

# An Analysis of Alternate Variance Estimation Methods for the American Community Survey Group Quarters Sample

Don Keathley, Alfredo Navarro, Mark Asiala  
U.S. Census Bureau, Washington, D.C., 20233

## Abstract

The American Community Survey (ACS) began producing estimates of the group quarters (GQ) population, for the full implementation of the survey, in 2006. The ACS uses a successive difference replication (SDR) method to compute variances for these estimates. Part of the SDR method involves the assignment of a set of replicate factors to each sample GQ person. In 2006 and 2007, the ACS assigned these sets at the person level. This assignment method resulted in the underestimation of GQ population variances, as they did not take clustering of persons in GQs into account. An alternate assignment method more accurately reflects the clustering effect that is associated with GQ sample persons. Alternative variance estimation methods (jackknife, random group) provide benchmarks with which to compare a preferred SDR method.

**Key Words:** American Community Survey, group quarters, variance

## 1. Background

The American Community Survey (ACS) started collecting annual sample data for persons in group quarters (GQ) in 2006. The statistical methods side of the data collection process includes the sample design (includes sample selection), weighting, and variance estimation. An examination of the variance estimation methodology that the ACS used in 2006 and 2007 suggested that the ACS may have underestimated GQ variances in these two years – an alternative variance estimation method could correct this situation.

This potential underestimation of GQ variances would be most noticeable at the lowest levels of geography for which the ACS produces GQ estimates – this is due to the ACS GQ sample being a state-based design, where sub-state estimates are not taken into account. For single-year GQ estimates in 2006 and 2007, sub-state estimates were published at the county level. The only single-year county-level GQ estimates that the ACS published for 2006 and 2007 were for the total GQ population.

### 1.1 Sampling

In 2006 and 2007, the ACS used a state-level stratified sample to select sample persons in GQs. There were two sampling strata per state. One stratum included GQs with expected populations of 16 persons or more and that were open on census day 2000 (large GQ

---

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

stratum). The other stratum included GQs with expected populations of 15 persons or less and GQs that were closed on census day 2000 (small GQ stratum; Hefter, 2005 and 2006).

The ACS used two stages of sample selection in the large GQ stratum. First, the ACS used the GQs' expected populations to form a string of expected persons within each state. It then proportionally reduced each string by a factor of 10 – this resulted in strings of clusters, where each cluster contained ten expected persons. The ACS then selected a systematic 1-in-40 sample of clusters within each state. Each selected cluster is referred to as a hit (i.e., a “hit” cluster). Each hit is associated with a GQ. The ACS goes to these GQs and selected a systematic sample of 10 persons for each hit, using the entire actual GQ population as the frame. GQs with expected populations of less than 400 persons were in sample with a probability proportional to their size; they are associated with either zero or one sample hit. GQs with expected populations of 400 persons or more were in sample with certainty, and could be associated with multiple sample hits.

The ACS selected sample in the small GQ stratum in two stages. The first stage consisted of selecting sample GQs; the second stage involved the selection of sample persons within sample GQs. In the first stage, the ACS selected GQs at a flat 1-in-40 rate, using two phases of sample selection. In the first phase, the ACS systematically divided the small GQ stratum sampling frame into five equal-sized groups, by state. These groups became the once-in-five year sampling frames for the small GQ stratum<sup>1</sup>. In the second phase, the ACS selected a systematic 1-in-8 sample of GQs from the given year's frame. The ACS selected sample persons in sample GQs in the second stage. If a sample GQ's actual population had 15 or fewer persons, then all persons were in sample. Otherwise, the ACS selected a systematic subsample of 10 persons from the GQ. Each group of sampled persons from a small stratum GQ represented a sampled hit.

## 1.2 Weighting

There were three stages of weighting for GQ sample persons (Keathley, 2006a and 2007a). The first stage calculated baseweights for all sample persons; these weights were functions of a person's probability of selection (initial baseweight), adjustments due to differences in actual and expected GQ populations, and a trimming procedure<sup>2</sup>. The second stage adjusted the baseweights to account for non-interviewed persons (non-interview adjusted (NIA) weights). The third stage adjusts the NIA weights in a post-stratification procedure, so that the state-level sums of the weights equal independent population controls by major GQ type. The weights that emerged from the post-stratification procedure were unrounded – the ACS rounded these weights to integers. These rounded weights are those that the ACS used in computing GQ estimates for 2006 and 2007, including estimates of county-level totals of the GQ population.

