

Comparison of Census 2000 and American Community Survey 1999-2001 Estimates for San Francisco and Tulare Counties, California

Background

The American Community Survey (ACS) is a Census Bureau initiative that emerged in the post-1990 period because of Congressional interest in having access more than once each decade to social and economic information for America's cities, towns and neighborhoods. The statistical design that resulted from these calls for more frequent data is a large, monthly, sample survey of addresses and a survey instrument much like the traditional census long form. These monthly samples, when pooled across time, are designed to provide annual intercensal estimates of social and economic data for all areas of the nation, including small geographic areas, such as villages and census tracts, and small subpopulations, such as race and ancestry groups. The idea for a rolling monthly survey in the U.S. had been discussed among statisticians for several decades. Congressional interest in the matter, however, was sufficient to prompt the Census Bureau in the early 1990s to commence serious work on the idea and to engage its various advisory committees in discussion on the topic. What began as discussion of continuous measurement matured during the decade into a formal American Community Survey initiative.

To ensure the wide acceptance of the American Community Survey throughout the data user community and Federal statistical system, the Census Bureau needs to conduct site-by-site comparisons of American Community Survey data to Census long form data for small areas and population groups to fully understand and assess any differences in data quality. In anticipation of the 2000 Census, and the opportunities this census would provide for comparing ACS results with census results from the long form sample, the Census Bureau identified 36 counties where ACS surveys would provide 3 years of pooled ACS survey data centered roughly on 2000.

The comparisons must reflect statistical calculations related to the data analysis. In addition, the comparisons must account for the analysis of specific population and housing variables such as income, poverty, and education for small areas. The Census Bureau must understand the explanations for the differences identified through such comparisons. Local knowledge of specific sites and locally available administrative records will contribute to a better understanding of these differences.

The development of the American Community Survey (ACS), which the Bureau plans to phase in,

with full implementation scheduled for 2005, will be aided through statistical analyses that meet rigorous technical standards and thereby enhance the credibility and acceptance of the American Community Survey as a replacement for the decennial census long-form sample.

ACS pilot testing began in late 1998 with a monthly sample, in most sites, sufficient in size that 36 months of pooled ACS surveys, centered on 2000, are believed to provide estimates of social and economic attributes with statistical precision roughly equal to the statistical precision in census long form estimates of these attributes. The recent release of SF3 data and the impending release of the pooled ACS data for the ACS sites counties now make a comparative analysis possible. Two counties in California are included in the ACS comparison sites: San Francisco and Tulare.

San Francisco County

Census 2000 enumerated 756,990 persons in households and 346,525 housing units within the 175 census tracts in San Francisco County, California. Located in northern California, the city and county of San Francisco is surrounded on three sides by the Pacific Ocean and San Francisco Bay. This cosmopolitan city/county hosts millions of visitors, conventioners and business travelers each year and attracts residents and students from around the world. The county's compact 46 square miles is home to a total of 776,733 people, a little more than two percent of the state's population. The total population includes the group quarter population of 19,743, which is not included in this study.

The population grew 7 percent since the 1990 census. Positive net migration due to a strong flow of foreign immigrants accounted for 74 percent of the growth. San Francisco County is projected to maintain the same population size in 2010 and decrease 4 percent by 2020. Nearly a third of the population is Asian. The Latino population has remained stable at 14 percent compared to a growing Latino population statewide -- now 32 percent. Almost half of the recent legal immigrants to San Francisco were born in China, Philippines, Hong Kong and Vietnam.

San Francisco County has a higher proportion of renter-occupied housing units, a lower proportion of vacant units and fewer persons per household than the state. Median income and educational attainment are higher and unemployment is lower in the county than in the state.

Tulare County

In the state's Tulare County, Census 2000 reported 361,970 household population and 119,640 housing units within 75 census tracts. Tulare County is centrally located in the Central Valley of California. Agricultural activities have allowed Tulare County to become the second-leading producer of agricultural commodities in the United States. Substantial packing and shipping operations, along with light and medium manufacturing plants, are increasing in number and becoming an important factor in the county's economy. This county of 368,000 represents a little more than one percent of the state's total population.

The population grew 18 percent since the 1990 census. Natural increase accounted for over 80 percent of that growth. Tulare County is projected to increase an additional 52 percent by 2020. Over half of the county's population is Latino, compared to 39 percent in 1990. Over 80 percent of the recent legal immigrants to Tulare are from Mexico. The large variety of about 250 seasonal agricultural crops also attracts unauthorized immigrants to the county.

