DATA ISSUE ALERT

American Community Survey Data for Communities with Group Quarter Facilities

---

Overview. Some people live in "group quarter" facilities such as prisons, jails, nursing homes, college dormitories, group homes, etc. Because of shortcomings in the Census Bureau's current methodologies, **ACS statistics tend to greatly overstate local group quarter populations in some years and to understate them in others**.

Therefore, until methodological improvements are implemented:

(1) The ACS should not be used as a source of information on group quarter residents below the statewide level.

The decennial census is the best source of local data on the number and characteristics of group quarter residents in census years. For inter-censal and post-censal years, the Census Bureau's population estimates program provides more reliable figures on total group quarter population than the ACS. However, ACS statistics on group quarter residents are suitable for use at the statewide level.

(2) Because most ACS statistics are based on the entire population of an area (including residents of group quarters), data users should check for anomalies in the group quarter figures whenever unexpected changes are observed in local ACS statistics.

Anomalies in ACS group quarter estimates can cause other ACS figures to deviate from a community's actual characteristics and fluctuate in unexpected ways from one year to another. Such problems should be expected to occur in some years for small communities with ANY group quarter residents and for larger communities with large or varied group quarter populations.

Figures on group quarter residents can be found in ACS tables B09016 and C09016. Until methodological improvements are implemented, the primary value of these figures will be to assess the likely adverse impact of group quarter anomalies upon reported characteristics for an area's total population.

Affected Data Items. In addition to affecting data on the number of group quarter residents, this issue affects data on most data items for which group quarter residents can differ from the general population such as age distribution, race, sex, marital status, educational attainment, school enrollment, nativity, year of entry, residence one year ago, citizenship, language, employment variables (labor force status, commuting, industry, occupation), veteran status, and amount and source of individual income.

Data on disability, health insurance status, and poverty are less affected by group quarter anomalies than the data items listed above. Because tabulations for disability and health

insurance are based on the non-institutional population, they are not affected by estimates for institutions such as correctional facilities, nursing homes, and hospitals. However, they are affected by estimates for non-institutional group quarters such as dormitories, military quarters, group homes, and shelters. Poverty statistics exclude residents of college dormitories and military quarters as well as persons in institutions.

Data on housing units, housing characteristics, households, household income, and other household characteristics are not affected by the shortcomings of group quarter data.

Nature of the Problem. The Census Bureau's approach to sampling group quarter facilities[1] should produce valid results at the statewide level, but it can produce unexpected and misleading results at lower levels of geography. The ACS group quarter figures are based on a sample of small facilities and a sample of individuals in large facilities. The survey results are weighted *at the statewide level* to reflect the state's estimated total number of residents in each type of facility. In years where a particular group quarter facility is not represented in the sample, its population is treated in the ACS as zero. In years where it is included in the sample, on the other hand, its number of residents can be greatly overstated because they must represent residents elsewhere whose facilities were not included that year.

The following examples illustrate the nature and implications of this issue.

- **Example 1**: Hypothetical village with 240 residents in households and 15 residents in a home for the aged.

  Only one out of every 40 facilities with an expected size of 15 or fewer residents is surveyed each year. Thus, 39 out of 40 small facilities are represented by zeros in any particular year, and residents of the 40th facility are weighted as if the facility had about 40 times as many residents as it actually does.

  Thus, the home for the aged in this village would be absent from all ACS tabulations in most years. However, it would be weighted as if it had approximately 600 residents in the year it was included in the sample.

  Figures for a village of this size are only published in the "five-year" ACS products, which combine data collected over the course of five consecutive years. (For example data for 2005-2009 are released in 2010, and data for 2006-2010 are released in 2011.) Under the current methodology, this community's home for the aged would be omitted from ACS five-year data releases for 35 years out of every 40. In the remaining years, the ACS tabulations would reflect four years in which the facility was treated as having no residents and one year in which it was treated as having 600 residents. Thus, the facility would be treated as representing about one-third of the village population in these

[1] See U.S. Census Bureau, *Design and Methodology*: *American Community Survey*, Publication ACS-DM1, issued April 2009, pp. 4-9 to 4-12.
http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf

years.[2] This would seriously skew the community's data with respect to age distribution, education, individual income, labor force participation, and most other "person characteristics" covered by the ACS. The reported characteristics of the community would change abruptly when the facility entered the sample and then change again five years later when it was no longer reflected in the five-year data product.

- **Example 2:** Hypothetical city with 20,000 residents in households and 200 residents in a prison.

  Facilities with an expected size of 16 to 399 residents are represented by 10 interviewees in some years but by no interviewees in other years. In years when such a facility is represented in the sample, its ten interviewees are weighted to represent approximately 400 group quarter residents.

  Thus, a prison with 200 inmates would be treated as having 400 inmates in about half of the years and as having no inmates at all in the remaining years. Data for communities with 20,000 residents are published in both the "three year" and "five year" data products. The prison would usually be included in one or two years out of every three, but it could sometimes be included for three years in a row or omitted for three years in a row. Thus, the weight assigned to prisoners in the three-year data products would gradually and irregularly fluctuate between 0% of the city's population and about 1.9% of the city's population.[3] Although such fluctuations are undesirable, they would probably not be noticed by most users of ACS tabulations for the community's total population. The fluctuations in ACS estimates of the local group quarter population, on the other hand, would be very large and almost totally spurious.

