RESEARCH REPORT SERIES
(Statistics #2011-05)

Developing Guidelines Based on CVs for when Three-Year Estimates Can Be Used Instead of Five-Year Estimates in the American Community Survey for Areas with Populations of 65,000 or More

Michael Ikeda

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: April 20, 2011

Developing Guidelines Based on CVs for when Three-Year Estimates Can Be Used Instead of Five-Year Estimates in the American Community Survey for Areas with Populations of 65,000 or More

Abstract:
A method is outlined for developing guidelines for acceptable use of three-year estimates rather than five-year estimates for areas with more than 65,000 total population in the American Community Survey (ACS). The method is based on the coefficents of variation of the estimates and is an application of the method outlined in Ikeda (2008). It is applied to estimates from the 2005-2007 ACS and the results are presented.

Key Words:
American Community Survey, Three-Year Estimates, Five-Year Estimates, Guidelines

1. Introduction and Methodology

For areas with more than 65,000, total population the three-year estimates are now available for the American Community Survey (ACS). Ikeda (2008) outlined a method for developing guidelines, based on the coefficient of variation (CV), for when one-year estimates can be used instead of three-year estimates for these areas. This report applies the method to developing guidelines for the use of three-year estimates rather than five-year estimates. One complication is that ACS first included Group Quarters data in 2006. Therefore, while there are three full years of data from the housing unit population in the 2005-2007 ACS, there are only two years of data from the Group Quarters population. This will tend to somewhat inflate the CV for estimates that include persons from Group Quarters. This should not usually have much of an effect on the calculations, since the Group Quarters population will usually be fairly small compared to the housing unit population.

For estimates that are basically estimated counts of people, households, or housing units, we seem to be able to obtain reasonable guidelines by setting cutoffs based on the *percentage that the estimate is of the total*. For three-year estimates, we can use the empirical bivariate distribution of "estimated percent of total" and "estimated CVs" to set cutoffs so that:

(a) most three-year estimates with percentages above the cutoff will meet a desired CV, and
(b) none or very few of these estimates will have CVs that exceed the desired CV by too much.

For a given target CV, if the percentage of total is greater than the cutoff, then the three-year estimate can be used for any CV larger than the target; otherwise the five-year estimate can be used. We divide the areas into several population categories and set separate guidelines for each category.

To assign cutoffs we will divide the "percent of total" distribution into ranges. For each target CV we find the "percent of total" range where the percentage of CVs greater than the target exceeds 10%. We then set the cutoff at the upper limit for that "percent of total" range.

For estimates that are not counts, we will subdivide by type of estimate (aggregate total, median, ratio, gini coefficient) and use the empirical CV distribution for each estimate type. We again expect to divide the areas into separate population categories and set separate guidelines for each category. However, there are some noncount estimates with high CVs even for very large areas, so any general guidelines would need to be used with caution.

2. Results

Table 1 gives cutoffs for three-year count estimates obtained by applying the methodology developed in Ikeda (2008) and outlined above to base table estimates from the 2005-2007 American Community Survey (ACS). The base tables (called the detailed tables in American Factfinder) provide all of the detailed data on basic characteristics for each geographic area. Estimates at the U.S., region, and division summary levels were excluded from the calculation,

as were geographic components (specified breakdowns such as urban, in metropolitan statistical area, etc.) within state. Overviews of ACS data are given in U.S. Census Bureau (2007, 2009). The file used for the research is an internal Census Bureau file that includes both estimates from published tables and estimates from tables filtered out due to variance concerns. A table is filtered out if the median CV for the table is greater than 0.61 (Baumgardner, 2006). The calculations in this paper did not take the results of variance filtering of the three-year tables into account. This would probably tend to err on the side of safety when it comes to published three-year estimates, as the estimates that are filtered out should tend to have high CVs. Note, however, that five-year tables are not filtered for variance. Tables 1a-1f expand the results in Table 1 to give explicit "Yes/No" recommendations for the use of three-year estimates for each population range. Based on feedback from the Decennial Statistical Studies Division of the U.S. Census Bureau, Tables 1a-1f would be the preferable format for illustrating the guidelines.

Cutoffs were also calculated separately for person estimates and household /housing unit estimates for each population category. The cutoffs for person estimates tend to be slightly lower and the cutoffs for household/housing unit estimates tend to be slightly higher than the overall cutoffs, but not so much that separate tables would be required.

