Multiple ACS Estimates: Pick a Number, Any Number!

Linda Gage
State of California

1. Background

The U.S. Census Bureau produced a series of multi-year estimates for evaluation for thirty-four comparison counties in the American Community Survey (ACS) where survey data have been collected since 1999. The demographic, social, economic, and housing data tables for the population residing in households include six years of single year estimates for 2000 through 2005 for geographic areas with populations of 65,000 or more, five years of 3-year estimates for 1999-2001 through 2003-2005 for geographic areas with 20,000 or more population, and three years of 5-year estimates for 1999-2003 through 2001-2005 for geographic areas under 20,000 population that also includes census tract data.

Annual estimates from the ACS have been published for areas with 250,000 or more population since 2003. Annual estimates for areas with 65,000 or greater population were first published in the summer of 2006 and the second set of annual estimates that includes the group quarters population data will be published this summer. Data for areas with populations from 20,000 to 65,000 will receive 3-year estimates, based on the monthly data collected during 2005 through 2007, in the summer of 2008. Data for communities with populations of 20,000 or fewer as well as data for census tracts will be pooled during the 2005 through 2009 period with the 5-year estimates scheduled for release in 2010. Data from the ACS comparison counties provide the opportunity to discover more about the nature of the various multi-year estimates, including the 5-year averages and census tract data that will not be available for all areas until 2010.

The goals of the Census Bureau's multi-year evaluation program are to address issues of data quality for small areas, stability and interpretation of multi-year estimates, methodological issues in producing the estimates, methods for comparing multi-year estimates, and the display and release of the data series. Discoveries made in analysis of the multi-year data can inform and improve the forthcoming ACS products. This research will focus on the usefulness and accuracy of ACS data for the San Francisco and Tulare county ACS test sites in California. It will consider issues that will arise in assisting data users with identification and interpretation of ACS estimates, stability and recommended uses of the various estimates series, and comparison of ACS estimates with available official administrative data sources.

The fourteen separate data series available for this study multiplied by the number of geographies included multiplied by the wealth of demographic, social, economic, and housing variables multiplied by the various quality measures represents a colossal data base. And so it will be for all future ACS data users. My initial approach was to select a very restricted set of geographies and variables and to attempt to assess these various data series with reference to the margins of error, the quality measures, and some available administrative records.

The amount of data collected and planned for annual dissemination is massive. There will be multiple annual or multi-year estimates for a specific variable that was collected in a given year for a single geographic area. For example, data collected in a single calendar year such as 2003 will be components of seven of the fourteen separate currently published estimates. In the future, those data will also be represented in two additional data series when 2002-2006 and 2003-2007 5-year estimates are released. Only the annual 1-year estimates series represents independent estimates. The overlapping data in the multi-year estimates moderates the annual changes in the ACS point estimates. While increasing the stability of the estimates, changes in the averaged data may or may not represent real change.

More geographic areas become available when more years are pooled to produce the ACS estimates. Areas such as county, county subdivision, census tract, block group, place, Native Home Land, Public Use Microdata Area (PUMA), zip code, and school district may be available for the 5-year estimates. County, county subdivision, place, and school district geographies are supported in the 1-year estimates. These geographies as well as Public Use Microdata Areas (PUMAs) are included for 3-year estimates.

As with the decennial census long-form before it, what we know about the ACS data and data quality may be largely what is revealed by the survey itself. Few administrative data sources are available to compare with the census or ACS estimates especially for small geographic areas and, except for variables like

population and housing that are used as survey controls, future decennial censuses will not provide new benchmarks for the "long-form" characteristics data now collected by the ACS.

2. ACS, Administrative Data and Population Controls

2.1 ACS and Population Controls

There are two ACS comparison counties in California: San Francisco and Tulare. San Francisco, with a household population exceeding 700,000 since 2000, has coterminous city and county boundaries; therefore for that county there is but one city and it has 1-, 3- and 5-year estimates. Tulare County's household population exceeded 350,000 since 2000. The county along with Visalia, its largest city, with a population over 100,000 qualifies for 1-year estimates as well as the 3- and 5-year estimates. The cities of Porterville and Tulare, with populations between 40 and 50,000 each have 3- and 5-year estimates. Five cities with populations below 20,000 have 5-year estimates: Dinuba, Exeter, Farmersville, Lindsay, and Woodlake.

