Comparing Social Characteristics Between Census 2000 and the American Community Survey

David A. Raglin, Theresa F. Leslie, and Deborah H. Griffin, United States Bureau of the Census

Abstract: This paper compares the results from Census 2000 to the results from the American Community Survey (ACS) in 2000, which was called the Census 2000 Supplemental Survey (C2SS), for various long-form social characteristics such as nativity, place of birth, language spoken at home, and ancestry. The comparisons not only show which differences are statistically different, but also where the differences are meaningfully large enough to change conclusions made from the data. The paper identifies differences that are systematic because of methodological reasons.

The ACS is a new survey being tested by the Census Bureau. It is designed to be a replacement for the Decennial Census long form. The long form asks questions on education, employment, income, ancestry, housing value, rent, and many other topics. The C2SS was designed in part to test how feasible it was to collect long form data at the same time as a Census was being conducted and it was the first large-scale national collection of data using the ACS process.

1. Introduction

The Census Bureau has proposed that the decennial census long form, sent to a sample of about one in every six households, be replaced by the American Community Survey (ACS). The ACS is designed to continuously collect information and release data for geographic areas with 65,000 or more people every year. One obvious question is how do data collected from the ACS compare to data collected from the decennial census long form?

The ACS has been under development since 1996, and in 2000, for the first time, data were collected for a nationally-representative sample using ACS methodology in a survey called the Census 2000 Supplementary Survey (C2SS). Collecting data nationally during 2000, a decennial census year, provides the ability to make comparisons between the census sample and the data collected in the C2SS (which we will refer to as "ACS data" for the rest of this paper). The Census Bureau is doing just that in a series of four reports, roughly analogous to the four basic data profile tables that the Census Bureau produces for data users: general/demographic, social, economic, and housing. Topics included in each table include:

- General/Demographic: sex, age, race, Hispanic origin, relationship, household type, housing unit occupancy/vacancy and tenure.
- Social: education, marital status, fertility, grandparents as caregivers, veterans status, disability, migration, citizenship, language, and ancestry.
- Economic: employment, commuting, industry and occupation, class of worker, income, and poverty.
- Housing: units in structure, year built, rooms, year moved in, vehicles available, utilities, value, mortgage, and rent.

This paper presents the initial work for the report comparing social characteristics between the ACS and Census 2000 long form sample data. Variables examined here include place of birth, citizenship, year of entry, language spoken at home, and ancestry. Other papers comparing social characteristics were presented at the 2003 Joint Statistical Meetings – Stern (2003) on disability, Boggess (2003) on education, and Dye (2003) on grandparents as caregivers.

The comparisons in this paper will help answer the following questions:

- Can we identify systematic differences in the distributions for selected social characteristic variables between the census long form and the ACS?
- What are possible causes of those differences that should be investigated?
- How might these differences affect data users during the transition from the decennial census sample distributions to the ACS distributions.

The word "distributions" is important. This paper compares the distributions of key items, not the point estimates themselves, because the ACS will not produce the official counts of the population or of housing, but will produce annual estimates of distributions of detailed social, economic, and housing characteristics of the nation. A myriad of data products will be produced by these characteristics. For example, the ACS will provide estimates of the percentage of children in poverty, data on levels of education by race, and information on the economic characteristics of the aging population. The ACS will collect these data over the decade allowing the tracking of change in these and other important demographic and socioeconomic distributions required for informed governance. Therefore, consistency in the collection of these data are fundamental to consistency in later data products.

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

2. Methodology

This section describes the method used to compare the ACS and Census 2000 results for place of birth, citizenship, year of entry, language spoken at home, and ancestry. The general approach was to determine if meaningful differences exist; and if so, to suggest reasons for these differences. This analysis uses the concept of "meaningful differences", which is discussed in great detail in Section 2.1.4 of this paper.

Tables which summarize percentage distributions of characteristics of the household population from Census 2000 and the ACS. The ACS estimates that differed beyond sampling error were identified and categorized as meaningful or not meaningful. These differences are of primary interest because they reflect the differences that are seen in the final published results.