Table 1: Cutoffs by Percent of Total by Desired CV and Population Range for Three-Year Count Estimates

	Required Coefficient of Variation (CV)			
Population Range	5%	10%	20%	35%
65,000 to 150,000	25%	9%	2.5%	0.8%
150,001 to 300,000	16%	5%	1.2%	0.4%
300,001 to 1,000,000	7%	1.7%	0.4%	0.2%
1,000,001 to 3,000,000	2.5%	0.6%	0.2%	0.05%
3,000,001 to 15,000,000	0.8%	0.2%	0.05%	0.02%
15,000,001 to 40,000,000	0.2%	0.05%	0.01%	0.005%

Estimates for U.S., Region, and Division were not included in calculations.

Table 1a: Table of Guidelines on Use of Three-Year Estimates for Geographies with 65,000 to 150,000 Population

	Required Coefficient of Variation (CV)			
Estimate as % of Total	5%	10%	20%	35%
less than 0.8%	No	No	No	No
0.8%-2.49%	No	No	No	Yes
2.5%-8.9%	No	No	Yes	Yes
9%-24.9%	No	Yes	Yes	Yes
25% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1b: Table of Guidelines on Use of Three-Year Estimates for Geographies with 150,001 to 300,000 Population

	Required Coefficient of Variation (CV)			
Estimate as % of Total	5%	10%	20%	35%
less than 0.4%	No	No	No	No
0.4%-1.19%	No	No	No	Yes
1.2%-4.9%	No	No	Yes	Yes
5%-15.9%	No	Yes	Yes	Yes
16% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1c: Table of Guidelines on Use of Three-Year Estimates for Geographies with 300,001 to 1,000,000 Population

	Required Coefficient of Variation (CV)			
Estimate as % of Total	5%	10%	20%	35%
less than 0.2%	No	No	No	No
0.2%-0.39%	No	No	No	Yes
0.4%-1.69%	No	No	Yes	Yes
1.7%-6.9%	No	Yes	Yes	Yes
7% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1d: Table of Guidelines on Use of Three-Year Estimates for Geographies with 1,000,001 to 3,000,000 Population

	Required Coefficient of Variation (CV)			
Estimate as % of Total	5%	10%	20%	35%
less than 0.05%	No	No	No	No
0.05%-0.19%	No	No	No	Yes
0.2%-0.59%	No	No	Yes	Yes
0.6%-2.49%	No	Yes	Yes	Yes
2.5% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1e: Table of Guidelines on Use of Three-Year Estimates for Geographies with 3,000,001 to 15,000,000 Population

	Required Coefficient of Variation (CV)			
Estimate as % of Total	5%	10%	20%	35%
less than 0.02%	No	No	No	No
0.02%-0.049%	No	No	No	Yes
0.05%-0.19%	No	No	Yes	Yes
0.2%-0.79%	No	Yes	Yes	Yes
0.8% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

Table 1f: Table of Guidelines on Use of Three-Year Estimates for Geographies with 15,000,001 to 40,000,000 Population

	Required Coefficient of Variation (CV)			
Estimate as % of Total	5%	10%	20%	35%
less than 0.005%	No	No	No	No
0.005%-0.009%	No	No	No	Yes
0.01%-0.049%	No	No	Yes	Yes
0.05%-0.19%	No	Yes	Yes	Yes
0.2% and greater	Yes	Yes	Yes	Yes

Estimates for U.S., Region, and Division were not included in calculations.

To illustrate how these tables can be used, suppose a researcher is interested in characteristics for Lake County, Illinois. Lake County currently has an estimated population in the range 300,001 to 1,000,000 (644,356 in the 2000 Census) so Table 1c would be used. Suppose the researcher is satisfied if the CV is about 20%. Looking at Table 1c, any population estimate expected to be greater than 0.4% of the total population will usually meet the CV cutoff. Thus, for example, if the researcher is interested in an estimate of the foreign born population (14.8% of the total according to the 2000 Census) the three-year ACS estimate can be used, while for estimates that make up a very small percentage of the county population (e.g., persons of Canadian ancestry) the three-year estimates would not be recommended.

Similar reasoning would be used for estimates of households, except that we would use the percentage of the total number of households. Thus, again using Lake County as an example, if the researcher is satisfied if the CV is about 20% then the one-year ACS estimate can usually be used for household estimates that are expected to be more than 0.4% of the total number of households such as families with income in the last 12 months below poverty level (just over 3% of the total number of households according to the 2005-2007 ACS) .