The U.S. Census Bureau and the California Department of Finance's Demographic Research Unit each produce independent state and county estimates for California. The U.S. Census Bureau's intercensal population and housing estimates serve as population controls for the American Community Survey. The most current estimates are for July 1, 2006. The State's independent estimate is 37,444,385. The Census Bureau's estimate for the State of California is 36,457,549. This is a difference of nearly one million persons, close to 3 percent lower, than the official estimates produced by the State of California demographers.

The discrepancy in San Francisco County is about 60,000, nearly 8 percent. In Tulare, the differences are more modest, around 6,000 persons, less than 1.5 percent. The ACS data in the multi-year study are for the household population. In this county, the group quarters population estimate, of less than 6,000 persons, that is used by the Census Bureau is provided by the State of California. The group quarters population does not account for the differences in the agencies estimates. The primary discussion of the ACS data in this paper will be centered in Tulare County to reduce questions of the effect of population controls on the ACS data.

2.2 Administrative Data

Some of the administrative data sets that are readily available for comparison with the ACS estimates include births from the California Department of Health Services, civilian labor force from the California Employment Development Department, number of registered vehicles from the California Department of Motor Vehicles, number of persons granted legal permanent resident immigrant status from the Department of Homeland Security, and public school enrollment from the California Department of Education.

Birth, civilian labor force, and number of registered vehicles data are available at the county level and were compared with ACS annual estimates for Tulare County at the 2006 Joint Statistical Meetings as were elementary school enrollments at the county level. In general, the ACS estimates were less linear than the administrative records but the administrative counts of births and elementary enrollment generally fell within the upper and lower bounds of the 2000 through 2004 ACS estimates. The ACS civilian labor force estimates reported by the California Employment Development Department were somewhat higher than the ACS upper bound estimates but the trends followed a similar pattern. The State data, reported monthly, showed substantial increases in the Tulare County civilian labor force during April through July when many migrants, documented and undocumented contribute to the crop harvesting activities. It was not possible to assess the success of the ACS in surveying these populations or view monthly ACS data for this variable. It should be noted that some portion of these seasonal agricultural workers are housed in group quarters housing. The number of registered vehicles reported by the California Department of Motor Vehicles as likewise higher than the ACS upper bound of the estimate of number of vehicles available. This is at least partially, if not wholly, due to the conservative assumption that only 3 vehicles was the number represented by households selecting the "3 or more vehicles available" response category. Again, the trends from 2000 through 2004 were similar for both data series. In future, the comparison of the number of persons granted legal permanent resident immigrant status from the Department of Homeland Security will be compared to the ACS estimates of those who resided abroad one year ago at the county level.

3. High School Enrollment Comparisons

3.1 Tulare Joint Union High School District

For this initial analysis, ACS high school enrollment estimates for the Tulare Joint Union High School

District and the Visalia Unified School District in Tulare County will be examined in relationship to official enrollment reported by those districts to the State of California's Department of Education to address two goals of the multi-year estimates research: issues of data quality for small areas and the stability and interpretation of multi-year estimates. There are over 1,000 public school districts in California enrolling over six million students. This paper looks at only two of the school districts and is restricted to high school enrollments.

The table below displays the ACS estimates from the 14 separate 1-, 3-, and 5-year series, along with their upper and lower bounds calculated from the margins of error provided with the estimates, and the official enrollment reports of the school district. In order to display these data it was necessary to separately access or calculate each value: 6 for the State Department of Education and 42 for the American Community Survey. There is a massive quantity of data and on-going discussions with data users will be critical to the future access and ultimate availability and usefulness of the ACS data. In addition to these data for the school district the ACS estimates and margins of error are also available for each of the district's twenty-one census tracts for the three separate 5-year estimates series.

