

American Community Survey Group Quarters Small Area Estimation

The ACS Group Quarters (GQ) sample was designed to produce state-level estimates of characteristics of the GQ population. However, GQ residents are a component of the total resident population, and characteristics of the total population are published for smaller areas. There is an insufficient GQ sample to produce estimates for these smaller areas. Data users had noticed large year-to-year fluctuations in estimates of counties (Beaghen and Stern, 2009), and a lack of representation of GQ sample across small areas had been documented (Asiala, Beaghen, and Navarro, 2011). In order to address these concerns, the GQ population sample has been supplemented by a large-scale, whole person imputation into not-in-sample GQ facilities starting with the 2011 ACS data release, which includes the 2011 ACS 1-year, 2009-2011 ACS 3-year, and 2007-2011 ACS 5-year estimates. Roughly as many GQ persons are imputed as interviewed. Increasing the size of the GQ sample was not an option due to fiscal constraints.

In Table 1, we see that about 46% (or 20,105 tracts) of the 44,157 tracts with GQ facilities on the ACS sampling frame from 2006-2010 did not have at least one GQ facility in the ACS sample. With this new Group Quarters Small Area Estimation (GQSAE) methodology, all 44,157 tracts with GQ facilities on the ACS GQ sampling frame from this time period have GQ person representation.

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

	Frequency	Proportion
Tracts with GQs	44,157	100%
Tracts with ACS GQ sample	24,052	54%
Tracts without ACS GQ sample	20,105	46%

For perspective, note that in the Census 2000 sample (long form) 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. Further, Census 2000 visited all GQ facilities and thus every facility potentially had persons in the long form sample.

Estimates of substate geographies of the total GQ population and of characteristics of the total resident population will be more reliable than earlier years with the previous methodology. The state-level estimates of characteristics of the GQ population will be relatively unchanged by the new methodology.

Regarding ACS data products, data users should notice little or no difference. In particular, the Census Bureau will continue to produce only state-level estimates of characteristics of the GQ population. We do not plan to release estimates of characteristics for substate areas for this population in the future. There is a break in series between the 2010 and earlier ACS data products and the 2011 ACS data. Namely, the 2010 and earlier ACS data products only have sample GQ person data, whereas the 2011 ACS data products have both the imputed and sample GQ person data for years 2007-2011. This is particularly relevant for substate areas where the GQ population is a significant portion of the total resident population.

The goal of the imputation methodology was two-fold.

1. The primary objective was to establish representation of county by major GQ type group in the tabulations for each combination that exists on the ACS GQ sampling frame. The seven major GQ type groups are defined by the Population Estimates Program and are given in Table 2.
2. A secondary objective was to establish representation of tract by major GQ type group for each combination that exists on the ACS GQ sampling frame.

Table 2: Population Estimates Program Major GQ Type Groups

Major GQ Type	Institutional / Non-Institutional
Correctional Institutions	Institutional
Juvenile Detention Facilities	Institutional
Nursing Homes	Institutional
Other Long-Term Care Facilities	Institutional
College Dormitories	Non-Institutional
Military Facilities	Non-Institutional
Other Non-Institutional Facilities	Non-Institutional

For all not-in-sample GQ facilities with an expected population of at least 16 but less than 400 persons, we imputed a number of GQ persons equal to 2.5% of the expected population of the GQ facility. For those GQ facilities with an expected population of fewer than 16 persons (small facilities), we selected a random sample of GQ facilities as needed to accomplish the two objectives given above. For those selected small GQ facilities, we imputed a number of GQ persons equal to 20% of the facility's expected population.

Interviewed GQ person records were then sampled at random to impute into the selected not-in-sample GQ facilities. An expanding search algorithm searched for donors within the same specific type of GQ facility and the same county. If that failed, the search included all GQ facilities of the same major GQ type group. If that still failed, the search expanded to a specific type within a larger geography, then a major GQ type group within that geography, and so on until suitable donors were found.

Table 3, in the Appendix, gives the number of interviewed and imputed GQ person records by state for the 2011 ACS data releases. Note that there were almost as many imputed person records as interviewed ones.

The weighting procedure made no distinction between sampled and imputed GQ person records (Asiala, Beaghen, and Navarro, 2011). The initial weights of person records in the large GQ facilities equaled the observed or expected population of the GQ facility divided by the number of person records. The initial weights of person records in small GQ facilities equaled the observed or expected population of the GQ facility divided by the number of records, multiplied by the inverse of the fraction represented on the sample frame of the small GQ facilities of that tract by major GQ type group combination. As was done in previous years' weighting, we controlled the final weights to an independent set of GQ population estimates produced by the Population Estimates Program for each state by each of the seven major GQ type groups (U.S. Census Bureau, 2009).

Standard errors for ACS estimates are calculated using a replicate-based methodology that takes into account the sample design and estimation procedures (U.S. Census Bureau, 2009). To account for the additional variance due to the imputations, we developed a method based on the use of design factors (Asiala and Castro, 2012) for application to the GQ person replicate weights.