## 1.3 Variances

The ACS uses a successive difference replication (SDR) method for computing variances. This method specifically takes systematic sampling into account (Fay, 1995). In 2006 and 2007, the procedure first sorted person-level records by sampling stratum,

<sup>1</sup> These groups are updated on a regular basis.

<sup>2</sup> The trimming procedure accounted for unexpectedly large baseweights. The ACS implemented this procedure in 2007; in 2006, the ACS used a capping procedure.

state, GQ order of selection (GOS), and person ID. GOS takes within-GQ hit order of selection into account. It then assigned each person a set of 80 replicate factors, where the factors were either 1.0, 0.292893221, or 1.707106781. Sample persons in the same hit received different sets of replicate factors in 2006 and 2007 (original assignment method). These assignments were done independently by state within sampling stratum. The procedure then multiplied every sample person's initial baseweight by each of the person's factors, resulting in 80 replicate weights. Each replicate weight then went through a modified version of the weighting process in the previous section, ending in 80 final replicate weights. The SDR method uses sums of the final and replicate weights to estimate variances (Fay, 1995; Starsinic, 2007; Keathley, 2006b and 2007b).

$$var(\hat{Y}_{cty,0}) = \frac{4}{80} \sum_{r=1}^{4k} (\hat{Y}_{cty,r} - \hat{Y}_{cty,0})^2 \quad (1)$$

where           cty = county  
                   r = replicate  
 $\hat{Y}_{cty,0}$  = estimated GQ population total in county cty  
 $\hat{Y}_{cty,r}$  = estimated GQ population total in county cty, replicate r

#### 1.4 Original, Alternate Assignment Methods

The variance estimation procedure in section 1.3 resulted in what appeared to be underestimates of the standard errors for county-level estimates of the total GQ population. A possible reason for this is that the original assignment method (assigning different sets of replicate factors to persons in the same hit) did not account for clustering of persons within GQs. Alternate assignment methods that could better account for clustering are to assign the same set of replicate factors to every person in a GQ or hit, i.e., assign sets of replicate factors at the GQ and hit levels, respectively. This assumes there is enough of a clustering effect so that the choice of assignment method matters.

## 2. Methodology

We conducted our research on a flow basis, where we proceeded from one bit of analysis to another. Our research started by determining whether there was enough evidence to support the examination of the SDR variance estimation methodology that the ACS was indeed underestimating GQ variances using the original assignment method (section 2.1). If the evidence was sufficient, then would assigning sets of replicate factors at the hit or GQ level result in more accurate estimates of the standard errors, i.e., is there enough of a clustering effect so that the choice of assignment method matters (section 2.2)? If the choice of assignment method did matter, then we wanted to determine which assignment method was better at taking the clustering of persons in GQs into account (section 2.3). Finally, we wanted to compare the standard errors from our preferred assignment method(s) to those from using other variance estimation methods (section 2.4) – do the comparisons support the use of the preferred assignment method(s)?

## 2.1 Is there Sufficient Evidence to indicate that the ACS Underestimated GQ Variances using the Original Assignment Method

To answer this question, we first obtained the estimates of GQ population totals and their margins of error<sup>3</sup> for 2006 and 2007 for each county for which the ACS published estimates in 2007. Then we computed the differences in these estimates for each county, along with the Z-score for each difference. We assumed that actual changes in county-level total GQ populations from 2006 and 2007 were small enough so that the expected distribution of the Z-scores should approximate the unit normal distribution. The more the actual Z-score distribution departs from that of the unit normal, the more likely it is that the ACS underestimated the standard errors. Table 2 in section 3.1 shows the actual and expected Z-score distributions.

## 2.2 Comparing Alternate Assignment Methods to the Original Assignment Method – Does the Choice of Assignment Method Matter?

If there was enough evidence to conclude that the ACS was underestimating GQ variances using the original assignment method, then we would want to determine if an alternate version of the assignment method would result in more accurate estimates of the standard errors, i.e., would they better account for the clustering of persons in GQs. The two alternatives we considered were to assign sets of replicate factors at the hit and GQ levels, where everyone in the same hit or GQ, respectively, would receive the same set of factors.