Table 1: Tulare Joint Union High School District Enrollment

	Estimate Period					
	2000	2001	2002	2003	2004	2005
	1999-01	2000-02	2001-03	2002-04	2003-05	
		1999-03	2000-04	2001-05		
Estimate Series						
ACS 1-Year Upper	n/a	6,478	6,436	8,001	7,462	6,609
ACS 1-Year Point	n/a	5,438	5,253	6,565	6,019	5,582
ACS 1-Year Lower	n/a	4,398	4,070	5,129	4,576	4,555
ACS 3-Year Upper	5,736	5,956	6,106	6,453	6,744	
ACS 3-Year Point	5,235	5,345	5,432	5,721	5,968	
ACS 3-Year Lower	4,734	4,734	4,758	4,989	5,192	
ACS 5-Year Upper		5,470	5,951	6,001		
ACS 5-Year Point		5,093	5,411	5,449		
ACS 5-Year Lower		4,716	4,871	4,897		
CA DOE	4,176	4,239	4,230	4,350	4,551	4,714

The population universe for the ACS enrollment question is the population 3 years and over and the question asked "At any time IN THE LAST 3 MONTHS, has this person attended regular school or college? Include only nursery or preschool, kindergarten elementary school, and schooling which leads to a high school diploma or a college degree. The next question asked "What grade or level was this person attending?" and there is a check box for "Grade 9 to grade 12". There may be some adults living in the school district who are enrolled in some high school classes or adult education classes at the high school. Although data for 2000 were not available for this district, the 3- and 5-Year averages are produced for the time periods that would include the 2000 data. The California Department of Education (CA DOE) data is a census or snapshot of enrollment reported by the school district to the California Department of Education as of October of each year. The data are for public school enrollment only, however, the addition of private school enrollment would not raise the official enrollment statistics much. The total private school enrollment in the high school grades for the entire county is around 400 per year. It is possible there is a modest amount of home schooling in the high school grades. The ACS estimates are generally higher than the enrollment reported by the state. Again, the California number captures most of the enrollment but is restricted to reported public school enrollment and would be expected to be somewhat lower than the ACS estimate. It is also a single measure of enrollment in October and would not capture any enrollment increase throughout the year or during the crop harvesting season that begins in April; however the ACS monthly sample could.

Except for 2003 when there was a noticeable spike in the ACS 1-year estimate, the California enrollment data was within the bounds of the ACS estimate towards the lower end. Adding regression trend lines to the ACS 1-year estimates and the California enrollments reveals that the overall trends are similar

and those of the ACS lower bound estimates and the reported enrollment are very close. If the trend lines were not employed, the 1-year ACS estimates appear unstable and do not appear to be capturing the steady upward trend of the school district's enrollment. Fall 2006 data are currently available from the California Department of Education but 2006 estimates are not yet available from the ACS. The enrollment gain between 2005 and 2006 is similar to the 2004 to 2005 growth and would not be expected from the two recent years of decreasing enrollments in the ACS estimates.