References

- 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.
- Asiala, M., and Castro, E., (2012). "Developing Replicate Weight-Based Methods to Account for Imputation Variance in a Mass Imputation Application", in preparation for submission in the proceedings of the 2012 Joint Statistical Meetings.
- 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.
- U.S. Census Bureau (2009). "Design and Methodology: American Community Survey". Issued April 2009. http://www.census.gov/acs/www/methodology/methodology_main/

APPENDIX

Table 3: Counts of Interviewed and Imputed Group Quarters Persons by State

	2011 Interviewed Persons	2011 Imputed Persons	2009-2011 Interviewed Persons	2009-2011 Imputed Persons	2007-2011 Interviewed Persons	2007-2011 Imputed Persons
United States (Excluding Puerto Rico)	148,486	150,052	440,150	398,046	728,592	658,414
Alabama	2,111	2,303	6,406	6,350	10,514	10,578
Alaska	825	486	2,122	1,086	3,109	1,761
Arizona	2,388	2,722	7,198	5,766	11,423	9,112
Arkansas	1,444	1,387	4,205	4,056	6,995	6,987
California	14,244	15,587	43,575	39,269	73,504	64,943
Colorado	1,923	2,006	5,793	5,667	9,705	9,482
Connecticut	2,255	2,267	6,638	5,745	11,077	9,404
Delaware	832	318	2,380	788	3,526	1,368
District of Columbia	748	480	2,082	1,168	3,396	1,995
Florida	7,258	8,153	22,200	19,984	35,936	33,192
Georgia	5,071	4,476	13,991	11,355	22,432	18,628
Hawaii	642	746	1,616	1,853	2,663	3,044
Idaho	856	615	2,068	1,691	3,155	3,032
Illinois	5,614	5,226	17,114	14,004	28,382	23,280
Indiana	3,570	3,415	10,564	9,497	17,320	15,696
Iowa	1,947	2,078	5,840	5,867	9,843	9,681
Kansas	1,558	1,628	4,492	4,601	7,385	7,618
Kentucky	2,304	2,243	7,031	6,578	11,518	10,902
Louisiana	2,578	2,085	7,327	6,742	11,973	11,561
Maine	746	908	2,130	2,315	3,398	3,812
Maryland	2,510	2,444	6,932	6,532	11,463	10,971
Massachusetts	3,827	4,704	11,623	11,673	19,669	19,041
Michigan	4,752	4,792	13,786	14,058	22,157	23,240
Minnesota	2,630	2,859	7,721	7,423	12,878	12,089
Mississippi	1,780	1,792	5,338	5,040	8,788	8,371
Missouri	3,206	3,750	9,535	9,511	15,931	15,507
Montana	725	633	2,269	1,568	3,396	2,653
Nebraska	1,091	1,104	3,060	3,168	5,084	5,329
Nevada	1,054	552	2,807	1,620	4,445	2,789
New Hampshire	858	782	2,650	2,057	4,181	3,418
New Jersey	3,793	3,204	11,138	10,082	18,545	17,173
New Mexico	808	777	2,200	2,088	3,576	3,582
New York	9,742	11,149	30,775	26,449	52,616	42,347
North Carolina	4,683	5,163	13,612	14,042	23,006	22,904

APPENDIX

Table 3 Continued: Counts of Interviewed and Imputed Group Quarters Persons by State

	2011 Interviewed Persons	2011 Imputed Persons	2009-2011 Interviewed Persons	2009-2011 Imputed Persons	2007-2011 Interviewed Persons	2007-2011 Imputed Persons
North Dakota	882	447	2,599	1,306	3,909	2,238
Ohio	5,900	6,051	17,330	16,295	28,670	26,836
Oklahoma	2,035	2,162	6,178	5,910	10,352	9,815
Oregon	1,444	1,843	4,260	5,269	6,987	8,773
Pennsylvania	8,082	7,591	22,715	22,118	37,851	37,167
Rhode Island	900	777	2,541	1,928	3,944	3,213
South Carolina	2,574	2,585	7,811	6,798	13,149	11,122
South Dakota	698	638	2,372	1,565	3,790	2,647
Tennessee	2,801	2,924	8,483	7,776	14,208	12,733
Texas	10,469	9,090	30,439	25,236	51,128	41,838
Utah	871	815	2,341	2,519	3,784	4,510
Vermont	805	547	2,401	1,373	3,591	2,261
Virginia	4,084	4,492	12,406	10,629	20,940	17,247
Washington	2,049	2,992	6,428	7,759	11,040	12,573
West Virginia	882	929	2,632	2,563	4,419	4,248
Wisconsin	2,796	3,082	8,759	8,677	14,670	14,606
Wyoming	841	253	2,237	632	3,171	1,097
Puerto Rico	965	1,674	2,227	2,789	3,604	4,227