Evaluations of Imputation Methods to Improve the American Community Survey Estimates of the Group Quarters Population for Small Geographies

Mark Asiala, Michael Beaghen, Alfredo Navarro, Lynn Weidman
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract
The Census Bureau has recently developed a new methodology to improve the American Community Survey (ACS) estimates of the group quarters (GQ) population for small areas. What motivated this work is that while the ACS GQ sample was designed to produce estimates at the state level, the estimates of the GQ population contribute to ACS estimates of the total resident population for substate areas such as counties and tracts. Consequently, there are small geographies which either do not have GQ sample or have GQ sample that is not representative of the area, which can lead to distorted estimates of characteristics and/or total population for these geographies. The approach taken is to impute whole person records (and weight them appropriately) to selected GQ facilities which appear on the sampling frame but were not selected into sample. This paper examines the results of this method using real ACS data, including comparing the results of the new methodology with those of the current ACS methodology.

Key Words: sample design, small area estimation

1. Introduction

The Census Bureau has undertaken a research program aimed at improving the American Community Survey (ACS) estimates of characteristics of the group quarters (GQ) population for substate geographies such as counties and tracts. The resulting new estimation methodology for GQ population will be implemented starting with the ACS estimates produced in 2012, that is, the 2007-2011 5-year, the 2009-2011 3-year, and 2011 1-year estimates. The development of this methodology was spurred both by limitations in the usability of the ACS data pointed out by ACS data users, and by long-term concerns from within the Census Bureau about the design of the ACS GQ sample and its weighting. At the heart of the matter was that the ACS sample design and weighting were designed to produce state-level estimates of characteristics of the GQ population, whereas estimates of the GQ population contribute to substate estimates of the characteristics of the total resident population. ACS estimates of characteristics of the GQ population are published for states and larger geographies, but not for substate geographies. However, ACS data products which include GQ population are released for substate areas as small as block groups.

We focused our resources on developing a new estimation methodology because we do not have other good alternatives at this time. No changes in the GQ sample design could be made quickly enough to remedy the problem for the ACS estimates produced in 2012. Further, the sample for GQ persons is fixed by budget constraints, and any changes to the sampling plan that increase the number of GQs requiring visits by interviewers would increase the cost of the survey. Also, while publishing estimates for only the household population for substate areas was an option, it was not appealing, as data users expect estimates for the total resident population, such as had been provided by the Census 2000 sample (long form) data. The approach we developed involved imputing GQ person records into facilities that are on the ACS sampling frame but not selected in sample.

The bulk of the paper describes the methodology and results of two evaluations. The first evaluation was a simulation based on Census 2000 decennial data. The second used 2006-2009 ACS data and the 2010 ACS sample frame to mock up 2010 ACS data. The first sections are introductory, starting with a description of the problem and the goals of the new methodology, and moving on to a description of the imputation methodology. While this introductory overview has appeared in other documents (Asiala, Beaghen, Navarro, 2011), it is repeated here as necessary background. In Section 2 we describe some aspects of the ACS GQ sampling and estimation processes and review evidence illustrating the concerns that motivate this research. In Section 3 we show gaps in representation of the ACS GQ sample across tracts and counties. Section 4 describes the general approach of the new imputation-based methodology. Section 5 discusses the evaluation study with simulated data, Section 6 discusses the improvements made to the method based on the results of this evaluation, and Section 7 discusses the second evaluation which uses ACS data. In Section 8 we give an overview of further evaluations.

2. ACS Sampling and Estimation for the Group Quarters Population

For a better understanding of the issues in this paper, some description of the ACS GQ sample design and estimation is needed. Of salience is that the sampling and estimation methodologies for the GQ population are designed to produce optimal state-level estimates, as it is only for states or larger geographies that estimates of the characteristics of the GQ population are produced. Only the estimates of the total GQ population are published for geographies smaller than the state. While the sample stratification includes type of GQ and geography, the sampling rates are such that many counties and tracts do not have sample for particular major types of GQ which nevertheless exist within them. Further, the GQ population estimates are controlled at the state level, whereas the ACS estimates of the total resident population are controlled at the level of county-based weighting areas.

The GQ sampling selects groups of GQ residents, not the GQ facilities themselves, in contrast to the HU address sample. The GQ frame is divided into two sampling strata within each state, a small GQ stratum and a large GQ stratum, each with different sampling methods. The small stratum consists of GQs with expected populations of 15 or fewer and GQs closed on April 1, 2000. Small stratum GQs are sampled systematically within each state, sorted by small versus closed on census day, new GQ facility versus previously existing, GQ type, and geographical order (county, tract, block, street name, and GQ identifier). The sampling rate varies by state, being higher for states with the smallest GQ populations, but was about 1-in-40 for many states in the 2008, 2009 and 2010 ACS samples[1] (Marquette, 2011). If there are 15 or fewer people found in a small stratum GQ, then everyone in the GQ is in sample. If there are 16 or more people found in a small stratum GQ, then ten people are systematically selected from the GQ. The large stratum includes GQs with expected populations of 16 or more. The primary sampling unit for large stratum GQ facilities is a group of ten people, not the facility itself. For each large stratum GQ selected to be in sample, one or more systematic samples of groups of ten people are taken to achieve the state sampling rate. All large GQ facilities in a state are sorted by GQ type and geographical order in the large GQ frame. On the 2007 GQ sampling frame, there were approximately 105,000 small stratum GQ facilities, 77,000 large stratum GQ facilities, and 3,000 facilities with an unknown population[2] (U.S. Census Bureau, 2009).

3. Representation of ACS Group Quarters Sample Across Tracts and Counties

The distribution of ACS sample GQ facilities across counties and tracts illustrates the limitations of the sample design with respect to producing small area estimates. Table 1 and Table 2 show the representation of the ACS sample across tracts in the years 2006-2010. In Table 1 we see that less than half of the tracts with GQ facilities did not have at least one GQ facility in the ACS sample from 2006-2010, that is, 20,105 of 44,157. The number of tracts and counties with GQs is determined from the ACS sampling frame, which is based on the 2000 decennial census. For perspective, note that in Census 2000 there was better coverage of GQ facilities. The Census 2000 long form was distributed to a sample of 1-in-6 persons residing in GQs, and further, the Census 2000 visited all GQ facilities and thus every one potentially had persons in the long form sample.

Table 1: ACS GQ Sample in Tracts in 2006-2010

	Frequency
Tracts with GQs	44,157
Tracts with ACS GQ sample	24,052
Tracts without ACS GQ sample	20,105

Table 2 shows the representation of ACS sample in tracts by seven major types of GQ facilities. The categorization by seven major types shown in the tables is used in assigning the weights and is a convenient categorization here.

[1] The GQ sampling rates by state were changed for the 2011 ACS GQ sample.
[2] The GQ facilities without estimates of population are sampled like the small stratum GQ facilities.

Major GQ type is relevant because people in different types of GQ facilities differ from each other in consistent, predictable ways. Table 2 shows that large numbers of tracts with GQs do not have ACS sample of the same major type of GQ. For example, of the 4,993 tracts with an adult correctional facility, 1,908 did not have any facilities in the ACS sample from 2006 to 2010.

Table 2: ACS GQ Sample in Tracts by Major Type of GQ in 2006-2010

Major Type of Group Quarters	Tracts with ACS Sample	Tracts without ACS Sample	Total Tracts with Type of GQ
(1) Adult correctional facilities	3,085	1,908	4,993
(2) Juvenile facilities	1,343	1,582	2,925
(3) Nursing/Skilled nursing facilities	10,859	5,775	16,634
(4) Other health care facilities	1,075	2,533	3,608
(5) College/university student housing	2,538	827	3,365
(6) Military group quarters	304	276	580
(7) Other noninstitutional facilities	11,805	23,611	35,416

Table 3 shows the analogous counts as Table 2 but for counties instead of tracts. Even at the county level there is a significant proportion of counties which lack representation by major type of GQ.

Table 3: ACS GQ Sample in Counties by Major Type of GQ in 2006-2010

Major Type of Group Quarters	Counties with ACS Sample	Counties without ACS Sample	Total Counties with Type of GQ
(1) Adult correctional facilities	1,879	865	2,744
(2) Juvenile facilities	734	463	1,197
(3) Nursing/Skilled nursing facilities	2,693	267	2,960
(4) Other health care facilities	581	741	1,322
(5) College/university student housing	1,018	143	1,161
(6) Military group quarters	225	173	398
(7) Other noninstitutional facilities	2,019	838	2,857

Further evidence is seen in large year-to-year fluctuations of county estimates of GQ population, and unrealistic estimates of persons per household, among other unexpected results. For an example, consider the estimates of GQ population for Harford County, Maryland, for 2006, 2007, and 2008, which were 2,897, 6,138, and 1,463. Another relatively extreme example is Benton County, Oregon, which had 6,129 and 2,709 according to the 2006 and 2007 ACS (the Population Estimates Program or PEP estimate of GQ population for these years was 4,280).

These year-to-year fluctuations highlight the point that which GQ facilities fall into sample has a disproportionate effect on the estimates of substate geographies. Thus we expect to see effects on the estimates of characteristics of the total resident population. Though the actual number of people residing in GQs is small, their effect on estimates of the characteristics of the total resident population can be disproportionately large for characteristics that are strongly related to GQ residence. Such characteristics include disability status, income, and variables derived from income, such as poverty. Beaghen and Stern (2009) document such concerns with estimates of poverty rates.