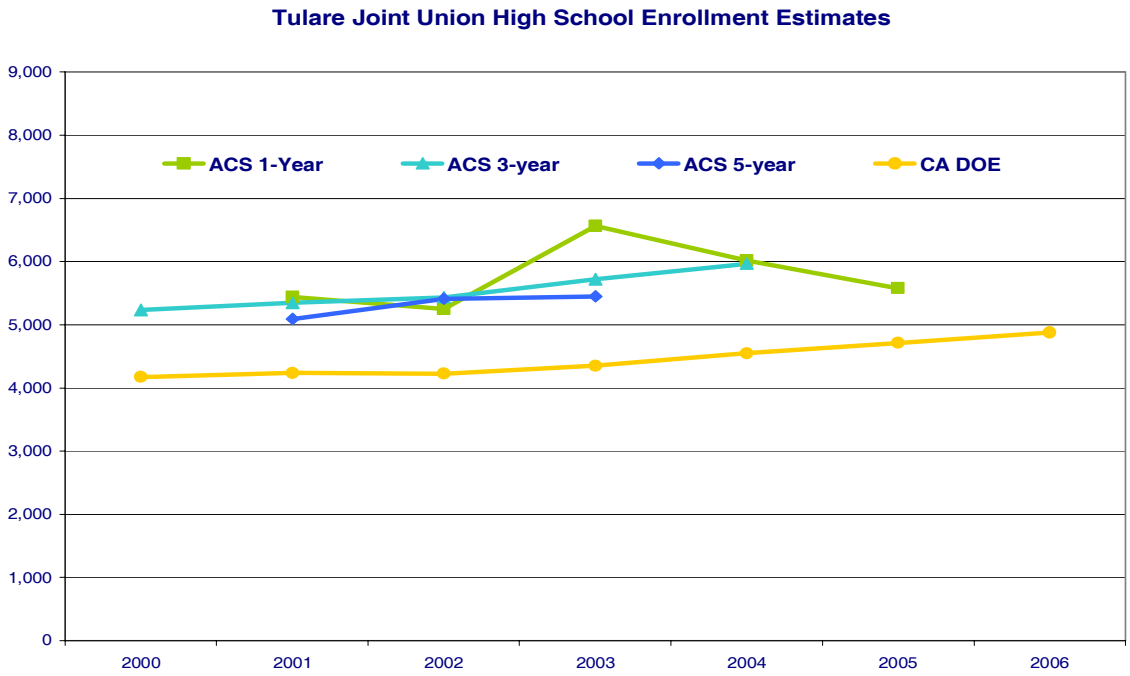

Figure 1: Tulare Joint Union High School Enrollment Estimates: ACS 1-, 3-, and 5-year and Reported Enrollments

The aggregation of 3 or 5 years of data adds stability to the estimates and more clearly indicates the upward trend of enrollments. A graph of the point estimates of the 1-, 3-, and 5-year estimates along with the California data would illustrate the need for a couple of suggested practices when using ACS data: caution about attributing real change from one estimate period to the next and pitfalls in reliance on the point estimate without consideration of the margin of error or regard for the data series.

While it's helpful to compare administrative records and ACS estimates it will generally not be possible for most of the data collected in the survey. Administrative records, when available, are most likely to be at relatively high levels of geography such as zip code, place, or county. Data users will have extremely few external references to judge the appropriate ACS data to select. In the example above, the range of the 1-year estimates for 2005 is 4,555 to 6,609. Most users would likely select the point estimate in between. Were there a perfect correlation between the ACS estimates and the administrative records the better choice would be the lower bound ACS estimate of 4,555 compared to the administrative data value of 4,714 students. In the absence of administrative records or other information the basic question about the 1-year estimates series is whether there is reason to suppose the enrollment fluctuates so much from year to year. A glance at the five years of 1-year data would call the 2003 estimate into question, especially after the 2004 or 2005 data were available.

There has been discussion about which multi-year estimates to compare to other multi-year estimates or to administrative records. Some users advocate comparing those multi-year estimates with a common mid-point while others suggest comparing those with a common end point. For example, using the mid-point approach, in 2003 the 1-year estimate is for calendar year 2003; the 3-year estimate is for the period 2002-2004; and the 5-year estimate is for the period 2001-2005. Likewise, in the option to compare estimates that have the same ending date, the 1-year estimate for 2005 is calendar year 2005; the 3-year estimate is for the period 2003-2005; and the 5-year estimate is for the period 2001-2005. In the case of the 3- and 5-year enrollment data for Tulare Joint Union High School

there is a very slightly better correspondence between the ACS and Department of Education data aligning the end-points.

3.2 Visalia Unified School District

Visalia Unified School District is a somewhat larger district in Tulare County that has complete data available for the 1-, 3- and 5-year ACS series. These are the ACS and California Department of Education data for the high school enrollment in this district.

Table 2: Visalia Unified School District Enrollment

	Estimate Period					
	2000	2001	2002	2003	2004	2005
	1999-01	2000-02	2001-03	2002-04	2003-05	
		1999-03	2000-04	2001-05		
Estimate Series						
ACS 1-Year Upper	11,470	9,602	9,018	9,046	9,711	10,415
ACS 1-Year Point	10,247	8,418	7,588	7,648	8,211	8,938
ACS 1-Year Lower	9,024	7,234	6,158	6,250	6,711	7,461
ACS 3-Year Upper	9,807	9,559	8,963	8,874	9,366	
ACS 3-Year Point	9,185	8,737	8,178	8,010	8,587	
ACS 3-Year Lower	8,563	7,915	7,393	7,146	7,808	
ACS 5-Year Upper		9,467	9,334	8,881		
ACS 5-Year Point		8,933	8,759	8,288		
ACS 5-Year Lower		8,399	8,184	7,695		
CA DOE	7,054	7,358	7,411	7,531	7,825	8,089

