Investigation of Anomalies in Derived Variances for Estimates from The American Community Survey Public Use Microdata File[1]

Sirius Fuller and Karen King
Decennial Statistical Studies Division
U.S. Census Bureau

Abstract
The American Community Survey releases Public Use Microdata Sample (PUMS) files annually for users to calculate their own estimates. PUMS contains individual housing unit and person records for a limited set of geographic areas. Two methods exist for users to calculate standard errors and variances for estimates: a generalized variance method (design factor) and a replicate weight based method.

Research conducted outside of the Census Bureau has shown that variance based on the replicate weights is much higher than the variance based on design factors for certain estimates at the national level. The Census Bureau is conducting research which will attempt to duplicate those findings and to look for possible causes of these results. It will look at the creation process for the replicate weights for PUMS records, focusing on the impact of PUMS sampling and weighting. It will examine the process to create design factors for PUMS data, focusing on the iterative linear regression used to create design factor candidates. This paper will show results and may offer some practical solutions to bring the two sets of standard errors and variances into better alignment.

Keywords: American Community Survey; Public Use Microdata Sample; Design Factors; Design Effects

1. Background and Research Questions

The American Community Survey (ACS) is a continuous survey which samples households in every county in the United States and municipio in Puerto Rico. It collects timely data on demographic, housing, social, and economic characteristics. In addition to housing units, group quarters (GQ) such as college dormitories, prisons, nursing facilities and military barracks are also sampled. The ACS replaced the long-form version of the Decennial Census. In 2011, the sample of the survey was expanded from roughly 2.9 million to about 3.5 million housing units addresses. A design change was implemented that reallocated sample into smaller geographic areas, thus improving the reliability for these small areas There was also a large-scale imputation of group quarters person records to improve the major group quarters estimates in every county to the sampling frame.

With the release of the ACS data products, Public Use Microdata Sample (PUMS) files are also released for data users to compute their own estimates. The PUMS files contain a subsample of ACS individual housing unit and person records for the nation, region,

[1] This paper is released to inform interested parties of research and to encourage discussion. Any views expressed on statistical and methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau

regional subdivision, state and Public Use Microdata Area (PUMA)[2]. Estimates created using PUMS should approximate, but may not match, the ACS estimates. Two methods are provided to data users to calculate standard errors (SE) and variances for their estimates from the PUMS files. The first is a replicate weight (RW) based successive difference replication (SDR) method. This method, used on the PUMS data, is the same methodology used to compute margins of error (MOE) for ACS data products published on American FactFinder (AFF)[3]. The second method for calculating SE with PUMS data is a generalized variance method which uses a simple random sample (SRS) standard error adjusted using a design factor (DF) from the published set of design factors. A DF is the square root of a design effect (DEFF) which was first proposed by Leslie Kish in 1956 and is used to adjust the variances of a complex sampling design.

The work presented in this paper is in response to research conducted by Michel Boudreaux, Peter Graven and Michael Davern using the 2009 1-year PUMS data files. Their paper, entitled “Design Effect Anomalies in the American Community Survey”, found an anomaly for selected PUMS characteristics where the SDR variance is higher than expected compared to the variance generated using the generalized variance methodology at the national level.[4] It went on to say that the difference is not as great for the average state. The anomaly will be explained in more detail below.

This paper will answer the following research questions:

> Can we replicate the results shown in the Boudreaux et al. paper using the more recent 2011 1-year PUMS data files?
>
> Does the process for creating PUMS dataset (specifically the replicate weights) contribute to the variance anomaly?
>
> Does the creation process of the PUMS generalized variance parameter (i.e. design effect) contribute to the variance anomaly?

The first question looks to see if the issues still exists. The other questions demonstrate that we are examining both sides of this issue: Is the presence of the anomaly caused by the replicate weights or the design factor methodology?

One point should be made clear. The data that Boudreaux, et al. used for analysis is the 2009 1-year PUMS data. Although PUMS is a subsample of the full ACS data, it is subject to additional processing that modifies, adjusts and perturbs the data and weights. Issues in the PUMS data files do not necessarily indicate issues or problems with the full ACS data. Investigating and examining the full ACS data is beyond the scope of this paper.

---

[2] PUMAs are areas of roughly 100,000 people created after each Census in order to provide data users with more granular data while still preserving confidentiality.
[3] The MOE is calculated by multiply the standard error by 1.645. The standard error is the square root of the variance.
[4] Their research was presented at the 2012 Joint Statistical Meeting in San Diego.

