Developing Replicate Weight-Based Methods to Account for Imputation Variance in a Mass Imputation Application*

Mark E. Asiala and Edward C. Castro†

Abstract
The Census Bureau is currently developing a new imputation-based methodology to improve the estimates of the group quarters population for small areas. As a part of that work, a new variance estimation methodology was needed that would properly account for the imputation variance component in addition to the sampling variance component. Furthermore, any methodology would need to be incorporated into the American Community Survey (ACS) replicate weight variance methodology in order to minimize the impact on production tabulation systems. A benchmark was established using multiple imputation techniques for a fixed sample. We then compared alternative methods that attempted to incorporate the full measure of variance which incorporates the imputation variance into the replicate weights. Details on how the replicate weights are computed are described. This work led to a recommendation of a variance estimation methodology used in the 2011 ACS data products.

Key Words: variance estimation, multiple imputation, replicate weights

1. Introduction

The group quarters (GQ) population is only a small component of the total population nationally at approximately three percent. However, at substate and subcounty geographies such as census tract, the GQ population can be a substantial proportion of the population. The current statistical design for the production of American Community Survey (ACS) group quarters estimates is to support the GQ profile data product that is released only at the state and national levels. However, the GQ data are also included in the production of estimates for the characteristics of the total resident population in substate areas. This conflict between the design and ultimate use of the GQ data can lead to substantial variances in the substate estimates of both the GQ population and the total resident population.

In response to this issue, the Census Bureau has conducted a research project exploring how this variance issue can be mitigated via estimation techniques since optimizing the sample design for GQs for small areas is not fiscally feasible. As a result of that project, the Census Bureau is in the process of implementing a new estimation methodology for the GQ data. This process involves a mass imputation application where the interviewed GQ data are imputed into the a subset of the not-in-sample GQs on the GQ sampling frame. As a result, the substate distribution of the GQ population in the estimates better reflects the distribution on the frame. Details of this methodology can be found in Asiala, et al. (2012).

It is the Census Bureau's policy to publish the margin of error with every estimate that is produced from the ACS. With this new methodology, the level of imputation approaches

*This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

†US Census Bureau, 4600 Silver Hill Rd, Washington, DC 20233

a near doubling of the number of records. Thus the total variance for estimates which are based on the new GQ estimation methodology will contain a non-negligible amount of imputation variance in addition to the sampling variance. This paper details the methodology that was developed for the 2011 ACS data products to account for this imputation methodology within the current variance estimation framework.

The remainder of the paper provides additional background on the GQ estimation methodology and the methodology used to estimate and incorporate the variance due to imputation in our final estimates of total variance.

2. Imputation and Weighting Methodology

2.1 Background

With the existing GQ sample design, there are no controls in place to ensure that, for a given substate area, a GQ will be selected from the frame to be included in the ACS sample. Furthermore, even if a GQ is selected in the area, the sampled GQ facilities may not reflect the variety of GQs present on the frame. For example, an area may contain a mixture of types of facilities including correctional institutions, nursing homes, or military facilities. Simply selecting a GQ of one of those types would not give an accurate portrayal of the mixture of GQ types present.

The existing weighting process is thus limited in its ability to form representative estimates for these substate areas. While the current process uses independent estimates as controls within cells formed by state cross by major type (see Table 1), it is clear that even if we had substate estimates for use as controls, we would still need to implement a substantial collapsing routine that could handle the areas where we either have no representation within the area by major type or no representation at all. The level of collapsing necessary, given the sample design, would cause biases that would essentially just translate our variance issue to an issue of substantial bias.

Table 1: Definition of Group Quarter Major Types

Major Type	Definition
1	Adult Correctional Facilities
2	Juvenile Facilities
3	Nursing Homes
4	Other Institutional
5	College Dormitories
6	Military
7	Other Noninstitutional

2.2 Basics of Imputation Methodology

It was within this context that the new methodology was developed. If we could impute records into the not-in-sample GQs present on our frame—at least a sufficient number to achieve some base substate representation—then we could make use of lower level constraints to improve the substate variances. The goal of the imputation methodology is two-fold:

1. The primary objective is to establish representation of county by major type of GQ in the tabulations for each combination that exists on the frame.