- **Example 3**: Hypothetical city with 65,000 residents in households, 10 residents in a group home, 100 residents in a nursing home, 500 residents in a dormitory, and 1000 residents in a prison.

  As explained in example 1, facilities with 15 or fewer residents are omitted from the sample in most years but weighted to about 40 times their actual size in other years. Thus, the group home would be omitted in 39 years out of 40 but weighted as if it had about 400 residents in the remaining year. The nursing home would be omitted in 3 years out of 4, but it would also be weighted as having about 400 residents in the remaining year. The dormitory would be included in the sample every year, and it would be represented by 10 residents in some years and by 20 in others. Thus, its residents would be weighted to represent either 400 or 800 people. The prison would be represented each

[2] Approximate number of GQ residents in the 5 year product = (15*40 + 0 + 0 + 0 + 0) / 5 = 120 where 15 = the actual number of GQ residents, 40 = the typical weighting factor for the year in which the facility is in the sample, 0 = the ACS estimate of GQ residents in years where the facility is omitted from the sample.

Proportion of estimated ACS population in group quarters = 120 / (240 + 120) = 0.333 where 120 is the approximate number of GQ residents in the 5 year product and 240 is the expected ACS estimate of the household population under ideal conditions.

[3] (400 + 400 +400) / (20,000 + 400 + 20,000 + 400 + 20,000 + 400) = .0196 = 2%

year by either 20 or 30 residents, and they would be weighted to represent either 800 or 1200 people.

Thus, the ACS group quarter figure for this community could fluctuate between a minimum of about 1200 and a maximum of about 2800 people (1.8% and 4.1% of total population) instead of the actual 1610 (2.4%). The law of averages should keep the figure toward the middle of this range in most years, and a "systematic" sampling method should also prevent too many nearby facilities of the same type (or too many cell-blocks from the same prison) from all being over-sampled or all being under-sampled in the same year.

.

The estimated *characteristics* of this community's group quarter residents would also fluctuate from one year to another, since one type of facility would be over-represented in one year and another type might be over-represented the next year. Although characteristics of group quarter residents are not explicitly reported in ACS products, they do affect the reported characteristics for the total population.

This example illustrates that data on characteristics of the total population will be imperfect yet not unreasonable in most years for large communities with several large group quarter facilities. However, it should be expected that some communities will have noticeable anomalies in some years. In addition, ACS figures on the number of group quarter residents in individual communities can be expected to show large fluctuations from one year to another that are largely spurious.

- **Example 4**: Actual fluctuations in ACS group quarter estimates for Michigan communities from 2006 through 2008.

  The table on the following page shows the ACS estimates of group quarter population for all 21 places in Michigan with 65,000 or more residents.

  As summarized below, a majority of these places had fluctuations of over 100% from one year to another, and a majority had absolute differences of over 1000 group quarter residents.

Percent Difference between Highest and Lowest GQ Estimate		Numeric Difference between Highest and Lowest GQ Estimate	
10% to 25%	2 cities	300 - 500	4 cities
26% to 100%	6 cities	501 – 1000	6 cities
101% to 400%	7 cities	1001 – 2500	6 cities
401%+	6 cities	2501 – 5000	5 cities

These very large differences primarily reflect variations in sample composition rather than changes in the actual number of group quarter residents in these communities.

ACS Estimates of Group Quarter Population for Michigan Communities with 65,000 or more Residents: 2006-2008

Area Name	2006	2007	2008	Range (difference between highest and lowest observed figures)	
Ann Arbor city	12,143	11,865	16,558	4,693	40%
Canton CDP	0	0	0	0	n.a.**
Clinton CDP	1,013	417	2,230	1,813	435%
Dearborn city	666	0	427	666	infinite
Detroit city	16,416	14,037	13,155	3,261	25%
Farmington Hills city	1,562	2,345	763	1,582	207%
Flint city	2,886	3,765	3,230	879	30%
Grand Rapids city	9,234	10,397	7,766	2,631	34%
Kalamazoo city	6,903	6,860	7,541	681	10%
Lansing city	371	1,769	0	1,769	infinite
Livonia city	542	2,881	3,111	2,569	474%
Pontiac city	2,398	3,689	1,357	2,332	172%
Rochester Hills city	1,518	652	774	866	133%
Shelby CDP	344	0	0	344	infinite
Southfield city	433	1,230	437	797	184%
Sterling Heights city	1,308	1,077	409	899	220%
Troy city	149	0	373	373	infinite
Warren city	1,226	1,202	890	336	38%
Waterford CDP	2,383	1,892	1,314	1,069	81%
West Bloomfield Twp. CDP	914	548	n.a.*	366	67%
Westland city	3,447	1,335	5,318	3,983	298%
Wyoming city	1,674	642	424	1,250	295%

* West Bloomfield Twp. CDP was not included in the 2008 data release because its estimated population was below 65,000 residents.

