EVALUATION OF SUBNATIONAL ACS FOREIGN-BORN DATA – BENCHMARKING REPORT 1

by

Steven Camarota, Ph.D. and Jeffrey Capizzano*
April 2005

* Steven Camarota, Ph.D. is the Director of Research at the Center for Immigration Studies and Jeffrey Capizzano is the Director of Public Policy and Research at Teaching Strategies. This work was developed under a subcontract with Sabre Systems, Inc. and utilized funds provided by the U.S. Census Bureau. The opinions expressed in this report are those of the authors and do not necessarily express those of either Sabre Systems, Inc. or the U.S. Census Bureau.

INTRODUCTION

Like most large national surveys, the American Community Survey (ACS) employs a complex sampling design to select the housing units that will be included in the sample. Instead of conducting a simple random sample, the U.S. Census Bureau chooses the ACS sample using a two-stage stratified sampling process (ACS Source and Accuracy Statement 2003).[1] Prior to selecting the sample, the geographic area of the United States is split into primary sampling units (PSUs).[2] The PSUs are then combined into sampling strata so that each stratum is as similar as possible across a host of social and economic characteristics. In drawing the ACS sample, pairs of PSUs are first selected from each stratum, and then a sample of housing units is drawn from the selected PSUs.

The two-stage stratified sampling process is used in the ACS to support county-level estimation. The ACS sampling method results in a sample of housing units from 1,240 of the country's 3,141 counties, allowing the U.S. Census Bureau to conduct the ACS in a much more efficient and cost-effective manner than if a simple random sample were used. Given that a non-trivial number of housing units in the ACS sample receive an in-person visit, the fact that sampled housing units are limited to roughly one-third of the counties in the U.S. makes the data collection process significantly easier to manage.

One drawback to the complex sampling design of the ACS is that it often suppresses the amount of variation one would observe if a simple random sample were used. Therefore, the

[1] Starting with the full implementation of the ACS in 2005, the ACS will no longer employ a two-stage sampling design. Instead, a direct, systematic county-based sample of housing units will be drawn for each county in the United States.
[2] PSUs are almost always comprised of either a metropolitan area, a large county, or a grouping of smaller counties, and are always contained within a state boundary.

standard errors that are derived from statistical packages that assume a simple random sample (e.g., SAS) understate the true standard error of an estimate, exaggerating the precision of ACS estimates (Kish 1992; Landis et al. 1982). Luckily, the effects of complex sampling are correctable through the use of methodological techniques that account for the "design effects" (DEFF)[3] associated with complex sampling.

The U.S. Census Bureau currently offers external researchers one set of approaches, called "generalized variance procedures," to approximate unbiased standard errors of estimates derived from the ACS Public Use Microdata Sample (PUMS). This set of approaches allows for the approximation of standard errors of such statistics as totals, means, medians, per capita amounts and ratios. However, while these adjusted standard errors can be used to conduct significance testing between two estimates, the "PUMS Accuracy of the Data (2002)" documentation does not provide the formula that can be used to conduct the significance testing or describe how it is to be used. Moreover, the U.S. Census Bureau currently does not provide variables such as replicate weight, PSU, or strata variables on the PUMS file that would allow for alternative methods of variance estimation.[4]

There may be important implications associated with the fact that the U.S. Census Bureau provides neither the documentation on how to conduct significance testing nor the variables necessary for alternative methods of variance estimation. First, and perhaps most importantly, the lack of documentation on significance testing may limit the utility of ACS data. A primary goal of the ACS is to allow policymakers to more quickly understand and respond to demographic and economic changes occurring at the national, state, and county levels. This goal

[3] The term design effect is used to describe the ratio of the variance of sample estimates for a particular sample design relative to the corresponding variance of a simple random sample with the same sample size.
[4] The U.S. Census Bureau is exploring ways to provide this additional information on the PUMS file and will provide additional sampling information (e.g., PSU and strata variables) by request.

can only be accomplished if researchers outside of the U.S. Census Bureau—particularly those employed by state and county governments—have clearly documented methods available to detect both differences between various sub-populations within their jurisdiction and changing demographic and economic circumstances over time.

In addition, the fact that there is no discussion in the PUMS documentation concerning the impact of ACS design effects on the standard errors of ACS estimates opens up the possibility for errors in the interpretation of survey estimates. As discussed, most statistical programs assume a simple random sample when calculating standard errors and tests of significance. If the complex sampling design of the ACS is not taken into account, significance testing done using these statistical programs will be incorrect. Without clear documentation of this issue and guidance on how to correctly conduct significance testing, a less sophisticated researcher may make the mistake of reporting results that do not take into account the complexities of the ACS sampling design.