For the noncount estimates we use a somewhat analogous method. We again break the areas into population size categories. However, because "percent of total" is not really relevant for noncount variables, we set a single CV cutoff for each type of estimate within each size category. We consider four CV cutoff points: 5%, 10%, 20%, 35%. For each type of estimate within each population size category we find the smallest cutoff point for which the percentage of CVs greater than the cutoff point is less than 10%. That cutoff point is our CV cutoff for the given type of estimate and population size category. Table 2 lists the CV cutoffs. The three-year estimates are usually acceptable for any target CV above the cutoff. As mentioned above, however, the cutoffs should be used with caution since there are noncount estimates with large CVs even for areas with large populations.

In contrast to Ikeda (2008), the "65,000 to 150,000" and "150,001 to 300,000" ranges are separated out because they have different cutoffs for the three-year estimates. Note that an "Aggregate" estimate is an estimated total of a numeric variable (such as income) over a given subpopulation. A Gini coefficient is a measure of the inequality of a distribution.

Table 2: Three-Year Noncount Estimates--CV cutoff by Type of Estimate and Population Range

	Type of Estimate			
Population Range	Aggregate	Median	Ratio	Gini
65,000 to 150,000	---	---	35%	10%
150,001 to 300,000	35%	35%	35%	5%
300,001 to 1,000,000	20%	20%	20%	5%
1,000,001 to 3,000,000	20%	20%	10%	5%
3,000,001 to 15,000,000	10%	10%	10%	5%
15,000,001 to 40,000,000	5 %	5%	5%	5%

CV cutoff not listed if 50% or greater. Estimates for U.S., Region, and Division were not included in calculations.

Again using Lake County as an example, for medians the three-year estimates are usually acceptable if the desired CV is greater than 20%. Thus, if the researcher is interested in median household income and has a desired CV of 10% the three-year estimate probably should not be used.

The above results provide some rough guidance on when the three-year estimates are likely to be reliable enough to meet a given target CV. We can generally say that when the three-year estimates are expected to be sufficiently reliable, then the five-year estimates can also be expected to be sufficiently reliable. However, we cannot necessarily assume the reliability of the five-year estimates in those situations where the three-year estimates are not expected to be sufficiently reliable.

3. Future Research

While the one-year estimates used in Ikeda (2008) are only available for areas above 65,000 in population, the three-year estimates are available for areas above 20,000 in population. The above analysis calculates cutoffs for the areas above 65,000 in population, using the same population categories used for the one-year estimates. We may now proceed to calculate cutoffs for areas between 20,000 and 64,999 in population, probably dividing this range into two population categories. After that, the obvious next step would be to conduct a similar analysis based on five-year base table estimates. Availability of five-year estimates is not generally limited by population. And even for areas that also have three-year estimates there will usually be more five-year tables published than three-year tables. Any such analysis of the five-year estimates would probably be divided into three stages: the first stage would cover areas above 65,000 in population; the second stage would cover areas between 20,000 and 64,999; and the third stage would cover areas below 20,000 in population.

Acknowledgements

The author would like to thank Michael Beaghen and Lynn Weidman for helpful comments on earlier drafts of this paper.

References

Internal Census Bureau Memorandum from S. Baumgardner to D. Hillmer (2006), "Final Report on "Research on Changes to the Current Base Tables Filtering Rules", U.S. Census Bureau, September 26, 2006.

Ikeda, M. (2008), "Developing Guidelines Based on CVs for when One-Year Estimates Can Be Used instead of Three-Year Estimates in the American Community Survey (ACS) for Areas with Populations of 65,000 or More," Study Series SSS2008/02, Statistical Research Division, U.S. Census Bureau, October 2008, <www.census.gov/srd/papers/pdf/sss2008-02.pdf>

U.S. Census Bureau (2007), "Quick Guide to the 2006 American Community Survey Products in American Factfinder," U.S. Census Bureau, September 2007, <factfinder.census.gov/home/saff/aff_acs2006_quickguide.pdf>

U.S. Census Bureau (2009), "Quick Guide to the American Community Survey Products in American Factfinder, 2008 ACS 1-Year Estimates, 2006-2008 ACS 3-Year Estimates" U.S. Census Bureau, October 2009, <factfinder.census.gov/home/saff/aff_acs2008_quickguide.pdf>