% | 64% | 0% |
| KY | Spencer County | 6,454 | 15% | 17% | 1.40 | 6% | 0% | 0% | 0% | 94% | 0% |
| MS | Stone County | 6,902 | 16% | 20% | 1.28 | 0% | 0% | 18% | 0% | 82% | 0% |
| VA | Floyd County | 6,929 | 16% | 18% | 1.30 | 4% | 0% | 0% | 0% | 96% | 0% |
| GA | Greene County | 7,210 | 12% | 11% | 1.25 | 14% | 13% | 16% | 0% | 57% | 0% |
| CO | Grand County | 7,248 | 10% | 10% | 1.25 | 22% | 24% | 11% | 0% | 42% | 0% |
| UT | Wasatch County | 7,809 | 12% | 13% | 1.36 | 8% | 0% | 0% | 0% | 92% | 0% |

Source: 2010 and 2011 Governmental Unit Measure of Size files, and American Community Survey housing unit sample files

Table 18. Counties (with $\sqrt{R} \geq 1.2$ ) from Table 16 with Largest Increase in CV from 2010 to 2011

| | | | | | | Percent of County Estimated Occupied HU by Substrata Groups | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| State | County | Occ. HU (2011) | CV 2010 | CV 2011 | $\sqrt{R}$ | 01-03 | 04-07 | 08-10 | 11 (BR) | 12-14 | 14-16 |
| GA | Bryan County | 11,232 | 12% | 17% | 1.60 | 0% | 8% | 0% | 25% | 0% | 67% |
| MS | Stone County | 6,902 | 16% | 20% | 1.28 | 0% | 0% | 18% | 0% | 82% | 0% |
| GA | Paulding County | 48,995 | 6% | 10% | 1.76 | 0% | 0% | 0% | 5% | 21% | 74% |
| NC | Hoke County | 17,424 | 11% | 14% | 1.55 | 0% | 0% | 0% | 26% | 17% | 57% |
| GA | Monroe County | 10,538 | 13% | 17% | 1.36 | 1% | 0% | 0% | 0% | 99% | 0% |
| GA | Forsyth County | 62,204 | 6% | 9% | 1.68 | 0% | 0% | 0% | 0% | 33% | 67% |
| TN | Sequatchie County | 5,664 | 17% | 20% | 1.26 | 0% | 0% | 29% | 0% | 71% | 0% |
| GA | Dawson County | 8,420 | 13% | 17% | 1.30 | 0% | 10% | 0% | 25% | 65% | 0% |
| NV | Lyon County | 21,084 | 9% | 12% | 1.57 | 1% | 0% | 9% | 0% | 27% | 64% |
| FL | Flagler County | 42,127 | 7% | 10% | 1.67 | 1% | 2% | 0% | 0% | 31% | 66% |
| GA | Putnam County | 9,653 | 14% | 17% | 1.23 | 0% | 0% | 0% | 35% | 65% | 0% |
| GA | Lumpkin County | 11,806 | 13% | 16% | 1.24 | 0% | 0% | 0% | 0% | 100% | 0% |
| VA | Fluvanna County | 9,457 | 13% | 16% | 1.31 | 0% | 0% | 21% | 23% | 56% | 0% |
| KY | Spencer County | 6,454 | 15% | 17% | 1.40 | 6% | 0% | 0% | 0% | 94% | 0% |
| SD | Lincoln County | 17,481 | 9% | 11% | 1.86 | 3% | 7% | 15% | 0% | 0% | 75% |

Source: 2010 and 2011 Governmental Unit Measure of Size files, and American Community Survey housing unit sample files

## V. Conclusions / Further Research

The original intent of the ACS sample reallocation was to allocate sample in such a way as to minimize the differences in estimate reliability for tract level estimates by size class. When completed in 2010, the research conducted used simulations to find the appropriate allocation scheme to be implemented in 2011. Results from this evaluation show that the production samples selected in 2011 achieved the intent of the reallocation. Using actual rates from 2011, Table 3 shows that tract level theoretical five year CVs have been brought more in line with each other across size categories. Using actual data from the 2011 sampling frame, we see in Table 2 sampling rates that match up with the simulated rates in the original study. Both of these results confirm that the allocation implemented in 2011 achieved the original goals.

Looking at places, we see a large improvement in estimate reliability for the smaller size categories. Table 5 shows a 15 percentage point drop in median CV for the smallest areas, which is consistent with what we expected to see for these areas. The reallocation increased the sampling rates for the smallest governmental entities, keeping them constant from year to year, and these results clearly show the improvement for these smaller areas. American Indian areas (AINDN) also show improved CV estimates for the smaller areas. The smallest size category for these areas in Table 7 shows a 17 percentage point improvement in the theoretical five year CV, while the other size categories also show improvement. Alaska Native Village Statistical Areas, which is a specific subset of AINDNs on which the ACS focuses, as well as unified school districts both show similar patterns of estimate improvement. For all of these design areas, size classes that

correspond to the fixed rate strata show the largest improvements in five year CVs. This is an expected result of the reallocation.

There was the expectation for additional weight variation on total variance, specifically at the county level. The expectation was that counties would have an acceptable small increase in total variance due to the additional strata. Table 15 shows that for most counties, the impact that the reallocation has on total weight variation is in line with what we saw in the previous research using simulated data, and is a positive result for the reallocation. There are some larger counties where the reallocation could increase the total weight variation, however these larger counties already have small CVs, and so we expect any additional weight variation to still produce relatively small CVs. However, there are many counties (1,589 counties) where the weight variation will decrease as a result of the reallocation. This translates into another component of improvement in estimate reliability. Many of the counties that have larger values of $\sqrt{R}$ have a more extreme distribution of valid addresses between the sampling strata, resulting in a large variability of the sampling rates within the county. Tables 17 and 18 show some of these counties, along with how the sample is distributed by sampling strata for these areas. The distribution of sample gives a reflection of how the governmental entities are arranged within each of these counties, and shows that many of them have only a few sampling rates and are often clustered in the smaller rate categories. This explains the large values of the $\sqrt{R}$, including the smaller counties within this subgroup.

Future research could include an examination of the benefits of further stratification, along with impact on design areas. A further analysis of the actual five-year CVs for tracts should be completed once the 2011-2015 ACS estimates are released in 2016.