4. Overview of the New Methodology

The objective of the new methodology is to improve the estimates of the GQ population for counties and tracts, thereby also improving estimates of the total resident population for counties and tracts. The limitations in the sample design can be viewed both in terms of high variances of estimates of the GQ population for substate geographies, as well as in a lack of representation of ACS sample in counties and tracts which are known to have GQ facilities. Though we approach the problem from the point of view of trying to have GQ person records, sampled or imputed, in the smallest geographies, a successful methodology should shrink the variances of small area estimates. The methodology described in this section is an overview, for more details see Asiala et al (2011).

4.1 The Basic Approach

The approach to the problem is to populate selected GQ facilities without ACS sample with person records copied from in-sample GQ facilities, with appropriate weighting adjustments. This imputation is a whole person imputation and not an item-level imputation. The whole set of person characteristics of the donor is copied to the recipient record (with the exception of geography-dependent variables, see Section 6, item 5). The recipient record maintains the recipient GQ type characteristics and current residence geography. Imputing to not-in-sample facilities has the important advantages for data processing that the imputed person records function as pseudo-sample and are transparent to the data processing and production of estimates. For this approach we identified the following key challenges.

- How do we construct the frame which we later populate with GQ person records?
- Which not-in-sample GQ facilities do we impute to?
- How many person records should be imputed to each GQ facility?
- How do we select GQ person records to serve as donors?
- How do we assign weights to imputed and sample GQ person records?

4.2 The Frame

The listing of GQ facilities to which we potentially impute is the ACS GQ sampling frame. In addition to the listing of GQs, an important feature of the frame is population counts, which are needed in determining how many GQ person records to impute to a given GQ and in the weighting.

We did consider several alternatives for use as the frame. The 2010 decennial census listing is more up-to-date than the 2000-based sampling frame, which is the basis for sampling frames through the 2011 ACS samples. Thus we considered a frame enhanced with the 2010 decennial census listing of GQs. This is the frame we used for the evaluation discussed in Section 7 (see Section 7.2 for more details on this approach and why we rejected it). We also considered just using the 2010 decennial census listing itself as the frame for 2010 imputations. However, we decided against this approach because it would have excluded GQ facilities confirmed to exist in the 2010 ACS interviewing.

4.3 Which Group Quarters Facilities to Impute to?

The next question was to determine which subset of not-in-sample GQ facilities on the frame to impute to. Imputing to all of them had the appeal that there would be GQ person representation in every geography for every detailed type of GQ (see Appendix C for the list of detailed types of GQ facilities in the ACS) where GQ facilities exist. However, it would have required imputing a prohibitively large number of records. Thus we impute to only a subset of GQ facilities, prioritizing which GQs to impute to as follows.

- 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.
- 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.

These priorities lead to a scheme where all large stratum GQs are imputed to, but only a sample of small stratum GQs are imputed to so that the second objective is met. Note that the second objective is relevant only to the 5-year estimates, for which we produce tract-level estimates. However, we use the same methodology for 1-, 3-, and 5-year estimates.

4.4 Identify GQ Facilities that Require Imputations and Determine How Many GQ Persons to Impute

The GQ selection procedure gives priority to obtaining representation for each major GQ type group in each county. Hence we refer to it as “county first”. Then facilities are selected to establish representation for each major type group at the tract level. How many imputed person records each not-in-sample GQ facility receives is a function of its population, which is either modeled or observed. For this determination we make a distinction between small and large stratum GQ facilities. A detailed outline of the procedure follows.

1. For each year and each large GQ not in sample, impute the number of records equal to 2.5% of the expected GQ population. This is roughly similar to the overall sampling rate of the GQ population.
2. For each year and for each combination of county and major GQ type on the year's frame that is neither in the year's sample nor in the year's imputes (from Step 1), randomly select a small GQ facility from the small GQ facilities in the county of the same major GQ type.
3. For each GQ selected in Step 2, impute the number of records equal to 20% of the expected GQ population or 1, whichever is larger.
4. Identify all combinations of tract and major GQ type that exist on any year's sampling frame but are not in any year's sample, nor in any year's imputed records.
5. For each combination identified in Step 4 and for each year that the combination exists on the sampling frame, select a small GQ facility with equal probability from the small GQ facilities in the tract of the same major GQ type.
6. For each GQ selected in Step 5, impute the number of records equal to 20% of the expected GQ population or 1, whichever is larger.

4.5 Select Donors – The Expanding Search Method

The donor selection method is referred to as the expanding search approach (Erdman and Nagaraja, 2010). Note that for each year, donors are selected only from that same year. The donor selection procedure chooses from within specific type when the donor to imputation ratio within the specific type is large enough for this to be feasible, and gives preference to donors from facilities that are geographically close. (A listing of specific types of GQs grouped by major type is given in Appendix C). Once GQ facilities have been selected for imputation, the donor pool for each facility is set to be the first combination of geography and GQ type in the following list in which there is at least one donor per five imputed records needed. Donors are recruited first in the lower ranking step starting with step 1. If a suitable donor is not found in a given step, then proceed to the next step.

1. County and specific type
2. County and major type
3. State and specific type
4. State and major type
5. Division[3] and specific type
6. Division and major type
7. Region[4] and specific type
8. Region and major type
9. Specific type
10. Major type

For example, suppose that in a particular county we wish to impute one hundred records into college dormitories. If at least twenty dormitory residents in the county have been interviewed, we sample these interviews for imputation, with replacement, one hundred times (we limit the number of times a donor can be used to five). If fewer than twenty dormitory residents in the county have been interviewed, we expand the geography of the donor pool to the state, division, region, or nation as necessary so that there are at least twenty records from which to sample.

4.6 Weighting

The new imputation methodology implies a new weighting scheme which makes a clean break from the old weighting design that was used for ACS estimates released prior to 2012. For details see Asiala (2011a), or Asiala,

---

[3] A census division is a grouping of states and the District of Columbia established by the U.S. Census Bureau for the presentation of census data. The nine divisions represent areas that were relatively homogeneous when they were established in 1910. The divisions are subdivisions of the four census regions.
[4] A census region is a grouping of states and the District of Columbia established by the U.S. Census Bureau for the presentation of census data. The four regions represent areas that were relatively homogeneous when they were established in 1910 and revised in 1950. Each region is divided into two or three census divisions.

Beaghen, and Navarro (2011). We will only point out two key features here. First, the weighting procedure is applied to the augmented data, that is, the data set containing both the sampled and imputed records. It makes no distinction between sampled and imputed GQ person records. Second, the weighting scheme ensures that when computing estimates for small areas, the weighted data for GQ persons only represent persons within that tract or county, depending on the length of the estimation period.

5. Evaluation with Simulated Data Based on Census 2000

Two evaluation studies of the imputation procedures have been completed. The first evaluation was a simulation study with Census 2000 100 percent GQ data (Weidman, 2011, and Erdman and Nagaraja, 2010). This study used an inventory of all residents of GQs on April 1, 2000 and their basic demographic information – sex, age, race, and Hispanic origin. Estimates of these characteristics were compared and evaluated on simulated samples for four imputation procedures and the design-based method. The purpose of this study was three-fold:

a. To operationalize the imputation software and ensure that it worked correctly.
b. To select between four different imputation procedures.
c. To analyze the differences between the imputation and sample only (design-based) estimates to determine if there are any major problems with the imputation procedures that need to be addressed or would cause us to discontinue this research and continue with the design-based estimates for the 2012 ACS production.

We summarize the results here. The study and the results were discussed in greater detail in the subsequent sections.

- The simulations found that nearly half of the augmented data were comprised of imputed records. In addition, the number of imputed records could far exceed the number of sampled records for some major GQ types.
- Most donors were found within the specific GQ type. It generally found donors within the state of the GQ to be imputed, and many times within the county.
- We found that the imputation-based method was systematically biased even at the state level for sex and certain age groups.
- The variances of imputation-based estimates were smaller than the design-based estimate variances. Comparisons of the mean squared error, mean absolute deviation, and percent absolute deviation gave mixed results.
- Some of the shortcomings of the imputation methods identified in the simulation study were mitigated by minor changes to the imputation methodology in the evaluation with ACS data (see Section 6). In particular, we have identified a list of single-sex GQ facilities to which we restricted donors to be of the correct sex.

5.1 Details of the Imputation and Weighting Methods in the Evaluation Based on the Simulation

Given the complexity of the GQ population, the procedure for sampling this population, and the level of coverage that is desired, careful consideration must be given to both the selection of GQ facilities for imputation and to the selection of donor records. As such, in the evaluation based on simulations we considered two methods for the selection of GQ facilities, and two methods for the selection of donors.

Both GQ selection procedures select all large-stratum GQ facilities for imputation. However, one procedure first selects facilities needed to produce non-zero 1-, 3- and 5-year county-level estimates, and then selects small-stratum facilities required to produce 5-year tract-level estimates; the second procedure reverses this order. In both GQ selection methods, the fraction of residents to be imputed is chosen not only to resemble ACS sampling rates but also to produce reasonable variance estimates. The details for both procedures are given in the following section.

Once facilities have been selected for imputation, we choose donor records to populate the facilities according to one of two methods, described in Section 5.2. Both donor selection procedures give preference to donors from within the same specific GQ type, and ensure that donors come from within the same major GQ type as the recipient GQ. However, one method focuses on finding donors from facilities which are geographically close (expanding search), while the other focuses on finding donors from facilities in geographies which are demographically similar (K-means clustering).

The imputed records are appended to the sampled records to form a complete augmented data set.

5.2 Selection of GQ Facilities for Imputation

The first GQ selection procedure gives priority to obtaining representation for each major GQ type group in each county. Then facilities are selected to establish representation for each major type group at the tract level. This is the 'county first' method described in detail in Section 4.4.

