UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

2012 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS12-RER-24
DSSD 2009 AMERICAN COMMUNITY SURVEY RESEARCH MEMORANDUM SERIES ACS09-R-03

MEMORANDUM FOR	ACS Research and Evaluation Steering Committee
From:	Alfredo Navarro ***signed 8/10/12*** Assistant Division Chief, ACS Statistical Design Decennial Statistical Studies Division
Prepared by:	B. Dale Garrett Variance Estimation and Statistical Support Branch Decennial Statistical Studies Division
Subject:	Improving Weighted Person-Level Estimates from the American Community Survey's Public Use Microdata Sample File

Research was conducted on using an expanded weighting matrix to improve consistency between estimates of the full sample American Community Survey (ACS) and its Public Use Microdata Sample (PUMS). An iterative proportional fitting technique similar to that used by the 2008 ACS was tested on the PUMS file using data collected during 2007. This technique sought to improve estimates of householders and spouses, as well as persons by race, Hispanic origin, sex and age group.

After preliminary review the expanded weighting matrix was applied to the 2009 and later PUMS files. The research presented shows some of the results of tests as well as the impact on the 2009 PUMS estimates.

If you have any questions concerning this report, please contact Dale Garrett (301) 763-1961.

Attachment

cc:
ACS Research and Evaluation Team
Deborah Griffin
Mark Asiala

American Community Survey Research and Evaluation Program
8/10/2012

Improving Weighted Person-Level Estimates from the American Community Survey's Public Use Microdata Sample File

B Dale Garrett
Decennial Statistical Studies Division

United States Census Bureau

Improving Weighted Person-Level Estimates from the American Community Survey's Public Use Microdata Sample File

B. Dale Garrett
Decennial Statistical Studies Division
U.S. Census Bureau

Presented at the August, 2011 Joint Statistical Meetings of the American Statistical Association

Keywords: ACS, weighting, public use microdata

Abstract: The Public Use Microdata Sample (PUMS) files contain records for a subsample of the housing units and persons of the American Community Survey (ACS) annual sample. A weighting process was introduced for the 2009 PUMS that expanded the raking matrix to include more demographic controls and family equalization with the goal of forcing more consistency between the PUMS and the ACS full sample estimates. This paper discusses the preliminary research, the trade-offs of doing the weighting at the state versus Public Use Microdata Area (PUMA) levels, and some of the impact on estimates of the new weighting procedure.

Introduction

The Public Use Microdata Sample (PUMS) subset of the American Community Survey (ACS) publishes actual survey data so that users can perform their own data analysis. The PUMS files are provided for periods that cover one, three and five years. The ACS program began releasing 1-year PUMS files in the year 2000 when the ACS was in its infancy. As the ACS program has expanded, so has the PUMS product. Beginning with the 2005 sample, the ACS moved to a full sample of 250,000 addresses per month and started publishing a full one percent of the ACS universe in the 1-year PUMS file covering 50 states, D.C. and Puerto Rico. For 2007, the ACS published two files, a 1-year PUMS file of data collected during 2007 and a 3-year file covering the three years of data collection 2005 through 2007. For periods ending in 2009, the ACS published a trio of PUMS files, including a 5-year file with data collected during the years 2005 through 2009.

For periods ending in 2009, the PUMS weighting process implemented an expanded weighing control matrix for its housing unit population with an iterative proportional fitting (raking) in order to better agree with the full sample ACS for estimates of spouses, householders and population demographics (race/Hispanic origin, age and sex). The PUMS files ending in 2008 used a simple ratio adjustment for the Public Use Microdata Area (PUMA) population by gender. This paper reviews the research that helped us choose the weighting methodology of the 2009 and future PUMS files, and some of the actual results of this change.

Confidentiality

The confidentiality of the respondent is protected by a number of disclosure avoidance procedures. Personally identifying fields, such as names, addresses and local geography are omitted from the PUMS files. Some housing units are swapped between geographies, and a small amount of noise is added through techniques such as top-coding and bottom-coding of open-ended variables, collapsing of detail for categorical variables, synthetic data and age perturbation. To protect confidentiality, the lowest level of geography on the PUMS file is the PUMA. The PUMA boundaries used through the 2011 PUMS products were formed to include about 100,000 persons or more as of the 2000 Census.