Given these potential implications, this paper briefly discusses the challenges and trade-offs from the researcher's perspective of the different techniques that may be used in calculating unbiased standard errors of ACS estimates. We begin by discussing both the publicly available and internal U.S. Census Bureau methods for calculating the standard error of ACS estimates. These include two methods under the "generalized variance" approach (design factors and parameters) and the replication method.[5] We then discuss the strengths and weaknesses of each

[5] Two methods not discussed here are the "direct" approach and the Taylor-series approximation. Using the "direct" method, an unbiased standard error is obtained using the published ACS estimate and confidence interval. One subtracts the lower bound of the estimate's confidence interval from the estimate, and then divides the product by 1.65 (for 90 percent confidence intervals). We do not discuss this approach because this method can only be applied to published ACS estimates and not to those that may be derived by a researcher external to the U.S. Census Bureau. The Taylor-series is not discussed because it has not been used by the U.S. Census Bureau to calculate unbiased standard errors. In the recommendations section of the paper, however, we do discuss the method as an option for

approach from the perspective of a researcher outside of the U.S. Census Bureau. We conclude by making recommendations concerning ways the U.S. Census Bureau can make it easier for an external researcher to calculate unbiased standard errors of ACS estimates, particularly those that can be used in significance testing.

CALCULATING ACS STANDARD ERRORS

Currently, there are at least three methods that have been used by the U.S. Census Bureau to obtain or approximate unbiased standard errors of ACS estimates. The first two methods fall under the category of "generalized variance" procedures, which are outlined in ACS " PUMS Accuracy of the Data" documentation. The first method in this category is called the "design factor" method. In order to approximate the unbiased standard error of an estimate, this method employs the use of a design factor produced by the U.S. Census Bureau that is applied to a "basic" standard error, which is the standard error that would have resulted had the ACS used a simple random sampling process. The second method in this category, called the "parameter" method, employs parameters created by the U.S. Census Bureau that are substituted into a formula to approximate an unbiased standard error. The final method is called the "replication" method, which can be used to calculate a precise standard error of any ACS estimate using replicate weights. Only the generalized variance procedures can be used by researchers outside of the U.S. Census Bureau to calculate the standard errors of ACS estimates. However, the replicate weights necessary for the replication method have not been publicly released. Below, we explain these approaches in greater detail and discuss their strengths and weaknesses from a researcher's perspective.

---

calculating standard errors and recommend that the variables needed for this approach be released on the ACS PUMS file.

The Generalized Variance Approach Using Design Factors

The design factor approach is a two-stage approach used to approximate unbiased standard errors of population totals and percentages derived from ACS data (See Appendix A). The first stage of the approach involves calculating a "basic" standard error of an estimate. This basic standard error is derived either by using a formula or by consulting a "look-up" standard error table. Both formulas and tables are provided by the U.S. Census Bureau. Next, the basic standard error is multiplied by a design factor, which reflects the extent to which the basic standard error is inflated due to the complex sampling design. The product of multiplying the design factor by the basic standard error results in an approximation of an unbiased standard error for that estimate. Design factors are provided by the U.S. Census Bureau for estimates of numerous characteristics (e.g., citizenship status or educational attainment).

The Generalized Variance Approach Using Parameters

The parameter approach is used to approximate unbiased standard errors of aggregates, total population and housing units, means, medians, per capita amounts and ratios (other than proportions). Two different parameter methods are available depending on the type of statistic for which a standard error is needed—the "a, b & c method" and the "a & b" method (See Appendix B). Formulas and parameters for specific characteristics are provided by the U.S. Census Bureau. Researchers interested in the standard error of a specific estimate can look up the parameters for that estimate, which can then be substituted into a formula. Solving the formula using these parameters results in the approximation of an unbiased standard error.

The Replication Method

The method that the U.S. Census Bureau currently employs for computing the standard errors of publicly released ACS estimates is a specific version of the replication method called the “successive difference replication approach.” Replication is perhaps the most conceptually intuitive of the three methods because it mirrors the theoretical approach of calculating standard errors by measuring variation of a statistic across multiple samples. Under this replication method, eighty factors strictly between zero and two are assigned to each housing unit.[6] The base weights are multiplied by these factors, and the weighting process is re-run for each of the eighty sets of replicates. Unlike replication methods like random groups or bootstrapping, no sampling of the collected data is performed. The full sample is used to produce each of the “replicate” estimates.

Each observation on the internal, person-level ACS data file contains the full-sample weight and the 80 replicate weights. To calculate the standard error of an estimate, the full-sample estimate is first calculated using the full-sample weight. Next, 80 “replicate estimates” are calculated by estimating the statistic using each of the replicate weights. The standard error for the full-sample estimate is derived using a formula that incorporates the difference between the full-sample estimate and each replicate estimate.