Tulare County has higher proportions of owner-occupied and vacant units and more persons per household than the state. Median income and educational attainment are lower and unemployment is higher in the county than in the state.

Project Data

The U.S. Census Bureau provided 82 demographic, 92 social, 93 economic and 97 housing characteristics by census tract in four profiles to allow comparison of these characteristics from the 2000 Census long-form sample and American Community Survey estimates (three years, 1999 through 2001, averaged). The demographic characteristics are in the broad categories of sex and age; race; Hispanic origin and race; relationship; households by type; housing occupancy; and housing tenure. Social characteristics include: school enrollment; educational attainment; marital status; grandparents; veteran status; disability status; place of birth, citizenship, and year of entry; region of birth of foreign born; language spoken at home; and ancestry. Economic categories are: employment status; commuting to work; occupation; industry; class of worker; income and benefits; and number below

poverty. Housing characteristics include: units in structure; year structure built; rooms; year householder moved into unit; vehicles available; house heating fuel; selected characteristics; occupants per room; value; mortgage status and monthly owner costs; monthly owner costs as percentage of household income; gross rent; and gross rent as percentage of household income.

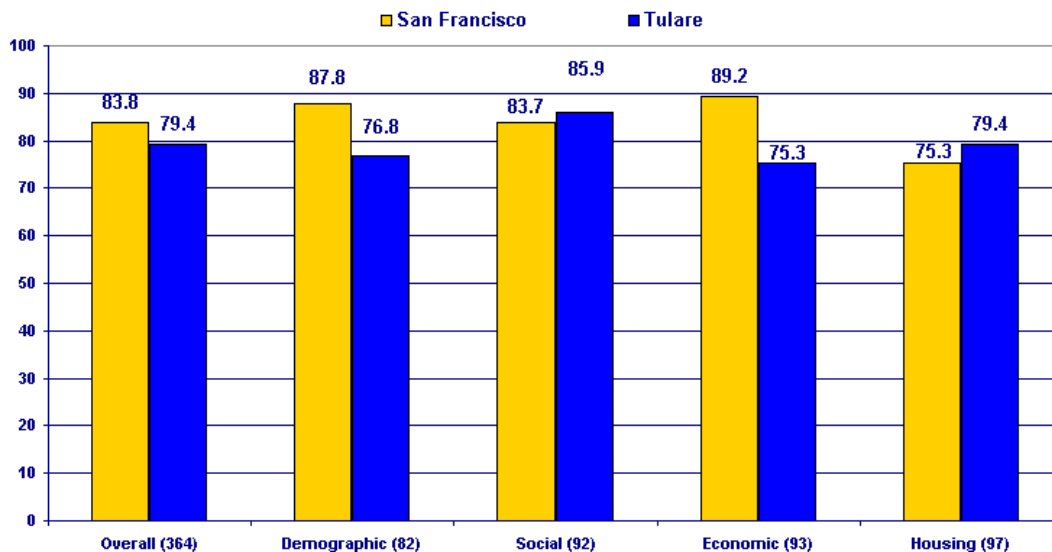
In addition to the census and survey data, several data quality measures are available including: standard errors, self-response rates, sample unit nonresponse rates, sample completeness rates, and self-response and interviewer-response population and housing item allocation rates. The Bureau also provided tests of statistical significance for most of the comparison characteristics and professional, technical and collegial support from ACS research staff.

Data Comparability and Analysis

In each of the four profiles for San Francisco and Tulare counties, the majority of data values obtained from the 2000 decennial census long-form sample and the average of the 1999-2001 ACS estimates were statistically comparable. Although there was a high degree of agreement, some variables were significantly different and have prompted additional research and analysis by Census Bureau staff. The most striking and universal differences occur in the collection of data on race, disability status, vacancy status, number of rooms in structure, and grandparents as caregivers. This was true in the California sites. Research on possible differences due to the collection instruments and collection methods have been reported at professional meetings and are available on the Bureau's website.

Though there were differences by profile, overall around 80 percent of the total variables were comparable. The somewhat lower agreement in the demographic and economic variables in Tulare County is understandable given the difference in the population during a Census snapshot on April 1 compared to averaged annual population data that may include a seasonal migrant labor force. Other differences may occur when comparing a variable like school enrollment collected on April 1 to averaged school enrollment including the summer months when many fewer children are enrolled. Data for "seasonal" variables and populations will be much more difficult to understand, more difficult to use, and for some data users not useful.