2. Methodology

This paper will focus on two characteristics mentioned in the Boudreaux et al. paper that we are most concerned about: the percent of people with health insurance coverage and the percent of people below poverty.[5] This paper will use three measures of variance and the published PUMS design factors, which are all defined below. SAS will be used to compute all estimates and percents presented in this paper.

A variance ratio is calculated by computing the SDR and SRS variances first and then dividing the SDR variance by the SRS variance.

$$Variance\ Ratio = \frac{SDR\ (RW)\ Variance}{SRS\ Variance} \tag{1}$$

This variance ratio mirrors the design effect reported in the Boudreaux et al. paper mentioned previously. The variance ratio is the actual design effect calculated using the actual PUMS data. This is different from the published design effect created by squaring the published design factor (pubDF).

The square root of the variance ratio may be considered a DF. The DF computed in this way can differ from the pubDF. The pubDF are created using a model-based approach described below. The pubDF squared multiplied by the SRS variance creates the DF variance, which should approximate the SDR variance.

$$(pubDF^2) \times SRS\ Variance \approx SDR\ variance \tag{2}$$

2.1. Computing the Replicate Weight Variance for Percent Estimates

The SDR variance for an estimate $(\hat{X})$ is computed using the standard formula provided in the 2011 1-year PUMS Accuracy of the Data.

$$SDR(RW)variance\,(\hat{X}) = \frac{4}{80}\sum_{r=1}^{80}\left(\hat{X}_r - \hat{X}\right)^2 \tag{3}$$

Here $\hat{X}$ can be any type of characteristic estimate (count, percent, ratio, etc.) computed using the PUMS weight. The $\hat{X}_r$ is the rth replicate estimate computed using the rth PUMS replicate weight. There are a total of 80 replicate weights provided in the PUMS dataset. This is the same methodology used for producing SDR variances for the ACS data. In this case, $\hat{X}$ is the percent, $\hat{p}$.

[5] The Boudreaux et al. paper examined mean person's earnings, percent of people with health insurance coverage, the percent of people below poverty, the percent of people who lived in housing units which are rented and the percent of housing units which are rented.

2.2. Computing the Simple Random Sampling Variance for Percent Estimates

The SRS variance for a percent estimate, $\hat{p}$, uses the formula found in the 2011 1-year PUMS Accuracy of the Data. The formula is:

$$SRS\ variance\ (\hat{p}) = \left(\frac{99}{B}\right)\hat{p}(100 - \hat{p}) \tag{4}$$

Where B is the base, or denominator, of the percent and $\hat{p}$ is the percent for the PUMS estimate for a specific characteristic. The 2011 1-year PUMS data is a one percent sample. The 99 is the appropriate finite population correction factor applied to a one percent sample.

2.3. Computing Model-Based Design Factors for Publication

The 2011 1-year PUMS Accuracy of the Data document includes a set of published DFs at the national and state level for about 66 DF subject groups. These DF subject groups are listed as "Characteristics" and can be for one variable (e.g. "Health Insurance") or for multiple PUMS variables (e.g. "Place of Birth, Year of Entry, and Citizenship Status"). The square of the pubDF also approximates a design effect (pubDEFF).

$$(pubDF)^2 = pubDEFF \tag{5}$$

For each DF subject group, pubDFs are created using a multistep process:

a. The full sample ACS estimates and variances from selected ACS detailed table(s) from AFF are used as inputs to a linear regression model. In some cases all of the estimates from an ACS table are used. In other cases, only a subset of the ACS table estimates is selected for inclusion in the model.
b. The model produces several design effect candidates[6].
c. The candidates are then used to create DEFF variances using the 2011 1-year PUMS data.
d. An algorithm[7] is used to pick the candidate with the DEFF variance which best approximates the PUMS replicate weight variance.

The DEFF is a generalized design effect from a model, and thus covers many different estimates. It will approximate the replicate weight variance better for some estimate than for others.

Using the 2011 1-year PUMS data, we will show whether the anomaly is still present by comparing the variance ratio for poverty and health insurance to their pubDEFF. Finally, we will examine if the final selected DF candidate may be impacted by adjusting the input to the modeling.

---

[6] During the creation of a DEFF candidate outliers are identified using specific predetermined criteria. The outliers are then removed from the data and the linear regression is rerun.
[7] The algorithm is the mean absolute relative difference. The relative difference is defines as the difference between the DEFF variance and the SDR variance divided by the SDR variance.