2. A secondary objective is to establish representation of tract by major type of GQ for each combination that exists on the frame, as is reasonably feasible.

To accomplish these goals, we decided that for all "large" GQs with an expected population of 16 or more persons, we would impute the equivalent of a 2.5% sample of GQ persons into the GQ. For those "small" GQs with an expected population of fewer than 16 persons, we would select a random sample subset of those GQs only as needed in order to accomplish the objectives listed above. For those small GQs that are selected, a 20% sample of GQ persons is imputed into the GQ. The primary reason for the differing treatment by size was that the large GQs have a more substantial impact on the estimates than the small GQs. Thus, if GQs from this set are missing from the sample, they are more noticeable. A second reason is that these larger GQs tend to be more stable than the small GQs which come in and out of existence more frequently.

Once the not-in-sample GQs are selected, interviewed GQ records are selected at random to impute into the selected GQs. The methodology is an expanding search algorithm that initially looks for donors within the same specific type (e.g., federal prison) and the same county. If that fails then the search includes all GQs of the same major type (e.g., all adult correctional facilities). If that still fails, then the search expands to a specific type within a larger geography, then major type within that geography, and so on until suitable donors can be found. Part of the methodology places restrictions on the minimum size of the donor pool as well as the maximum number of times a specific donor can be used. This restriction is a major reason why the search may need to expand beyond the county or the specific type.

2.3 Basics of Weighting Methodology

The weighting methodology for the new imputation method makes a clean departure from the design-based method previously used. The new methodology contains three steps: the assignment of initial weights that also account for any nonresponse, a series of constraint adjustments that calibrate the weights to the substate frame totals by major type, and lastly, a poststratification adjustment to a set of independent GQ population estimates at the state by major type level. In the weighting, no distinction is made between sampled or imputed records; all are treated exactly the same.

The initial weights are constructed such that all GQ records are equally weighted within a GQ. Records attached to large GQs are self-representing of that GQ and are simply equal to the expected or observed population of the GQ divided by the number of records. For the small GQs, the initial weights are formed to make the collection of small GQs to be, in expectation, self-representing within the tract by major type. Their weights are simply equal to inverse of the fraction of the small GQs on the frame that are represented in the data file (either sampled or imputed).

These initial weights are then adjusted by a tract-level constraint formed by summing the populations from the frame for all tract by major type combinations. The primary purpose of this constraint is to correct for the extent to which the total weight of the small GQs does not match the frame. These adjustments tend to be small. There are also county-level and state-level constraint adjustments. These adjustments correct for GQs that were not selected from the frame for the imputation process because there is no tract code assigned to them. Since all GQs that were a part of the decennial census process are geocoded to a tract, this is a small issue currently. However, we receive a small number of updates to the frame each year that are not geocoded so this step helps to protect against losing frame coverage as a result of those updates.

Finally, we control the final weights to an independent set of GQ population estimates produced by the Population Estimates Program. These controls are for state by major type and are formed by making updates to the decennial census counts that come from the states. Once these controls are applied, we have our final weights.

3. Research Questions

Under the previous design-based weighting, we used successive difference replication (see Fay & Train, 1995) to estimate the sampling variance. Applying this method naïvely to the set of sampled and imputed records would result in a significant underestimate of the total variance. For this reason, we sought ways on how we could adjust this variance estimate to account for the imputation variance. We are presently limited to a system of variance estimation that is replicate weight based due to our tabulation system. Given that limitation and our available development time, we decided to employ a relatively simple process of inflating the naïve variance estimate using a set of design factors. In the context of repeated or multiple imputation, this naïve sampling variance is known as the within imputation variance. The between imputation variance is the variance among the resulting estimates stemming from several, repeated imputations (see Rubin, 1996). With those items defined, we set the following research questions:

1. How does the estimate of total variance (within and between imputation variance) for the imputation-based estimates compare to the design-based estimates?
2. How does the ratio of the estimate of total variance to the within imputation variance vary as a function of geography, major type, and characteristic?

