Group Quarters Sampling and Estimation Full Implementation Plans for the American Community Survey

Mark E. Asiala, US Census Bureau
2023-2 DSSD, Washington DC 20233-7613, mark.e.asiala@census.gov

Key Words: American Community Survey, ACS, group quarters, GQ, sampling, estimation

1 Background

The American Community Survey (ACS) is part of the Census Bureau's plans for a re-engineered 2010 Census. The ACS will collect long-form (sample) data on an annual basis in order to produce single and multi-year estimates which are comparable to the long-form estimates traditionally produced after each decennial census.

The Census Bureau classifies the living quarters for the total population into two broad categories for sampling and estimation purposes: housing units and group quarters (GQs). Housing units can be a single-family house, townhouse, mobile home or trailer, apartment, group of rooms or a single room that is occupied or intended as a separate living quarters. In constrast, group quarters is a living quarter in which unrelated people live or stay other than the usual house, apartment, or mobile home. For example, nursing homes, prisons, college dormitories, military barracks, group homes, and shelters are all examples of group quarters. The Census Bureau also divides the group quarters into two subcategories: institutional and non-institutional. Nursing homes and prisons are two examples of institutional group quarters. The other types of GQs given previously are examples of non-institutional group quarters.

The ACS was intially implemented in four test counties in 1996. Only housing units were sampled in 1996. 1997 included a small-scale test of field operations in one ACS site. A second test was conducted in the last half of 1998 in eight test counties. In 1999, a full scale GQ sample was selected and interviewed in the 36 ACS test counties. In 2001 a similar GQ sample was selected and interviewed but, in 2000, the GQ sample was not selected to avoid interference with the Census 2000 operations. For budgetary reasons, GQs have not been included in the ACS since 2001.

As the ACS moves towards full implementation, it will include a national GQ sample. The ACS will sample 2.5% of the GQ population in order to produce estimates for the total population and also a limited set of estimates for the GQ population.

The design for creating this sample has posed several challenges which are outlined in this paper including frame development, sampling and estimation issues.

2 Frame Development

The starting point for the sampling frame for the ACS GQ sample is the special place / group quarters (SPGQ) file from Census 2000. A special place is defined as the administrative office governing the GQ. It may or may not be in the same physical location as the GQ. For example, all state prisons for a state may report to the same administrative office located in that state's capital. This frame has had several enhancements through various clerical operations and additional data sources in order to improve the coverage. They included:

- Unduplication
- Refinement of GQ counts
- New data sources

The operational details of these operations are contained in Loudermilk, 2004.

2.1 Unduplication

The initial frame obtained from the Census 2000 GQ operation contained a number of duplicate GQ records. Some of these duplicates occurred because of multiple visits to one location by different Census operations and because of the difficulty in distinguishing between GQs within a special place.

A combination of computer and clerical matching conservatively removed duplicate records from the

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

file. A computer matching algorithm identified special places which appeared to be duplicates based on address, name and/or phone number. A clerical operation then evaluated the sets of flagged potential duplicates to judge whether or not the set of special places identified really did appear to be duplicates. In some of these cases, the duplicate special places had overlapping GQ records and in other cases the special place appeared to be a duplicate but simply had its GQs assigned to two different special places. The clerical operation linked these identified duplicates and combined the distinct GQ records under one special place. The number of identified duplicates is relatively small affecting about 0.2% of the GQs in the sampling frame.

2.2 Refinement of GQ counts

A second operation refined the Census 2000 population count based on the observed population during the 2001 ACS GQ operations. These updates, however, are available only for GQs which were in sample in 2001 in the 36 test counties. While this refinement is limited in scope for the first year of full implementation, the sampling after the first year will have the ability to use the observed population from GQs in sample from previous years in all counties. The specifics of how the observed counts will be used in conjuction with the Census 2000 counts has not been finalized. Averaging the observed population over time may provide a more stable estimate to be used for the sampling measure of size. This is being studied further.

The SPGQ file also includes all Census 2000 Count Question Resolution motions filed before the 2003 deadline which resulted in modifications to the file. These are appeals by state, local, or tribal governmental entities that challenge counts based either on boundaries, geocoding or coverage errors. Examples of these types of corrections include GQs which were assigned to the incorrect block or GQs identified in the Census 2000 operation which were erroneously included or excluded in an area.

2.3 New data sources

The ACS will not have any GQ listing operations independent of the the 2010 Census activities. Thus, most updates to the sampling frame will come from administrative sources.