STRENGTHS AND WEAKNESSES OF EACH APPROACH

The Generalized Variance Methods

There are strengths and weaknesses to using the generalized variance methods to approximate the standard error of an ACS estimate. The fundamental strength of this approach is that it is straightforward and does not require the use of a statistical program to calculate the unbiased standard error. In both the design factor and parameter methods, standard errors can be calculated by hand if necessary, or can be calculated quickly and easily using a spreadsheet.

[6] These factors are based on a Hadamard matrix.

One weakness of the generalized variance approach is that the method by which researchers would use the adjusted standard errors to conduct significance testing is currently not offered in the PUMS documentation. For example, if a researcher wished to understand whether the percentage of foreign-born individuals living in the United States with a high school diploma was statistically different from that of native-born individuals, it is unclear how this test would be done. In addition, given that only one design factor/set of parameters is given for each question in the ACS, it is unclear how good of an approximation the standard errors (and corresponding significance tests) are for specific sub-groups of respondents. For example, the design factor for the characteristic "age"—the factor that inflates the standard error of an estimate to adjust for the design effects of the ACS—is 1.1 in the 2002 PUMS documentation. This design factor is to be used when examining age regardless of whether the level of analysis is at the national, state or county level. However, it is unclear whether this design factor would make the appropriate adjustment to the standard error of an estimate, for example, of the percentage of three- and four-year olds in Rhode Island.

The PUMS documentation is also not clear regarding how one should calculate standard errors of sub-populations with specific characteristics. The PUMS documentation notes that when an estimate is the product of two characteristics, the largest design factor for the combination should be used (with the exception of estimates involving race or Hispanic origin). However, if you subset the data to look at only the foreign-born and estimate the percentage of foreign born with a high school diploma, it is unclear whether the educational attainment design factor should be used, or the larger of the educational attainment and citizenship status design factors. Moreover, the approach of using the larger of two or more design factors for estimates involving more than one characteristic raises questions about the accuracy of the standard error that is derived.

The Replication Method

There are also strengths and weaknesses of the replication method. Perhaps the greatest strength of replication over the generalized variance procedures is that the procedure does not approximate the standard error, but generates a standard error that is specific to any given estimate. In addition, the method of calculating the standard error of an estimate using replication is the same regardless of the characteristic or the sub-population examined.

A fundamental weakness of the replication approach is that calculating the standard error involves some complex computation issues. Calculating a full-sample estimate, 80 replicate estimates and then using these estimates in a formula to derive the unbiased standard error cannot be done by hand. Moreover, most statistical software packages, including SAS, are not equipped to easily handle replicate weights. Therefore, in order to calculate an unbiased standard error in SAS using replicate weights, a macro must be created that allows the program to cycle through the 80 replicates to create 80 replicate estimates, and then these estimates must be substituted into the standard error formula. In addition, given the large sample size of the ACS, the computation time to derive a standard error for each estimate can be quite long, even when using a mainframe computer.

CONCLUSIONS AND RECOMMENDATIONS

As with most large national surveys, the complex sampling process of the ACS can cause issues regarding the calculation of unbiased bias standard errors. Failure to take the ACS sampling design into account can cause misreporting of standard errors and the results of significance testing. This paper has briefly discussed three methods of approximating or deriving unbiased standard errors of ACS estimates. Unfortunately, only the generalized variance procedures are available to researchers outside of the U.S. Census Bureau. We have discussed the strengths and weaknesses of these different approaches should the U.S. Census Bureau decide to further assist researchers in calculating unbiased ACS estimates. Below, we offer recommendations on what should be done to make a greater range of methods available for calculating unbiased standard errors.

Recommendations

1. **The U.S. Census Bureau should more fully articulate the effect of the ACS's complex sampling design on standard errors and confidence intervals in the ACS PUMS "Accuracy of the Data" statement.**

While the ACS PUMS "Accuracy of the Data Statement" discusses the ACS sampling strategy, issues of sampling error, and provides instructions for the calculation of standard errors using the generalized variance procedures, there is not an overt discussion of the effect that the complex sampling design of the ACS has on the standard errors. While this issue may be inferred through the design factor discussion, nowhere does the documentation caution ACS users about how the complex sampling design can bias the standard errors. Such a discussion is important to avoid the potential misinterpretation of ACS estimates, particularly for those researchers who may use statistical software packages like SAS that do not account for complex sampling designs.

2. **The U.S. Census Bureau should provide better instructions on how the design factors and parameters should be used.**

Major points that need to be covered include: what to do if a design factor for a particular characteristic is not available; and, what do if one wants to calculate a standard error of a specific characteristic for a specific sub-population (e.g., low-income, the foreign born, women, etc.). In addition, instructions should be offered on how the adjusted standard errors derived from the design factor and parameter approaches can be used to conduct significance testing.