4. Research Methodology

The general methodology for this research was to first create a set of variance estimates for the within and between imputations by creating 40 imputations. While this is not practical in production, it provided a benchmark for which to generate and evaluate a method that could be used in the production of the official ACS estimates of total variance. From these data, a single set of inflation factors were created that would be used to adjust the variances for all characteristics. Finally, these approximated variances were compared to the benchmark to evaluate their effectiveness.

4.1 Create Forty Imputed Datasets

The basic methodology to prepare our datasets was to first parameterize the random seeds for each step in the imputation module where a random draw is made. We then varied the random seeds for each of the 40 imputations. As a result, each imputation contains a potentially different set of GQs selected for imputation and a different set of donors imputed into each not-in-sample GQ. The expected size of the GQs does not vary.

Once the 40 imputations are complete, we create the edited microdata for each imputation and then fully weight each of the resulting datasets. In addition, a set of replicate weights are created for each record using successive difference replication which treats both the sample interviews and the imputed records as though they are all sample records. These replicate weights form the basis for estimating the within imputation variance of the next section.

4.2 Creating Estimates and Variance Estimates

We then calculated estimates and variance estimates using each dataset for a broad set of characteristics at the county by major type level. The table S2601A is a profile of a variety of characteristics for the GQ population which is published at the state level. From this table, we extracted all estimates of counts excluding estimates of percentages, means, or medians. We then redefined any overlapping estimates so as to estimate mutually exclusive characteristics. For example, instead of forming estimates in the categories for educational attainment of high school or higher and bachelor's degree or higher, we formed estimates for (1) high school or higher but below bachelor's degree and (2) bachelor's degree or higher. This yielded a total of 70 estimates per county by major type combination divided into 15 characteristic groups. For a complete set of estimates, see Table 7.

Once the estimates and variance estimates were calculated, we then use the replicated imputation formula (see Rubin, 1996)

$$\operatorname{Var}\left(\bar{\theta}\right) = \hat{W} + \left(1 + m^{-1}\right) B = \frac{1}{m}\sum_{j=1}^{m} \operatorname{Var}\left(\theta_j\right) + \frac{m+1}{m}\sum_{j=1}^{m}\frac{1}{m-1}\left(\theta - \bar{\theta}\right)^2 \qquad (1)$$

where the first term is the average within variance and the second term (inside the summation) is the variance between imputations. In our case $m = 40$. The first term estimates the within imputation variance using the successive difference replication methodology. The second term estimates the between imputation variance by calculating an S^2 type variance among the imputed estimates.

In a true multiple imputation application the imputations would be a model of the true population. Since our application is a mass imputation taken from the sample records, our imputations are conditional on the sample selected and not the population as a whole. Thus, this may understate the true variance in the population to the degree that our nonresponse is not missing at random.

4.3 Creating Generalized Design Factors

Due to constraints of our tabulation system, we cannot adjust for the variance due to imputation differently for every characteristic or combination of characteristics. Therefore our goal is to estimate one overall adjustment for the increase in variance due to imputation per state by major type. The cell definitions were chosen based on the observation that there was considerable variation in the degree of imputation by state and major type and the fact that the data could be relatively thin at lower levels of geography such as county.

To accomplish this, we separated the problem into two steps: first, create a design factor for each characteristic at the state by major type level across counties and second, create an average design factor across the characteristics.

4.3.1 Characteristic level Design Factors

For a given characteristic, state and major type of GQ, all county-level estimates for data items within the characteristic which were greater than 40 were included in the data set as input to the generalization. The rationale for this was that the average weight for a GQ record was approximately 20, so by using a minimum threshold of 40 we would expect that the estimate was formed on average by two or more records. This restriction was critical given the sparseness of the GQ population particularly when further broken out by major type. A no-intercept regression line was then fitted to the data where the independent variable was the estimated within imputation variance and the dependent variable was the estimated total variance including the between imputation variance. The design factor was then defined as the slope to the fitted regression line. Since the within imputation variance is also a component of the numerator, this slope is always greater than or equal to one. The degree that the slope exceeds one is determined by the size of the between imputation variance relative to the within imputation variance.