Federal prisons A major new source of data came from the Bureau of Prisons. The Bureau of Prisons is responsible for all of federal prisons and detention centers and keeps accurate, up-to-date counts of prisoners in each of its facilities. The Census Bureau worked with the Bureau of Prisons to obtain these counts for updating our sampling frame. These counts also allowed us to reflect any changes to the federal prison inventory from new prisons or prisons which have closed in addition to having the updated counts for existing prisons..

State prisons Some updates to the state prison inventory have also been made using publically available state government websites. These updates were used to add, delete, and to update existing population counts for state prisons in all 50 states in a manner similiar to the federal prisons.

GQs closed on Census Day Census 2000 removed all records which were scheduled to be closed on Census day, April 1, from the SPGQ file. The records were kept, however, on a preliminary file. Those records that were designated closed on Census day were added back to the master inventory file.

Master Address File One last source of data, albeit a small one, are updates to GQ records that come from the Master Address File, the Census Bureau's corporate address list for Census operations. The master address file may have addresses currently classified as housing units be reclassified as GQs or vice versa. These updates would add or subtract addresses from the GQ inventory respectively. These updates typically come from various field operations including some demographic survey listing operations.

3 Sampling

The sampling frame for GQs has three sampling strata (operational details in Asiala, 2004). The first stratum contains small GQs with a Census 2000 or updated population count of 15 persons or fewer (but had a Census 2000 population which was not zero). The second stratum contains large GQs with a Census 2000 or updated population count of 16 persons or more. The third stratum contains those GQs which were closed on Census Day (April 1) in 2000. Examples of GQs in this stratum include seasonal GQs which are open only in the summer months, etc.

The ACS will exclude certain GQ types from its sampling because of the operational difficulties in sampling, collecting, and estimating for these groups. These GQ types include:

1. 703—Shelters for abused women.
2. 704—Soup kitchens.
3. 705—Mobile food vans.
4. 706—Non-sheltered outdoor locations.
5. 900—Crews of vessels.
6. 910—Transient locations (campgrounds, fairs, carnivals, marinas, etc.).
7. 913—Dangerous encampments.

The

For sampling purposes, the first and third strata are sampled using the same methodology and the second stratum uses a separate methodology. Both are described below.

3.1 Small GQs

The Census Bureau has made a commitment that no mailing address in the housing unit frame will be in sample for ACS more than once during a rolling five-year period. The purpose of this commitment is to reduce concerns of respondent burden because the ACS is an annual survey. In the spirit of that commitment for housing units, a similar sampling design was created for small GQs since some GQs may be similar in size to large housing units.

The sampling for small GQs has two stages of selection: a sample of GQs is taken in the first stage and a sub-sample of the first-stage sample is taken in the second stage. sample and a sub-sample. The first-stage sample is a uniform 1-in-5 systematic sample of GQs. For the initial year, each small GQ is assigned to one of five samples: the current year first-stage sample or one of four "backsamples" assigned to one of the preceding four years. These backsamples are never used for the final sample but are used for the 5-year unduplication. Subsequent years will treat the oldest backsample as the new first-stage sample. For example: the 2001 backsample becomes the 2006 first-stage sample, the 2002 backsample becomes the 2007 first-stage sample, and so on. In this manner, any small GQ can only be in the first-stage sample once every five years. A small GQ in the 2006 first-stage sample will not appear in another first-stage sample until 2011. Since the second-stage sample is selected from the first-stage, this guarantees that the small GQ can be interviewed at most once every five years.

The second-stage sample is a 1-in-8 systematic sub-sampling of the first-stage sample. This sub-sampling yields the target sample rate of 2.5%. Small GQs selected in the second stage are sent to the field for data collection. Those not selected will not be sent for data collection but will still remain ineligible for being in the first-stage sample for five years. For small GQs, the ACS interviews all residents of the GQ.

3.2 Large GQs

The second stratum, Large GQs, are sampled using a Probability Proportional to Size (PPS) design. No attempt is made to unduplicate the current year's sample from previous years so large GQs will be eligible for selection each year. Unlike the small GQs, the fundamental sampling units for large GQs are GQ persons. For efficient field operations, the ACS will select clusters of 10 persons at a time. Each cluster of 10 persons is based on the Census 2000 population or an updated measure of size. The sampling is done with replacement so it is possible for a GQ to have more than one cluster selected from it. Since the sampling rate will be 2.5%, the same as the small GQs, any GQ with 400 persons or more will have at least one cluster selected. In addition, if the GQ has 401 or more persons, it can have more than one cluster of 10 persons selected.