We selected all possible samples from the 2007 GQ sample frame; we used the methodology in section 1.1 to select our samples. The result was 40 samples in each of the two sampling strata, amounting to 1,600 total samples in each county.

We simulated estimates of the total GQ population for each of the 1,600 samples in each county. This was done by summing the baseweights for the sample persons in small GQ stratum  $s$  + the baseweights for the sample persons in large GQ stratum  $l$ , where  $s \in \{1, \dots, 40\}$  and  $l \in \{1, \dots, 40\}$ . We used baseweights because none of the samples were implemented, e.g., we did not send questionnaires to any person in any of the samples. Consequently, we assumed all sample persons in all samples were interviews, thereby eliminating any non-interview adjustments. We also assumed each GQ's actual population was equal to its expected population, thereby making actual within-GQ sampling rates equal to expected within-GQ sampling rates. Finally, we assumed the summation of baseweights to the state -by- major GQ type level would be our estimates of totals, thereby avoiding the post-stratification adjustment.

Then we simulated variance and standard error (SE) estimates for all 1,600 total GQ population estimates in each county. We computed these variance and SE estimates using the original and both alternate replicate factor assignment methods. Then we computed the average SEs, across all samples, for each method in each county.

---

<sup>3</sup> Margins of error = standard error  $\times$  Z-score. The ACS uses 90 percent confidence intervals, so the Z-score is 1.645)

As a benchmark, we computed the variance and standard deviation for each county-level total GQ population estimate, across all samples in a county. Equations (2) and (3) show the variance and standard deviation equations, respectively, that we used:

$$S_{cty}^2 = \frac{1}{1,600-1} \sum_{smp=1}^{1,600} (\hat{Y}_{cty,smp} - \hat{\bar{Y}}_{cty})^2 \quad (2)$$

where  $smp$  = sample  
 $\hat{Y}_{cty,smp}$  = estimated total GQ population in county  $cty$ , sample  $smp$   
 $\hat{\bar{Y}}_{cty}$  = average estimate of the total GQ population in county  $cty$ ,  
 across all samples  
 $S_{cty}^2$  = variance of the estimated total GQ populations,  
 across all samples

$$S_{cty} = \sqrt{S_{cty}^2} \quad (3)$$

where  $S_{cty}$  = standard deviation of the estimated total GQ  
 populations, across all samples

Finally, we produced national-level quantiles of both the average county-level SEs from using the three assignment methods as well as the standard deviations. We omitted counties with zero GQs ( $N=0$ ) from the quantiles. Comparisons of the quantiles between the three methods themselves and the standard deviations would indicate which, if any, of the assignment methods best takes clustering of persons in GQs into account; they would also indicate any potential bias in the three methods' SEs. Table 3 in section 3.2 shows these distributions.

### 2.3 Comparing Alternate Methods using Sample Data

If the analysis in 2.2 showed that one or both alternate assignment methods were better at taking clustering of persons in GQs into account than the original assignment method, then we wanted to determine if either alternate method was better than the other in taking clustering into account. To do this, we first computed standard errors for the actual 2006 and 2007 county-level estimates of the total GQ population using both alternate assignment methods as well as the original method. Next, we performed statistical comparisons for each county's 2006 versus 2007 estimates for all three assignment methods; we used the 10-percent level of significance, which is the Census Bureau standard. We then computed national level proportions of the counties that had statistically significant differences between their 2006 and 2007 totals, for each method. We omitted counties with estimates of zero total GQ population in both years from these proportions. These proportions, along with the GQ sample design (how the ACS selects GQ samples), would allow us to make a preference of assignment method. Table 4 in section 3.3 shows these proportions.

### 2.4 Comparing the Preferred Alternate-Assignment SDR Method to other Variance Estimation Methods

After selecting our preferred alternate-assignment SDR method (preferred SDR method) in section 2.3, we wanted to compare this method against other variance estimation

techniques, to see how similar or diverse the average standard errors (SEs) would be across methods. The methods we chose were a 20-group jackknife and a 25-group random group technique (Wolter, 1985). We chose these numbers of groups in an attempt to minimize the mean squared errors of the resulting variances. In addition to estimates of overall totals, we wanted to compare the average SEs from these three methods for county-level estimates of demographic totals as well. We used 2000 Census data to do this.

