

Evaluation of Data from the American Community Survey and the Current Population Survey

Stephen Wenck, Sameena Salvucci, Synectics for Management Decisions,
Stephen Wenck, Synectics, 1901 North Moore Street, Suite 900, Arlington, VA 22209

Introduction

The Bureau of the Census is introducing a new survey to collect the detailed demographic data traditionally collected on the decennial census long form. This survey, the American Community Survey (ACS), offers a new approach for collecting accurate, timely information that provides consistent measures for all areas. When fully implemented, the ACS will collect data from 3 million households a year so that detailed characteristics about the entire United States are updated every year, instead of once every ten years. With the ACS, community leaders and other data users will have timely information for planning and evaluating public programs for everyone from newborns to the elderly. The ACS will be able to identify changes in an area's population and give an up-to-date statistical picture when data users need it. Communities will be able to use the data to track the well-being of children, families, and the elderly; determine where to locate new highways, schools, and hospitals; show a large corporation that a town has the workforce the company needs; evaluate programs such as welfare and workforce diversification; and monitor and publicize the results of their programs. Moreover, as an on-going survey, the ACS is a flexible vehicle, capable of adapting to changing customer needs. Once it is fully implemented, the potential is there to add questions of national policy interest or specialized supplements to help identify the situations of special population groups.

The ACS went through a demonstration period from 1996 to 1998. It is now being thoroughly tested. Since 1999, under the Continuous Measurement Program, the ACS questionnaire has been administered in 31 comparison sites. Since 2000, a "Supplemental" ACS questionnaire has also been administered to households nationwide, although to a smaller sample than used by the Continuous Measurement Program. The Census 2000 Supplementary Survey (C2SS) was conducted as part of Census 2000 operations to demonstrate the operational feasibility of collecting long form information at the same time as, but in a separate process, from the decennial census. The 2001 and 2002 Supplemental Surveys are being used to measure the quality and usability of estimates based on averaging data collected over several years.

Full, nationwide implementation of ACS is planned to start in 2003 (pending Congressional funding). The ACS will then collect data from households located in every county, American Indian and Native Alaskan area, and Hawaiian Homeland, as well as in Puerto Rico. Beginning in 2004, ACS data will be available every year for areas and population groups of 65,000 or more. It will take five years to accumulate a large enough sample to provide estimates with accuracy similar to the decennial census for small areas and population groups of 20,000 or less, so updated information for areas such as neighborhoods will be available starting in 2008.

Researchers, of course, have not simply waited for a new survey like ACS, or relied only on Decennial Census data. Many use data from the Current Population Survey (CPS). Government policymakers and legislators use CPS data as important indicators of the United States' economic situation and for planning and in evaluating many government programs. The press, students, academics, and the general public also use the data.

The CPS is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The CPS was originally designed to measure unemployment each month, and it is the primary source of information on the labor force characteristics of the U.S. population. The survey has been conducted for more than 50 years. The sample is scientifically selected to represent the civilian non-institutional population. Respondents are interviewed to obtain information about the employment status of each member of the household 15 years of age and older; however, published data focus on those ages 16 and over. The sample provides estimates for the nation as a whole and serves as part of model-based estimates for individual states and other geographic areas. The estimates obtained from the CPS include employment, unemployment, earnings, hours of work, and other indicators, and they are available by a variety of demographic characteristics, including age, sex, race, marital status, and educational attainment, as well as by occupation, industry, and class of worker. Supplemental questions are often added to the regular CPS questionnaire, producing estimates on such topics as school enrollment, income, previous work experience, health, employee benefits, and work schedules.

Given that the ACS questionnaire and the CPS questionnaire cover many of the same topical areas, and estimates of the same operationalized concept can be produced by both surveys, it is natural for the Census Bureau and researchers all over the country to be interested in seeing how the estimates differ between the surveys, and to see how ACS estimates compare with the Decennial Census data. This paper presents the results of such a comparison.

Approach

To get the best possible estimate across surveys, it was decided to compare the ACS data available from the 31 comparison site combined with the C2SS against the 2000 CPS data.

The C2SS used the questionnaire and methods developed for the ACS to collect demographic, social, economic, and housing data from a national sample of 700,000 households. Group quarters were not included in the sample. (They are not included in the CPS sample either.) The C2SS was conducted in 1,203 counties with monthly samples of about 58,000 housing units. The C2SS was designed to be used in combination with data from the ACS comparison site tests to produce annual estimates. Estimates are available for the United States as a whole, the 50 states, and the District of Columbia, as well as large counties and cities.

The C2SS and ACS data were combined at the state level to produce complete state-level datasets. Additionally, all C2SS and ACS data were combined into one large national dataset. The C2SS/ACS datasets were restricted to the variables of interest in order to reduce processing time.