Once these hits have been selected, an additional subsampling will occur at the time of interview. When the GQ is contacted, the field representative will collect the current population of the GQ. The field instrument will then automatically calculate a take-every and start-with for the field representative to use in order to obtain exactly 10 persons in sample for each cluster selected for the GQ. This take-every is fed back along with the data to allow the proper calculation of the weights.

3.3 Assigning sample dates

One goal of the ACS for both its housing unit and GQ frames is to capture any seasonality of characteristics of the population. This is done by randomly assigning sample interview dates throughout the year. For GQs, each small GQ or cluster from a large GQ is assigned a random sample date. In the case of a large GQs with more than one cluster designated for sample, the clusters may be in different months. The exception is federal prisons and state and local jails. All federal prisons are collected during the same month by request of the Bureau of Prisons to reduce the burden on prison officials and staff of conducting the ACS interviews. Since the federal prison population is fairly stable throughout the year, concentrating the interviews into a one month period does not compromise the goals of the

ACS to capture the characteristics of the seasonal population.

For state prisons and local jails, all sample cases for the facility are done in the same month but different prisons or jails may be assigned to different months. This method of assignment allows the ACS to reduce the burden on the staff administering the questionaires while attempting to capture seasonality by doing different facilities at different times of the year.

The assignment of interview months to sample hits for all other GQ types is done randomly throughout the year distributing a roughly equal workload by month. Very large GQs may have sample hits occuring in every month or even more than one hit per month. For these GQs, multiple sets of 10 GQ persons need to be selected within a single month. Field procedures will ensure that no person duplication exists across the GQ hits within a single month so all persons designated to be interviewed will be unique. No procedures are in place to unduplicate which persons in a GQ will be in sample across months, however.

3.4 Sort Order

Internal research determined the sort order for the sample selection that met the design goals of the sampling. The sampling is at the national level using a state, GQ type, followed by county sort order. At issue was whether to have GQ type or county be the second item of the sort. Two goals were initially at odds with one another: the county-level design of the ACS versus the desire to produce good estimates by GQ type. The solution ultimately was driven by the decision that the GQ estimates for the first year of implementation would be primarily state or national level estimates. More information concerning the nature of the planned data products is discussed in the next section.

4 Estimation

4.1 Weighting

The weighting has three basic components: the unbiased weights, the non-interview adjustment, and an adjustment for coverage. The unbiased weights must represent both the initial probability of selection and also the sub-sampling done at the time of interview, if any. The unbiased weight is 40 (the inverse of 2.5%) if the observed population of the GQ is the same as the population in the sampling frame. If the GQ is larger than expected, however, the additional sub-sampling at the time of interview will make the weights larger and conversely if the GQ is smaller than expected. The sum of the unbiased weights for the sample persons in the GQ will always equal the observed population.

The details of the non-interview adjustment are still being devised. One method under consideration to perform the non-interview adjustment within the GQ. This, however, has two drawbacks: spreading the non-interview adjustment over a relatively small number of interviews (less than 10) would cause large weights and negatively affect the variances and also cannot accomodate the situation where there are no interviews for an entire GQ. Another method would perform the non-interview adjustment within GQ type and/or geography. This would spread the adjustment over more interviews which is advantageous from a reliability point of view. This would also still allow some control over adjusting within GQ type and geography, both of which are important in order to produce accurate estimates by GQ type for states and the nation as planned. In the 1999 ACS GQ estimation, the non-interview adjustment was done within county and also within method of data collection where possible (Liu, 2000).

The details of the adjustment for coverage also have not been finalized. Two primary methods have been considered: control the household population and GQ population separately or simply control for the total population with the two universes combined. The first method can be problematic because for many counties there may be good population estimates for the household population but since the GQ population is small the GQ population estimates may be rather weak. A more robust estimate of the GQ population may be possible by combining some or all counties in a state but this leads to differences between the total population by county from the ACS as compared to the official intercensal population estimates. In addition, this creates additional incosistencies between the estimation and estimation areas for housing unit and group quarters. On the other hand, if we control for total population by combining the GQ and household populations, then we no longer have a controlled universe for any household only or GQ population only estimates. The latter option was the method used for the 1999 estimation (Dahl, 2000). This option is more conservative since it only relies on the strength of the total population intercensal estimates and not the GQ population estimates. We continue to study this issue but the latter option may be favored because it is more conservative.