2.1 Creation of Comparison Tables

Before conducting the comparisons, two factors had to be taken into account. The first factor was that unlike Census 2000, the ACS did not include the group quarters population in 2000. To allow legitimate comparisons to be made, the data for the group quarters population were removed from the Census 2000 files, providing tables that included only the household population. The concept of group quarters is covered in Section 2.1.2 of this paper.

The second factor was that since both the ACS and the census long form sample were subject to sampling error, the comparisons had to take into account that error. The weights used for both the ACS and the Census long form sample data were the final weights after all population control adjustments. The ACS sampling errors were calculated using standard ACS variance estimation techniques. The census sampling errors were derived by calculating the simple random sample variance and then applying a design factor. The design factors used were from the 1990 Census because the factors for Census 2000 were not available at the time these calculations were made, and the impact of using the Census 2000 factors is not known. That is a limitation of this study.

Tests for statistical significance were conducted and the results shown in the tables. The combination of the magnitude of the differences and the statistical significance were studied to determine differences that were meaningful from a practical–as compared to a statistical– viewpoint.

2.1.1 Geographic level of comparison

The tables in this paper compare the distributions at the national level. This is a limitation given that the ACS, as the name implies, will produce distributions at much more detailed levels of geography. Tables at the national level may mask systematic methodological differences that affect subgroups of the population.

Therefore, future research will include comparisons similar to those in this paper for the 30 ACS test sites. These sites are counties or groups of counties which have been part of the ACS since 1999. The sites are both geographically and demographically diverse. They include urban counties such as San Francisco, CA, and Bronx, NY, suburban counties like Lake, IL (near Chicago) and Broward, FL (near Miami), and rural counties like Oneida and Vilas, WI, and Starr and Zapata, TX. The test sites were also oversampled compared to other counties in the ACS. This allows for analysis of key subgroups.

2.1.2 Group quarters population

The Census Bureau classifies all people not living in housing units as living in group quarters. A housing unit is defined as a house, apartment, a mobile home or trailer, a group of rooms or a single room occupied as a separate living quarters or, if vacant, intended for occupancy as a separate living quarters. There are two types of group quarters: institutional (for example, correctional facilities, nursing homes, and mental hospitals) and non-institutional (for example, college dormitories, military barracks, group homes, and shelters). The Census Bureau did not want to burden group quarters with duplicate data collection during 2000 by having the ACS include group quarters in the sample. Group quarters were collected in the 30 test sites in 2001 and will be part of the ACS in the future.

2.1.3 Tables

The published profile tables were chosen as the key tables for analysis. This section describes the contents of those tables, how they were produced, and how they should be interpreted. An example table is shown at the top of the next page.

The first row of the table shows the estimate of the number of people in the population of interest. For example, we collected nativity and place of birth for all persons, so the population of interest is the household population. To contrast, we collect language spoken at home only for those people five years of age and older, so people five or more years old living in households made up the population of interest.

The rest of the lines show the percentage of the population of interest that fell into the appropriate category or subcategory. For example, categories under nativity and place of birth include "native" and foreign-born", and a subcategory under foreign-born is "naturalized citizen". The categories and subcategories shown are based on those in the Social Characteristics profile table produced for both the census and the ACS. The percentages are shown to one decimal place–the level of accuracy used in profile tables.

Those lines also show the difference between the ACS and the census estimates, with an asterisk denoting differences that are statistically significantly different between the ACS and the census estimates at the 90 percent confidence level, the level used by the Census Bureau. Adjustments were made when appropriate for multiple comparisons using the Bonferonni method.

Example Table: Variable, Comparison of Distributions, In Percent

Category	Census	ACS	ACS - Census
Population of interest	xxx.x million	xxx.x million	---------
Category 1	xx.x	xx.x	xx.x*
Subcategory 1a	xx.x	xx.x	xx.x
Subcategory 1b	xx.x	xx.x	xx.x*
Category 2	xx.x	xx.x	xx.x*
Subcategory 2a	xx.x	xx.x	xx.x*
Subcategory 2b	xx.x	xx.x	xx.x

* Difference is statistically significant at the $\alpha = 0.10$ level (90 percent confidence level).
Note: The difference column may not be the same as the ACS column minus the Census column due to rounding.