Percentage of Attributes with Similar Estimates by Profile



Although there was a high degree of agreement, some variables were significantly different and have prompted additional research and analysis by Census Bureau staff. The most striking and universal differences occur in the collection of data on race, disability status, vacancy status, age of housing unit, number of rooms in structure, persons employed, median household income, number of persons with a high school diploma and no college experience, and grandparents as caregivers.

This was true in the California sites. Research on possible differences due to the collection instruments and collection methods have been reported at professional meetings and are available on the Bureau's website.

The Census Bureau supplied single year data for the 1999, 2000, and 2001 ACS as well as the three-year data averaged on 2000 to assist in identifying possible sources of disagreement, such as an unusually high or low value in one year that would affect the average, between the census and the averaged survey results. The single year data do not contain a sufficient sample size to approximate the quality of the census long-form sample at the tract level. However, after graphing several variables in each county, the ACS data appeared fairly stable on an annual basis, yielding results that followed the general pattern of the census responses.

In Tulare County, three age groups were found to be significantly different. Many more examples of the distributions of variables in each county are in the accompanying graphs at the end, many show some level of difference between the census and the ACS. In many cases, even when statistical tests identified

differences as significant, the ACS data generally appeared useful and usable. Simply observing a statistically significant difference provides no guidance as to which data are better. It is easy and perhaps natural to assume that the census data are superior; but that bears examination. For practical purposes, it appears that most of the ACS data could, on an annual basis, be used in place of the census data and should provide a more current measurement, especially as the census count ages and remains static throughout the decade.

Tables and graphs worked well to examine the county level data. Maps were more effective in providing comparison of data at the census tract level. All examples in this paper are for Tulare County. San Francisco County had a reduced sample size that does not support tract level analysis. Tracts in that county were aggregated to neighborhoods and will be analyzed in subsequent work.

In Tulare County the Census estimated 51.3 percent of the population as Hispanic or Latino and the ACS estimated 51.4 percent. This is not a significant difference. The Census reported 185,170 Latino persons compared to the ACS estimate of 186,543. This difference is not statistically significant; this difference of 1,373 persons may or may not be meaningful to a local planner. Four of the county's 75 tracts did show significant differences; however this could be important for needs analysis and program planning. The maps show overall similarity in the distribution of the Latino population in the county.

There were significant differences in the percentage of the foreign-born population. The Census

estimated 22.9 percent foreign-born compared to the ACS estimate of 21.5 percent. This translates to 82,800 compared to 78,000 persons. The distribution in the Census and ACS are very similar, 80 percent of the census tracts did not have significant differences in the estimates of the foreign-born.

There were also significant differences in the estimation of median household income. In Tulare County, the Census reported a value of \$33,983 compared to the ACS estimate of \$31,467. This is consistent with Census Bureau research in other ACS sites that generally found lower income values reported in the ACS and is theoretically consistent with an annualized income figure including seasonal population with lower earnings.

Why Would Differences Occur?

Several reasons for differences between the results of the 2000 census and the American Community Survey have been discovered: questionnaire wording and layout, data collection mode (mail back, telephone or personal visit follow up), residence rules, allocation rates, controls and weights, ACS detecting real change, seasonal populations, and the experience of enumerators.

One factor significantly dampens the differences between the census and the ACS. The Census Bureau's post-2000 population estimates are used as population and housing controls for the ACS. The Census Bureau's post-2000 population and housing estimates have been benchmarked to the 2000 census. Population estimates, without the benefit of the control to census 2000 data would generate many more differences between the census and ACS result. We remain concerned about the quality of the Bureau's intercensal estimates.

Data Quality Measures

The measures provided to assess the relative quality of the ACS and census long-form sample estimates included self-response rates (households sending back the form), housing unit and household population sample completeness rates (how well the sample represented the universe), various nonresponse rates (incomplete answers that were allocated or "filled in"), and unit nonresponse.

Self-Response Rates

The Census, with a national media campaign and extensive partnerships, had much higher self-

response rates than the ACS overall and in the California test sites. Why is self-response a quality measure? Traditionally, questionnaires mailed in by participating households were more complete than those supplied by census interviewer or enumerators who visited households that had not sent back the census form. The Census achieved a self-response rate of 65.7 in San Francisco and 63.4 in Tulare, compared to ACS rates of 57.9 and 50.1 respectively. The darker shades indicate lower levels of cooperation. The large white area in San Francisco is Golden Gate Park.