4.2 Data Products

The general design of the ACS is to provide state-, county-, and place-level estimates annually for areas of 65,000 total population or more; county- and place-level three-year average estimates for areas of 20,000 total population or more; and five-year average estimates for all areas down to the tract- and block-group-level.

The framework for the data release of separate GQ estimates is still under development. Initially it was thought that county-level GQ estimates would be a goal for the first year of implementation. This possibility was a driving factor in the experimentation with the sampling sort order used for the systematic sample. Later, however, it became clear that, at least for the first year, separate GQ estimates could be released only at the state or national level.

A brief description of the current data products proposal is given below to aid in the understanding of how the data products requirements influence the sampling and weighting methodologies.

Base Tables Estimates which prior to the incorporation of GQs had been released for household population only will now be released for total population as appropriate at all levels. Some tables which concern household characteristics will remain household population only. For example, age and sex characteristics will be given for the total population. Tables like household income will still be given for the household population only since the defined universe for this table is all households.

ACS Narrative and Tabular Profiles The GQ population will be added to ACS profile lines where the Census 2000 profile tables reported for total population for all levels for which profile tables are produced. The lines giving estimates of total GQ population and GQ population by institutional / non-institutional will not appear in the profile tables the first year.

Public Use Microdata Sample File The Public Use Microdata Sample File for the ACS will contain GQ information once GQ data is collected.

Proportion of total population in GQs To aid in the interpretation of the estimtates, a proportion of total population in GQs will be given for each county. These proportions are planned to be given in ranges rather than precise percentages.

Subpopulation Profiles Currently there is a system in place for producing profiles for subpopulations (e.g., 65 and older, Hispanics, et al.) which have a total population of 65,000 or more. Currently under consideration is to do subpopulation profiles for the GQ population in a similar manner.

Special Tabs Special tabs may be possible for the GQ population subject to disclosure avoidance and data quality of the GQ data. With the first-year data, an evaluation is planned to investigate the quality of the GQ estimates by GQ type at the state and national level. Depending on the outcome of those evaluations, additional data product releases for subsequent years may be considered. In addition, as data product requirements change, the sampling and weighting methodologies will need to change accordingly in order to support the data product requirements.

5 Conclusion

The full implementation plans for including the group quarters population in the ACS sample universe has posed many new challenges which have not needed to be addressed in previous tests. Updates to the frame on a national scale required more coordination with other government agencies than the past. Updates to the frame are increasingly important as the time since its creation for Census 2000 grows.

Most of the sampling process for GQs has been tested in previous tests. One notable improvement is the automation of the subsampling process used by the field staff at the time of interview to sample the cluster of 10 GQ persons. The automation should make it easier for both the field staff and the headquarters staff because it is simpler to use and is more flexible which will decrease the need for questions to be handled by headquarters. This automation will also more reliably report back the sub-sampling take-every for use in the weighting procedures.

The estimation poses some of the greatest challenges. GQ data has only been incorporated with the household data in one of the years, 1999, when the GQ tests have been conducted. In 1999, both the housing unit and GQ sample were only done in the 36 test counties. Thus, many of the estimation challenges which are particular to the national sample have not been developed or tested. The issue of how the non-interview adjustment will be applied is the lesser of the two challenges. The more difficult

challenge is to find the best method to apply the correction for coverage for both the GQ and household population and potential for using different estimation areas for the two universes. These challenges will remain closely tied to the data products requirements of the ACS GQ data.

References

Asiala, M. (2004). "ACS Memorandum ACS-S-39: Specifications for selecting the American Community Survey 2005 Group Quarters Sample". Internal Census Memorandum.

Dahl, S. (2000) "ACS Memorandum ACS-W-2: Specificiations for Weighting the 1999 American Community Survey Housing Unit Sample". Internal Census Memorandum.

Liu, J. (2000) "ACS Memorandum ACS-W-2B: Specifications for Weighting the 1999 American Community Survey Group Quarters Sample (before Person Weighting)". Internal Census Memorandum.

Loudermilk, C. (2004) "ACS Memorandum ACS-UC-19: Creating the Group Quarters Universe for the American Community Survey for Sample Year 2005". Internal Census Memorandum.