The CPS data were already combined to the national level. State datasets were created for the sub-national comparisons with C2SS/ACS. As with the C2SS/ACS data, the CPS data were restricted to the variables of interest in an effort to reduce processing time.

The reliability of the C2SS/ACS and CPS data were systematically compared across various socioeconomic characteristics. Additionally, 2000 Decennial Census figures were used for comparison.

The comparisons were performed at the national level. Additional tests were performed on subpopulations such as Hispanic, foreign-born, Arab ancestry, and selected states. Variables with no or negligible difference in wording and intent were used in the comparison.

Rigid statistical testing was not required for the reliability testing; instead simple percent differences were calculated. Since neither C2SS/ACS nor CPS are considered the accepted value, we used the percent difference to determine the similarity of the estimates. This is found by dividing the absolute difference of the two measured values by their average, shown in the formula below.

$$\text{Percent Difference} = \left(\frac{|measure_1 - measure_2|}{\left(\frac{measure_1 + measure_2}{2} \right)} \right) * 100$$

Results

No egregious percent differences were found between the C2SS/ACS, CPS, and 2000 Decennial Census. Please note that only the percent difference numbers are presented. At the time the presentation and paper were prepared, the C2SS/ACS data were still embargoed.

Population-level characteristics

Results are discussed for four population-level socioeconomic characteristics: sex, relationship, race, and Hispanic origin. As was mentioned earlier, the group quarters population is included in neither the C2SS/ACS nor the CPS estimates. However, the group quarters population is included in the 2000 Decennial Census figures. In many instances, it was possible to separate out the group quarters population from the non-group quarters population for the 2000 Decennial Census figures. Where it was possible, the comparisons between C2SS/ACS and Census were made. If it was not possible to remove the group quarters population from the 2000 Decennial Census figures, then the percent differences for C2SS/ACS vs. the 2000 Decennial Census were made, but exceptions due to the group quarters population were noted and set aside.

The table below shows the percent differences for sex and relationship. There is no difference between C2SS/ACS and the 2000 Decennial Census at the grand total level because the C2SS/ACS estimates were controlled to the 2000 Decennial Census numbers. However, C2SS/ACS estimates other than the grand total were not controlled to the 2000 Decennial Census.

There is very little difference between C2SS/ACS and CPS at the grand total level. The differences

increase slightly when looking at breakdowns by sex or relationship, but not enough to cause concern.

Population - Sex and Relationship

Subject	Percent Difference	
	CPS	Census
Total population	0.16%	0.00%
Sex		
Male	0.29%	0.01%
Female	0.04%	0.01%
Relationship		
Householder	2.09%	1.42%
Spouse	1.86%	0.37%
Child	5.69%	1.83%
Own child under 18 years	2.03%	1.29%
Other relatives	13.79%	6.75%
Under 18 years	15.01%	5.70%
Nonrelatives	7.54%	6.31%
Unmarried partner	28.49%	2.43%

The next table shows the percent differences for age groups. Here it was not possible to remove the group quarters population from the 2000 Decennial Census figures. The group quarters population obviously has a large impact on the 85 years and over age group, as would be expected.

Population - Age

Subject	Percent Difference	
	CPS	Census
Total population	0.16%	0.00%
Age*		
Under 5 years	2.98%	0.75%
5 to 9 years	0.98%	1.81%
10 to 14 years	1.85%	0.87%
15 to 19 years	7.60%	8.19%
20 to 24 years	5.16%	7.96%
25 to 34 years	2.47%	2.95%
35 to 44 years	0.48%	1.25%
45 to 54 years	1.64%	1.18%
55 to 59 years	3.13%	1.44%
60 to 64 years	2.04%	0.65%
65 to 74 years	1.69%	1.59%
75 to 84 years	0.36%	5.98%
85 years and over	3.89%	25.99%

*Census 2000 figures include group quarters population

The final population table shows race and Hispanic origin breakdowns. Here again, the group quarters population is included in the 2000 Decennial Census figures. This population has a significant effect on the differences between C2SS/ACS and the 2000 Decennial Census for American Indians and Alaska Natives, Asians and Pacific Islanders, and the Other Hispanic or Latino group. This was not unexpected as these groups tend to have higher rates of group quarters residence.

Some of the larger differences between C2SS/ACS and CPS in the Puerto Rican and Other Hispanic or Latino subgroups can be attributed to the small sample size for these groups in the CPS.