2.1.4 Meaningful differences

Due to the large sample sizes for the ACS and the census long form sample, most differences in this report–no matter how small–were statistically significant. Therefore, just because a difference is statistically significant does not necessarily mean that there is a difference in a practical or meaningful sense.

To determine whether differences are meaningful, it is important to look at the magnitude of the difference and to examine findings reached from Census 2000 to assess if the same conclusions would have been made based on the ACS results. If the differences are so small that the differences would not cause data users to change their conclusions, then there would be no practical or meaningful differences between the ACS and Census 2000 results, even if the differences may be statistically significant. It is therefore important that the statistical significance and the actual difference be considered in combination.

An example of this can be found by looking at Table 1a in section 3.1. Census 2000 estimated that 4.5 percent of the household population were naturalized citizens, compared to the ACS estimate, which was also 4.5 percent. Based on those figures, the difference between the ACS and the Census is shown as -0.1 percent. However, to two significant digits, the difference is -0.053 and the standard error of the difference is 0.025. This produces a p-value of 0.035, which is less than the 0.10 needed for statistical significance at the 90 percent confidence level. Therefore, while the difference seems small, it is statistically significantly different because the large sample sizes produce a very small sampling error.

However, if a difference of -0.053 percent does not change a data user's conclusions, then there is no meaningful difference between the Census and ACS results. We know we are making our own judgement about meaningful differences in this paper, but we are also providing the differences so that the readers can make their own judgements.

2.2 Survey and Census Methods

A systematic review was undertaken to inventory ACS and Census 2000 methods and assess their execution. The fundamental differences in purpose between the ACS and Census 2000 led to critical differences in the choice of methods including questionnaire design, enumeration and interview procedures, and data processing techniques.

Some methods reflected a conscious decision to measure a concept in a different manner. For example, the Census 2000 residence rules, which determine where people should be enumerated, are based on the principle of usual residence and are centered on the reference date of April 1st. The ACS residence rules count people who are living or staying at a residence for more than two months, or who do not have another usual place to stay. The reference date is the date the data are collected. These rules are designed to collect representative data throughout the year.

A survey's design and implementation methodology result in nonsampling error that may affect the survey's results. This paper identifies possible methodological reasons for meaningful differences that are found as well as areas for future research.

Coverage error is minimized in both the ACS and the census long form sample by adjusting weights so that estimates for key demographic variables match those for known totals from the whole census. Ideally, it would be helpful to compare unadjusted data to isolate coverage issues. This is possible for ACS, which conducted the unit nonresponse adjustment and the adjustment to totals in separate steps; however, for the census long form, the adjustment was one step, so the adjustment to the known totals cannot be separated out.

3. Results

Using the methodology described above, tables presented in this section compare the distributions based on published ACS data for 2000 to distributions based on the Census 2000 household population.

3.1 Nativity, Place of Birth, and Citizenship

Tables 1a and 1b show the comparisons between the Census and ACS for nativity, as well as place of birth and citizenship for foreign-born residents. There are several statistically significant differences, shown by the asterisks in the last column, but we do not see any of those

Table 1a: Nativity and Place of Birth, Comparison of Distributions, In Percent

Category	Census	ACS	ACS -Census
Total population	273.6 million	273.6 million	--------
Native	88.8	88.9	0.2*
Born in the United States	87.5	87.7	0.2*
State of residence	60.1	59.8	-0.3*
Different state	27.5	27.9	0.4*
Born outside United States	1.2	1.2	-0.0
Foreign born	11.2	11.1	-0.2*
Entered since 1990	4.8	4.8	0.0
Naturalized citizen	4.5	4.5	-0.1*
Not a citizen	6.7	6.6	-0.1*

* Difference is statistically significant at the $\alpha = 0.10$ level (90 percent confidence level).
Note: The difference column may not appear to be the same as the ACS column minus the Census column due to rounding.