The dataset was then evaluated and all records whose studentized residual was greater than four or whose relative absolute deviation between the predicted total variance and the estimated total variance was greater than 30 percent were removed and the process of fitting the regression line was repeated. This was done to remove extreme values that could skew the average. This loop was repeated a total of four times in order to improve the robustness of detecting the extreme values in the dataset. In general, most extreme values were identified in the first or second pass. The third and fourth passes tended to identify only a few marginal cases.

This process was also performed at the national level across all counties. This national level characteristic design factor was used in those rare cases where the removal of the small estimates for the county by type group resulted in no counties contributing towards the calculation of the design factor. In those instances, the national-level design factor was used for that characteristic.

The entire process was then performed for each characteristic, within each major type within each state.

4.3.2 Overall Design Factors

The goal for the overall design factors was to be able to make one adjustment that would approximate the characteristic-level design factors while being slightly conservative to avoid significantly underestimating the variances for some characteristics. For this reason, we decided that the average design factor, across characteristics, for a given state / major type combination may not be the best choice given that it would tend to underestimate the variances for approximately half of the characteristics.

The overall design factor thus was calculated using two additional considerations. The first was that the average was computed as a weighted average based on the number of data items within the characteristic. This was done to provide some relative weight based upon the detail of the characteristic as used in S2601A (U.S. Census Bureau, 2011). The second consideration was combining this weighted average with the maximum characteristic design factor by using $\min(\text{weighted average} + 0.2, \text{maximum design factor})$. This allowed the overall design factor to deviate from the weighted average up to smaller of 0.2 or the maximum characteristic design factor. The distance of 0.2 covered the difference between the maximum design factor and the weighted average approximately 95% of the time. The result was that overall design factor for each state / type combination was greater than approximately 90% of the characteristic design factors.

4.4 Integrating Design Factors into Replicate Factors

The method used to integrate the overall design factor into the replicate factor assignment is one that has been used for integrating the finite population correction factor in the ACS replicate weighting since the 2005–2009 ACS 5-year weighting and the Census 2000 long form before that (see, Gbur & Fairchild, 2002 and Starsinic, 2011).

In the successive difference replication, replicate factors $f_{i,j}$ ($i =$ replicate number, $j =$ sample unit number) are defined as

$$f_{i,j} = 1 + 2^{-1.5} a_{i1,j} - 2^{-1.5} a_{i2,j} = \begin{cases} 1 \\ 1 + 1/\sqrt{2} \\ 1 - 1/\sqrt{2} \end{cases} \tag{2}$$

where $a_{i1,j} = \pm 1$ and $a_{i2,j} = \pm 1$ are cells taken from the Hadarmard matrix. While in the application cited above, they applied the finite population correction factor to decrease the variance, we used the design factor to increase the variance by defining new factors, $f^*_{i,j}$, as follows

$$f^*_{i,j} = 1 + \left(2^{-1.5} a_{i1,j} - 2^{-1.5} a_{i2,j}\right) \times DF_{s,g} \tag{3}$$

where $DF_{s,g}$ (s=state, g=major type) is the design factor between the estimate of total variance (including the imputation variance) to the within imputation variance (the naïve sampling variance). Starsinic (2011) then shows, under certain assumptions, that the resulting variances are adjusted, in this case, by the value of $DF_{s,g}$ as though it had been applied to the final variance estimate.

There is one limitation to this method, however. If the value of $DF_{s,g}$ is greater than the square root of two then $f^*_{i,j}$ would be less than zero. Thus, we performed a check to determine if the overall design effect was greater than this value and, if so, we limited the maximum value to be 1.4.