The second GQ selection procedure described below gives priority to obtaining representation for each major GQ type group in tracts that are not sampled in the 5-year period ('tract first'). Then facilities are selected to produce non-zero 1- and 3-year estimates for each major type group and county combination. Details of this procedure are as follows. Note that based on the evaluation we rejected this approach in favor of the 'county first' approach described earlier.

1. For each year and each large GQ not in sample, impute the number of records equal to 2.5% of the expected GQ population.
2. Identity all combinations of tract and major GQ type that exist on any year's sampling frame and are not in any year's sample, nor in any year's imputed records (from Step 1).
3. For each combination identified in Step 2 and for each year that the combination exists on the sampling frame, select a small GQ facility at random.
4. For each GQ selected in Step 3, impute the number of records equal to 20% of the expected GQ population or 1, whichever is larger.
5. Identify all combinations of county and major GQ type which are on the year 5 sampling frame but not in the year 5 sample, nor in the year 5 imputed records.
6. For each combination identified in Step 5, and for each year in the range 3-5 that the combination exists on the sampling frame and is not in the year's sample nor in the year's imputed records, select a small GQ facility at random.
7. For each GQ selected in Step 6, impute the number of records equal to 20% of the expected GQ population or 1, whichever is larger.

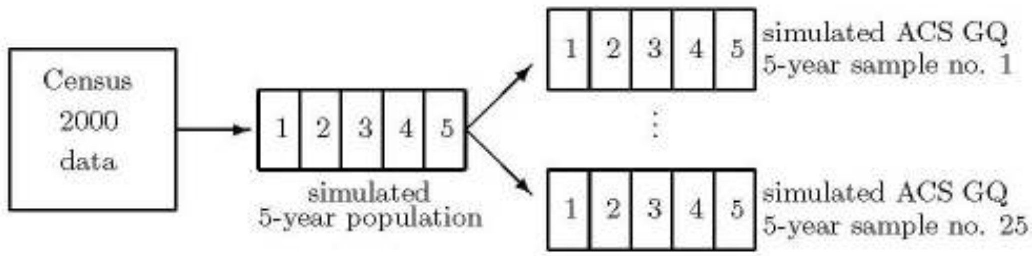
5.3 Alternative Imputation Methods

The two methods we investigated were the expanding search and the cluster-based imputation. We concluded that the expanding search worked better. All subsequent evaluations use the expanding search approach. Since it is described in detail in Section 4.5 we only describe in this section the alternative, the cluster-based imputation. For this approach tracts were grouped by similarity on twelve characteristics from the Tract-Level Planning Database (Bruce and Robinson, 2007), a collection of household, demographic, and socioeconomic variables assembled from Census 2000 to help in determining characteristics related to the percentage of questionnaires returned. Eight distinct clusters of tracts were identified to help design a marketing campaign for Census 2010 (Bates and Mulry, 2008) without any consideration of geography. Clusters were used to guide donor selection in the following manner. Once GQ facilities have been selected for imputation, we first match facilities on a combination of cluster and specific GQ type. For each combination, if there is at least one donor per five imputations needed, donors are selected at random from within cluster and specific type. If there are not enough donors, clusters are collapsed together in a specified manner until the donor ratio is attained. In the rare case that the ratio cannot be attained, the donor pool is expanded to major GQ type.

5.4 The Simulation Study

For this study, ACS GQ samples were drawn from a population simulated using the Census 2000 short-form data for each state and the District of Columbia. To generate these sampled records, we followed the ACS sampling procedure outlined in by the U.S. Census Bureau (2009) and the ACS group quarters sampling specifications (Williams, 2008). We made some simplifications but attempted to make the process as realistic as possible. To this end, we allowed the population size of each GQ to vary across years and we simulated a limited form of GQ closings. We simulated one 5-year population set using the Census 2000 data and drew 25 sets of 5-year samples from that population. Figure 1 is a diagram of this procedure.

Figure 1: Simulation Procedure



To evaluate estimates made from both the current and proposed methodologies, we compared them to the corresponding parameters of the simulated population values. To evaluate 1-year estimates, we used the fifth year of each five year sample, for 3-year estimates, years three through five were used, and, for 5-year estimates, all five years of each sample were used.

Since we were attempting to mimic the actual ACS sampling and data collection process in our simulation, the ACS sampling methodology was used to choose samples for both small and large stratum GQs (Section II) but GQ types which are not sampled in the ACS were omitted. To reduce the complexity of the simulation, we did not include person non-response adjustments to the weights; rather, we assumed that all sampled persons were survey respondents.

5.5 Analysis

A simulation study such as ours offered two distinct advantages: (a) we knew the "truth" because we simulated the population and (b) we can directly compute measures such as bias and mean squared error since we have drawn multiple, independent samples. The results for 1-, 3-, and 5-years were similar and our discussion covers all three periods. All graphs can be found in Appendix A.

To evaluate the results of the imputation methods, we analyzed the estimates from the five methods: the design-based method and four imputation-based methods. These are summarized below.

1. Design-based: estimates were computed using the sampled data only (denoted as "sample" in plots).
2. Expanding Search-County Imputation: Expanding search, county-level coverage handled first (denoted as "expand county").
3. Expanding Search-Tract Imputation: Expanding search, tract-level coverage handled first (denoted as "expand tract").
4. K-means Search-County Imputation: K-means search, county-level coverage handled first (denoted as "k-means county").
5. K-means Search-Tract Imputation: K-means search, tract-level coverage handled first (denoted as "k-means tract").

We produced counts of basic demographic characteristics from the Census 2000 short form data. We divided these counts into the following categories for comparison purposes:

1. total population,
2. sex: male; female,
3. age: 0-17; 18-34; 35-64; 65 years and older,
4. ethnicity: Hispanic/not Hispanic, and
5. race: white, not Hispanic; black, not Hispanic; and other, not Hispanic.

For each characteristic, geography, and major type group, we had five sets of estimates and a true population value.

5.6 Summary Measures

We drew 25 independent samples from the simulated 5-year population. For a given characteristic estimate of the true simulated population value for a given combination of (period length, major GQ type, geography, imputation method), we were able to directly compute the following comparison measures for the four imputation methods and the design-based method.

Bias = Mean estimate across samples – population value

Standard Deviation = Square Root{[Sum (sample estimate – mean estimate across samples)**2] / 24}

MSE[5] = [Sum (sample estimate – population value)**2] / 25

MAD[6] = Sum(absolute value (sample estimate – population value)) / 25

%MAD = MAD * 100 / population value

We found that the imputation-based methods were all more biased than the design-based method at each geography level: tract, county, state. In Figure 2, the bias at the tract, county, and state levels is plotted for the 5-year estimates of the number of persons who are white but not Hispanic, broken down by the seven major GQ types (refer to Tables 2 and 3 or Appendix C for the major GQ types). The box plots can be interpreted as follows. On the x-axis, each group of box plots represents a GQ major type. The y-axis indicates the bias. Each individual box plot within a group represents a method and represents the range of bias values across all tracts/counties/states for the demographic variable and major GQ type combination. In each plot, we see that the range of bias is larger for the imputation-based methods than for the design-based method (in gray) for all geography levels.

A second, more specific example can be seen in Figure 3. The absolute bias[7], variance, root mean squared error, and mean absolute deviation are plotted for the 5-year estimates of the number of females who live in an adult correctional facility (GQ type 1) in a county. For each of these plots, the estimate using the "expand county" method of a measure was plotted on the x-axis. The design-based values of that measure are plotted on the y-axis. Each point represents a specific state in the U.S. or the District of Columbia.

To interpret these plots more easily, we added three features. First, the 45° line was plotted (dotted line). If a point was above the 45° line, then the design-based method had a higher value of the measure indicating the imputation-based method performs better in the geography represented by that point; if a point is below the 45° line, then the reverse is true. On each plot, the total number of states which are above and below the 45° line are printed (points directly on the line are excluded here). Finally, because many points may overlap, areas which contain more points are shown in a darker color than those with fewer points.

Let us focus on the plot for absolute bias (on the top-left). A lower value of absolute bias is more desirable, regardless of the method applied. Therefore, knowing which method (between the design-based and the "expand county" methods) has a lower value of a measure is informative. From this plot we can see that most of the points are below the 45° line indicating that the "expand county" method produces more biased estimates. Upon further investigation, we find that the imputation procedures are imputing more males into adult correctional facilities than females. As a result, we systematically undercount the number of females (negative bias) and systematically over

---

[5] MSE stands for mean squared error.
[6] MAD stands for mean absolute deviation.
[7] Bias is plotted as absolute value of bias, or |bias|. This is because bias can be any real number which causes problems when comparing values. For instance, if the bias for the design-based method is 10 and the bias for the imputation-based method is -20, doing a direct comparison would indicate that since -20 < 10, the imputation-based method has a lower bias. However, this is not the case. |-20| > |10| and therefore the imputation-based method has a higher bias although in the opposite direction. In this analysis, we only care about magnitude not direction because bias in any direction is undesirable.

count the number of males (positive bias) in adult correctional facilities. One way to mitigate this problem would be to identify which adult correctional facilities are male-only, female-only, or mixed and find donors accordingly.

If we refer back to Figure 3**,** we see that the variances (top-right plot) are generally smaller for the imputation-based estimates when compared to the design-based estimates. That is, most of the points on the graph are above the 45° line. This seems to be the case across all combinations of variable, GQ type, period length, and imputation method. The results, however, are mixed for the MAD and MSE analyses. We can also see this in Figure 3 (bottom two plots) because the points are scattered randomly around the 45° line. In general, for both of these measures, due to use of the controls, the imputation-based methods tend to perform better across counties for characteristics which "behave" like total population such as gender, white (not Hispanic), and total population. For the MSEs, the design-based estimates tend to perform better across both counties and tracts for characteristics which measured "subpopulations" such as the individual age groups and race other (not Hispanic). These relationships do not always hold for either the MAD or MSE at the state level, however.

