Can the American Community Survey Replace the Census Long Form?

Deborah H. Griffin, Susan P. Love, Sally M. Obenski
U.S. Bureau of the Census
Washington, DC 20233

Presented at the annual conference of the American Association for Public Opinion Research
Nashville, TN
May 15 - 18, 2003

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.

Introduction

The American Community Survey (ACS) is crucial to successfully re-engineering the decennial census design and moving away from the massive and nearly overwhelming effort to collect demographic, social, housing, and economic data once every 10 years from one-in-six households. Rather than occurring as part of the decennial census, this detailed and dynamic activity will be ongoing throughout the decade. Dividing this huge national workload into manageable pieces over a longer time frame provides unique opportunities that are expected to lead to overall improvements in data quality.

This paper discusses why managers are confident that the ACS can be implemented as designed. In addition, it addresses the issue of survey quality–will the ACS deliver high quality data that will meet 21st century demographic data needs? The paper will describe why the Census Bureau has concluded that the ACS is a sound replacement for the decennial census long form.

Background

The American Community Survey

Over 10 years ago, in response to congressional and other stakeholder demands for more timely and relevant data, the Census Bureau began examining a new approach for gathering detailed demographic, housing, and socioeconomic data. Research on the feasibility of collecting these data as part of an ongoing survey began in 1994. Since then, the program has evolved in preparation for full implementation. Initially tested in four sites, the ACS program expanded to a total of 30 test sites in 1999, and in 2000 to a national sample. The Census 2000 Supplementary Survey (C2SS) was designed to demonstrate the operational feasibility of using ACS methods in a national setting.

When fully implemented, the ACS will allow housing units in every county in the U.S. a chance to be selected to participate in the survey through an annual sample of about three million housing units. Puerto Rico will also be included. The ACS will produce information on content items similar to the decennial long form for communities across the country, including small areas such as census tracts, small towns, American Indian Reservations, Native Alaskan villages, and rural areas. The first data for communities of 65,000 people or more will be available after a single year of data collection. Those data will be updated annually. The first data for communities with between 20,000 and 65,000 people will be available after three years of data collection. The smallest areas and groups (less than 20,000 people) will require five years of data. These three and five year data products will also be updated annually.

The ACS is conducted continuously on independent monthly samples of addresses. The data for each sample are collected over a three month time period and the ACS design relies on optimizing three modes of data collection. Initial attempts are made to collect the data using mailout/mailback techniques, with information on the questionnaires returned by mail keyed and the data reviewed for completeness. Incomplete mail response records are followed up by telephone. Addresses that do not respond to the mailout are interviewed either by a Computer Assisted Telephone Interviewing (CATI) operation, or by Computer Assisted Personal Interviewing (CAPI) conducted on a subsample of the remaining nonresponsive addresses. The ACS divides a huge nationwide workload into manageable pieces over a multi-year time frame. The data are continuously collected, captured, processed and tabulated, reviewed, and released every year.

Feasibility and Quality

Operational Feasibility

Operational feasibility refers to the soundness of methods and the ability to successfully implement those methods. Of critical importance are assessments of whether major operations can be executed on schedule and within budget. Feasibility also encompasses productivity measures, staffing and equipment allocations, and the development of accurate workload projections.

ACS methods were designed, assessed for operational feasibility, and revised over several years of implementation. Test sites were chosen to provide an opportunity to evaluate ACS methods in a variety of settings. The test sites represent a spectrum of factors including areas with high growth, areas known to be difficult to interview, areas with highly seasonal populations, and areas with diverse racial, ethnic, occupation, and industry representation. Based on successful survey-taking experience in these sites, ACS managers were confident of the ability to expand to a national sample. To reduce the operational risks of expanding from 36 counties in 1999 to all counties nationwide, it was proposed that a national sample of over 800,000 addresses in over 1,200 counties take place in 2000. If operational feasibility could be demonstrated with that sample, expansion to a sample of 3 million addresses was thought less risky. This led to the recommendation that a national test of ACS methods be undertaken in 2000. The C2SS is that national test of operational feasibility.

Survey Quality

It is important that both the producers and consumers of survey data become knowledgeable about the elements of survey quality. According to the Federal Committee on Statistical Methodology, survey quality has four key dimensions: relevance, accessibility, timeliness, and accuracy (Office of Management and Budget, 2001). This report will discuss timeliness and accuracy.

Timeliness refers to either the length of the data collection's production time or to the frequency of the data collection. Timely data are current data. Data are considered timely if a minimal amount of time separates the event described by the data and the data availability. Data from recurring surveys, such as the ACS, produce current data while periodic or one-time survey data may quickly become obsolete.