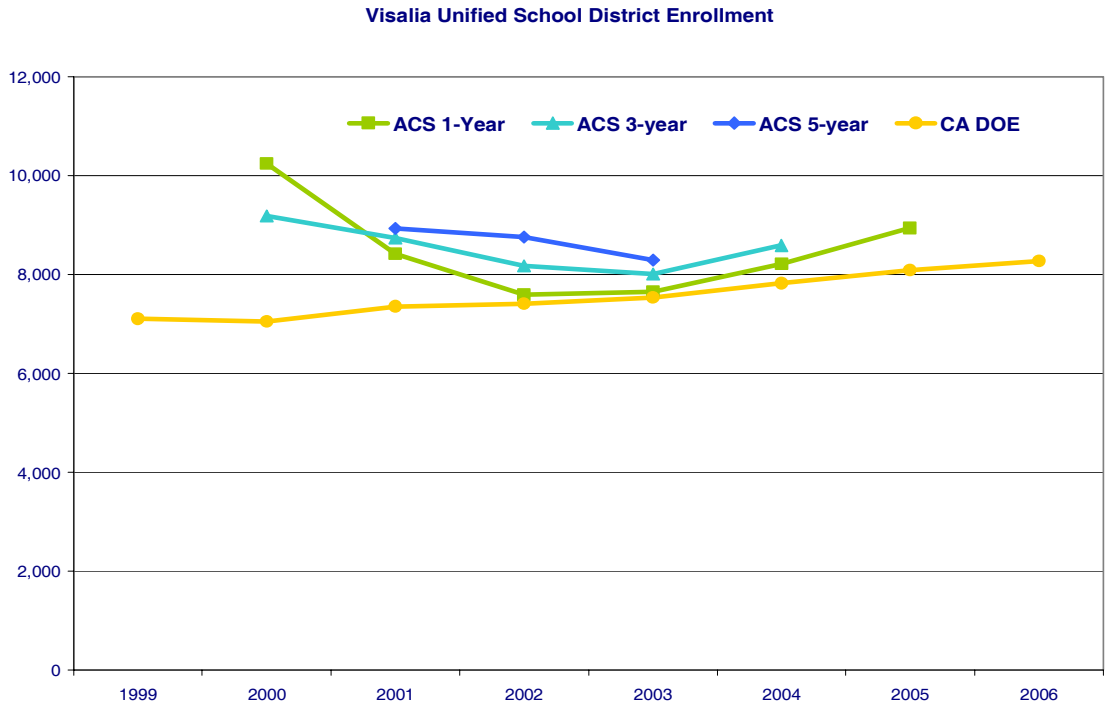

Figure 2: Visalia Unified High School District Enrollment Estimates: ACS 1-, 3-, and 5-year and Reported Enrollments

For this school district the State of California enrollment data fall easily within the bounds and quite close to the point estimates for many years. A very high ACS estimate for 2000 also affects two of the five

3-year averages and two of the three 5-year averages. In the 1-, 3-, and 5-year estimates the enrollment trend line through the 2000-2005 period is negative unlike the consistently positive growth in enrollment reported to the state. The large discrepancy in the 2000 ACS data, along with the 2001 and 2002 1-year estimates could, at the time of their release, have affected some local decisions concerning future high school enrollments and the need for facilities, staff and programs.

4. ACS Quality Measures

In addition to the margin of error for each estimate, the ACS quality measures include sample size, coverage rates, response rates, and item allocation rates.

4.1 Margin of Error

Margin of error estimates are expressed for various levels of geography. Margins of error at the census tract level, a low level of geography, can be expected to be larger than for larger, more populous geographies such as an entire school district, a city, a county, or a state.

Users should have confidence that the true value has a 90 percent chance of falling within the estimate range described by adding and subtracting the margin of error to the point estimate. A user might have less confidence in a current estimate for a census where the estimate is 46 plus or minus 42 than in an estimate that was 46 plus or minus 5.

4.2 Sample Size

There is a measure of the number of addresses initially selected for the ACS; however, throughout the survey process some address are found to be nonexistent or non-household, some addresses are not contacted during the non-response follow-up phase, and persons at some addresses refuse to participate in the survey. There is an additional measure of the number of final interviews that are conducted either by response by mail, phone, or to an interviewer. Approximately one-in-three households that do not respond by mail or telephone are selected for interviewer follow-up.