5.7 Assumptions about the Expected Counts

Among the GQs which are not in sample there are two important, but unknown, attributes: (1) whether or not the GQ exists, and (2) how many people live in the GQ. Both are updated generally only for GQs *in sample*. Nevertheless, we rely heavily on both facts for imputation and weighting. To illustrate the potential effects of each element, we examine three counties in the state of Ohio which contain college dormitories (major type 5).

In Table 4, the true number of GQs in each county for each year is shown along with the expected number. Note that the year 0 row represents the Census 2000 value. In each of these cases, the actual number of GQs differs from the expected number. This is simply because there is no formal mechanism to update GQ closings if a GQ has not been in sample. For County A, we never find out that in years 3-5 there are no college dormitories. Since the GQs in the county are not sampled, their population count is not updated and we impute records into a county which should have a count of 0 for both the three and the 5-year estimates. This is an example of attribute (1).

Table 4: True and Expected Number of GQs by Year

Year	County A		County B		County C	
	True	Expected	True	Expected	True	Expected
0	2	–	6	–	19	–
1	2	2	6	6	15	19
2	1	2	4	6	15	19
3	0	2	4	6	12	18
4	0	2	4	6	11	17
5	0	2	4	6	11	17

Element (2) can be partially examined in Table 5 which lists, for each year, the true total population of the county along with the expected total population. There are clearly large discrepancies between each pair of columns. This occurs because, in addition to some GQs closing, other GQs can vary in size from year-to-year, and these changes can be extreme. Again, as with attribute (1), the counts of a GQ are not updated on the sampling frame if it is not sampled. The tract and county constraints depend heavily on these expected counts and not having current counts available has a negative impact on the imputation-based estimates.

Table 5: True and Expected County Population Counts by Year

	County A		County B		County C	
Year	True	Expected	True	Expected	True	Expected
1	38	38	834	843	1,527	684
2	38	38	656	808	1,518	825
3	0	38	713	863	1,796	726
4	0	38	687	931	687	931
5	0	38	700	938	1,502	596

5.8 Comparing Methods

There was little difference in the results between the “county” methods and the “tract” methods. That is, “expand county” and “expand tract” were similar and “k-means county” and “k-means tract” were similar. These two pairs of methods differed only by which GQs are chosen for imputation, not where the donors were found. The number of imputations varied little between the “county” and “tract” methods which indicated that most of the difference across methods resulted from the choice of donor.

The expanding search methods and K-means methods found donors in very different locations. If physical proximity is more important to obtaining better donors, then the expanding search methods should perform better than the K-means methods. If the characteristics of the surrounding housing population are of greater value, then the reverse should be true.

We have found that in general, the K-means methods performed no better, and often worse, when compared to the expanding search methods. This was the case when comparing bias and to a lesser extent for MSE, MAD, and %MAD. In some cases, however, the K-means methods performed better in terms of lower variance.

One example is shown in Figure 4 (in Appendix A) for estimates of Hispanic persons (counties, 5-year estimates). The top plot graphs the bias across counties for each GQ type. Each individual box plot represents a method (design-based included). The bottom plot shows standard deviation in a similar way. The K-means methods tend to have a much higher bias compared to the expanding search methods especially for GQ types 1 (adult correctional facilities), 5 (college dormitories), and 6 (military facilities). However, the variance tends to be a bit lower for the K-means methods.

5.9 Conclusions of Simulation Study

Nearly half of the augmented data were comprised of imputed records, regardless of the imputation method applied. In addition, the number of imputed records could far exceed the number of sampled records for some major GQ types. Most donors were found with the specific GQ type across all imputation methods. While the K-means methods generally found donors outside of the state, the expanding search methods generally found donors within the state, or many times within the county of the GQ to be imputed.

We found that the imputation-based methods were systematically biased even at the state level. The variances were smaller than the design-based estimate variances and comparisons of the MSE, MAD, and %MAD gave mixed results. The two ways of choosing GQs to impute into yielded similar results; however, the K-means methods seemed to perform less well than the expanding search methods. Nevertheless, some of the downsides of the imputation-based methods could be mitigated by minor changes to the imputation methodology.

Each of the seven major GQ types house people who have very different characteristics from each other and from household residents. Therefore, especially for small areas, obtaining information on GQ residents is vital to producing estimates of the total resident population. The imputation-based approaches discussed here allow us to do exactly that.

6. Changes to the Methodology

We identified from the simulation analysis three areas where tangible improvement in the imputation procedure could be made. These changes, along with a fourth change implemented as a special edit, are described in this section. All four changes were implemented in the evaluations with ACS data discussed in Section 7 and will be implemented in future ACS methodology.

1. *Taking account of the sex of residents within a GQ.* The major and specific GQ type designations do not indicate whether a GQ has residents of a single or both sexes. However, the simulation study showed that estimates for the number of female residents and the number of male residents were biased, especially in adult correctional facilities. If auxiliary information could be used to determine the sex compositions of individual GQs, the information could be used to select donors for imputed GQs which reflect the sex composition of the imputed GQ. For instance, a GQ thought to be all female would have only female donors imputed into it, not male donors. Such a procedure would help reduce the bias of the estimates. Such auxiliary information is compiled from sources such as historical ACS sample records and census records.

2. *Adjusting the expected GQ populations.* A problem with the current methodology is that the expected number of residents in a GQ is not a good approximation of the true, but unobserved population size. These expected values do not account for either the changes in population size over time or GQs that close; however, we depend on them in both the imputation and weighting steps. Thus we constructed an algorithm from which a more accurate population size can be computed from the expected population values (Joyce, 2011).

3. *Limiting the number of times a person can be used as a donor in a tract to five.* A more technical issue is that of repeating donors. For the simulation, the number of times a donor could be used was limited overall. Applying this rule still, however, allowed for the possibility that a donor was used multiple times within a tract. In such a situation, the donor's characteristics are highly concentrated in one area. This is undesirable, especially if the chosen donor is atypical. Therefore, we limit donor repetition within a tract while still maintaining an overall, more global, maximum. Note that this bound must be chosen with care. The goal of diluting the influence of an individual donor must be balanced against the goal of finding donors who are geographically nearby.

4. *Geography-dependent characteristics.* Certain geography-dependent characteristics such as migration (residence one year ago) and journey to work present particular challenges for the imputation based method, as assigning the geography of the donor can lead to unreasonable results for certain geography-dependent characteristics. For example, say we assign a donor person from Smith County to a recipient GQ listing in Springfield County. Further, say the donor indicates he lived in Smith County the previous year. If we assign Smith County as the residence a year ago, it will seem as if the respondent had moved between counties over the year, although the respondent had indicated they had not moved. An analogous quandary exists with journey to work; do we want the imputed respondent to indicate that he commuted to the neighboring county? Koerber (2011) found that for characteristics which are geography-dependent, such as Place of Work and Migration, always taking the donor values leads to a poor distribution of estimates of the characteristics. Koerber found, however, that replacing the donor geographies for these characteristics with the current residence geography of the recipient for certain cases leads to markedly improved distributions of these characteristics in the sense they are more like the distribution based on the sample. These replacements or special edits amount to an additional processing step to be conducted after the GQ small area imputation.

7. Preliminary Evaluation with ACS Data

The purpose of the second evaluation was to establish the reasonableness and to confirm the usability of the imputation methodology, while also allowing an assessment of the methodological changes made after the simulation (Asiala, 2011b). This second evaluation used ACS data so that estimates of the diverse characteristics produced by ACS could be analyzed (the Census 2000-based simulation considered only demographic characteristics). We refer to this evaluation as preliminary because we planned to conduct more thorough analyses with ACS data.

The analyses in this evaluation were based on the following premises.

- It is impossible to establish that the imputation estimates are superior to the design-based with traditional statistical measures such as bias, variance, MSE, or other loss functions. This is because, unlike in the simulation, the true population characteristics are unknown so we have no measures of bias. Nor did we have a sound variance estimator for the imputation methodology in place at the time of the research.

- We believe the design-based estimates of the GQ population are essentially unbiased at the state level. The ACS state-level estimates of the GQ population have been generally accepted to be sound. Thus any large differences between the design-based and imputation estimates at the state level would suggest problems with the imputation methodology.

The results reported in this section are based on the analysis of the differences between the design-based and imputation-based state level estimates. While we expected that the state-level estimates would differ some between the two approaches, for a favorable evaluation of the new methodology we required those differences to be generally of little practical importance. We took two approaches to establishing practical importance. First, we sought guidance from our subject matter experts to help gauge whether a difference in the estimates should be considered to be meaningful or not. Second, we determined if differences were within the sampling variability of the design-based estimates (a difference outside of the sampling variance would be considered to be of practical significance).

7.1 2010 ACS Data for the Evaluation

The 2010 ACS data with interviews were not available at the time of this research, although the 2010 ACS GQ sample was. Thus we used the 2006–2009 ACS interview data and the sample for the 2010 ACS data to construct a 2010 ACS dataset for this evaluation. If a GQ in the 2010 ACS sample had previously been in sample from 2006-2009, we used those interview results as the 2010 interview results. If a 2010 sample GQ was not previously in sample, GQ person data was used from the same detailed type of GQ facility from the same state. (In a third evaluation currently underway we use the actual 2010 ACS results – see Section 8). We used simulated 2010 ACS data rather than 2009 ACS data because we wanted to test the creation of an enhanced frame using 2010 Census data. This enhancement was only applicable to the imputation frame for the 2010 data year.