Accuracy refers to the closeness between estimated and true (unknown) values. This is probably the most important aspect of survey quality. Accuracy can be measured by studying sources of survey errors. High levels of survey error can lead to incorrect conclusions by data users. Although survey errors can be summarized in different ways, they fall into two broad categories—sampling and nonsampling errors.

Sampling error refers to the variability that occurs by chance because a sample, rather than an entire population, was surveyed. Lessler (1992) describes sampling errors as being present by design, resulting from a conscious choice to study a subset rather than the whole population. She further notes that sampling errors are not the result of mistakes per se, although mistakes in judgement when designing a sample may cause larger errors than necessary.

Nonsampling error refers to all other errors that occur in a survey, such as nonresponse (missing or incomplete information from the sample), coverage (missing or duplicate units or persons), measurement (data collection errors), and processing errors. Lessler (1992) notes that nonsampling errors are often thought of as being due entirely to mistakes and deficiencies during the development and execution of the survey procedures. She states that a perfect design, perfectly implemented would be free of nonsampling errors. Although any error can result in poor survey data, the main sources of potential error should be explored to provide a clear picture of the combined effect on the resulting data.

Generally speaking, survey designers must make trade-offs between not only sampling and nonsampling error but also among the other three dimensions of survey quality--relevance, accessibility, and timeliness. The trade-offs are determined through analyses of cost, schedule, and required performance. For example, the decennial long form sample has been the standard provider of small area socioeconomic data. However, concerns about the timeliness of census sample data and levels of nonsampling error have prompted managers to adopt the ACS as a replacement.

Methodology

The implementation of the C2SS was required to demonstrate that national implementation of the ACS is operationally feasible. Despite competition from Census 2000 for resources and lack of experience in the field with a nationwide workload, staffing was sufficient, major operations were carried out as anticipated, and expected levels of response were met. Given the large increase in workload over 1999, the C2SS operations provided insight into activities needing improvement or revision in preparation for full implementation. Based on the results of the C2SS, managers are confident of the ability to successfully conduct the ACS in all counties with independent national samples of 3 million addresses each year. Detailed operational feasibility assessments can be found in U.S. Census Bureau (2001).

This report focuses on selected measures of survey accuracy to assess the viability of replacing the decennial long form sample with the ACS. Accuracy encompasses both sampling and nonsampling errors. Sampling error occurs in the ACS data because, like the long form data, it is based on a sample. Sampling errors are not discussed in this paper Nonsampling errors include errors due to nonresponse, coverage, measurement, and processing reasons.

Nonresponse is a well-known source of nonsampling error. There are two main types of nonresponse – unit nonresponse and item nonresponse. Nonresponse exists in surveys and affects survey estimates to varying degrees. The exact amount of bias due to nonresponse is rarely known. Nonetheless, proxies for nonresponse bias, which measure specific aspects of survey accuracy, are critical to informing data users of the usefulness and limitations of the data. This paper uses survey response rates to assess the potential for unit nonresponse bias. Item allocation rates were calculated to measure the potential for estimate errors due to item nonresponse. Survey response and item allocation rates do not provide complete measures of nonresponse error. They represent the first of two components, the second being the difference between responding households and nonresponding households. The amount of nonresponse error introduced is a function of both the unit nonresponse and the item allocation rates, and the extent of the differences in the characteristics of nonrespondents and respondents. Groves and Couper (1998) provide additional details on this issue. This paper focuses on the incidence of unit and item nonresponse. Research is currently underway to assess how well ACS nonrespondents compare to ACS respondents.

Unit Nonresponse

Survey response rates are calculated regularly for most surveys. They are commonly used in assessing the potential for unit nonresponse error and are traditionally calculated as the ratio of interviewed cases to the sum of interviews, eligible noninterviews, and (when applicable) an estimate of the cases with unknown eligibility that are noninterviews. Note that cases that are not eligible to be interviewed are excluded from the denominator and considered out of scope. This is the basic definition endorsed by the American Association for Public Opinion Research (AAPOR) and documented in AAPOR (2000).

Starting with the collection of data from the first ACS test sites in 1996, survey response rates have been produced annually for the ACS. The AAPOR definition is used. Addresses found to be commercial or nonexistent are declared out of scope, and all other addresses are classified as either interviews or noninterviews. Estimates are weighted by the initial probabilities of selection dictated by the sample design and the subsampling applied prior to personal visit interviewing. In the ACS we assume that eligibility is known for all sample cases (i.e., if a case has an unknown eligibility, it is assumed eligible.)