Sample Completeness Rates

This is not a familiar measure. Basically, it assesses how well the census or survey sample represents the area's population and housing counts. Technically, "the Census 2000 sample completeness rates are based on the comparison of the number of long form sample data defined units and their population weighted by their probabilities of selection to the 100 percent housing unit and household population count" while the ACS rates were "based on the comparison of the initially weighted (without the population and housing controls and removing the noninterview adjustments) total housing and household population estimates to the final ACS estimates of total housing and household population." The ACS outperformed the Census relative to housing units but was comparable relative to household population.

Nonresponse Rates

The occupied unit nonresponse rate is a measure of the percentage of occupied housing units for which there was not enough information obtained to be considered an interview. The census nonresponse rate was considerably higher than the ACS rate. In San Francisco, the census had an occupied nonresponse rate of 12.0 compared to the ACS rate of 6.4. The difference was even greater in Tulare County where the census rate was 10.1 compared to the ACS rate of 3.9. In the maps, deep shading indicates high nonresponse rates.

Housing units with a "complete" interview may still have missing data for one or more of the questions. When there was no response an imputation was made to complete the information. Typically, in the Census self-response questionnaires are more complete than interviewer-response questionnaires. This was true in the 2000 Census.

Average Population Allocation Rates by Collection Mode

	<u>San Francisco</u>		<u>Tulare</u>	
	ACS	Census	ACS	Census
Total Item Allocation Rate	7.5	12.3	7.1	13.2
Self-response Item Allocation Rate	8.8	10.3	9.8	12.0
Interviewer-Response Item Allocation Rate	6.0	17.6	4.9	15.7

The Census had higher allocation rates in both counties in every mode. When ACS interviewers collected the information from the household the allocation rates were significantly lower than those for census interviewers. The ACS interviewers got such complete responses that the allocation rates for their questionnaires were substantially lower than the allocation rates on the self-response questionnaires.

Some data are more easily obtained than others. There is great variation in the allocation rates by item. Some population items, like gender, had very low allocation rates. Items such as income or grandparents status as caregiver often had allocation rates above 20. Self-response item allocation rates for the ACS and census were very similar in both of the California sites. The upper range of the interviewer-response allocation rates was notably higher in the census compared to the ACS.

The pattern of item allocation rates was similar for both the ACS and the census in both California ACS sites. Items difficult to collect in the census were still difficult to collect in the ACS. For EVERY variable, the ACS had lower allocation rates than the census. This improvement in data quality was gained in the non-response follow-up phase where the better-trained and more experienced enumerators got more complete interviews. Each housing unit that is enumerated in non-response follow-up will represent itself and at least two other non-responding households. It is important to maintain high a high mail response AND high quality interviews.

The response rates were substantially improved for some of the most difficult items like income. The variables below and in the following tables of population and housing item nonresponse rates are in alphabetical order of the Census Bureau's acronym for the item.

Comparison of Allocation Rates (Item Nonresponse) for Population Items

	Total Item Nonresponse			
	San Francisco County		Tulare County	
	Census Allocation Rates	ACS Allocation Rates	Census Allocation Rates	ACS Allocation Rates
Total Allocations in Census 2000 and the Averaged 1999-2001 American Community Survey				
Population Items				
Difficulty going out	11.7	5.0	10.8	4.0
Mental difficulty	9.0	4.3	9.2	3.3
Self-care difficulty	9.5	4.4	9.6	3.4
Difficulty working at a job	12.9	5.2	12.3	4.2
Age	4.1	2.5	3.6	2.1
School enrollment	8.6	4.1	8.8	2.9
Carpool size	13.1	6.0	13.0	5.0
Citizenship	3.0	1.6	1.8	0.5
Transportation to work	10.4	4.6	10.5	3.6
Class of worker	18.1	8.1	21.6	6.8
Commuting time	12.8	8.7	15.0	11.7
English ability	8.0	4.9	8.2	3.0
Employment status recode	12.8	5.8	14.4	4.9
Grade attending	15.4	7.3	11.9	5.5
Grandchildren living in home	5.8	3.5	6.4	3.5
Educational attainment	10.3	5.5	10.8	4.8
Months responsible for grandchildren	21.9	16.3	20.3	15.0