PUMS Sampling and Weighting Overview

Sampling

The PUMS sample is selected for each 1-year PUMS from the full sample of ACS housing and person interviews. The ACS person sample since 2006 has been composed of two groups of persons – those living in housing units (HUs) and those living in Group Quarters (GQs). When PUMS is selected from ACS, the sampling intervals are designed to yield a PUMS sample of one percent of each universe: persons from HUs and persons from GQs within each state. After sampling HUs for PUMS, all persons in a selected HU are retained in the PUMS sample and are assigned the same PUMS sampling interval. The 3-year and 5-year PUMS files are formed by combining the samples from several 1-year PUMS files.

PUMS Weighting before 2009

PUMS weights are formed as the product of the final weight from the full ACS sample, the PUMS subsampling interval and a ratio-adjustment factor. The ratio-adjustment factors bring the PUMS estimates into better agreement with the full sample for selected parameters.

For PUMS files having periods ending in 2008 and earlier, the ratio-adjustment factors were formed as follows: HUs were controlled within PUMA by occupied and vacant, the persons from GQs were controlled at the state level by institutional, noninstitutional and gender, while persons from HUs were controlled at the PUMA level by gender. This paper is concerned with only the persons from HUs weighting methodology. Table 1 shows the 2008 weighting cells used for PUMS ratio adjustments for persons from HUs. The controls were the sum of ACS person weights across all ACS sample persons within each cell.

Table 1. 2008 PUMS Ratio Adjustment Cells for Persons from HUs

	Number of Males	Number of Females
By PUMA		

Source: 2008 PUMS Accuracy of the Data

For the PUMS files having periods ending in 2009, the weighting process implemented a new raking procedure which is the subject of this paper.

Proposed PUMS Raking Methodology

The proposed PUMS raking procedure was similar to one used by the full sample ACS for persons from HUs for several years. It involved three steps to be repeated in an iterative process. The factor from the previous step would be applied to weights before doing the ratio-adjustment of the next step. The final ratio-adjustment factor was equal to the product of the ratio-adjustments calculated for all steps through convergence to a specific tolerance level. The matrix for the first step used three cells: householders with spouses, spouses, and others. The matrix for the second step used two cells: householders and others. The third step used a matrix formed by crossing six race and Hispanic origin categories by 13 age categories by gender. The three steps in the raking were to be repeated within each weighting area until the spouse/householder cells met a specified tolerance up to a maximum of 40 repetitions. At the end of each round, the spousal/householder cells were tested to see if the current estimate agreed with the cell controls within a tolerance of less than 0.0001 through round 19 and 0.001 for rounds 20 - 40. Any weighting area which converged at the end of a round was removed from the ongoing raking process. Any weighting areas which did not converge by the end of the 40^{th} round, kept the factors from the 40^{th} round.

Tables 2 and 3 show the matrix for steps one and two and explain the controls used by PUMS.

Table 2. Step One of Raking - Spousal/couple Adjustment

Categories:	Householder with spouse	Spouse	Other
Controls:	Sum ACS HU weight for couple households	Sum ACS HU weight for couple households	Sum all ACS person weights for all persons, then subtract householders with spouses and spouses

Source: American Community Survey Accuracy of the Data (2008)

Table 3. Step Two of Raking - Householder Adjustment

Categories:	Householder	Other
Controls:	Sum ACS HU weight for occupied households	Sum ACS person weights for all persons, subtract householders

Source: American Community Survey Accuracy of the Data (2008)

Demographic cells of step three were formed by crossing the categories of Tables 4 and 5. Controls were formed from summing ACS person weights within each cell.

Table 4 - Step Three of Raking, Part I

Hispanic/race categories	
Hispanic Origin:	Race Categories:
NonHispanic	White
	Black
	American Indian / Alaska Native
	Asian
	Native Hawaiian or Pacific Islander
Hispanic	

Source: American Community Survey Accuracy of the Data (2008)

Table 5 - Step Three of Raking, Part II

Age and Gender Adjustment		
Age Category:	Male	Female
0 – 4		
5 – 14		
15 - 17		
18 – 19		
20 – 24		
25 – 29		
30 – 34		
35 – 44		
45 – 49		
50 – 54		
55 - 64		
65 - 74		
75 or more		

Source: American Community Survey Accuracy of the Data (2008)

The ACS uses the successive difference method of replicate weights to estimate standard errors. The PUMS files contain adjusted ACS replicate weights. Each replicate weight is run through the same PUMS weighting process as the full PUMS weights, so that each PUMS replicate weight total agrees better with the replicates of the full ACS sample.