3. Results

The following sections present results of the research.

3.1. Examining Whether the Anomalies Exist Using 2011 1-year PUMS Data

The Boudreaux et al. paper used 2009 1-year PUMS data files. This paper uses the more current 2011 1-year PUMS data files. An important first step is to establish whether the anomaly at the national level is present in the 2011 1-year PUMS data as there have been several changes. Beginning in June of 2011, the sample size of the ACS increased. There was also a large-scale imputation of group quarters to improve small area estimates[8]. Both of these may have had an impact on PUMS, which is a subsample of the full ACS data.

We test for the anomaly using the PUMS variance ratio compared with the pubDEFF. As can be seen in Table 1, the anomaly is still present for the percent with health insurance coverage and the percent of persons below poverty for both 2009 and 2011 1-year data at the national level. In 2011, the variance ratio for the percent with health insurance coverage is about four times the pubDEFF. The variance ratio for the percent of persons below poverty is about 2.8 times the pubDEFF.

Table 1: Demonstrating the National-Level PUMS Anomalies by Comparing the Variance Ratio to the PubDEFF for 2009 and 2011

Characteristic	2009 1-year PUMS data		2011 1-year PUMS Data	
	Nat'l Var Ratio	pubDEFF	Nat'l Var Ratio	pubDEFF
Percent of Persons Below Poverty	10.24	2.89	8.30	2.89
Percent With Health Insurance Coverage	7.05	1.69	6.65	1.69

Source: 2009 and 2011 1-Year PUMS Data

Notice that the published DF have not changed since 2009 which demonstrates the stability of the design factors. Also, notice that the variance ratios have changed between the two years. The 2009 PUMS variance ratios are larger than the 2011 variance ratios. Changes to the design and sampling size of the survey may be the reason for these differences.

We can see below in Table 2, the variance ratio at the national level is greater than the simple mean[9] of the variance ratios at the state level. Again, this shows that the anomaly persists in the 2011 1-year PUMS data with the difference between the variance ratio and pubDEFF at the national level being larger than the average for the states.

[8] A brief explanation of the imputation process may be found in the 2011 1-year ACS Accuracy of the Data document located at http://www.census.gov/acs/www/data_documentation/documentation_main/.

[9] The range of the variance ratios for the state percents of people below poverty goes from 3.29 to 7.29. The range for the variance ratio for the percent of people with Health Insurance ranges from 1.96 to 4.69.

Table 2: Comparing the National-Level Variance Ratio to the Mean State PUMS Variance Ratio for 2011 1-year PUMS

Characteristic	Nat'l Var. Ratio	National pubDEFF	Mean State Var. Ratio	Mean State pubDEFF
Percent of Persons Below Poverty	8.30	2.89	4.55	2.98
Percent With Health Insurance Coverage	6.65	1.69	2.99	2.71

Source: 2011 1-Year PUMS Data

3.2. Investigating the Impact of the PUMS Processing on the Variance Anomaly

In order to discuss PUMS processing and its impact on PUMS variances, we need to give some details on the ACS processing. As stated earlier, ACS selects a sample of the housing unit and group quarters population. The probability of selection of an address becomes the ACS initial weight. A separate sequence of eighty replicate factors are assigned to each sample record in a complex process based on the eighty row Hadamard matrix. The 780 sequences are assigned repeatedly through the sample records. It is the expectation that each sequence is assigned roughly the same number of times. Here we will call this the "balanced state". The initial weights for the eighty replicate weights are formed by multiplying the replicate factor to the case's ACS initial weight. The replicate weights are processed in the same way as the full weights.

Returning our focus to the PUMS creation, there are various stages (subsampling, weight and estimation) and each contributes to the impact on the variance.

3.2.1. Examining the Creation Process of the PUMS Data Files

For PUMS, a subsample of ACS records is selected to make up roughly one percent of the housing unit universe. Subsampling is done at the state level. When a sample record is selected for PUMS, its responses, its full sample weight, and its sequence of eighty replicate weights are also selected. With the subsampling, there is the possibility that the sequences are now unevenly spread across the PUMS sample records. In a sense, becoming "unbalanced". This would have a negative impact on the replicate weight variance of an estimate.