We then applied both the design-based and the imputation-based methodologies to this simulated 2010 ACS data. The weighted data created using each methodology were then tabulated for a broad set of characteristics for the GQ-only population including: age, race, education, marital status, migration, foreign born, employment, occupation, and place of work as appropriate. The characteristics included corresponded to those found in the GQ subject table, S2601A, with some additions including world region of birth, place of work, and the sex ratio. See Appendix B for the full set of characteristics tabulated.

For this report we focused on the state-level GQ estimates for the entire GQ population. For each characteristic, percent distributions were calculated for comparison purposes so that a change in a base estimated count did not cause changes for all the characteristics for that base. Our analysis focused on only the percent distribution data contained in the GQ subject table.

7.2 Construction of the Enhanced Frame for the Imputation Methodology

As mentioned in Section 4.2, we explored the construction of an enhanced frame that incorporates both the ACS sample and the 2010 decennial census GQ listing. For this evaluation the enhanced frame was constructed from two pieces. The first piece resulted from a match between the ACS 2010 sample and the GQ Enumeration (GQE) file, which was constructed for conducting Census 2010 (reference). Note that since the 2010 Census GQ facility population values were unavailable at the time of the evaluation, we had to construct synthetic Census 2010 population values. These were constructed from information on facilities capacity values obtained in the GQE and from ACS 2006-2009 data. The second part of the enhanced frame consisted of GQs which were present within the 2010 ACS sample but failed to match to the Census GQ listing. Thus the enhanced frame consisted of all 2010 ACS GQ sample and all Census 2010 GQE GQ listings valid for the ACS.

We point out that the resulting GQ listing was likely limited by duplication, as the match may not have identified some ACS and GQE listings which referred to the same GQ facility. This could arise for several reasons. Perhaps the most common cause was that some ACS sample GQs were deleted by decennial census operations and added again, possibly with a new address, a new name, or a new geocoding. Overall, 60.1% of the ACS 2010 sample GQs matched to this 2010 listing. The proportion of the nonmatching 39.9% which was duplicated is hard to assess, but it was potentially a large portion of it. In the estimation and weighting the effects of this potential duplication were limited by the county- and tract-level controls. In the implementation of the imputation methodology, the enhanced frame will be constructed differently than described in this section because of lessons learned in this evaluation process. There will be no attempt to augment the 2010 frame with the GQE or 2010 decennial data.

7.3 Analyses Methodology

The primary question this analysis was to answer was whether imputation-based estimates differed in a meaningful way from the design-based estimates at the state-level. Working with our subject matter experts at the bureau, the following criteria were established. First, we established differences of such a magnitude that they could be considered (nominally) important.

Nominal Criteria

Nominally Small Difference: The difference between the two estimates is less than 2 percentage points AND less than 5 percent of the design-based estimate

Nominally Large Difference: The difference between the two estimates is larger than 5 percentage points OR larger than 10 percent of the design-based estimate.

Indeterminate Difference: If a difference does not meet one of the above categories then it depends on the characteristic and geographic type and size as to whether it is important or not important.

Note that for the relative differences, we required the design-based estimate to be at least 1 percentage point in size so that the relative change would be noticeable in our published estimates which show only one decimal place for the percentage. To establish important differences, we added the requirement for practical significance.

Importance Criteria

Not Important Difference: The difference between the two estimates is not *practically* significant OR not nominally large.

Important Difference: The difference between the two estimates is *practically* significant AND nominally large.

Indeterminate Difference: If a difference does not meet one of the above categories then it depends on the characteristic and geographic type and size as to whether it is important or not important.

As mentioned earlier, a practical difference was one greater than the sampling error of the design-based estimate. We did not have in place a test of statistical significance. In order to obtain an accurate test of statistical significance for the difference, we would need to have both a good estimate of the variance of the design- and imputation-based estimates and a good estimate of the covariance between the estimates from the two methods. We expected the covariance to be high between the design- and imputation-based estimates since they both draw upon the same data. However, at the time of the research we were in the development stages of designing a good variance estimator for imputation-based estimates, the lack of which limited our ability to produce an accurate variance estimate. Furthermore, our initial attempts to capture the covariance in our variance estimate for the difference have also produced limited results. For that reason, we apply a test of practical significance where we compared the size of the difference to the margin of error for the design-based estimate with a 90% confidence level.

If a more refined variance estimation methodology were to confirm our expectation that the state-level estimates using the two methods (design-based and imputation-based) are highly correlated, then we may flag a high percentage of characteristics as having statistically significant differences. For this reason, a meaningful analysis may, in fact, need still to rely on a test of practical significance if the estimated variance of the difference is very small.

7.4 Results - General

Before analyzing by the three classifications of importance given above, we first present some basic summary statistics. Tables 6–11 provide a general picture of the overall properties of the differences between the two methods across all characteristics and across all states. (There were 76 characteristic percent distribution estimates in total). A more in-depth examination of the data follows this general synopsis in Section 7.5.

Table 6 shows that for characteristics, a given estimate was of practical significance in 5.5 of the 51 states including DC (averaged over the 76 characteristics). Further, the estimates in the profiles were above 1 percentage point for

48.0 states (averaged over the 76 characteristics), so most characteristics were large enough to have the percent relative change test applied. However, there were still some characteristics where small estimate sizes were common and so the minimum size criterion was useful in analyzing them.

Tables 7 and 8 show the results for the nominal criterion alone in the categorization of the differences. They show that, in general, it was the criterion of nominal absolute differences that identified estimates of characteristics with large changes more so than the criterion for percent relative differences. Without the minimum size criterion, we did observe several characteristics which had large relative differences that were of no meaningful importance since the rounding used for the publication of the data would have reduced the difference to zero.

Table 6: Mean Number of States Where Difference is of Practical Significance (Averaged over 76 Characteristics)

Mean Number of States	5.5

Table 7: Mean Number of States Where the Absolute Differences Would be Considered Small, Large, or Indeterminate (Averaged over 76 Characteristics)

	Nominal Size Based Absolute Differences		
	Small (less than 2%)	Large (greater than 5%)	Indeterminate
Mean Number of States	38.8	4.3	7.6

Table 8: Mean Number of States Where the Relative Differences Would be Considered Small, Large, or Indeterminate (Averaged over 76 Characteristics)

	Nominal Size Based Percent Relative Differences (minimum estimated size 1%)		
	Small (less than 5%)	Large (greater than 10%)	Indeterminate
Mean Number of States	50.3	0	0.3

Table 9 provides a summary of the classifications of the differences considering both the nominal difference criteria as well as the practical significance difference tests. Table 9 shows that on average, the number of states where the difference is considered to be large is very small and that applying the practical significance test leads to even fewer which have important differences. Table 9 indicates, however, that there are a fair number of differences which are classified in the “Indeterminate” category which necessitates further review to determine which characteristics this impacts.

Table 9: Mean Number of States Where the Difference is Considered Important, Not Important, or Indeterminate, Using the Nominal Difference Criteria in Combination with Practical Criteria (Averaged over 76 Characteristics)

	Nominal	Nominal & Practical
Difference is Large/Important	4.3	1.2
Difference is Not Important	38.5	47.7
Difference is Indeterminate (neither Important nor Not Important)	7.8	1.7

7.5 Results - Detailed

Of the 76 estimates given in percent in the profiles, 37 had differences for at least one state which were “large” based on the nominal size of the differences. All of these were due to large absolute differences in the estimates. The full list of characteristic groups that had any states with nominally large differences along with the number of states is found in Table 10 below. As can be seen in the table, many of these characteristic groups had fewer than five states with nominally important differences. Of the 19 characteristics that had fewer than five states that that were nominally important, most dropped to one or two states that were also of practical significance (not shown in the tables). For the remaining 18 characteristics, all but two dropped to fewer than five states when the practical significance criterion was also applied (not shown in the tables). The two exceptions were the estimated percentage of male under 18 years of age and female under 18 years of age, where seven states continued to have differences classified as important when including the practical significance criterion. When we investigated which states had the important differences, we found that most of them were small states: DC, HI, MN, ND, PA, SC, and UT. The noted exception was PA. The source for this appears to originate from the GQ type categories "Other Health Care Facilities" and "Other Noninstitutional Group Quarters".

Table 10: List of Characteristic Groups with States with Nominally Large Differences

Characteristic	Number of States with Nominally Large Differences
Age categories:	
18–24 years	4
85 years and older	1
Age by Sex categories:	
Under 18 by Male / Female	30
65 or older by Male / Female	5
Marital Status categories:	
Widowed	1
Never married	2
School Enrollment categories:	
Nursery through 12th grade	3
College or Graduate School	3
Educational Attainment categories:	
High school graduate or higher	3
Bachelor's degree or higher	1
Nativity by Sex categories:	
Foreign born by Male / Female	14
Naturalized U.S. Citizen by Male / Female	26
Not a U.S. Citizen by Male / Female	18
Year entered U.S. categories:	
2000 or later	18
1990 to 1999	9
Before 1990	18
Employment Status categories:	
In Labor Force	1-3
Civilian Labor Force	1-3
Armed Forces	1-3
Not in Labor Force	1-3
Percent of civilian labor force unemployed	6
Occupation categories: all six categories	2-4
World Region of Birth: all four categories	6-14
Place of Work category: Worked outside state of residence	1

We also looked at the state data when broken out by institutional / non-institutional. The list of characteristics that had at least one state showing a nominally large difference was about the same as the list in Table 10. The exceptions were the race categories "White" and "American Indian and Alaska Native" which were nominally large in one state. It was not, however, of practical significance when compared to the margin of error of the design-based estimate. The characteristics that were both nominally large as well as practically significant in a minimum of five states are presented in Table 11 below.