The proposed methodology is more complicated than the one used for the earlier PUMS files. Questions asked were "was the improvement worth the effort to implement" and "how should we implement so that we maximize the improvement?"

Overview of Research

Three options were considered in preparation for the future PUMS weighting process.

1. Keep the weighting methodology the same as was used for the 2008 and earlier PUMS files.
2. Expand the weighting cell matrix and implement the new raking procedure at the state level.
3. Expand the weighting cell matrix and implement the raking procedure at the PUMA level.

Tests were run on these options using 2007 1-year PUMS data. Options were examined for reasonability, and compared using a set of 257 person characteristic estimates from the popular ACS Profile tables on the American Factfinder. Reasonability checks examined the number of cells required to collapse, the distribution of the final ratio-adjustment factors, the distribution of final weights, and the number of cells which converged within 40 iterations of the raking.

For the analysis, the full ACS sample estimates were considered to be the expected value for PUMS conditioned on a fixed ACS sample. In order to measure improvement, the absolute value difference between PUMS and the full sample ACS was computed. PUMS estimates which moved closer to the full sample ACS estimate after raking were considered positive improvements. PUMS estimates which moved farther from the full sample ACS estimates were considered to be negative changes.

Metrics to measure change in estimates included Z-scores of differences, relative percent differences, as well as the Wilcoxon signed-rank test. The distributions of Z-scores of differences between PUMS and ACS estimates for each weighting option were examined to see if the number of significant differences changed between options. The signed-rank test was used to evaluate the trend for improvement for each of the 257 Profile characteristic estimates separately at the state and PUMA levels.

Data used to create the Profile estimates were ACS data. PUMS estimates were calculated using the ACS characteristics of PUMS sample cases instead of their actual PUMS data to avoid any changes introduced by PUMS edits such as rounding, top-coding, collapsing of detail categories and age perturbation. Also, any sampling and nonsampling error present in ACS estimates was ignored since we are only testing the effects on the PUMS weighting procedure.

Research Methodology

Evaluation of Changes in Estimates

We began by creating a standardized change variable (C1) which measured how the difference between ACS and PUMS estimates would change if PUMS used an alternative weighting methodology. Differences were examined for estimates at the national, state and PUMA levels.

The formula for the change variable C1 is shown below. The numerator is the difference between two absolute value differences. The denominator is the standard error of the numerator. In the formula, think of the absolute value of the difference as a distance. The first distance (d1) is the absolute value of the difference between the full sample estimate and a PUMS estimate using one of the weighting methods. The second distance (d2) is the absolute value of the difference between the full sample and the same PUMS estimate using a different weighting method. The standard error in the denominator was calculated using the successive difference method of replicate weights. The difference (d1-d2) was calculated for each of 80 replicates in order to estimate the standard error of the difference.

Note that the standardized measure C1 is positive whenever the second weighting method used in d2 brings the PUMS estimate closer to the full sample value than the first method used in d1. C1 is negative whenever the second method moves the PUMS estimate farther from the full sample value than the first method.

$$C_1 = \frac{(d1 - d2)}{SE(d1 - d2)}$$

where:

d1 – absolute value of ($ACS - PUMS_1$)
d2 – absolute value of ($ACS - PUMS_2$)
ACS - the value of the ACS estimate
$PUMS_1$ - the PUMS estimate using one set of weights
$PUMS_2$ - the PUMS estimate using a second set of weights
SE(d1-d2) - the replicate weight standard error of (d1-d2)

The purpose of standardizing this change variable was to form a metric which measures the size of a change that can be ranked in a meaningful way for the signed-rank test. The change variable C1 is also approximately a Z-score.