The initial PUMS weights for a selected sample record are the full and replicate ACS weights multiplied by the state level subsampling factor. A series of adjustments are made to bring PUMS weighted estimates into closer agreement with the published ACS weighted estimates for several key demographic characteristic at the PUMA level. Resulting factors from a series of ratio-estimate adjustments are applied to individual sample records base on their age, sex, race, Hispanic origin, marital status, and relationship to householder. More detail on the impact of these adjustments will be given below.

Table 3: National Variance Ratio Computed Using Weights at Various Stages of the Creation of the 2011 PUMS

Characteristic	Initial Weights	Post-Raking Weights	Published (Rounded) Weights
Percent of Persons Below Poverty	7.28	8.32	8.30
Percent with Health Insurance Coverage	5.28	6.65	6.65

Source: 2011 1-Year PUMS Internal Data

The results presented in Table 3 show that the weighting process for the person weights has an impact on the variance ratio at the national level for these two characteristics. There is some increase in the variance ratio from the initial weighting stage to the post-raking weighting stage. This is somewhat counter-intuitive since, in general, the post-raking weighting step tends to either reduce the variance of estimates or not increase them by much. This may be explained below. As a note, the creation of the ACS weights was also investigated and a similar pattern was observed.

3.2.2. Examining the Variance Ratios for Various Demographics

After cases are selected from ACS to be included in the PUMS data files the weights are adjusted to agree with certain independent estimates. This includes raking the data to create the spouse equalization/householder equalization raking factor and the demographic raking factor.

The first raking factor adjusts the weights to make sure that the estimate of married householders is approximately equal to the number of spouse. The second factor adjusts the person weights so that the weighted sample counts equal ACS population estimates by age, race, sex, and Hispanic origin at the PUMA level. Because of collapsing of groups in deriving this factor, only total population is assured of agreeing precisely with the published ACS population estimates at the PUMA level.

Table 4 examines the variance ratios for the percent with health insurance and percent in poverty by selected relationship statuses. The reference person, spouse and unmarried partner have tighter constraints placed upon them during the raking process in the creation of the PUMS data files. People with a relationship status other than the three mentioned above have looser constraints. It can be seen that the high variance ratios for both health insurance and poverty are largely driven by people with a relationship status other than reference person, spouse and unmarried partner.

Table 4: Variance Ratios by Selected Relationship Statuses for the 2011 PUMS

Characteristic	Ref. Person	Spouse	Unmarried Partner	All Other Rel. Statuses
Percent of Pers. Below Poverty	1.80	1.68	2.08	7.92
Per. with Health Ins. Coverage	2.98	1.88	1.90	4.41

Source: 2011 1-Year PUMS Data

The largest group of people who are not a reference person, spouse or unmarried partner have the relationship status of biological son or daughter. They tend to be under 18 years old. Thus, the large variance ratio seen in Table 4 for the all other relationship statuses is presumably driven by those who are under the age of 18.

The next table goes on to examine the observation more closely. Table 5 examines the variance ratios for specific age groups. As expected, for the people below poverty the variance ratio was largest for the age range of 5 to 17. Interestingly, the variance ratio for ages zero to four was lower than the variance ratio for ages 18 to 64.

For the percent with health insurance, the largest variance ratio was for ages 18 to 64. This was not the expected result. The likely reason for why the variance ratios are so small for the younger and older populations is that almost all people below 18 and 65 or over have health insurance either from a public health insurance or through their parent's health insurance possibly due to the recent health insurance law going into effect.

Table 5: Variance Ratios for Selected Age Ranges for the 2011 PUMS

Characteristic	Ages 0 to 4	Ages 5 to 17	Ages 18 to 64	Ages 65 and Over
Percent of Persons Below Poverty	3.48	5.95	3.88	1.28
Percent with Health Insurance Coverage	2.32	3.33	5.92	2.24

Source: 2011 1-Year PUMS Data

3.2.3. Examining the Mode of Response for Replicate Weights Variability

An Individual Respondent Variance (IRV) is defined for each respondent on the housing and person data files as:

$$IRV_i = \frac{4}{80}\sum_{r=1}^{80}(w_{ir} - w_{i0})^2 \qquad (6)$$

This is similar to what the Boudreaux et al. paper calls a mean squared error (MSE). In their paper, they reported certain records with very high MSE, which corresponds in this paper to a very high IRV.