3. The U.S. Census Bureau should consider releasing the replicate weights on the ACS public use file.

Given the strengths of the replicate weight approach for calculating standard errors, researchers should have the opportunity to use this approach. Therefore, the replicate weights should be released on the ACS PUMS file along with documentation outlining how they should be used. In addition, the U.S. Census Bureau should provide statistical programs in SAS, SPSS, STATA and other programming languages that create the standard errors using replicate weights. In addition, programs should be provided that allow researchers to conduct simple significance tests between two sub-populations using the replicate weights. These programs should be well documented and user friendly, allowing to the extent possible, researchers to simply substitute variables they wish to estimate for the sub-population of interest.

4. The U.S. Census Bureau should consider providing sample design variables (PSU and strata) on the ACS PUMS file to allow researchers to use a Taylor-series approximation or other related linearization methods to calculate unbiased standard errors

One method of calculating standard errors not employed by the U.S. Census Bureau for ACS estimates (and therefore not discussed in this paper) is the Taylor-series approximation. This method of calculating standard errors has certain distinct advantages over the types of methods currently used by the U.S. Census Bureau. Most importantly, both SAS and STATA statistical software packages have commands that can employ this method. To use this method, variables related to the sampling method—the PSU and STRATA variables—must be made available on the PUMS data file. While ACS staff have indicated that these variables will be made available by request, the U.S. Census Bureau should consider making these variables, along with the replicate weights, available on the ACS PUMS file.

5. The U.S. Census Bureau should provide the resources to allow the ACS staff to implement the recommendations outlined in this paper.

The ACS staff noted that they are aware of the need to expand on the ACS documentation. However, ensuring the quality and timeliness of the ACS data have to take priority over making changes to improve the documentation. Therefore, if the US Census Bureau would like to see the recommendations in this paper implemented, it will need to provide the ACS staff with additional personnel and other resources as necessary.

APPENDIX A:
Design Factor Method[7]

The design factor method can be used to approximate the standard errors of most sample estimates of *totals* and *proportions* using the following formula (for totals) and the design factors provided by the U.S. Census Bureau:

$$se(estimate) \approx 1.2 * DF * \sqrt{142 * Y(1 - Y / N)}$$

Where: DF = Design Factor, N = Size of Geographic Area, and Y= Estimate of Characteristic

The following formula is used for percents:

$$se(estimate) \approx 1.2 * DF * \sqrt{\frac{142}{B} * p(1 - p)}$$

Where: DF = Design Factor, B = Base of Estimated Percentage, and p = Estimated Percentage

[7] For more information on this method, see U.S. Census Bureau. 2003. *Accuracy of the Data (2003).* http://www.census.gov/acs/www/Downloads/ACS/accuracy2002.pdf.

APPENDIX B:
a, b, & c and a & b Parameter Methods[8]

The U.S. Census Bureau provides formulas and parameter tables to approximate the standard errors of aggregates, total population, and housing units, as well as means, medians, per capita amounts, and ratios (other than proportions). To approximate the standard errors of aggregates, total population, and housing units, use the following a,b,& c Parameter Method formula:

$$se(estimate) \approx 1.2 * \sqrt{a + b * estimate + c * estimate^2}$$

Tables are provided in the "PUMS Accuracy of the Data" documentation in order to look up the appropriate a, b, & c parameters for the characteristic that is being estimated.

To approximate the standard error of means, medians, per capita amounts, and ratios (other than proportions), use the following a & b Parameter Method formula:

$$se(estimate) \approx 1.2 * \sqrt{a + b * Log(N)}$$

Where: LOG is the natural log function and N is the universe count for medians, or the denominator of the estimate for means, per capita amounts and ratios for the geography(ies) you are interested in.

The one exception to using this method is when the denominator of a mean, per capita amount, or other ratio is zero.

[8] For more information on this method, see U.S. Census Bureau. 2003. *Accuracy of the Data (2003).* http://www.census.gov/acs/www/Downloads/ACS/accuracy2002.pdf.

REFERENCES

Kish, L. “Weighting for Unequal Pi.” 1992. *Journal of Official Statistics*, 8, 183-200.

Landis, R.J., Lepkowski, J.M., Eklund, S.A., and Stehouwer, S.A. 1982. “A Statistical Methodology for Analyzing Data from a Complex Survey: The First National Health and Nutrition Examination Survey.” *Vital and Health Statitistics.* 2:92.

U.S. Census Bureau. 2002. *Accuracy of the Data (2002)*. http://www.census.gov/acs/www/Downloads/2002/AccuracyPUMS.pdf

U.S. Census Bureau. 2003. *Accuracy of the Data (2003)*. http://www.census.gov/acs/www/Downloads/ACS/accuracy2002.pdf