The change variable was used for the following comparisons:

1. the original PUMS weighting versus the state level raking
2. the original PUMS weighting versus the PUMA level raking
3. the state level raking versus the PUMA level raking

To test whether or not the changes tended to be positive or negative, the Wilcoxon signed-rank test was then computed on each of the 257 change estimates. For each item the signed-rank test was applied across changes of state estimates and PUMA estimates. The Wilcoxon signed-rank statistic S is computed by SAS as:

$$S = \Sigma r_i^+ - n(n+1)/4$$

where:

r_i^+ = the rank of | C_i | after discarding values of C_i=0
i = 1 to 52 for state level or 1 to 2099 for PUMAs
n = the number of nonzero values
Σ = the sum is over the values of $C_i > 0$

The null hypothesis for the signed-rank test was that the expected median change in C1 for an item was zero across a level of geography. Using $\alpha = 0.1$ for a two tailed test, whenever the probability of exceeding an absolute value of S was less than 10 percent, the alternative hypothesis was accepted that the change for that item was not zero. The number of negative and positive significant changes was counted to determine which option had the greatest number of estimates with positive changes. Purely random changes to the weights would be expected to yield a small number of negative and positive changes. When this test is applied to a set of 50 or more paired estimates, S is expected to be normally distributed[1]. Using $\alpha = 0.1$, we expected a small number of significant changes (possibly 5 percent in each tail) due to random noise. The option having the greatest number of positive changes and the least number of negative changes

[1] Mendenhall, Scheaffer, Wackerly. "Mathematical Statistics with Applications, Second Edition", Boston, MA: Duxbury Press, 1981, section 15.4.

would be chosen as the best if the number of positive changes was substantially greater than the number of negative changes.

Evaluation of Changes in Standard Errors

To test whether or not the raking tends to give us smaller standard errors, the signed-rank test was also to be applied to the change in standard errors (C2). A reduction in standard errors due to the raking process can be interpreted as an indication that the PUMS weights are in better agreement with the full sample since each of 80 replicate weights was adjusted to agree with the full sample for spouses, householders, race/Hispanic origin, age and sex.

The measure of change for standard errors was:

$$C_2 = \frac{(SE_1 - SE_2)}{ACS}$$

where:

ACS = the value of the full sample ACS estimate
SE_1 = the replicate weight standard error of the original PUMS estimate
SE_2 = the replicate weight standard error of the optional PUMS estimate

C2 is standardized by dividing the differences in the standard errors by the ACS estimate. C2 will be positive whenever the raking decreases the size of the standard error estimate. The conditional expected value of C2 is equal to the difference in the coefficients of variation for the two PUMS options. Both raking options were expected to cause a decrease in the standard errors.

Results of Research:

The reasonability checks and analysis of the changes in Profile estimates are given here from the research using 2007 PUMS data.

Summary of Reasonability Checks of Raking Options

The reasonability checks helped us to understand the effects of the raking options on the PUMS file and clarified some of the differences.

Here are tables to show the reasonability checks on the distribution of ratio-adjustment factors and the frequency of collapsing of race/Hispanic origin cells.

Distribution of overall raking factors and the effect on weights

Table 6 shows the distribution of overall factors applied to PUMS initial weights by the raking process across fifty states, the District of Columbia and Puerto Rico. The distribution is based on the count of cells in which the factor was applied. The cells for the original 2007 PUMS were male or female by PUMA. The cells for the raking were the collapsed demographic cells split by spousal and householder cells within either state or PUMA.

Table 6: Distribution of Ratio-adjustment Factors Applied to PUMS Person Weights

	Maximum	99th percentile	95th percentile	Median	5th percentile	1st percentile	Minimum
2007 PUMS original method	1.095	1.059	1.038	0.999	0.965	0.947	0.911
state-based raking	2.683	1.222	1.077	0.999	0.926	0.834	0.495
PUMA-based raking	3.345	1.514	1.271	0.993	0.810	0.724	0.064

Source: 2007 American Community Survey Special Tabulation

As expected, the raking process produced more variation in the ratio-adjustment factors than the original 2007 PUMS method. In particular the PUMA based raking had factors as large as 3.345 and as small as 0.064. The smallest factor of 0.064 occurred in a PUMA which did not converge during the 40 rounds of raking.

As a check on the effect of the extreme factors, Table 7 shows the distribution of person weights resulting from the three weighting methods using 2007 PUMS data across fifty States, the District of Columbia and Puerto. The distributions of the weights are similar in spite of the differences in the factors. The largest weight occurred in the original PUMS. The smallest final weight occurred for the PUMA based cells. The weights from the two raking tests are shown unrounded, however, in a production run, the weights would be systematically rounded and the smallest weights would be rounded up to 1.