We hypothesized that the larger IRV is due to the mode of data collection. The mode is the method of data collection used to obtain a response from the housing unit. In 2011, ACS housing unit addresses received a questionnaire via mail[10]. If they do not respond and a telephone number was available, they were then contacted by telephone. This mode is called the computer assisted telephone interview (CATI). Finally, if they do not respond in CATI or a telephone number was not available, then a subsample of these cases was visited for the computer assisted personal interview (CAPI)[11]. In PUMS, the CATI and CAPI modes are grouped together.

GQs are handled differently from households. All GQs receive a personal interview. As a reminder, the majority of GQs are college dormitories, prisons, military barracks and

[10] Beginning in 2013, a new internet mode went into production. Households receive instructions on how to respond via the internet prior to receiving a questionnaire through the mail.
[11] For some areas, such as Remote Alaska, a slightly different mode collection process is used. Please see the 2011 1-year ACS Accuracy of the Data document located at http://www.census.gov/acs/www/data_documentation/documentation_main/ for more details.

nursing homes. GQs are listed separately from the housing modes in Table 6 due to fact that they can have very different demographic characteristics from housing units. For example, the majority of people living in college dormitories are between the ages of 18 and 24.

Table 6 shows the mean, median and 95th percentile values for the IRV computed for each person record by mode. The results for the PUMS housing weights showed similar results.

Table 6: Individual Respondent Variances by Mode for the 2011 1-year PUMS

Mode	Median	Mean	95th Percentile
HU: Mail	5,077	9,616	32,188
HU: CATI/CAPI	12,767	32,635	122,688
Inst. GQ	3,071	5,066	16,090
Non-Inst. GQ	3,229	6,731	22,467

Source: 2011 1-Year PUMS Data

The IRV gives insight into the variability of the replicate weight compared to the full weight for a record. The higher the IRV the more variable are the replicate weights. In Table 6 it can be seen that, as expected, the IRV for HU CATI/CAPI cases are larger than corresponding HU Mail and GQ cases in the same percentile. The 95th percentile for the CATI/CAPI cases is roughly ten times the value of the median. In contrast, the 95th percentile for the mail mode is only about six times the median. For institutionalize GQ, the 9th percentile is roughly five times the median value, and for non-institutionalize GQs the 95th percentile is approximately seven times larger. Thus, we can see that the CATI/CAPI cases at or above the 95th percentile are more volatile than the other modes.

3.2.4. Using the Mean of the Replicate Weight Estimates to Calculate an Alternate Variance

An alternative method for computing the variance is to use the mean of the replicate weight percent estimates as opposed to the percent estimate calculated using the full weights. This is touched upon by Sul as well as Rao and Shao. Thus, $\hat{\hat{X}}$ from the variance formula for the replicate weight variance in Equation 3 is replaced with Xbar.

$$ALT\ SDR\ (RW)Variance\left(\hat{X}\right) = \frac{4}{80}\sum_{r=1}^{80}\left(\hat{X}_r - \bar{X}\right)^2 \qquad (7)$$

Where Xbar (or $\bar{X}$) is the mean of the replicate weight percent estimates.

$$Xbar = \bar{X} = \frac{\sum_{r=1}^{80}(\ddot{X}_r)}{80} \qquad (8)$$

This idea comes from the origin for the SDR, which is a variation on the Jackknife method for computing variances. As mentioned in Wolter, Quenouille introduced the Jackknife as a method for reducing bias and Tukey later mentioned that it could be adopted for variance estimation. Quenouille uses the mean of the replicates as an unbiased estimate of the parameter of interest. The variance formula is similar to Equation 9 and uses xbar.

It is generally assumed that the estimate using X_0 and Xbar will be approximately the same. However, slight differences between the two can have a large impact in the variance estimation. This can be seen in Table 7 below where the variance ratio decreases from 6.65 to 2.52 for the percent with health insurance coverage and from 8.30 to 4.81 for the percent of persons below the poverty threshold. Both are now closer to the pubDEFF.

This is primarily due to the majority of the replicate estimates being larger or smaller than the estimate created using X_0. Xbar is more likely to have an equal number of replicate estimates above and below it, especially if there are no extreme replicate estimates. This is the case with both the percent of people below poverty and the percent of people with health insurance.