ASA Section on Social Statistics

Interest, dividend, etc. income	21.1	12.5	22.3	9.1
Other income	18.7	9.6	20.3	8.4
Public assistance	18.8	9.5	21.0	8.5
Retirement income	19.2	9.8	20.9	8.7
Self-employment income	11.1	6.2	12.2	5.4
Social security/railroad retirement	19.6	10.3	21.6	9.4
Supplemental security income	19.4	9.3	21.3	8.0
Some income allocated	28.1	22.3	32.2	23.3
Wages & salary income	18.7	15.1	21.5	15.9
Industry	16.5	9.5	19.1	7.7
Language spoken	10.9	5.6	9.7	3.9
When last worked	12.8	6.2	14.6	5.1
Time of departure	15.7	10.6	18.2	17.5
Physical difficulty	9.2	4.6	9.3	3.7
Mobility status	7.7	4.0	8.5	3.0
Migration – county	10.2	12.8	10.3	12.0
Migration – place	10.3	12.8	10.7	12.5
Migration – state	10.1	12.0	10.5	11.2
Periods of military service	12.3	13.5	12.1	13.6
Served in armed forces	9.5	4.6	9.9	3.7
Years of active duty	12.4	7.2	11.9	6.3
Marital status	4.4	2.2	4.1	1.4
Occupation	16.2	9.4	19.1	7.9
Place of birth	10.9	5.5	10.9	4.1
Place of work - county	10.9	5.4	12.2	5.1
Place of work – place	11.2	5.6	13.1	5.8
Place of work – state	10.7	5.0	12.0	4.7
Race	5.1	2.5	9.6	4.5
Relationship	3.8	1.8	3.9	2.1
Responsible for grandchildren	18.3	23.8	17.4	21.9
Vision or hearing difficulty	8.6	4.5	8.8	3.7
Sex	1.6	0.7	1.8	0.4
Hispanic	5.8	3.4	4.3	2.4
Non-English language	6.4	3.7	6.4	3.2
Hours worked each week	18.6	8.8	21.5	9.1
Weeks worked last year	20.4	9.3	24.9	9.8
Year of entry	14.1	6.8	18.9	13.3

Comparison of Allocation Rates (Item Nonresponse) for Housing Items

Total Allocations in Census 2000 and the Averaged 1999-2001 American Community Survey	Total Item Nonresponse			
	San Francisco County		Tulare County	
	Census Allocation Rates	ACS Allocation Rates	Census Allocation Rates	ACS Allocation Rates
Housing Items				
Agricultural sales	36.1	10.0	21.9	8.7
Bedrooms	11.5	6.5	19.8	6.5
Bedrooms - Vacant	22.3	9.2	31.7	30.5
Business on property	11.4	3.4	9.9	2.3
Business on property - Vacant	12.7	77.9	5.2	95.5
Complete kitchen	3.6	0.8	4.3	0.8
Complete kitchen - Vacant	7.1	6.3	11.0	13.4
Complete plumbing	3.6	0.8	4.4	0.8
Complete plumbing - Vacant	7.1	5.8	10.8	13.6
Electricity cost	15.8	6.4	18.4	8.2
Gas cost	20.9	9.3	19.6	9.2
Heating fuel	9.2	2.9	10.7	2.1
Lot size	18.6	5.7	13.3	3.4

ASA Section on Social Statistics

Lot size - Vacant	10.7	1.3	10.3	4.5
Meals in rent	7.1	3.7	9.9	3.3
Meals in rent - Vacant	25.9	9.1	39.7	13.5
Monthly rent	15.5	5.0	16.3	3.5
Monthly rent - Vacant	42.5	33.6	53.0	50.0
Mortgage	7.0	2.3	7.8	2.1
Mortgage payment	23.0	12.3	21.3	11.3
Number of vehicles	6.0	1.2	8.2	1.2
Other fuel cost	25.7	10.6	30.2	12.8
Payment includes insurance	17.2	9.9	18.9	10.7
Payment includes property taxes	16.8	5.4	18.5	6.2
Rooms	7.3	3.4	8.7	2.4
Rooms - Vacant	21.8	9.0	30.7	32.7
Second mortgage payment	29.0	26.9	25.6	24.0
Telephone	4.2	0.8	5.7	0.8
Tenure	3.8	1.0	5.5	1.0
Total cost on mobile home	90.4	66.3	63.4	50.4
Units in structure	4.8	1.5	5.8	1.1
Units in structure - Vacant	2.2	3.1	2.6	1.4
Vacancy Status - Vacant	1.1	2.7	0.1	4.2
Value	13.0	11.6	14.6	11.7
Value - Vacant	23.3	10.9	25.7	39.7
Water and sewer cost	16.7	6.3	19.8	7.2
Year built	13.1	14.1	15.3	16.5
Year built - Vacant	16.2	18.1	36.3	44.3
Year moved in	6.3	3.2	7.9	3.1
Yearly property insurance	32.8	23.0	42.3	34.5
Yearly real estate taxes	28.4	20.2	38.7	30.3