The sample in 1999 through 2001 was augmented in the ACS comparison counties so the 3-year average data could be compared to the 2000 decennial census. The large drop in the initial addresses selected, the potential sample size, in 2002 is consistent with the sampling methodology designed for full implementation across the country. The percent of the housing units interviewed, calculated by dividing the number of final interviews by the ACS estimate of total housing units, also dropped in 2002. That drop is also partially due to budget constraints that caused the loss of data for a sample month. These decreases could raise some questions about data quality, they certainly underscore the importance of sufficient on-going budgets and sufficient sample sizes for the ACS.

The initial addresses selected for the 5-year estimates in the 2001-2005 period in Tulare County census tracts ranged from 88 to 790 based on the number of housing units. The percent of final interviews from those initial addresses ranged from 45 to 75 percent.

4.3 Coverage Rates

Coverage rates are the comparison of the weighted ACS population estimate of an area or group to the Census Bureau's independent estimate for that area or group. These rates are calculated only at the county level for the total population and by gender. Coverage rates for the Tulare County 1-year estimates range from 89 to 96 percent, 90-93 percent for the 3-year estimates and 90 to 92 percent for the 5-year estimates. Recall that the follow-up for non-responding households overall is about 1-in-3.

4.4 Response Rates

The ACS response rates are the comparison of the weighted estimate of interviews and the weighted estimate of units that were eligible to be interviewed. Response rates for the Tulare County 1-year estimates ranged from 92 to 98 percent and 95 to 96 percent for the 3-year and 5-year estimates.

Response rates by mode of data collection are helpful in understanding the final response rates as the following recent performance measures demonstrate. An area’s response rate is the sum of the mail, phone, and weighted personal interview rates. The final column is the non-response rate.

For example, in the table of response rates by data collection mode Tulare County has an overall response rate of 98.3%. That is the sum of the mode rates: the 38.9% mail response rate, the 14.1% phone follow-up response rate, and the 45.3% weighted non-response personal visit follow-up.

Unlike Census 2000, ACS enumerators do not visit 100% of the non-responding households. In general, there is an overall personal visit follow-up of 1-in-3 households. In this case, the 45.3% personal visit response rate is based on the 1-in-3 sample being

weighted to approximate a 100% follow-up. The overall percent contributed by completed personal interviews is 15.1 or fewer. These interviews are weighted to achieve the 45.3 percent rate. An unweighted measure of the response rate is closer to 68.1%. Since respondents in this mode are weighted by a factor of three or more, these completed interviews have the potential to substantially affect the demographic, social, economic and housing characteristics reported in the ACS especially in non-homogeneous areas with low mail response rates and high proportions of non-response follow-up.

Table 3: Response Rates by Mode

Area	% Mail	% Phone	% Personal Visit	Total Response	% Non-Response
Nation	47.9	9.4	40.5	97.8	2.2
California	44.8	8.9	43.6	97.3	2.7
Tulare	38.9	14.1	45.3	98.3	1.7
San Francisco	50.5	7.8	38.1	96.4	3.6

Pre-release estimates: 2006 ACS 4th quarter performance measures summary.

4.5 Allocation Rates

It is easier to collect some data items than others. Most people understand, know the answer, and respond to questions about their gender, age, or housing tenure. Allocation rates, the processing of "filling in the blanks", are low for these items. At the other end of the scale are items such as yearly real estate taxes, yearly property insurance or year of entry to the United States are harder to collect. In Tulare County these items have allocation rates around 30 percent. These rates can be very helpful in assessing the quality of specific variables in the ACS as well as in the decennial census data.

For the data discussed earlier on high school enrollment, two allocation rates are available: school attendance and grade level attending. The allocation rates for school attendance vary somewhat between 1 and 4 percent in the 1-year estimates for each school district.

The allocation rates for grade level attending are somewhat higher for each school district. If there is an 8.4 percent allocation rate for grade level attending as there was for Tulare Joint Union High School District in 2003, this can be interpreted as "8.4 percent of the population age 3+ enrolled in school had their grade level allocated." However, from the published data it cannot be determined what percentage of that 8.4 percent also had their school enrollment allocated. Overall, 1.6 percent of the number enrolled in school had data allocated.