As in section 2.2, we would take all possible samples of hits in each state. Since we were comparing replicate methods only using a non-ACS sampling frame, we adjusted the sampling procedure in section 1.1 by using GQ size as an implicit stratification variable<sup>4</sup>, instead of an , whereas GQ size . This resulted in forty possible systematic samples of hits in each state. We selected only one sample of ten persons per hit. Every sample person had a baseweight of 40. Also as in section 2.2, there was no implementation of the samples, so the baseweights are what we used to compute county-level estimates of totals (see above).

We chose sex, age, and race as our demographic variables. We computed county-level weighted totals, for each sample, for each of the following categories.

**Table 1. Demographic Variable Categories**

Sex	Age	Race	
Male	0-17	White only	Asian only
Female	18-64	Black only	Native Hawaiian/Pacific Islander only
	65+	American Indian/Alaskan Native only	Some other race only
			Two or more races

We computed variances and standard errors as before, using the preferred assignment method (assigning the same set of replicate factors to all persons in either the same GQ or hit). Using the same sort used in sampling, we assigned persons in either the same GQ or hit to the same group for both the jackknife and random group methods. Then we computed the variance estimates for the jackknife and random group methods using equations (4) and (5), respectively.

$$var(\hat{Y}_{jk,cty,smp,dvc}) = \frac{19}{20} \sum_{k=1}^{20} (\hat{Y}_{k,cty,smp,dvc} - \hat{Y}_{cty,smp,dvc})^2 \quad (4)$$

$$var(\hat{Y}_{rg,cty,smp,dvc}) = \frac{1}{25(24)} \sum_{k=1}^{25} (\hat{Y}_{k,cty,smp,dvc} - \hat{Y}_{cty,smp,dvc})^2 \quad (5)$$

where jk = jackknife  
 rg = random group  
 k = group k  
 dvc = demographic variable category  
 = blank for overall totals

<sup>4</sup> Stratum was the first sort variable. The remaining sort variables mimicked, as closely as possible, the sorting scheme the ACS uses when selecting its GQ samples.

$$\begin{aligned}
\hat{Y}_{k,cty,smp,dvc} &= \text{estimated total GQ population in group } k, \text{ county } cty, \\
&\quad \text{sample } smp, \text{ demographic variable category } dvc \\
\hat{Y}_{cty,smp,dvc} &= \text{estimated total GQ population in county } cty, \text{ sample } smp, \\
&\quad \text{demographic variable category } dvc \\
\hat{\bar{Y}}_{cty,smp,dvc} &= \text{average estimate of the total GQ population in county } cty, \\
&\quad \text{sample } smp, \text{ demographic variable category } dvc, \text{ across} \\
&\quad \text{all groups} \\
var(\hat{Y}_{jk,cty,smp,dvc}) &= \text{jackknife estimator of the variance for the estimated total} \\
&\quad \text{GQ population in county } cty, \text{ sample } smp, \text{ demographic} \\
&\quad \text{variable category } dvc \\
var(\hat{Y}_{rg,cty,smp,dvc}) &= \text{random group estimator of the variance for the estimated} \\
&\quad \text{total GQ population in county } cty, \text{ sample } smp, \text{ demo-} \\
&\quad \text{graphic variable category } dvc
\end{aligned}$$

As in section 2.2., we computed the variance and standard deviation for each county-level estimate, across all samples in a county, using equations (2) and (3), as benchmarks. We also produced national-level quantiles of both the average county-level SEs for the three variance estimation methods as well as the standard deviations, for all estimates. We omitted counties with zero GQs ( $N=0$ ) from the quantiles as well. Comparisons of the quantiles between the three methods themselves would indicate the level of consistency between the three methods, in terms of the average SEs. Comparisons of the quantiles between the three methods and the standard deviations would indicate any bias in the three methods' SE estimates. Table 4 in section 3.4 shows these distributions.

### 3. Results/Analysis

This section presents summary results and analyses for the four subsections in section 2, in the same order of presentation.

#### 3.1 Were the Original Method Standard Errors Underestimates?

There were 799 counties and municipalities for which the ACS published single-year GQ estimates of the total GQ population in 2007. We computed the differences between these estimates versus those from 2006 for each of these counties. We then computed the Z-score for each difference, using the standard errors from the SDR technique that used the original replicate factor assignment method. Table 2 shows ranges of the Z-scores, along with the number and percent of counties whose Z-scores fell into a particular range. It also shows the expected percent distribution of Z-scores, assuming ignorable changes in GQ populations between adjacent years.