Table 7: National Variance Ratio Using SDR Variance and the Alternative SDR Variance

Characteristic	Percent Estimate using X_0	Percent Estimate using Xbar	Variance Ratio	Alternative Var. Ratio	pubDEFF (pubDF squared)
Percent of Persons Below Poverty	15.8691	15.8486	8.30	4.81	2.89
Percent with Health Insurance Coverage	84.6263	84.6469	6.65	2.52	1.69

Source: 2011 1-Year PUMS Data

3.3. Examining the Creation Process for the PUMS Published Design Factors

The pubDF are created using a model-based approach explained earlier in this paper. Three aspects of this process were investigated. The first was to adjust some of the detailed national level geographies going into the model. The second was to modify the modeling process to be more tolerant of outliers. The third was to change the selected ACS estimates being used in the model. The first two had only a minimal effect on changing the design effect and thus will not be discussed further.

It is important to recall that the pubDF is the square root of the DEFF. The DEFF applies to the variance, while the pubDF is used with the SE. The pubDF is the parameter created using modeling and will be discussed below. It is also the parameter published in the PUMS Accuracy of the Data document. The DEFF is not published, but may be obtained by squaring the pubDF.

3.3.1. Using Alternative Detailed Tables and Estimates for Inclusion in the Linear Regression Model.

In order to create a DF candidate, selected estimates from one or more ACS detailed tables are identified for inclusion in the model. While changing the model by adjusting certain national level geographies or ignoring outliers has no appreciable effect on the DF candidate, altering the detailed table or the estimates from a specific detailed table does have a noteworthy effect.

The results of some alternative DF candidates created by restricting the estimates included in the modeling process for poverty are shown in Tables 8 below.

Table 8 shows the results for restricting the estimates input into the model for creating the Poverty DF Candidates. The largest DF candidates are produced using all estimates from the table which are below poverty and excluding all of more detailed estimate for sex by age. Recall that the square of the DF is the DEFF. The DEFF approximates the variance ratio which is the actual DEFF. The largest DF candidates are about 2.6. Squaring the DF of 2.6 yields 6.76. Although this is less than the variance ratio for the percent of people below poverty (8.30) it is still closer than the currently published DF of 1.7, which yields a pubDEFF of only 2.89.

The column for the Selected DF candidate chooses the "best" DF candidate from each model. The previous year's published DF (1.7) is also included in the selection process. The Selected DF is rounded to one decimal place as it would appear in the PUMS Accuracy of the Data document. The DEFF is the selected DF candidate. As can be seen, sometimes the previous year's DF is selected as the 'final' published DF.

Table 8: Alternate Design Factor Candidates for Poverty Using Detailed Table B17001: Poverty Status by Sex by Age

Model No.	Detailed Table and Estimates Used to Create DF Candidate	NAT Cand.	NAT ALL Cand.	Selected DF Cand.	DEFF
Current	All Estimates excluding Total (Current Model)	1.65	1.71	1.7	2.89
1	Male and Female Estimates only	1.66	1.82	1.8	3.24
2	All Estimates which are Below Poverty	2.61	2.42	2.6	6.76
3	All Estimates which are At or Above Poverty	1.46	1.46	1.7	2.89

Source: 2011 1-Year ACS Detailed Tables

Alternate DF candidates were also created for health insurance. The NAT candidates ranged from 0.73 to 2.03 and the NATALL candidate ranged from 0.62 to 2.02 depending on which estimates were included in the modeling process. However, due to the selection process for the Health Insurance DF, the previously published DF (1.3, or a DEFF of 1.69) was chosen regardless of the NAT and NATALL candidates. Additional research is needed to address this issue.

3.4. Other Observations

3.4.1. Searching for Anomalies in Other Percentage Estimates

Table 9 shows a variety of percent estimates using the 2011 1-year PUMS data. They were created to replicate the ACS percents published for various Data Profiles on AFF. The PUMS percent estimate is presented along with the PUMS variance ratio and the pubDEFF. The pubDEFF vary relative to the PUMS variance ratios but in most cases are comparable. The anomaly detected by Boudreaux does not seem to be widespread among other ACS estimates.

Table 9: Selected National Percent Profile Estimates using 2011 1-year ACS and PUMS Data

Characteristic	Percent Estimate (PUMS)	Variance Ratio (PUMS)	pubDEFF ($pubDF^2$)
Family Households (Families)	66.21	1.55	2.89
Population in Households: Spouse	18.29	2.15	2.89
With a Disability	12.10	1.73	1.44
Foreign Born	13.00	2.17	4.00
Speak English less than Very Well	8.66	2.55	2.25
Unemployed	6.51	1.70	1.44
Unemployed with no Health Insurance	46.08	2.02	1.69
Owner Occupied Housing Units	64.68	4.58	2.89
Utility Gas	49.02	1.05	2.25
$150,000 <= Value or Owner Occupied Housing Unit < $200,000	15.33	1.09	1.96
Two or More Races	2.80	4.57	4.84