Table 7: Distribution of Person Weights Using Three Methods of Weighting

	Maximum	99th percentile	95th percentile	Median	5th percentile	1st percentile	Minimum
2007 PUMS Original method	2032.0	364.0	249.0	83.0	24.0	15.0	1.0
State-based Raking	1941.1	364.0	249.5	83.1	24.4	15.3	1.1
PUMA-based Raking	1687.7	366.0	248.3	82.9	24.3	15.2	0.1

Source: 2007 American Community Survey Special Tabulation

There did not appear to be much difference in the variability of the weights between the three methods, although the PUMA-based cells did reduce the largest weights.

Convergence Criteria

During testing, about 99.9 percent of the PUMAs converged within 40 rounds for PUMA-based raking and all converged within 20 rounds for state-based raking. This seemed reasonable for each method.

Collapsing of race/Hispanic origin cells

The effectiveness of the raking process to improve an estimate depends partly on having enough records to construct a cell that can stand alone. Cells were required to collapse with another cell if it contains less than 10 persons, had an adjustment factor greater than 3.5, or had an adjustment factor less than 1/3.5 before starting the raking process. For this concern we noted that the state-based weighting areas had less collapsing than the PUMA-based weighting areas.

In the 2007 PUMS, the smallest race group, the Native Hawaiian or Pacific Islander (NHOPI) had sample in only 842 PUMAs and 48 state level areas. Eleven state and 762 PUMA weighting areas would be required to collapse the NHOPI with other generally larger race groups. The age/sex cells constructed within each race/Hispanic origin group would also have less collapsing at the state level. For this reason, the raking effect should be expected to produce better state level estimates of small race groups, such as American Indian Alaska Native (AIAN) and NHOPI, and estimates of several age categories by sex if state level weighting areas were used. The age/sex cells constructed within each race/Hispanic origin group would also have less collapsing at the state level.

Both sign and signed-rank tests were performed. The results were similar although the signed-rank gave slightly more significant estimates than the sign test due to the influence of many of the larger estimates moving in the right direction. For brevity, only the signed-rank results are shown here.

Summary of Signed-Rank Tests Using 2007 Data

The following tables show the results of applying the signed-rank test to differences in 257 estimates caused only by changing the weighting method. The comparisons were the original PUMS weighting method versus the state-based raking, the original PUMS weighting method versus the PUMA-based raking, and state-based raking versus PUMA-based raking. Data were taken from the 2007 PUMS weighting files and 2007 ACS data files.

Table 8 summarizes the application of the signed-rank test separately for two sets of Profile estimates, state-level estimates and PUMA level estimates. The signed-rank test was applied to individual estimates across all fifty states, the District of Columbia and Puerto Rico. Since we are using $p < 0.1$ for significance (a two tailed test), we would expect randomly to find about five percent of estimates to get worse and five percent should get better. After applying the signed-rank test separately for each of the 257 estimates, Table 8 shows the percentage in which increasing the number of cells brought PUMS significantly closer to the full sample values. Results for the state-level estimates are shown in columns one through three, and results for the PUMA-level estimates are shown in columns four through six. Using $\alpha = 0.1$, the first and fourth columns shows the percentage significantly worse, while the third and sixth columns shows the percentage of the 257 estimates that moved closer to the full sample values.

On the first row of Table 8, the state-based raking improved 42 percent of the estimates at the state level, while only 4.7 percent got worse. Since we might expect five percent to be classified as significantly worse due to random noise, it appears that the state level raking makes solid improvements over the original 2008 PUMS method. However, when looking at PUMA level estimates, the state –based raking made about 50 percent of the estimates significantly worse and about 10.9 percent improved. On the second row, the PUMA-based raking did improve 30 percent of the state-level estimates while the number which got significantly worse (6.6 percent) was still in the ballpark of random noise. The second row also shows that a full 60.3 percent of the PUMA level estimates improved, while only 7.8 percent significantly declined.

The third row shows that when compared to state-based raking, the PUMA-based raking did not improve many state-level estimates (only 2.3 percent), and 23.4 percent got worse. However, PUMA-based raking clearly improved the PUMA level estimates (68.9 percent) when compared to state-based raking.