**Table 2. Distribution of Counties with Published 2007 County-Level Estimates of GQ Population Totals, by Z-score of the Differences in 2006 versus 2007 Estimates**

Z-score Ranges	Counties	Actual Percent	Expected Percent
<b>Total Counties</b>	<b>799</b>		
(-∞, -4.00]	203	25.4	< 0.1
(-4.00, -1.96]	85	10.6	≈ 2.5
(-1.96, 1.96)	212	26.5	≈ 95.0
[ 1.96, 4.00)	83	10.4	≈ 2.5
[ 4.00, ∞)	216	27.0	< 0.1

Only 212 (26.5 percent) of these counties had a Z score of the difference between -1.96 and 1.96 – around ninety-five percent of the differences should have been in this range. Moreover, over half of the Z scores had values greater than or equal to | 4.00 |. These results were a strong indication that the ACS had underestimated the variances for the 2006 and 2007 estimates of the GQ county-level total population. This was sufficient evidence to examine the two alternate replicate factor assignment methods.

### 3.2 Comparing Alternate Assignment Methods to the Original Assignment Method

Table 3 shows quantiles of the average county-level SDR SEs from using the original and two alternate assignment methods. Average are across all 1,600 samples in a county. It also shows the quantiles for the county-level standard deviations. These statistics are based on the 2007 ACS GQ sample frame. The quantiles omit counties that had zero GQs in the universe in 2007.

**Table 3. Quantiles of Average County-Level SEs for Estimates of the GQ Population Total – 2007 ACS GQ Sample Frame<sup>5</sup>**

Method	Total Counties	Quantiles						
		Min	p10	q1	Med	q3	p90	Max
Original AM <sup>A</sup>	3,094	1.0	9.8	21.0	46.5	69.8	99.3	233.8
Alt. AM, GQ-level	3,094	1.0	85.7	193.6	456.3	756.1	1,308.7	6,918.7
Alt. AM, hit-level	3,094	1.0	86.0	193.4	450.0	668.9	926.3	2,319.0
Stand. Dev. (S)	3,094	6.3	167.5	215.1	300.2	420.4	580.2	1,535.8

<sup>A</sup> AM = assignment method

The average SEs from using the original assignment method are all much smaller than their corresponding S values – this reinforces the results we saw in Table 2. Additionally, the average SEs from using the two alternate assignment methods are generally closer to their corresponding S values than those from using the original method. These two

<sup>5</sup> See the attachment for definitions for column headings.



observations were sufficient evidence to conclude that the two alternate assignment methods were both better than the original method in taking the clustering of persons in GQs into account.

Table 3 also shows that, from the median to the maximum quantiles, the average SEs from using the hit-level assignment method are closer to their S counterparts than those from using the GQ-level assignment method. This indicates the hit-level method may be better at taking clustering into account than the GQ-level method.

There is also an indication that the two alternate assignment methods might be overestimating variances. This can be seen in the median to maximum quantiles, where the average SEs for the two alternate methods are noticeably higher than their corresponding S values.

### 3.3 Comparing Alternate Assignment Methods using Sample Data

Table 4 shows the number and percent of counties that had statistically significant differences in their 2006 versus 2007 estimated GQ population totals, from using all three assignment methods. These include counties for which the ACS did not publish estimates, i.e., counties with a total population of less than 65,000. Additionally, we only included those counties with a positive (non-zero) GQ population estimate in at least one of the two years.

**Table 4. Number and Percent of Statistically Significant Differences between 2006 versus 2007 County-Level GQ Population Estimates, by Replicate Factor Assignment Method**

Assignment Method	Total Counties	Number SD <sup>B</sup>	Percent SD <sup>B</sup>
Original	2,499	2,038	81.6%
Alternate – GQ-level Assignments	2,499	181	7.2%
Alternate – Hit-level assignment	2,499	220	8.8%