Source: 2011 1-year PUMS Data

3.4.2. Searching for Anomalies in the Detailed Estimates for Poverty and Health Insurance

Tables 10 through 13 show some detailed percent estimates published for poverty and health insurance.[12] Poverty is a family-based estimate since it is the income and size of the family that determines poverty status. Estimates of poverty for families are given in Table 10 below. Here, the pubDEFF is larger than the PUMS variance ratio, which means it is more conservative. Families are closely associated with households. The household weight and the person weight of the householder are closely linked. The householder's person weight has constraints placed on it during the raking process, which in turn causes the variance ratio for the families to remain low.

Table 10: Percentage of Families With Income in the Past 12 Months is Below the Poverty Level Threshold

Characteristic	Percent Estimate (PUMS)	Var. Ratio (PUMS)	pubDEFF ($pubDF^2$)
All families	11.70	1.71	2.25
With related children under 18 years	18.56	1.88	
With related children under 5 years only	19.47	1.48	
Married couple families	5.81	1.74	
With related children under 18 years	8.76	2.08	
With related children under 5 years only	7.51	1.38	
With related children under 18 years	40.53	1.77	
With related children under 5 years only	47.68	1.47	

Source: 2011 1-Year PUMS Data

[12] Note that the pubDEFF shown in these tables are only for poverty or health insurance. The 2011 1-year PUMS Accuracy of the Data document advises data users to use the highest published DF when creating estimates that cross different DF subject groups. The actual pubDEFF used for more detailed percent estimates may be different from the one shown.

In contrast, when examining the percent of people in poverty in Table 11, we see the pubDEFF is smaller than all but one estimate for the PUMS variance ratio. In particular, the percent estimates that include children such as “all people”, “children under 18”, and “people in families” had PUMS variance ratios at least twice the size of the pubDEFF. For the percentages that include adults 18 and over, the variance ratios and the pubDEFF are much closer.

Recall in Table 4, that the variance ratio for children in poverty was relatively high due to looser constraints placed on the “Other Relationship Statuses” category. We see this in the variance ratio for the percent of people in families is 8.46. This includes children in families and other relatives. The lack of control in the raking process results in higher variance ratios than seen in Table 11.

Table 11: Percentage of People Whose Income in the Past 12 Months is Below the Poverty Level Threshold

Characteristic	Percent Estimate (PUMS)	Variance Ratio (PUMS)	pubDEFF ($pubDF^2$)
All people	15.87	8.30	2.89
Under 18 years	22.33	7.21	
Related children under 18 years	22.02	7.22	
Related children under 5 years	25.67	3.48	
Related children 5 to 17 years	20.65	5.96	
18 years and over	13.83	3.85	
18 to 64 years	14.78	3.88	
65 years and over	9.33	1.28	
People in families	13.33	8.46	
Unrelated individuals 15 years and over	26.97	1.97	

Source: 2011 1-Year PUMS Data

For health insurance coverage, there is a wide range of variance ratios for PUMS data. In Table 12, all the variance ratios are higher than the pubDEFF. However, for the public coverage estimate the difference between pubDEFF and the variance ratio is much smaller.

Table 12: Percentage of People With Health Insurance for Civilian Non-institutionalized population

Characteristic	Percent Estimate (PUMS)	Variance Ratio (PUMS)	pubDEFF ($pubDF^2$)
With Health Insurance Coverage	84.86	6.79	1.69
With Private Health Insurance	65.22	9.95	
With Public Coverage	30.55	2.39	
Under 18 with Health Insurance Coverage	92.52	3.51	

Source: 2011 1-Year PUMS Data

For Table 13 below, we are looking at percent estimates for health insurance coverage for by labor force status. What is interesting about these is that many of the variance ratios are more reasonable in size compared to the pubDEFF. The exceptions are the categories associated with the percent estimates for employed.