Table 11: Characteristics Nominally Large and Practically Significant in at Least Five States

Characteristic	Number of States
Institutional	
Age category: Under 18 by Male/Female	10
School Enrollment categories: Nursery–12th Grade, College or Graduate School	9
Nativity category: Naturalized Citizen by Male/Female	8
Noninstitutional	
Age categories: 18–24 years	8
Age by Sex categories: Under 18 by Male/Female and 65 or older by Male/Female	6
Employment Status: Percent employed in Armed Forces	5

For the institutional GQ population, the first two differences originated mainly from differences in the two smallest GQ types, "Juvenile Facilities" and "Other Health Care Facilities" (the breakdown by GQ type is not shown in the Tables). The nativity differences, however, appear to originate mainly from the GQ types "Nursing Homes" and "Adult Correctional Facilities".

For the noninstitutional GQ population, most of the differences originated from the "Other noninstitutional facilities" GQ type group. The "Military" type group also contributed to the differences found in the 18–24 population and in the percent employed in the Armed Forces estimate. It is worth noting that the largest noninstitutional type group, college dorms, had no practically significant differences for these characteristics.

Lastly, we summarize the results by state instead of by characteristic (not shown in the Tables). For the nominal-only differences, we found the number of nominally large differences ranged from 0-17 characteristics out of a maximum of 76 per state. There were 38 states which had nine or fewer characteristics flagged as nominally large. Further, those states with the most differences were all smaller states: AL, AK, CT, DE, HI, ID, ME, MT, NH, ND, WV, and WY. Adding the practical significance criterion led to a much smaller number of characteristics with important differences. The maximum number of characteristics with important differences for a state was 8, although 20 states had no such characteristics (including all of the largest states) and 47 states had four or fewer. Those states with the highest number of characteristics with important differences were DE, ID, ME and WY.

7.6 Discussion and Summary

Our analysis of the state level estimates show that overall the number of nominally important differences was relatively small. There was some clustering by characteristic but most of these differences were smaller than the margin of error of the design-based estimates. In that context, only one characteristic showed important differences in more than five states, under 18 male/female, with seven states. We did not see widespread impacts that affect either a large number of states nor did we see the largest states impacted.

Our closer inspection of the data by institutional/noninstitutional categories did identify more characteristics with differences than we had when analyzing the total GQ population. The common thread to these characteristics was that they were driven mostly by the contributions from the smaller type groups. In particular the "Other Health Care" and "Other Noninstitutional" groups, which had a higher imputation rate than the other type groups, showed a more consistent impact on the estimates.

Our analysis by state showed that generally it was the smallest states which have the most characteristics whose differences are flagged as important. The largest states like California, New York, and Texas showed very few nominal differences and no differences that were also of practical significance. The results indicated that there were no states which have a high concentration of important differences - only four small states had more than five characteristics were flagged.

This initial review of these data did not show any substantial impacts on the state-level data that would necessitate major changes to our imputation methodology. An additional conclusion was, as discussed earlier, the decision not to attempt to match the GQ sampling frame with the 2010 Census listing of GQs to enhance the 2010 ACS sample listing of GQ facilities.

8. Additional Evaluations

As the evaluation of ACS data did not reveal any flaws that would lead us to halt the new methodology or force revisions, Census Bureau demographic analysts undertook a more extensive series of evaluations which examined the estimates produced by the GQ small area imputation. These evaluations were based fully on real ACS data. Subject matter analysts from the Census Bureau's Social, Economic, and Household Statistics Division and Population Division studied these estimates for their reasonableness. The evaluations showed that the GQ small area estimation generally had distributions of population for counties closer to the 2010 Census without detriment to national- and state-level estimates. Importantly, Smith (2011) found that the new method produced estimates of total population of counties by major GQ type which were more consistent with the 2010 Census than the design-based estimates, while finding no detriment in national-level age and sex estimates. Further, Jones (2012) determined the new estimates of race to be only slightly different for the nation and state. Likewise, Ramirez (2012) found only slight differences between the two methods for national and state estimates of Hispanic origin. In contrast to these favorable results, Rapino (2012) noted some deleterious effects of the new methodology in estimates of geography-dependent characteristics for place-level estimates (see Section 6, item 4 for discussion on how the limitations of the imputation with geography-dependent characteristics were mitigated).

In an evaluation that is underway at the time of this writing, we compare the design- and imputation-based 2006-2010 5-year ACS results to the 2010 Census. This evaluation aims to confirm and assess improvements for tract-level estimates, with comparisons to the 2010 Census serving as a benchmark.

Acknowledgements

The research and development of the GQ imputation for small areas was a large project with contributions from staff throughout the Census Bureau. In particular, Chandra Erdman developed the imputation methodology, Chaitra Nagaraja made important contributions to the simulation research, Patrick Joyce developed the model for the expected population sizes, and Edward Castro contributed to the weighting methodology.

References

Asiala, M. (2011a). "Description of the 1-, 3-, and 5-year GQ Weighting for 2010". U.S. Census Bureau Memorandum.

Asiala, M. (2011b). "Comparison of State-Level GQ Estimates using the Design-Based and the Imputation-Based Methodologies". U.S. Census Bureau Memorandum.

Asiala, M., Beaghen, M., and Navarro, A. (2011). "Using Imputation Methods to Improve the American Community Survey Estimates of the Group Quarters Population for Small Geographies". *2011 Joint Statistical Meetings: Proceedings of the Survey Research Methods Section.* American Statistical Association.

Bates, N. and Mulry, M. (2008). "Segmenting the Population for the Census 2010 Integrated Communications Program. *C2PO 2010 Census Integrated Communications Research Memoranda Series*, pp. 1-28.

Beaghen, M., and Stern, S. (2009). "Usability of the American Community Survey Estimates of the Group Quarters Population for Substate Geographies". *2009 Joint Statistical Meetings: Proceedings of the Survey Research Methods Section.* American Statistical Association.

Bruce, A. and Robinson, J. (2007). "Tract Level Planning Database with Census 2000 Data". U.S. Government Printing Office, Washington, DC.

Erdman, C. and Nagaraja, C. (2010). "Imputation Procedures for the American Community Survey Group Quarters Small Area Estimation". Research Report (Statistics #2010-09), Statistical Research Division, U.S. Census Bureau, Washington, DC.

Koerber, W. (2011). "Issues Paper – Place of Work, Migration, and Place of Birth Geography for GQ Imputed Records". Internal U.S. Census Bureau Memorandum

Jones, N. 2012. "Evaluation of 2010 ACS GQ Synthetic Race Data". Internal Census Bureau report.

Joyce, P. (2011). "Determining and Adjusting ACS Group Quarters Unit Level Population for Implementation with the Imputation-based Estimation Framework." Draft Report, October 2011.

Marquette, E. (2011). "2011 Group Quarters Initial Sampling Results for the American and Puerto Rico Community Surveys". Census Bureau Memorandum: DSSD 2011 American Community Survey Memorandum Series #ACS11–S–18.

Ramirez, R. 2012. "Evaluation of 2010 ACS GQ Synthetic Data". Internal Census Bureau report.

Rapino, M. 2012. "Synthetic GQ Data Review". Internal Census Bureau report.

Smith, A. 2011. "Review of 2010 ACS GQ Data". Internal Census Bureau report.

U.S. Census Bureau (2009). "Design and Methodology: American Community Survey". Issued April 2009. http://www.census.gov/acs/www/methodology/methodology_main/

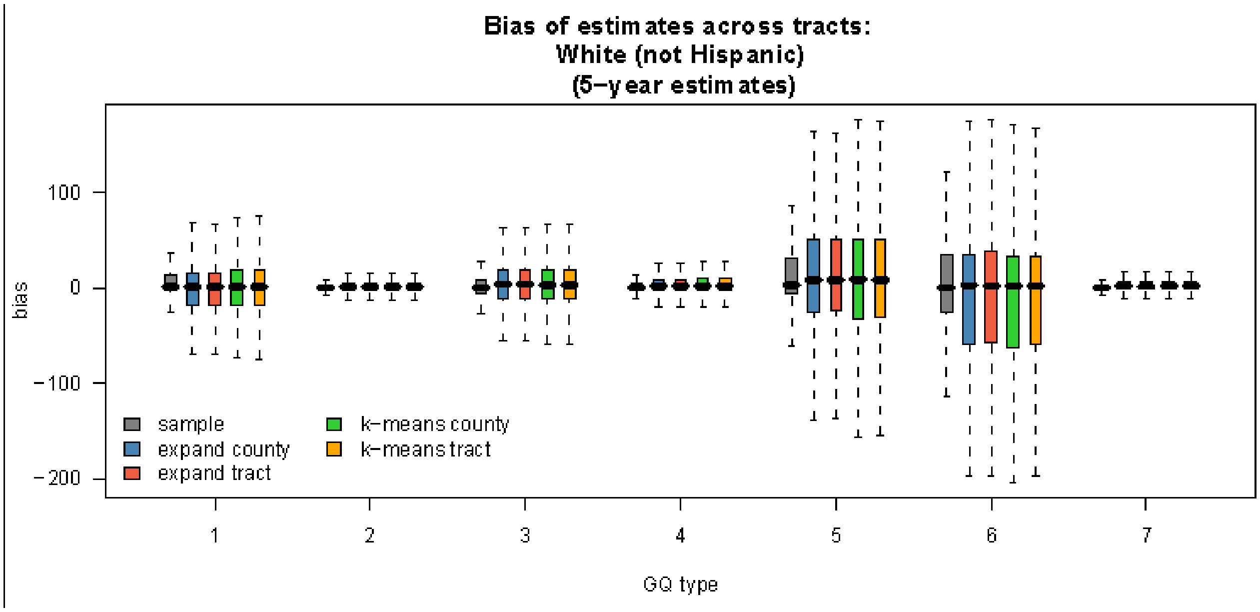
Weidman, L. (2011). "Research to Improve American Community Survey Group Quarters Estimates for Small Areas". Paper presented at the February 16, 2011 Committee on National Statistics Meeting on "ACS Group Quarters".