5. Results

5.1 Comparison of Imputation-based Variances to Design-based Variances

In Table 2, we summarize the median ratio of the state-level variance of the average imputed estimate to the variance of the design-based estimate grouped by data item group and major type. While, generally, the imputation based estimates had a smaller estimated variance it does vary within characteristic group and within type. Age and sex showed strong reductions in particular which is expected given that certain GQs types are dominated by a certain age category or gender. For example, college dorms are concentrated in the 18–24 range and adult correctional facilities are generally sex specific. In contrast, citizenship and year of entry tend to not show deep reductions in variance across all types as concentrations of these characteristics are more type-specific.

This part of our analysis was mostly informational. While we expected some decrease in variances, we did not have a clear expectation of the degree of reduction. Generally, we expected the ratio of the variances to be less than one at the state level given the use of the updated frame as covariate information. With the bulk of the ratios between 0.50 and 1.00 with some explainable exceptions, we generally considered these to be reasonable.

Table 2: Median Ratio of State-level Variance of Average Imputed Estimate to Variance of Design-based Estimate

Characteristic	Major Type						
Group	1	2	3	4	5	6	7
Age	0.93	0.61	1.00	0.62	0.66	0.49	0.49
Sex	0.38	0.26	0.80	0.30	0.34	0.44	0.25
Race	0.84	0.94	1.07	0.85	0.71	0.73	0.57
Hispanic Origin	0.86	0.86	0.89	0.69	0.80	0.69	0.58
Marital Status	0.89	0.79	0.95	0.56	0.91	0.52	0.43
School Grade Attending	0.87	0.37	0.62	0.28	0.78	0.43	0.45
Educational Attainment	0.86	0.71	0.88	0.51	0.47	0.53	0.36
Disability	0.71	0.41	0.62	0.32	0.73	0.54	0.23
Citizenship	0.97	1.16	1.23	0.86	0.78	0.75	0.73
Year of Entry	0.97	1.13	1.25	0.58	0.74	0.76	0.80
World Region of Birth	1.02	1.15	1.21	0.75	0.78	0.78	0.88
Language Spoken at Home	0.88	0.77	0.83	0.68	0.67	0.61	0.50
Employment Status	Not Applicable				0.58	0.40	0.30
Occupation Category	Not Applicable				0.71	0.38	0.47
Wages	Not Applicable						0.31

5.2 Ratio of Total Variance to Within Imputation Variance

5.2.1 Ratio of total variance to within variance by geography, major type, and characteristic

With the first subsection focusing on placing the estimated total variance in some context, our next step was to evaluate how the ratio of the total variance to the within imputation variance compare across geographic level, major type, and characteristic.

For this analysis, we focused on national- and state-level estimates of variances. Taking the median across estimates within a characteristic group and across geographies (as appropriate), we constructed a table of median ratios for cells defined by characteristic group and major type. We used an ANOVA test to determine which dimension best explained the variation. At both the national- and state-level, the characteristic group was not a significant source of the variation whereas major type was consistently significant. At the state level, the R^2 for the characteristic group dimension was 0.117 whereas the R^2 for the major type dimension was 0.815.

The result of this analysis helped solidify the importance of using major type in our final overall design factors. It should be noted, however, that when this same test was performed at the county-level the R^2 values were considerably lower, which suggests room for further improvement.

5.2.2 Ratio distribution generalized across county within state by major type

At the county level, the within imputation variance is significantly higher than at the state level. Thus the increase in variance due to the imputation is much smaller relative to the within imputation variance at the county level. The generalization process of obtaining