** Canton CDP had 86 group quarter residents in the 2000 Census, but it did not have any residents in the GQ sample from 2006 through 2008. In years when one of its facilities is included in the sample, its residents are likely to be weighted to represent several hundred people and its percentage range will be infinite.

Potential Solutions

The Census Bureau is considering the possibility of populating some of the GQ facilities that are not represented in the sample for a particular year with records from sampled facilities.[4] Such a change could be implemented as early as the 2011 release of ACS data for 2010. Although this would be a large improvement relative to the current approach, is has significant shortcomings of its own:

1) Since even one small facility can cause serious problems in the data for small communities, it is necessary to find a solution that produces reasonable data for *every* facility.

2) The characteristics assumed for residents of any particular GQ would change arbitrarily from one year to the next depending on characteristics of the "donor" records used for that year. Since there would be no way for data users to know which year (if any) reflects the "correct" characteristics for the community's facilities, there is no advantage to having a facility's own data in some years rather than using relatively stable statewide characteristics for all years.[5]

3) Because of the small number of survey records involved, the characteristics assumed for local group quarter residents would not be publishable. Thus, it would not be possible to tabulate characteristics of a community's household population by subtracting data on group quarter residents from data on the total population.

The following additional ideas should therefore be considered:

1) The best available estimate of total residents should be used for each facility.

    a) The Census Bureau's Population Estimates Branch has annual information on some facilities through the Federal-State Cooperative for Population Estimates (FSCPE). This information is based on administrative records and it generally indicates the number of residents in each facility on approximately July 1. Figures are also available for certain groups of related GQ's within units of government, such as the total number of dormitory residents for individual colleges in each community.

    b) Counts of residents are obtained one or more times during the year by ACS staff for some small group quarters and for all group quarters that are expected to have 400 or more

[4] Alfredo Navarro, U.S. Census Bureau Decennial Statistical Studies Division, "American Community Survey: Improving Estimates of the Group Quarters Population for Small Geographies," paper for presentation at the April 22, 2010 Meeting of the Census Scientific Advisory Committee.

[5] Over a 40 year period, a GQ with 15 or fewer residents would be represented by its own characteristics in one year and by characteristics of 39 different clusters of residents from other facilities in the remaining years. Thus, its estimated characteristics could change considerably from one year to the next. A facility with fewer than 400 people would be represented by ten of its own residents in some years and by a similar number of residents from other facilities in the remaining years. A facility with 400 or more people would be represented by a small and varying number of its own residents in all years.

residents.[6] This data is more reliable than data obtained through the FSCPE in some cases, but FSCPE data may be preferable in other cases due to more consistent reference dates.

c) For most other group quarters, the best available data may be from the latest decennial census

d) In future years, the total number of residents should also be collected from any facilities for which the data described above are unavailable or inadequate.

2) ACS counts of group quarter residents should be incorporated in the Census Bureau's population estimates in those cases where they are superior to figures available from other sources. (The confidentiality barriers to using such data in population estimates should be no greater than for comparable data collected by the decennial census.)

3) Characteristics of the group quarter population should be estimated for each facility by applying relevant statewide characteristics to the facility's estimated total number of residents.

a) This would involve computing all relevant ACS tabulations (age distribution, educational attainment, etc.) for each type of facility at the statewide level. (In other words, a statewide "geographic component" for each type of facility would be added to the list of summary levels for which tabulations are prepared for internal use by the Census Bureau.) The possibility of publishing these statewide tables should also be explored.

b) Tabulations for each individual facility could then be computed for internal use by multiplying cells of each statewide group quarter tabulation by the local area's estimated share of statewide residents in that type of facility.[7]

c) Publishable tables for the total population could then be prepared for each geographic area by adding the local tabulations for each relevant facility to the area's tabulation for residents of households.

[6] Each year, ACS staff determine the actual number of residents in 2.5% of the facilities that are expected to have 15 or fewer residents and in all facilities that are expected to have 400 or more residents. The percentage falls between 2.5% and 100% for facilities that are expected to have 16 to 399 residents, depending on the expected size of the facility.

[7] This computation should be made to several decimal places and then rounded to whole numbers. The rounding threshold should be allowed to vary for each cell in order to maintain consistency with the relevant statewide totals and with the estimated size of the facility, e.g. decimal components might round upward for one cell of a table if they exceed .487339, and decimal components for another cell might round upward if they exceed .541873.

It should also be noted that a few population characteristics are specific to certain geographic areas, such as residence one year earlier, place of birth, and place of work. These characteristics pose special challenges for any imputation method, and local tabulations for these items may need to be limited to the household population..

4) One of the most common and most important uses for group quarter data is to subtract it from relevant totals to produce data for the household population. The next redesign for ACS products should include a full range of tabulations for the household population as well as (or instead of) tabulations for the total population. The confidentiality concerns normally posed by such pairs of closely related tables should be alleviated by using statewide distributions rather than local data to estimate the characteristics of group quarter residents..

Please send comments or suggestions regarding this data alert to:

Kenneth Darga, State Demographer
Michigan Department of Technology, Management, and Budget
dargak@michigan.gov