Williams, A. (2008). "Specifications for Selecting the American Community Survey Group Quarters Sample". Census Bureau Memorandum: 2009 American Community Survey Sampling Memorandum Series #ACS09-S-6.

APPENDIX A

FIGURES

Figure 2: Bias at the Tract, County, and State Levels (5-year estimates) by the Seven Major GQ Types



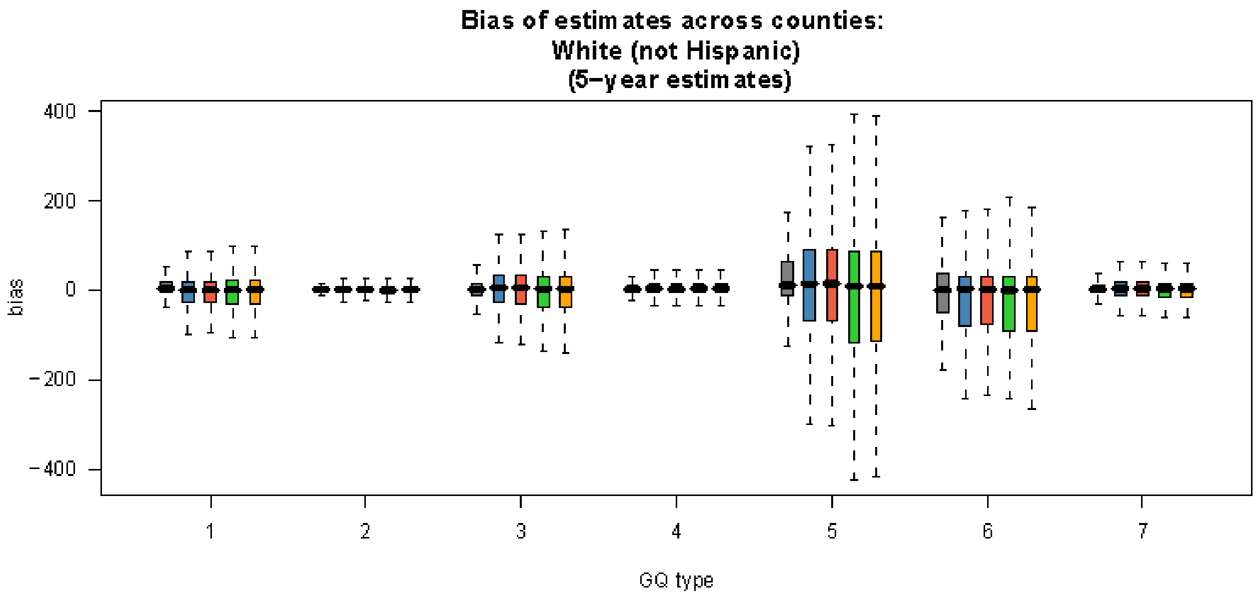


Figure 2 Continued: Bias at the Tract, County, and State Levels (5-year estimates) by the Seven Major GQ Types

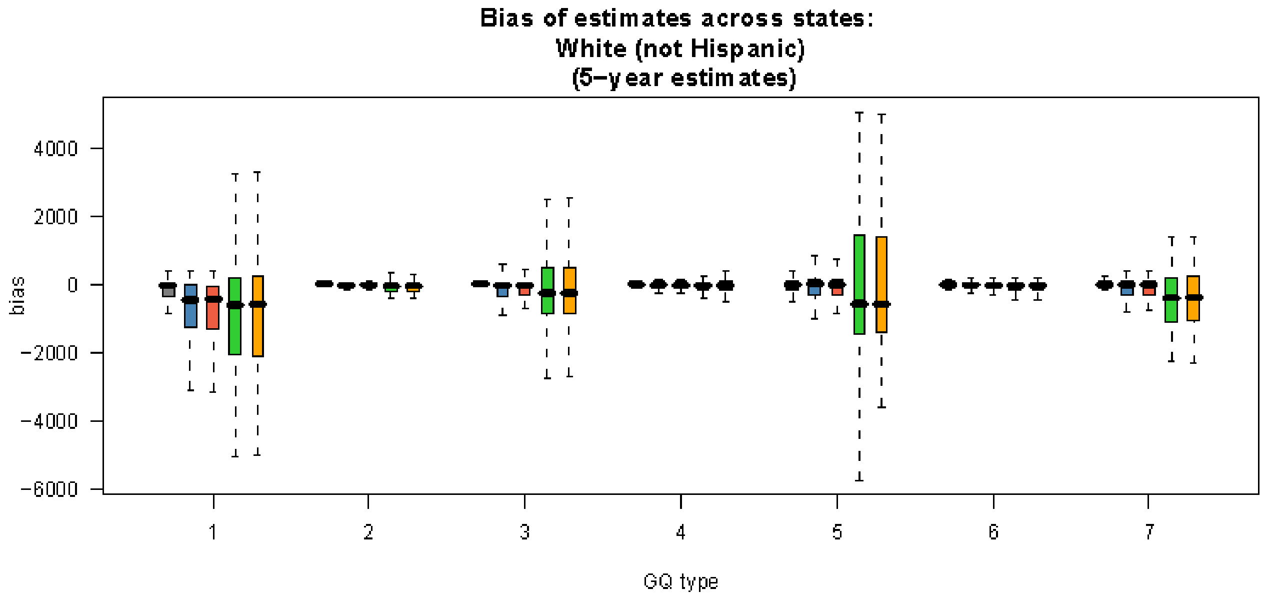


Figure 3: Design-based Estimates Against Imputation-based Estimates

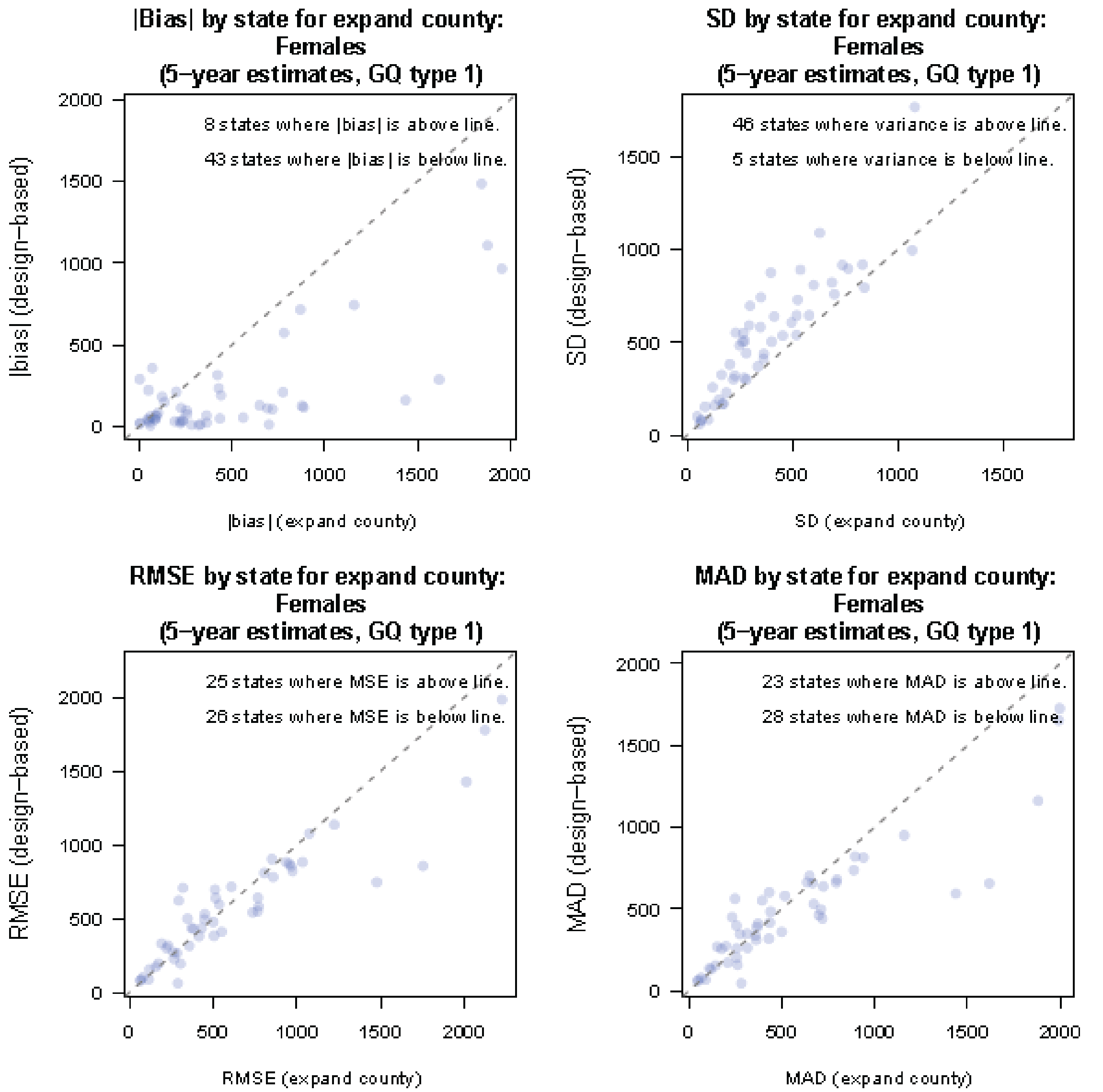
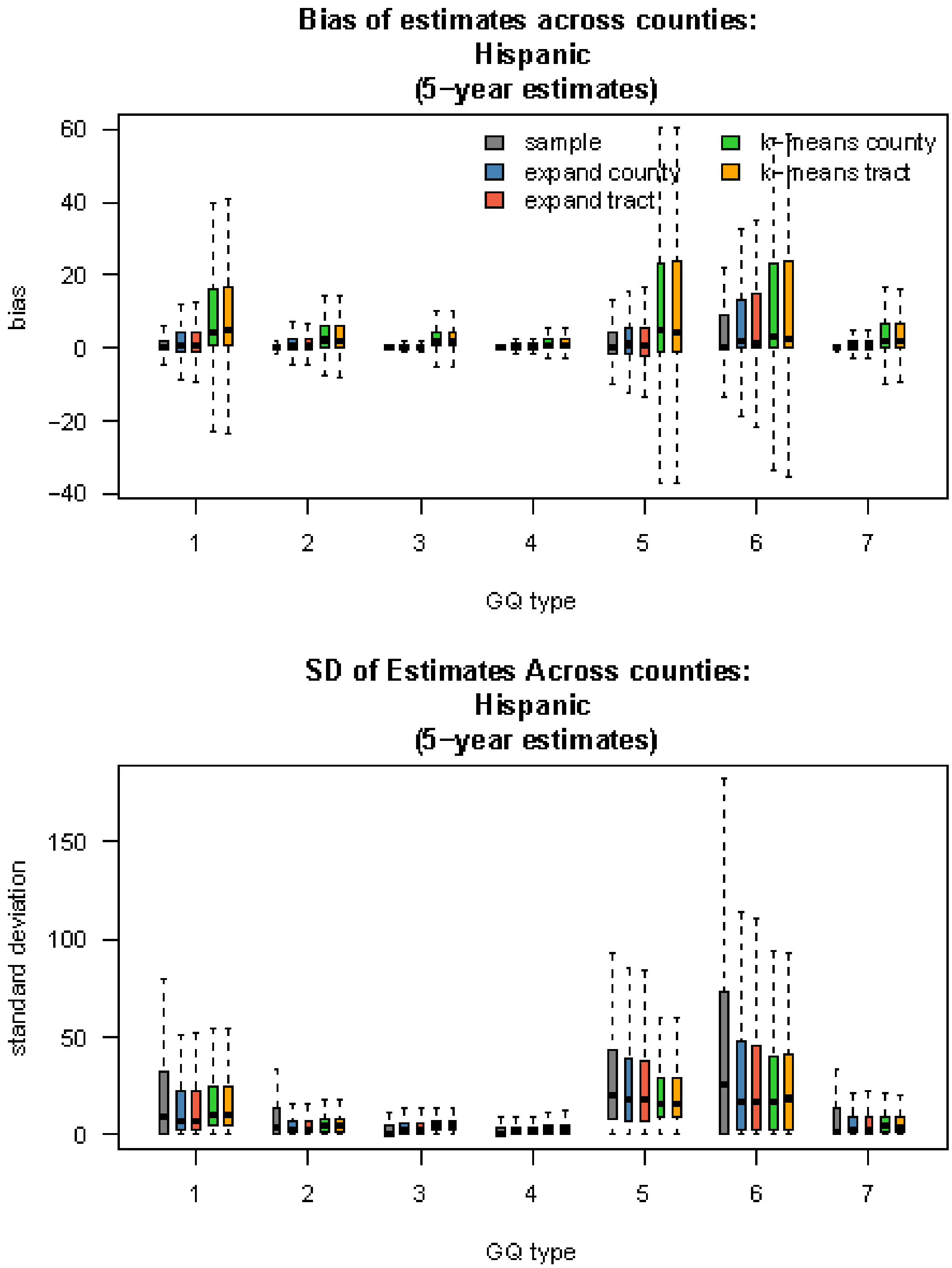