Table 1b: Region of Birth of Foreign Born, Comparison of Distributions, In Percent

Category	Census	ACS	ACS -Census
Total (excluding born at sea)	30.7 million	30.2 million	--------
Europe	15.7	15.7	-0.0
Asia	26.4	27.3	0.9*
Africa	2.8	2.8	0.0
Oceania	0.5	0.6	0.0
Latin America	51.8	50.8	-1.0*
Northern America	2.6	2.8	0.2

* Difference is statistically significant at the $\alpha = 0.10$ level (90 percent confidence level).
Note: The difference column may not appear to be the same as the ACS column minus the Census column due to rounding.

Table 2: Language Spoken at Home, Comparison of Distributions, In Percent

Category	Census	ACS	ACS -Census
Population 5 years and over	254.6 million	254.6 million	--------
English only	82.0	82.5	0.5
Language other than English	18.0	17.5	-0.5*
Speak English less than “very well”	8.2	7.6	-0.7*
Spanish	10.8	10.5	-0.3*
Speak English less than “very well”	5.3	4.9	-0.5*
Other Indo-European languages	3.8	3.7	-0.1*
Speak English less than “very well”	1.3	1.2	-0.1*
Asian and Pacific Island languages	2.7	2.7	-0.0
Speak English less than “very well”	1.4	1.3	-0.1*

* Difference is statistically significant at the $\alpha = 0.10$ level (90 percent confidence level).
Note: The difference column may not appear to be the same as the ACS column minus the Census column due to rounding.

differences to be meaningful. The ACS has a higher proportion of native-born residents than did the Census (87.7 percent versus 87.5, respectively), especially people who were born in a different state (27.9 percent versus 27.5). Among the foreign-born population, the ACS had a higher proportion of people born in Asia (27.3 percent versus 26.4) and a lower proportion of people born in Latin America (50.8 percent versus 51.8).

However, while the differences are not meaningful at the national level, the differences in the percentage of people born in Asia and Latin America deserve study at the site level to identify any systematic differences by area.

3.2 Language Spoken at Home

Table 2 shows the comparisons between the Census and the ACS for language spoken at home, breaking it out for people who did not speak English "very well". There is a trend for the ACS to have slightly lower reporting of languages spoken other than English, especially Spanish-speakers who did not speak English very well. One possible reason for that might be that Census 2000 had paper forms in five languages other than English (Spanish, Japanese, Chinese, Korean, and Tagalog) and language guides in over 40 languages, while the ACS paper instrument was available only in English and the followup instruments in only English and Spanish.

The differences do not seem to be meaningfully large on an individual basis, but the trend of ACS reporting a smaller percentage of people speaking a non-English language does bear more investigation, including looking at the site-level data and the data by mode of collection.

3.3 Ancestry

Unlike the first two categories discussed, there are both statistically significant and meaningful differences in the reporting of ancestry, as Table 3 shows. Census 2000 had consistently less reporting of almost all ancestries shown than the ACS. Ancestry is a unique variable in that if a person does not report any ancestries, then no ancestries are tabulated for that person. For most variables, if they are not reported, the item imputation process imputes a value. That does not happen for ancestry. If no ancestry is reported, then no ancestry is tabulated.

An ancestry was reported for a higher proportion of people in ACS than in the Census. In the Census, 81.0 percent of long forms had at least one ancestry reported, while in ACS, 88.3 percent had a reported ancestry.

Because of this differential reporting of ancestry, Table 3 differs in format somewhat from the previous tables. The first set of columns, labeled "All People" are the percentages based on the published data, and they show the percentage of all people who reported the given ancestry. The second set of columns use as their base the number of people who reported at least one ancestry. The percentages show the percentage of people that reported at least one ancestry that reported the given ancestry. This allows us to determine if the lower percentages of people reporting the specific ancestries in the Census is largely due to the fact that fewer people reported ancestry in general in the Census.

The table includes ancestries in the profile table that make up one percent of the population or more. Because the ACS sample in 2000 included only about 40 percent of the counties, and because ancestry can be very concentrated geographically, we did not feel comfortable comparing ancestries for smaller groups, and even for these groups, this is a limitation. (The full ACS will be in every county.) Note that all of the countries shown are European countries. The next step should be to look at this data for the 30 ACS sites to see if the same holds true for non-European ancestries that have concentrations in those sites.