The measurement of educational attainment and language spoken at home in the Census and ACS showed significant differences in some of the response categories in each of the California counties.

In San Francisco and Tulare counties, there were significant differences in several of the educational attainment categories between the Census and the ACS. The census portrayed lower proportions of persons as high school graduates or as holder's of a bachelor's degree. The total allocation rates for education attainment may cause some of the differences. Allocation rates in the ACS were around 5 percent while the rates in the Census exceed 10 percent. The same pattern holds in the allocation rates for language spoken at home. In San Francisco the Census allocation rate of 10.8 was reduced to 5.6 in the ACS and in Tulare County the Census rate of 9.7 was reduced to 3.9.

Strategies for Analyzing and Using ACS Data

The data available to make your own assessment of the comparability, quality usefulness, and potential benefits of the American Community Survey is initially overwhelming. The data, quality measures, and geography make analysis a challenge. Statistical measures like the differences, standard errors, z-scores and p-values can help to quickly look for significant differences but some statistically

significant differences may not be meaningful differences in the world of the data user. In general the ACS appears to be measuring the same things in much the same ways as the census and getting similar results. There is still much to learn about data comparability, reasons for differences and whether "different" is better, worse or just different. There are differences between the Census and the ACS, some statistically significant differences. These may ultimately be welcome differences if the ACS data are consistent, more current and of higher quality than data from the decennial census long-form sample. A few suggestions as you proceed to use the ACS data:

- As you do your own analysis, don't try to analyze all the data all at once even if you use all the items or must supply them to others.
- Concentrate on the data items that you already use in your work or frequently. Compare those items with the census data.
- Don't assume the census picture is more accurate. Check the quality measures.
- Compare ACS and census data to administrative records that you may have available.
- Consider whether the data make sense.
- Learn to use and provide standard errors supplied with ACS data.
- Communicate your findings with the Census Bureau and others evaluating the ACS data. This will improve the survey as it matures.

What about the Differences:

First, are the statistically significant differences meaningful differences? If so, and the data are different - is one better than the other? How valuable is more current data? Can we improve or explain the differences? Can we now move beyond the differences between the census and the ACS and begin to evaluate an ACS time series? This research was based on the 3-year averaged ACS data for 1999, 2000, and 2001. With the recent release of 2003 ACS data, the Census Bureau could provide additional 3-year averages: 2000 to 2002 and 2001 to 2003. We could then move beyond comparing ACS data to census 2000 and begin to evaluate and use ACS as an annual data source.

Prospects and Predicaments

The American Community Survey has been designed to collect and provide more complete and more current demographic, social, economic and housing information between censuses and to replace the Census 2010 long-form. The success of this endeavor depends upon continuous and adequate funding, sufficient sample sizes, and a current and accurate Master Address File. Shortfalls in any of these areas could reduce data quality. The decennial Census is subject to the same perils.

The ACS faces additional challenges as it continues to evolve and improve:

- Including the population residing in facilities like prisons and dormitories (group quarters).
- Improving the Census Bureau's population estimates that are used as the population controls for the ACS.
- Assisting data users to use series of averaged data and data for small jurisdictions and seasonal areas.

There is a high level of agreement between census 2000 and ACS data. The American Community Survey, in its infancy, can be modified and improved. This is an exciting and promising survey. All of the evaluation data are now available on the Census Bureau's website, along with current ACS data, and numerous evaluation studies. I encourage you to become more familiar with the ACS data and perform your own evaluations. It is critical that those who use Census data evaluate the ACS survey data and communicate any major issues to the Census Bureau. The result will be more timely, more accurate data for all communities.