<sup>B</sup> SD = significantly different

The ACS uses 90-percent confidence intervals, so we expected about ten percent of the differences for any of these methods to be statistically significant (assuming ignorable changes in GQ populations between adjacent years). Over eighty percent of the differences were statistically significant from using the original assignment method – this agrees with the general result in Tables 2 and 3, in that using the original assignment method underestimated the variances. Conversely, 7.2 and 8.8 percent of the differences were statistically significant using the alternate GQ-level and hit-level assignment methods, respectively. These results are consistent with what we saw in Table 3, in that the hit-level method is somewhat better at taking the clustering of persons in GQs into account. They also show the potential for the overestimation of variances from using the two alternate assignment methods

We preferred assigning replicate factors at the hit level. Our reasons for making this selection were:

- the ACS selects hits (not GQs) for sample
- the percent of statistically significant differences in Table 4 for this method was 8.8 percent, which was closer to the expected ten percent than the 7.2 percent statistic for GQ-level assignments

### 3.4 Comparing the SDR Estimator using the Hit-Level Assignment Method (Preferred SDR Method) to other Variance Estimators

Table 5 shows quantiles of the average county-level standard errors for the jackknife, random group, and preferred SDR methods for totals as well as their corresponding S values. Averages are across all 40 samples for a county. The statistics are based on 2000 Census GQ data. Tables 6, 7 and 8 show these quantiles for selected demographic characteristic totals. The quantiles omit counties that had zero GQs in the 2000 Census.

The four lower quantiles for the preferred SDR method in Table 5, for totals, are comparable to those for the jackknife and random group methods. From the median to the maximum, they are noticeably smaller than their jackknife and random group counterparts. The average SEs for the preferred SDR method are also closer to their corresponding S values than the average SEs for the jackknife and random group methods for the median through maximum quantiles as well. There is stronger evidence of variance overestimation than in Table 3, however, in that the average SEs for the preferred SDR method are larger than their corresponding S values starting with the 25<sup>th</sup> percent quantile (q1).

**Table 5. Quantiles of Average County-Level SEs for Estimates of the GQ Population Total – 2000 Census Data<sup>6</sup>**

Method	Total Counties	Quantiles						
		Min	p10	q1	Med	q3	p90	Max
Jackknife	3,089	1.0	85.1	192.9	471.6	835.5	1,250.5	3,362.2
Random Group	3,089	1.0	85.0	193.0	473.4	856.0	1,284.9	3,297.9
Preferred SDR	3,089	1.0	85.6	192.4	447.2	655.0	872.6	1,797.1
Standard Dev. (S)	3,089	6.2	137.3	172.3	255.7	369.7	500.3	1,282.3

**Table 6. Quantiles of Average County-Level SEs for Estimates of the GQ Population Total for Males – 2000 Census Data<sup>6</sup>**

Method	Total Counties	Quantiles						
		Min	p10	q1	Med	q3	p90	Max
Jackknife	3,089	0.0	31.6	80.0	242.8	606.2	900.4	2,368.6
Random Group	3,089	0.0	31.6	80.0	245.6	613.9	938.2	2,951.5
Preferred SDR	3,089	0.0	31.7	79.7	240.4	453.6	651.6	2,265.9
Standard Dev. (S)	3,089	0.0	63.2	111.0	186.6	297.2	442.8	1,765.8

<sup>6</sup> See the attachment for definitions for column headings.

**Table 7. Quantiles of Average County-Level SEs for Estimates of the GQ Population Total for Persons 18-64 – 2000 Census Data<sup>7</sup>**

Method	Total Counties	Quantiles						
		Min	p10	q1	Med	q3	p90	Max
Jackknife	3,089	0.0	15.0	53.4	244.2	689.3	982.5	2,801.2
Random Group	3,089	0.0	15.0	53.4	246.7	703.9	1,037.0	2,560.1
Preferred SDR	3,089	0.0	15.0	53.4	247.1	479.8	674.1	2,294.8
Standard Dev. (S)	3,089	0.0	47.4	108.2	183.0	277.0	402.7	1,772.3

**Table 8. Quantiles of Average County-Level SEs for Estimates of the GQ Population Total for Blacks only – 2000 Census Data<sup>7</sup>**

Method	Total Counties	Quantiles						
		Min	p10	q1	Med	q3	p90	Max
Jackknife	3,089	0.0	0.0	4.0	49.2	244.7	480.9	1,900.0
Random Group	3,089	0.0	0.0	4.0	49.4	245.6	496.7	1,951.3
Preferred SDR	3,089	0.0	0.0	4.0	48.6	211.8	349.8	1,422.5
Standard Dev. (S)	3,089	0.0	0.0	13.4	64.4	152.2	252.2	1,632.5

Tables 6, 7, and 8 show a cross-section of quantiles for various demographic characteristics. They show the same general results as what is shown in Table 5. The main difference is the point where the S values become smaller than the corresponding average SEs for the preferred SDR method, e.g., they become smaller at the 90<sup>th</sup> percent quantile (p90) in Table 8. Quantiles for the other demographic categories, e.g., females, are similar to those shown here.