Population - Race and Hispanic Origin

Subject	Percent Difference	
	CPS	Census
Total population	0.16%	0.00%
Race*		
White	4.93%	1.19%
Black or African American	5.56%	3.14%
American Indian and Alaska Native	6.02%	19.94%
Asian and Pacific Islander	5.15%	11.58%
Hispanic Origin*		
Hispanic	4.95%	2.40%
Mexican	0.45%	4.56%
Puerto Rican	15.55%	1.52%
Cuban	5.19%	0.61%
Other Hispanic or Latino	17.70%	20.28%
Not Hispanic or Latino	0.88%	2.86%

*Census 2000 figures include group quarters population

Household-level characteristics

Results are discussed for six household-level socioeconomic characteristics: household type, tenure, and sex, age, race, and Hispanic origin of householders. When dealing with household-level data, the problem of group quarters population goes away. However, there was a different problem: Decennial Census data were not available for all the breakdowns for which we had C2SS/ACS and CPS data.

The first household-level table below shows households by type and tenure status. Here C2SS/ACS and CPS barely differ at all. C2SS/ACS and the 2000 Decennial Census also differ by a negligible percentage.

Household - Type and Tenure

Subject	Percent Difference	
	CPS	Census
Total households	0.04%	0.63%
Households by type		
Family households (families)	1.48%	1.15%
Married-couple family	4.02%	2.53%
Female householder, no husband present	4.18%	2.49%
Nonfamily households	3.52%	0.47%
Householder living alone	3.37%	1.50%
Householder 65 years and over	0.22%	0.39%
Housing tenure		
Owner-occupied housing units	2.73%	1.93%
Renter-occupied housing units	5.48%	1.87%

The following table shows households by sex and age of the householder. We were not able to obtain Decennial Census figures for sex of the householder. The differences for the comparisons that could be made are not out of the range that was expected.

Household - Sex and Age

Subject	Percent Difference	
	CPS	Census
Total households	0.04%	0.63%
Sex of householder		
Male	1.52%	N/A
Female	1.99%	N/A
Age of householder		
15 to 24 years	2.01%	8.10%
25 to 34 years	0.46%	1.37%
35 to 44 years	0.12%	0.10%
45 to 54 years	0.11%	1.59%
55 to 64 years	1.65%	3.05%
65 years and over	0.99%	2.71%
65 to 74 years	0.49%	2.04%
75 to 84 years	2.11%	3.15%
85 years and over	0.44%	4.40%

The final table shows households by race and Hispanic origin of the householder. There are some large differences between C2SS/ACS and CPS for some of the smaller Hispanic origin subgroups and the American Indian/Alaska Native race group, but this can be attributed to the small sample size for these groups in the CPS.

Typically when dealing with the smaller race or Hispanic origin groups in CPS, 3-year averages are employed. While not shown here, this test was performed and the results were much closer to the C2SS/ACS estimates and 2000 Decennial Census figures.

Household - Race and Hispanic Origin

Subject	Percent Difference	
	CPS	Census
Total households	0.04%	0.63%
Race of householder		
White	3.29%	1.34%
Black or African American	6.06%	0.40%
American Indian and Alaska Native	30.90%	40.52%
Asian and Pacific Islander	5.52%	8.74%
Hispanic Origin of householder		
Hispanic	0.03%	1.10%
Mexican	6.16%	N/A
Puerto Rican	7.85%	N/A
Cuban	13.12%	N/A
Other Hispanic or Latino	14.74%	N/A
Not Hispanic or Latino	0.04%	0.79%

Next Steps

The comparison of the C2SS/ACS to the CPS was only the beginning of the project. The real purpose of the project was to convert the CPS Population Paper Listing (PPL) table packages for various sub-populations to use the ACS data. A Population Paper Listing is a standard set of demographic and socio-economic tables the Census Bureau publishes every year for various sub-populations. After the testing of C2SS/ACS vs. CPS was completed, the CPS Hispanic origin, foreign-born and Arab ancestry PPL table packages were converted to work with 2000 C2SS/ACS data.

There are a number of additional steps that are necessary in order to complete the project:

- Running the existing ACS PPL table packages (Hispanic origin, foreign-born, and Arab ancestry) for the 2001 data once it becomes available.
- Creating PPL table packages for other populations such as: Asian, Black, and other racial groups; other ancestry groups; as well as other special populations like the elderly.
- Expanding the PPL table set for each population to include some items not possible previously due to CPS sample size and questionnaire limitations.
- Publishing the static tables on the web as the means of dissemination.
- Developing a web-based table generator so that analysts can see the data in the ways that interest them. The table generator would have automatic cell suppression for small cell sizes.

Implications

The results of the testing have made the Census Bureau comfortable using ACS for purposes previously only served by CPS, but how does this help the Census Bureau?

With the larger sample sizes in ACS, it will be possible to create PPL table packages for subpopulations (for instance many of the ancestries and also Hispanic subgroups) that were not possible with CPS. It will no longer be necessary to use 3-year averages for many of the smaller subpopulations. Much richer data analyses are possible with the larger and more reliable sample of the ACS.

There are many more implications for the Census Bureau, but this is the major one in relation to the project we undertook.