Table 3: Mean State-level Generalized Characteristic Design Factor by Major Type

Characteristic Group	Major Type 1	2	3	4	5	6	7
Age	1.10	1.17	1.27	1.18	1.02	1.21	1.25
Sex	1.00	1.07	1.18	1.10	1.09	1.16	1.12
Race	1.02	1.10	1.02	1.07	1.03	1.19	1.04
Hispanic Origin	1.06	1.19	1.09	1.09	1.08	1.22	1.12
Marital Status	1.07	1.02	1.25	1.11	1.01	1.17	1.14
School Grade Attending	1.01	1.13	1.00	1.04	1.01	1.15	1.04
Educational Attainment	1.06	1.04	1.22	1.13	1.02	1.13	1.17
Disability	1.04	1.14	1.12	1.12	1.02	1.12	1.15
Citizenship	1.02	1.03	1.04	1.03	1.04	1.15	1.04
Year of Entry	1.02	1.03	1.04	1.03	1.04	1.15	1.05
World Region of Birth	1.01	1.02	1.03	1.03	1.03	1.11	1.04
Language Spoken at Home	1.03	1.09	1.06	1.06	1.05	1.17	1.06
Employment Status	Not Applicable				1.12	1.12	1.12
Occupation Category	Not Applicable				1.19	1.21	1.18
Wages	Not Applicable						1.02

a single factor per characteristic group for each state by major type across all counties within the state also eliminates the extreme values that occur for certain counties. Table 3 shows the mean across all states of these state-level generalized characteristic design factors. Similar to what was seen in Table 2, major type continues to be a strong predictor whereas the effect of a characteristic group tends to be localized to certain major types rather than being high for all major types.

5.3 Summary of Overall Design Factors

5.3.1 General Distribution

In Table 4, we see the distribution of the overall state-level design factor by major type (which include DC and PR). The distribution of the overall design factors is roughly symmetric and has a mean of 1.27 and a median of 1.3. Of the 82 state-major type combinations whose overall factor was 1.4, 29 of those had the value of 1.4 because their value was capped due to the limitation of our adjustment method for the replicates. For these 29, generally they were based on relatively few observations which led to high imputation variances. For example, some cases were based on only one GQ and for these cases our within imputation variance was relatively small. In our assessment, the use of a cap in these instances was reasonable.

5.3.2 Evaluation of Overall Design Factor to the Characteristic Design Factors

As can be seen in Table 5, in 91.5% of the state by major type combinations no characteristic design factor was higher than the overall design factor (they would round to the same value). In the remaining cases, most had only one or two characteristic design factors that were higher. In some rare cases where the sample size was small, the effect of capping the

Table 4: State-level Design Factor by Major Type

Major Type	1.1	1.2	1.3	1.4	Total
1	25	22	4	1	52
2	1	26	14	11	52
3	1	2	34	15	52
4	6	20	15	11	52
5	2	31	15	4	52
6	4	17	9	22	52
7	1	9	24	18	52
Total	40	127	115	82	364

Table 5: Number of Characteristic DFs Higher Than the Overall DF

Above	Frequency	Percent
0	333	91.5
1	8	2.2
2	9	2.5
3	2	0.6
4	2	0.6
5	2	0.6
6	1	0.3
7	3	0.8
10	1	0.3
11	1	0.3
13	2	0.6

overall design factor at 1.4 lead to a majority of characteristic design factors being higher than the overall factor.

To understand these situations more, we broke out these 31 combinations by major type to determine if there was any pattern. In Table 6, we see that major type 6 (military) had the highest concentration of these cases. The states where these occurred typically had only one or two military GQs in them and was consistent with the general pattern of high calculated design factors for cases with small sample sizes. Most of the remaining instances were spread out over the types 2–4 and involved similar small sample situations.

Finally, we also evaluated whether certain characteristics tended to have higher characteristic design factors than others for these combinations. The characteristic that stood out was age, which was present in 17 out of the 31 combinations. In most situations this came from a few age ranges that had estimates of high variance that were outside of the bulk of the distribution. For example, the estimate of persons outside of the 18–24 age group in military GQs tended to have higher variance and in some cases this lead to a higher characteristic DF for that state for the military GQs.