#### 4. Conclusions

The results in section 3 indicate that a replacement of the original replicate factor assignment method with an alternate method, when computing SDR variances for estimates of GQ population totals, was in order. The alternate method we preferred assigns replicate factors at the hit level. This method compared favorably to the jackknife and random group methods in Tables 5 through 8.

These results also suggest that both alternate assignment methods somewhat overestimate the variances, when compared to the benchmarks in Table 3 as well as the percent of statistically significant differences in Table 4. Tables 5 through 8 mostly reinforce the suggestion of overestimation of variances, from using the hit-level assignment method only. This potential overestimation, however, is nowhere near as marked as the underestimation from the original SDR method. And, the potential overestimation would err on the conservative side when conducting statistical tests.

<sup>7</sup> See the attachment for definitions for column headings.

## 5. Limitations

The main limitations stem from the analysis done with the Census 2000 data. One is that the jackknife and random group variance estimators tend to be biased when used with systematic sample designs (Tables 5 through 8 show evidence of this), which may make comparisons to these methods less than ideal. A second limitation is that although we examined variances for demographic totals using the census data, our focus was on estimates of the total GQ population. Our conclusion was for estimates of the total GQ population only as a result. A third limitation is that we did not select all possible person samples within GQs using the census data.

## 6. Future Research

Additional research could focus on by how much and why the preferred SDR method overestimates variances. There is some bias inherent in the SDR method, but how much of the overestimation is due to this bias (Fay, 1995)? It could also investigate methods for taking the reason for overestimation into account, either through adjusting the preferred SDR method or using a different variance estimator. An issue with using a different variance estimator would be to find a way to generate variances for estimates of the total population (combining GQ plus housing unit totals) – housing unit variances employ the same SDR technique.

Future research could look at variance estimates for demographic characteristics using ACS sample data, at either the county or state levels. Variance estimates at the state level could include those for proportions.

## 7. References

Fay, R.E. and Train, G.F. (1995), “Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties,” presented at the American Statistical Association Conference in Orlando, FL August 13-17, 1995

Hefter, S. 2005 “Specifications for Selecting the American Community Survey Group Quarters Sample.” 2006 American Community Survey Sampling Memorandum Series #ACS06-S-6

Hefter, S. 2006 “Specifications for Selecting the American Community Survey Group Quarters Sample.” 2007 American Community Survey Sampling Memorandum Series #ACS07-S-6

Keathley, D. 2006a “American Community Survey: Specifications for Calculating the Weights for the 2006 GQ Sample” 2006 American Community Survey Weighting Memorandum Series #ACS06-W-6

Keathley, D. 2006b “American Community Survey: Specifications for Calculating the Replicate Weights for the 2006 GQ Sample” 2006 American Community Survey Weighting Memorandum Series #ACS06-W-8

Keathley, D. 2007a “American Community Survey: Specifications for Calculating the Weights for the 2007 GQ Sample” 2007 American Community Survey Weighting Memorandum Series #ACS07-W-9

Keathley, D. 2007b “American Community Survey: Specifications for Calculating the Replicate Weights for the 2007 GQ Sample” 2007 American Community Survey Weighting Memorandum Series #ACS07-W-13, draft memorandum

Starsinic, M. (2007) “Specifications for Calculating Direct Standard Errors and Confidence Intervals for 2007 ACS Data Products” 2007 American Community Survey Variance Memorandum Series #ACS07-V-1

Wolter, K.M. (1985), *Introduction to Variance Estimation*, New York : Springer-Verlag

**Table A. Definitions for Column Headings in Tables 3, 5, 6, 7, and 8**

<b>Column Heading</b>	<b>Definition</b>
Min	Minimum
p10	Ten-percent quantile
q1	Twenty-five percent quantile
Med	Median
q3	Seventy-five percent quantile
p90	Ninety-percent quantile
Max	Maximum