Given the fact that the ACS had ancestry reported much more often, it is not surprising that for nine of the twelve ancestries shown, the percentage of people reporting the ancestry was higher in the ACS than in the Census, and for the other three, the difference was not statistically significant. However, it is a different picture if only people with reported ancestries are considered. Then, five ancestries have a higher percentage in the ACS than in the Census, but four have a higher percentage in the Census than in the ACS.

Table 3: Ancestry, Comparison of Distributions, In Percent

Category	All People			All People With Ancestry Reported		
	Census	ACS	ACS-Census	Census	ACS	ACS-Census
Total population	273.6 million	273.6 million	------	221.6 million	241.7 million	------
Dutch	1.6	1.9	0.3*	2.0	2.2	0.1*
English	8.8	10.3	1.5*	10.9	11.7	0.8*
French (exc Basque)	3.0	3.6	0.6*	3.7	4.1	0.4*
German	15.4	17.0	1.6*	19.0	19.2	0.2
Irish	11.0	12.1	1.1*	13.6	13.7	0.1*
Italian	5.6	5.8	0.2*	7.0	6.6	-0.4*
Norwegian	1.6	1.7	0.0	2.0	1.9	-0.1*
Polish	3.2	3.3	0.1	4.0	3.7	-0.2*
Scotch-Irish	1.6	1.9	0.3*	1.9	2.2	0.2*
Scottish	1.8	2.0	0.2*	2.2	2.2	0.1*
Swedish	1.4	1.6	0.1*	1.8	1.8	0.0
United States/American	7.5	7.3	-0.3	9.2	8.3	-0.9*

* Difference is statistically significant at the $\alpha = 0.10$ level (90 percent confidence level).
Note: The difference column may not appear to be the same as the ACS column minus the Census column due to rounding.

Data users need to understand that if one is looking at the percentage of people in each ancestry group, we expect that in a fully implemented ACS that there might be increased percentages of people identified for most ancestries, due to the increased reporting of ancestry in general in the ACS.

4. Discussion and Conclusions

Since ACS is designed to replace the decennial census long form, data users will want to know if there will be systematic differences in the results from ACS compared to the Census 2000 long form. This research, the start of a report that compares ACS and Census 2000 results for profile social characteristics, shows a variety of results.

For the nativity and place of birth tables, we found no differences that we saw as meaningful. For language, we found differences that were not meaningful on an individual level, but the trend of less reporting of the amount of foreign languages spoken at home should be investigated further at the site level.

For ancestry groups, we found large, meaningful differences, mainly resulting from the fact that ACS respondents were more likely than Census 2000 long form sample respondents to report any ancestry. This work focused on the largest ancestries, which happen to be from European ancestries. Additional analysis of site-level data to determine if this effect holds for other ancestries, but for data users, the important thing for data users to know is that there very well might be more reporting of ancestry and higher percentages of specific ancestries in a full ACS than there was in Census 2000.

This report is the first step of the social characteristics comparison report the Census Bureau will be producing. That report needs to investigate the differences at the site level and conduct further investigation into the reasons for the differences found in this report. The Census Bureau is also conducting a study at the site level of the 1999-2001 three year averages for the ACS compared to Census 2000, which could also uncover systematic differences that data users will see under a full ACS compared to Census 2000.

References

Boggess, S., and Graf, N. (2003). "Measuring Education: A Comparison of the Decennial Census and the American Community Survey", in *American Statistical Association Proceedings of the Survey Research Methods Section*, 2003, on CD-ROM.

Dye, J. (2003). "Comparing Grandparents' Response Items on Census 2000 and the Census 2000 Supplementary Survey/American Community Survey 2000", in *American Statistical Association Proceedings of the Survey Research Methods Section*, 2003, on CD-ROM.

Stern, S. (2003). "Counting People With Disabilities: A Comparison of Disability Measurement in Census 2000 and the American Community Survey", in *American Statistical Association Proceedings of the Survey Research Methods Section*, 2003, on CD-ROM.