Table 6: Distribution by Major Type of State-Major Type Combinations Where at Least One Characteristic DF is Greater Than the Overall DF

Major Type	Frequency	Percent
1	0	0.00
2	5	16.13
3	4	12.90
4	6	19.35
5	0	0.00
6	15	48.39
7	1	3.23

6. Conclusions and Future Research

The resulting design factors developed in the work were ultimately incorporated into the ACS variance estimation for 2011. While we noted that there are improvements which can be made to our variances estimation, this new method gives us a good start in our efforts to estimate the total variance including the imputation variance for our new GQ estimation method. Some areas which we would like to improve upon is tailoring the adjustment so that it can take into account the degree of imputation in a given substate area (e.g., county or tract). This should improve our variance estimates particularly when an area may be based entirely on sampled data and should not need to account for the additional variance due to imputation since there is none.

Another area we wish to explore is to find a variance estimation method that is better suited to our application where our imputations are conditional on the sample. In this respect, our imputation is akin to a hot-deck imputation except that we impute whole records rather than simply item imputation. Methods described in Shao (2002) and, more recently, Shao and Tang (2011) show considerable promise if these methods can transfer to our application.

REFERENCES

Asiala, M., Beaghen, M., Navarro, A., and Weidman, L. (2012). "Evaluations of Imputation Methods to Improve the American Community Survey Estimates of the Group Quarters Population for Small Geographies". *Proceedings from the 2012 Federal Committee on Statistical Methodology Research Conference.*

Fay, R., and G. Train. (1995). "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties". *Proceedings of the Section on Government Statistics.* Alexandria, VA: American Statistical Association, 154–159.

Gbur, P. and Fairchild, L. (2002). "Overview of the U.S. Census 2000 Long Form Direct Variance Estimation", *Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association, 1139–1144.

Rubin, D. (1996). "Multiple Imputation After 18+ Years", *Journal of the American Statistical Association*, 91, 473–489.

Starsinic, M. (2011). "Incorporating a Finite Population Correction Factor into American Community Survey Variance Estimates", *Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association 3621–3631.

Shao, J. (2002), "Replication Methods for Variance Estimation in Complex Surveys with Imputed Data", *Survey Nonresponse*, edited by Groves, R., Dillman, D., Eltinge, J., and Little, R., published by John Wiley & Sons, Inc.

Shao, J. and Tang, Q., (2011), "Random Group Variance Estimators for Survey data with Random Hot Deck Imputation", *Journal of Official Statistics*, Vol. 27, No. 3, 2011.

U.S. Census Bureau. (2011). "2010 ACS 1-year estimates, S2601A", American Fact Finder, factfinder2.census.gov, retrieved 09/27/2012.

Table 7: Characteristic Groups

Characteristic	Estimate Categories
Age	0–14, 15–17, 18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 74–84, 85+
Sex	Male, Female
Race	White Alone, Black Alone, AIAN Alone, Asian Alone, NHPI Alone, SOR Alone, Two or More Races
Hispanic Origin	Hispanic, Not Hispanic
Marital Status	Age less than 15, Married, Widowed, Divorced, Separated, Never Married
School Grade Attending	Not in school, K–12, College or Graduate School
Educational Attainment	Less than high school, High School or higher but less than Bachelor's Degree, Bachelor's degree or higher
Disability	No disability ; With a Disability, age 3–17 ; With a Disability, age 18–64 ; With a Disability, age 65 or older
Citizenship	Yes, born in the US; Yes, born in Puerto Rico, etc.; Yes, born abroad of American parent(s) ; Yes, naturalized ; Not a citizen
Year of Entry	Not in universe ; 1886–1989 ; 1990–1999 ; 2000 or later
World Region of Birth	Not Foreign Born (or born at sea) ; Europe ; Asia ; Latin America ; Other
Language Spoken at Home	English Only ; Language other than English but Speaks English "Very Well" ; Language other than English but Speaks English less than "Very Well"
Employment Status	Age less than 16 ; In labor force ;Civilian labor force ; Employed ; Unemployed ; Armed Forces ; Not in labor force
Occupation Category	Management, professional, and related occupations ; Service occupations ; Sales and office occupations ; Farming, fishing, and forestry occupations ; Construction, extraction, maintenance, and repair occupations ; Production, transportation, and material moving occupations
Wages	Age less than 16 ; With Earnings