5. Recommendations to ACS Data Users and the U.S. Census Bureau

These are continuing issues with the successful evolution of the ACS that the Census Bureau is well aware of but they are worth mentioning: evaluation of the recent addition of the group quarters population data to the ACS; improvement of the Bureau's intercensal population and housing estimates used as controls for the ACS; maintenance of sufficient sample size for the survey; assistance to data users in understanding, interpreting, and using the ACS; and continued development of effective and efficient ACS data dissemination.

5.1 ACS Data Users

First, use the data and assess its face validity, does it make sense? Are there any other ACS variables or other data sources that can be used to validate or corroborate your interpretation of the ACS data? Second, be very cautious about using a single data point out of context, look at the data series and assess how stable that variable is over time and whether fluctuations are reasonable. If an annual or 3-year average estimate fluctuates significantly consider using an estimate with a longer time frame. Look at the margin of error to assess how much confidence you should have in the point estimate. Also, look behind the data by assessing the quality measures that are provided for the ACS data. Was there a notable change in the sample size in the year you are using? Were the coverage rates stable in the time period you're assessing or were there changes in the coverage rates that could have an impact on your analysis? Was there a dramatic change in response rates that could affect data quality? Was the item allocation rate for the variable under study less than 2 percent or greater than 22 percent? The Census Bureau provides the confidence intervals and quality measures along with

the data so users can make informed decisions about the quality and stability of the estimates. And finally, realize that the American Community Survey holds great promise and is in a formative stage. The Census Bureau continues to improve the ACS based on the collective experience we have with the survey. It remains important to communicate successful uses of the ACS data, raise any questions that arise, and provide suggestions for data products.

It is important to begin, as early as possible, to assess the data for the geographies and variables that you use regularly to develop an understanding of the various ACS data available. There will be no comparable detailed population or housing characteristics data collected in the 2010 decennial census to benchmark or control the ACS data series beyond the basic "short form" data.

5.2 U.S. Census Bureau

The Census Bureau has successfully fielded an innovative survey that will provide detailed data about the characteristics of the residents of the United States and Puerto Rico at small levels of geography. Maintaining a successful program of continuous measurement requires a program of continuous research and evaluation and continuous dialogue with primary stakeholders and data users to maintain the accuracy and usefulness of the data.

- Institute continued internal and independent research, evaluation, and dialogue about the ACS content, methods, data collection, use, and products. In addition to staff research, it is critical to have on-going independent evaluation of the ACS methods by the National Academy of Sciences; independent research such as this study addressing issues of the effects of survey controls, examination of unweighted and weighted survey results, and effects of the fractional sample subject to non-response follow-up on data quality and stability; and regular discussion of the most effective ways to present ACS data to the public, policy-makers, and researchers. As ACS data will soon be available for all communities it is necessary to expand efforts to educate new and future users of the ACS.

- Design ACS date products that will allow users to easily access the multi-year estimates for their jurisdiction of interest. Many more data users are interested in "everything" about their community than about "everything" about what was collected in 2003-2005.

- Enhance the availability of quality measures by expanding them to sub-state areas and adding reports of response by mode.

- Evaluate the adequacy of the annual sample size of 3 million housing units during full implementation of the ACS. This sample size has not increased since the early planning stages of the survey. The housing stock in the United States grew by 13.6 million housing units between the 1990 and 2000 censuses. Midway between the 2000 and 2010 censuses, the housing stock increased by 8.6 million units, on pace to exceed the growth in the last decade. It is important to assess how a static sample size for the ACS will affect the quality of the estimates and the margins of error.

- Address the issues of providing useful data for large geographic areas such as Chicago, New York, or Los Angeles that receive annual ACS data for the large jurisdiction but no sub-city, census tract/block group data until 5-year aggregations.

- Provide guidance, if only descriptive rationales, to data users to help them decide when and whether to compare multiple multi-year estimates by aligning the mid- or end-year points in the series.

6. Conclusions

This initial effort was to assess specific ACS estimates with available administrative records. Census data users are accustomed to use a single data set every ten year and to see linear comparisons between censuses. The strength of the ACS is the wealth multiple annual data sets. Our challenge is to assess these multiple data sets and the quality measures which we're less familiar with to use those data effectively. There is a serious hope that in addition to many other uses, selected ACS variables such as vacancy rates, persons per household, and the question on residence one year ago can be used to inform and improve the State of California's population estimates.

A more complete paper with additional tables and charts is available from the author.