Figure 4: K-means Versus Expanding Search Donor Selection Methods



APPENDIX B

LIST OF CHARACTERISTICS TABULATED IN THE EVALUATION WITH ACS DATA

Profile Line	Profile Line Description	Type of Estimate (Count/Percent/Ratio/NA)
1	Total population	Count
2	Male	Percent
3	Female	Percent
4	Under 15 years	Percent
5	15 to 17 years	Percent
6	18 to 24 years	Percent
7	25 to 34 years	Percent
8	35 to 44 years	Percent
9	45 to 54 years	Percent
10	55 to 64 years	Percent
11	65 to 74 years	Percent
12	75 to 84 years	Percent
13	85 years and over	Percent
14	Under 18 years	Count
15	Male	Percent
16	Female	Percent
17	65 years and over	Count
18	Male	Percent
19	Female	Percent
21	One race	Count
22	White	Percent
23	Black or African American	Percent
24	American Indian and Alaska Native	Percent
25	Asian	Percent
26	Native Hawaiian and Other Pacific Islander	Percent
27	Some other race	Percent
28	Two or more races	Count
29	Hispanic or Latino (of any race)	Count
30	Not Hispanic or Latino	Count
31	White alone, Not Hispanic or Latino	Count
32	Population 15 years and over	Count
33	Now married, except separated	Percent
34	Widowed	Percent
35	Divorced	Percent
36	Separated	Percent
37	Never married	Percent
38	Population 3 years and over enrolled in school	Count

Profile Line	Profile Line Description	Type of Estimate (Count/Percent/Ratio/NA)
39	Nursery school through 12th grade	Percent
40	College or graduate school	Percent
41	Population 25 years and over	Count
42	High school graduate or higher	Percent
43	Bachelor's degree or higher	Percent
44	Civilian population 18 years and over	Count
45	Civilian veteran	Percent
46	Total population	N/A
47	With a disability	N/A
48	Population under 18 years	N/A
49	With a disability	N/A
50	Population 18 to 64 years	N/A
51	With a disability	N/A
52	No disability	N/A
53	Population 65 years and over	N/A
54	With a disability	N/A
55	Population 1 year and over	Count
56	Same address	Percent
57	Different address in the U.S.	Percent
58	Same county	Percent
59	Different county	Percent
60	Same state	Percent
61	Different state	Percent
62	Abroad	Percent
63	Total population	Count
64	Native	Count
65	Male	Percent
66	Female	Percent
67	Foreign born	Count
68	Male	Percent
69	Female	Percent
70	Naturalized U.S. citizen	Count
71	Male	Percent
72	Female	Percent
73	Not a U.S. citizen	Count
74	Male	Percent
75	Female	Percent
76	Entered 2000 or later	Percent
77	Entered 1990 to 1999	Percent
78	Entered before 1990	Percent
79	Population 5 years and over	Count

Profile Line	Profile Line Description	Type of Estimate (Count/Percent/Ratio/NA)
80	English only	Percent
81	Language other than English	Percent
82	Speak English less than "very well"	Percent
83	Population 16 years and over	Count
84	In labor force	Percent
85	Civilian labor force	Percent
86	Employed	Percent
87	Unemployed	Percent
88	Percent of civilian labor force	Percent
89	Armed Forces	Percent
90	Not in labor force	Percent
91	Civilian employed population 16 years and over	Count
92	Management, professional, and related occupations	Percent
93	Service occupations	Percent
94	Sales and office occupations	Percent
95	Farming, fishing, and forestry occupations	Percent
96	Construction, extraction, maintenance, and repair occupations	Percent
97	Production, transportation, and material moving occupations	Percent
98	Individuals	N/A
99	Per capita income (dollars)	N/A
100	Male	N/A
101	Female	N/A
102	Male	N/A
103	Female	N/A
106	With Food Stamp/SNAP benefits	N/A
107	All people	N/A
108	18 years and over	N/A
109	18 to 64 years	N/A
110	65 years and over	N/A
111	Foreign-born population excluding population born at sea	Count
112	Europe	Percent
113	Asia	Percent
114	Latin America	Percent
115	Other	Percent
116	Workers 16 years and over	Count
117	Worked in county of residence	Percent
118	Worked outside county but in state of residence	Percent
119	Worked outside state of residence	Percent
120	Civilian Noninstitutionalized Population	N/A
121	No health insurance coverage	N/A
122	Sex Ratio	Ratio

APPENDIX C

GROUP QUARTERS FACILITIES IN THE ACS BY SPECIFIC TYPE CODE AND GROUPED BY SEVEN MAJOR GQ TYPES

	(1) Correctional Institutions
(101)	Federal Detention Centers
(102)	Federal Prisons
(103)	State Prisons
(104)	Local Jails and Other Municipal Confinement Facilities
(105)	Correctional Residential Facilities
	(2) Juvenile Detention Facilities
(201)	Group Homes for Juveniles
(202)	Residential Treatment Centers for Juveniles
(203)	Correctional Facilities Intended for Juveniles
	(3) Nursing Homes
(301)	Nursing Facilities/Skilled Nursing Facilities
	(4) Other Long-term Care Facilities
(401)	Mental (Psychiatric) Hospitals/Psychiatric Units in Other Hospitals
(402)	Hospitals with Patients Who Have No Usual Home Elsewhere
(403)	In-Patient Hospice Facilities
(404)	Military Treatment Facilities with Assigned Patients
(405)	Residential Schools for People with Disabilities
	(5) College Dormitories
(501)	College/University Housing
	(6) Military Facilities
(106)	Military Disciplinary Barracks and Jails
(601)	Military Quarters
(602)	Military Ships
	(7) Other Non-institutional Facilities
(701)	Emergency and Transitional Shelters for People Experiencing Homelessness
(801)	Group Homes Intended for Adults
(802)	Residential Treatment Centers for Adults
(901)	Workers Group Living Quarters and Job Corps Centers
(902)	Religious Group Quarters