Table 13: Percentage of People With Health Insurance for Civilian Non-institutionalized population Aged 18 to 64

Characteristic	Percent Estimate (PUMS)	Variance Ratio (PUMS)	pubDEFF ($pubDF^2$)
Employed with Health Insurance	82.18	5.39	1.69
Employed with Private Health Insurance	77.63	5.19	
Employed with Public Coverage	6.53	1.83	
Unemployed with Health Insurance	53.50	2.08	
Unemployed with Private Health Insurance	33.93	2.21	
Unemployed with Public Coverage	21.82	1.72	
Not in Labor Force with Health Insurance	77.46	2.81	
Not in Labor Force with Private Health Insurance	49.85	2.54	
Not in Labor Force with Public Coverage	34.40	2.19	

Source: 2011 1-Year PUMS Data

3.5. Examining the Coefficient of Variation for Poverty and Health Insurance

The coefficient of variation (CV) is one measure we use to evaluate the reliability of an estimate. The CV is defined as the SE divided by the estimate. The reliability of an estimate is inversely proportional to the size of the CV. An estimate with a small CV is considered more reliable when compared to an estimate with a larger CV given a fixed sample size. There are no exact rules for what constitutes a 'good' CV, however, the smaller, the better. A CV greater than 0.61 implies that a 90 percent confidence interval includes zero, and thus the estimate is unreliable.

Table 14: Coefficient of Variation at the National Level for Poverty and Health Insurance

Characteristic	Percent Estimate	Replicate weight MOE[13]	Design Factor MOE[14]	Replicate Weight CV	Design Factor CV
Persons Below Poverty	15.87	0.0987	0.0576	0.0038	0.0022
With Health Insurance Coverage	84.63	0.0855	0.0428	0.0006	0.0003

Source: 2011 1-year PUMS Data

In table 14 above, the MOEs as well as the CVs are displayed. The MOE is displayed instead of the variance or SE so that comparisons may be made to MOEs on AFF. We see that the replicate weight MOE for both the percent of people below poverty and the percent of people with health insurance coverage is small. Traditionally, for percent estimates on AFF, MOEs are rounded to the nearest tenth. Here, both round to 0.1.

The design effect MOE are about half the size of the replicate weight MOEs, however, the MOE for poverty would naturally round to 0.1 while the MOE for health insurance would be forced to round up to 0.1 since it cannot round down to zero. Thus, the

[13] The MOE is 1.645 times the SE.
[14] The Design Factor MOE was calculated by multiply the SRS SE by the pubDF and 1.645.

differences between the replicate weight MOEs and the design effect MOEs are indistinguishable when reasonably rounded.

Examining the CV, we see that although the replicate weight CV is approximately twice the design effect CV, both are very reliable with standard errors well below 0.01 or 1 percent of their respective estimates. Therefore, we can conclude that despite the replicate weight variance being larger than the design factor variance, the difference in the reliability (based on the CVs) is negligible. This provides some context to the scope of the issue.

4. Conclusions

The anomaly first reported by Boudreaux et al. still exists using the 2011 1-year PUMS data. The sampling and weighting process used to create the PUMS dataset may have a role in the creation of the anomaly. In particular, the raking to control for the spousal equalization causes people with a relationship status other than the reference person, spouse or unmarried partner to have a higher variance ratio. The Replicate Weight variance can be driven by certain cases with a high variability in the weights. These cases are usually from housing units with data collected by the telephone or personal visit interviewing mode. Using the mean value of the replicate weight percent estimates in the calculation of the Replicate Weight variance for the percent of persons below poverty and with health insurance coverage reduces the size of the anomalies. Altering the detailed table(s) and/or the estimates which are used as inputs into the model to create DF candidates can impact the DF candidates and the selected published DF. Careful and thoughtful consideration to which detailed tables used in the process to create DF could bring the DF variance into closer approximation to the RW variance. In summary, for most analysis, the conclusions drawn from using replicate weight based variances or standard errors would not be substantively different from those using design effect based variances or standard errors.

5. References

1. U.S. Census Bureau, 2011. "PUMS Accuracy of the Data Document 2011." Washington, D.C., 2011

2. Lee, Eun Sul, Analyzing complex survey data, 2nd ed. (2006) Sage Publications, pp 22-39

3. Rao, J.N.K. and Shao, J., Modified balanced repeated replication for complex survey data Biometrika (1999) 86 (2): 403-415 doi:10.1093/biomet/86.2.403

4. Boudreaux, M., Davern, M. and Graven, P. (2012) Design Effect Anomalies in the American Community Survey. JSM presentation, Section on Survey Research Methods, VA: American Statistical Association.

5. Kish, L., Survey Sampling, (1965), Wiley, New York

6. U.S. Census Bureau, 2010, ACS Design and Methodology, Chapter 12, Revised

7. Wolter, K. (2007) Introduction to variance estimation, Springer-Verlag, New York