Table 8 –Summary of Signed-rank Tests by 2007 Estimate

	Comparison Using State-Level Estimates			Comparison Using PUMA-Level Estimates		
	Significantly worse	Inconclusive	Significantly Better	Significantly worse	Inconclusive	Significantly Better
State-based Raking versus Original Method	4.7 %	53.3 %	42.0 %	50.2 %	38.9 %	10.9 %
PUMA-based Raking versus Original Method	6.6 %	63.4 %	30.0 %	7.8 %	31.9 %	60.3 %
PUMA-based Raking versus State-based Raking	23.4 %	74.3 %	2.3 %	3.9 %	27.2 %	68.9 %

Source: 2007 American Community Survey Special Tabulation

Table 9 repeats the signed-rank tests on the standardized change in SE, using SEs from the 257 Profile estimates. The PUMA cells reduced standard errors when compared to original PUMS for both for state-level estimates and PUMA level estimates, and also beat the state cells for reduction in SEs. The second row of Table 9 shows that PUMA cells reduced SEs for 89.1 percent of the state level estimates, and 94.2 percent of the PUMA level estimates. The third row shows that when directly compared to the state cell option, the PUMA cells made consistent reductions in standard errors.

Table 9 – Summary of Signed-rank Tests of 2007 Standard Errors

	Standard Errors of State-Level Estimates			Standard Errors of PUMA-Level Estimates		
	Significantly worse	Inconclusive	Significantly Better	Significantly worse	Inconclusive	Significantly Better
State-based Raking versus Original Method	2.7 %	25.7 %	71.6 %	63.8 %	18.3 %	17.9 %
PUMA-based Raking versus Original Method	1.6 %	9.3 %	89.1 %	2.0 %	3.9 %	94.2 %
PUMA-based Raking versus State-based Raking	18.3 %	25.7 %	56.0 %	1.2 %	5.8 %	93.0 %

Source: 2007 American Community Survey Special Tabulation

Tables 8 and 9 indicate that both raking methods are expected to improve the PUMS estimates more often than not.

Here is a summary of how the above evidence was considered when deciding on which method to pursue for the 2009 PUMS weighting.

1. The state-based raking was better than the PUMA-based raking when reviewing the 13,364 state-level estimates. About 42 percent of the estimates saw improvements over the 2007 PUMS versus about 30 percent for PUMA-based. See the third column of Table 8, rows one and two.

2. The state-based raking was found to have a tendency to deteriorate when reviewing the 500,000 PUMA-level estimates. About 50 percent of the estimates saw deterioration over the 2007 PUMS versus about eight percent for PUMA-based raking. See the fourth column of Table 8, rows one and two.

3. The PUMA-based raking was better than state-based raking for the 500,000 PUMA-level estimates. About 60 percent saw improvements over the 2007 PUMS versus about 11 percent for the state-based raking. See the sixth column of Table 8, rows one and two.

4. The PUMA-based raking improved standard errors as compared to both the 2007 PUMS and the state-based raking. This improvement was substantial for both state and PUMA level estimates. See columns three and six of Table 9.

We also looked briefly at the magnitude of relative changes in each of these tables. When subtracting out the smallest relative changes (changes smaller than 0.5 percent) we found similar results also indicating that both raking options improved the estimates over the 2007 PUMS original method. State-based raking was somewhat better for state level estimates and PUMA-based raking was much better for PUMA level estimates. However it was noted that the magnitude of relative changes was much smaller for state level estimates.

After considering the number and magnitude of the significant changes, it was clear that the PUMA-based raking would improve far more estimates and make larger relative improvements in those estimates than the other options. The raking process was implemented in 2010 for the PUMS files ending in 2009.

Evaluation of PUMA-Based Raking Using the 2009 PUMS Sample

As a validation check on the raking implementation, comparisons have been run at the national and state levels to see how PUMS estimates compares to the full ACS sample estimates. The previous 2008 weighting program was run using 2009 PUMS data and compared to the 2009 PUMS production weights. The predicted changes of the raking versus the previous method are being checked here using the same analysis of a standardized change statistic by state and PUMA using the signed-rank test. Following these results, a table that shows the median of the percent differences of the changes is shown for 29 categories of person estimates.

The first row of Table 10 compares 266 Profile estimates, and the second row compares the standard errors.

