An Evaluation of the Consistency of Subcounty Estimates of Population Based on the American Community Survey and the Population Estimates Program

J. Gregory Robinson, U.S. Census Bureau

Background:

As with other household surveys, the American Community Survey data are controlled so that the population estimates defined by age, sex, race, and Hispanic origin agree with the Census Bureau's official population estimates. Such adjustments are important to correct the survey data for non-sampling and sampling errors. One source of non-sampling error is the potential under-representation of hard-to-enumerate demographic groups. The use of the population controls results in ACS estimates that more closely reflect the level of coverage achieved for those groups in the preceding census.

The use of the population estimates as controls also corrects for the inherent sampling variability in the ACS data that is based on a sample of the population rather than a full count. This variability can be expected to be larger for areas with a smaller ACS sample size. These sample errors can introduce biases in the demographic composition of the survey results in any given year and distort the implied year-to-year change in the estimates. The imposition of controls reduces these compositional and time series sources of error.

In the American Community Survey, the population controls by age, sex, race, and Hispanic origin are applied at the county level. At the subcounty level (e.g., cities and towns), however, the ACS estimates are not directly controlled. The geographic components of a county (e.g., Cincinnati in Hamilton County, Ohio) are forced to agree with the county total, so places basically share the same "adjustment factors" as the county, regardless of their size or demographic composition. One reason the ACS estimates for subcounty areas have not been controlled is that the corresponding population estimates are produced only for the total population, with no detail by age, sex, race, or origin.

Given that the ACS has now matured and is nearing full implementation (full sample size began in 2005, inclusion of group quarters in 2006, 3-year multiyear estimates published in 2008), we are examining methodological enhancements that improve the quality of the ACS estimates. The inclusion of population controls for subcounty areas is one research topic, thus the interest of this study.

Research Objective:

In this paper, we compare the ACS estimates for subcounty places to the population estimates. The differences in the two sets of estimates are systematically analyzed, in terms of percent difference by size of place, the pattern of differences (whether the ACS estimate is higher or lower than the population estimate and the characteristics of the places), and the variability in the implied change based on a time series of the estimates.

This evaluation helps address the following sets of questions:

1. For the place level estimates, what is the magnitude of difference between the two sets? How many places differ by more than 5 percent? Do the places with the largest discrepancies between the estimates tend to be smaller areas, which would be expected if sampling variability is a major factor?

2. Are a roughly equal number of ACS estimates above and below the population estimates, which would also suggest that random sampling variability is contributing to the differences?

3. Do the ACS estimates tend to be lower than the population estimates for some places, suggesting that differential survey undercoverage (for the place relative to the county) is a possible source of the discrepancy? Do the places with these estimate shortfalls share characteristics that would be consistent with this observation, such as relatively large minority population and renter-occupied housing (both of which have lower survey coverage and higher undercount rates at the national level)?

4. Can we detect places where the discrepancies between the two sets of estimates may be attributable to biases in the population estimate rather than sampling or non-sampling errors in the ACS? What criteria do we use to determine this?

5. How consistent is the implied annual change (e.g., 2006 to 2007) based on the ACS place level estimates with the change measured by the population estimates?

Data Sets:

Two sets of published estimates are used for this study—the annual ACS estimates of the resident population (2006-2007) and the annual estimates of the resident population derived from the population estimates program. For the ACS, both the annual numbers for places with a population of 65,000 or more and the "new" multiyear estimates for places of 20,000 to 65,000 population are evaluated against the population estimates. Household population estimates (2000-2005) from the ACS are also used for illustrative purposes. ACS and census data are used to identify the characteristics profiles of places.

Initial Results:

While the full paper will evaluate the consistency of the ACS and population estimates with the above data sets, the 2006 ACS data are currently available. The text and table below presents some preliminary findings, framed in response to the above questions.

1. The difference of the ACS and population estimates are of much greater magnitude for smaller places. As shown in the table below, the two sets of estimates for 2006 differ by more than 5 percent for over 40 percent of cities between 65,000 and 100,000 population; discrepancies of this magnitude are found in about 20 percent of cities over 100,000.

Size Category	Number of Cities	Difference of ACS and Population Estimate >5 Percent	%ACS is lower
Total	455	30.1	55.4
>250,000	59	20.3	71.2
100,000 to 249,999	179	20.7	58.1
65,000 to 99,999	217	40.6	48.8

2. Sampling variability is likely a dominant source of the discrepancies for the smaller cities, as an equal number of ACS estimates are below or above the population estimate. In the 217 places between 65,000 and 100,000 population, the 2006 ACS estimate was lower than the population estimate in 49 percent of the cases and higher for 51 percent. However, for larger cities the ACS estimate is more likely to be below the population estimate. For cities over 250,000, the ACS is lower over 70 percent of the time—this percent is 58 percent for cities from 100,000 to 250,000.

3. For cities of 250,000 or more population, where ACS estimates have been available since 2000, the shortfall of the ACS number relative to the population estimate is often persistent (the time series of the differences will be show for some cities). The paper will systematically provide characteristic data for the cities to infer if we are observing a bias attributable to differentially greater survey undercoverage in the cities that is not fully accounted for in the control to the county estimates. Most of the big cities in this category (e.g., Atlanta, Miami, Cleveland, Phoenix, Cincinnati) do share common characteristics such as (1) large minority populations, and (2) relatively high proportions of renters. For example, the 2006 ACS estimate of total population for Cincinnati is 8.9 percent below the published population estimate. Cincinnati is 48 percent minority and 61 percent of the housing is renter-occupied, while these shares are 15 percent and 25 percent for the balance of Hamilton County.

4. Common attributes are not as easy to identify for cities where the ACS estimates are persistently higher. The paper will explore if some of these may be cases where the population estimates might be systematically low, especially in high growth areas.

5. With the availability of the 2007 ACS estimates, the consistency of the implied change from 2006 to 2007 based on the ACS can be compared to the change measured by the population estimates.

In summary, there are some distinct patterns in the data when the ACS and population estimates are compared. This paper advances our understanding about the reasons for the discrepancies and informs additional research to address the use of population estimates as ACS controls for subcounty areas. One answer does not fit all (e.g., sampling variability and survey undercoverage in the ACS, estimates bias), so the explanations can be several. Even the difference in residence rules can be a factor, especially in areas with a large number of seasonal units.