Table 10. Summary of Signed-rank Tests for Changes in 2009 Profile Estimates and Standard Errors

	State-Level Estimates			PUMA-Level Estimates		
	Significantly worse	Inconclusive	Significantly Better	Significantly worse	Inconclusive	Significantly Better
2009 Estimates: Production vs the Previous method	3.8 %	67.3 %	29.0 %	6.0 %	32.0 %	62.0 %
2009 Standard Errors: Production vs the Previous method	0.8 %	11.7 %	87.6 %	0.8 %	3.0 %	96.2 %

Source: 2009 American Community Survey Special Tabulation

Row 1 of Table 10 shows that when 2009 PUMS production data is compared to 2009 data weighted with the older 2008 methodology, the results are similar to those as seen in row 2 of Table 8. About 29 percent of the state-level estimates improved, while only 3.8 percent deteriorated. When considering PUMA-level estimates, about 62 percent improved. The percent counted significantly worse for both state and PUMA estimates were in the ballpark of the five percent expected to decline due to random noise. Standard errors also improved considerably as predicted.

Table 11 shows the median percent improvements in the 2009 PUMS person-level estimates due to the raking procedure for several characteristics at the PUMA level. Improvements are estimated by using the ACS estimate as the base of the percent difference between ACS and PUMS. Positive values in this table indicate that the raking reduced the absolute relative difference between ACS and PUMS estimates. Negative values mean that the raking increased the difference.

Table 11: Improvements due to the Raking Used in the 2009 PUMS

Characteristic	Median Improvement of a PUMA-level PUMS Estimate as a Relative Percent of the ACS Estimate
Relationship to Reference Person	0.27 %
Marital Status	0.38 %
Fertility	-0.01 %
Grandparents	-0.20 %
School Enrollment	0.72 %
Educational Attainment	0.09 %
Veteran Status	0.46 %
Disability Status	0.37 %
Movers	0.00 %
Place of Birth	0.00 %
Citizenship and Year of Entry	0.34 %
Region of Birth for Foreign Born	0.51 %
Language Spoken at Home	0.36 %
Ancestry	0.00 %
Labor Force	0.36 %
Commuting	0.01 %
Occupation	0.00 %
Industry	0.00 %
Class of Worker	0.02 %
Income for Persons	0.10 %
Health Insurance	0.04 %
Poverty – Persons	0.00 %
Sex and Age	0.70 %
Race – single race	0.42 %
Race – other	0.93 %
Hispanic	1.83 %
Nonhispanic one race	0.93 %
Nonhispanic 2 races	0.39 %

Source: 2009 American Community Survey Special Tabulation

The fact that this table shows mostly positive changes fits with our expectation that many types of estimates have some positive correlation with characteristics being controlled. For this reason, the raking tended to bring PUMS estimates closer to the full sample ACS estimates.

Conclusion:

Both state-based and PUMA-based raking methodology showed the expected improvement in the weighted estimates, especially for demographic characteristics. PUMA-based raking did result in more collapsing of race and Hispanic origin groups than the state-based raking, and did result in greater variability in potential ratio-adjustment factors. However, the expected weight distributions were similar between the options.

The PUMA-based raking was chosen for the 2009 and future PUMS processing despite the additional collapsing and increased variability because it maximized the improvement of more PUMS estimates than the state-based raking. The PUMA-based raking made substantial improvements in standard errors across all estimate types as well.

After the implementation of the methodology and the production of the 2009 PUMS files, analysis confirmed that the expected improvement did occur.

References:

Mendenhall, Scheaffer, Wackerly. "Mathematical Statistics with Applications, Second Edition", Boston, MA: Duxbury Press, 1981. (Section 15.4 describes the Wilcoxon Signed-rank test.)

Bureau of the Census: Design and Methodology of the American Community Survey, 2009, http://www.census.gov/acs/www/methodology/methodology_main/

Bureau of the Census: PUMS Accuracy of the Data (2007), http://www.census.gov/acs/www/data_documentation/pums_documentation/

Bureau of the Census: PUMS Accuracy of the Data (2009), http://www.census.gov/acs/www/data_documentation/pums_documentation/

Bureau of the Census: Accuracy of the Data (2008), http://www.census.gov/acs/www/data_documentation/