Survey response rates have not historically been produced for the decennial census long form sample. Schindler et al (1992) recognized that about 10 percent of the 1990 Census long forms did not include sufficient data to qualify as sample interviews. This measure is quite similar to a survey nonresponse rate for the census sample.

The census sample can be examined using the same basic principles and standards recommended for sample surveys. As in all sample surveys, not all addresses selected to be in the sample are actually enumerated (i.e. interviewed). Long form addresses visited during the census are considered not eligible to be interviewed (enumerated) when they are determined to be commercial or nonexistent, or if they represent a duplicate of a unit already enumerated. These addresses are considered out of scope and are removed from the census inventory. Census sample noninterviews are units whose long form questionnaires have less than the minimal amount of information required to be included in the census sample. They can result from refusals, noncontacts, or from the collection of incomplete data. Unlike sample survey noninterviews, these units are not classified directly as noninterviews. Rather, they are converted from long forms to short forms and never tallied as noninterviews. This process has been discussed in a 1990 Census evaluation report (Schindler et al, 1992) and, more recently in comparative analyses of the quality of ACS and 1990 Census data in the Bronx (Salvo and Lobo, 2002).

Production of sample response rates for Census 2000, therefore, required a slightly different approach. A Census 2000 sample unit nonresponse rate was defined that considers the long form interviews and the expected census sample size. Probabilities of selection for the Census 2000 sample – the units enumerated on long form questionnaires – were determined at the block level.[1] Using the area sampling fractions, the expected number of census sample units for each block were determined from the total number of housing units actually enumerated in that block in the census.

Survey response rates for the ACS and for the Census 2000 sample were calculated for the nation and for 21 of the ACS test sites for which the Census Bureau has released yearly data. These ACS test sites are shown individually because they have sufficient sample sizes to produce reliable rates.

Item nonresponse

Item nonresponse occurs when a respondent fails to answer all required questionnaire items or fails to provide valid responses for questions. Statistical Policy Working Paper 31 suggests that item nonresponse rates be calculated as the ratio of the number of eligible units not responding to an item to the number of responding units eligible to have responded to the item. Responding units are the only units eligible for the computation of item nonresponse rates. In many surveys, missing data items are compensated for by using imputation methods. Information from items that were answered are used to impute values for items that are missing. Two commonly used methods of imputation methods are assignment and allocation. Assignments involve logical imputation where a response to one question implies the value for a missing response to another question. For example, first name can be used to assign a value to sex. Allocation, on the other hand, involves using statistical procedures, such as within-household or nearest neighbor matrices, to impute for missing values.

Item allocation rates were chosen as a measure of item nonresponse in the ACS and in Census 2000. An item allocation rate for questionnaire items can be computed in a method similar to that proposed in Office of Management and Budget (2001) as the ratio of the number of housing units or people for which a value for a specific item was allocated to the number of housing units or people for whom a response to that specific item was required.

Allocation rates for individual items in the ACS and in Census 2000 were computed, but because of the magnitude of these data, summary allocation measures were derived for this paper. These rates summarize completeness across all data items for occupied units (households) and are the ratio of all population and housing items that had values allocated to the total number of population and housing items required to have a response. These composite measures provide a summary picture of the completeness of all data. The allocation rates shown in this paper are fully weighted according to the procedures used to produce ACS and Census 2000 sample estimates.

As mentioned earlier, the ACS uses three modes of data collection - mail, computer-assisted telephone, and computer-assisted personal visit. For this paper we chose to look at the completeness of data for units interviewed by mail versus units interviewed by computer-assisted means - ACS's version of nonresponse follow-up. Telephone and personal visit units were therefore combined. To benchmark these levels of item nonresponse, comparisons were made to the Census 2000 sample data. The census data on allocations were similarly divided by

---

[1] Addresses in each block were systematically sampled according to the sampling fraction - ½, 1/4, 1/6, and 1/8. Sampled addresses that identified housing units were enumerated on long form questionnaires. The long form contained all the questions that appeared on the short form, along with many population and housing questions related to education, labor force participation, income, utility costs, mortgage payments, and so on.

mode, separating item completeness for units enumerated by mail and units enumerated in nonresponse follow-up.

Coverage Error

Survey undercoverage and unit nonresponse lead to the same problem—the exclusion of certain groups of people or households from the survey. Under-representation will occur if the characteristics of the housing units and population not included in the survey differ from the characteristics of those that are included. In the ACS, low levels of survey undercoverage and high survey response rates are needed to ensure that survey estimates accurately represent all population groups. The practical effect of both survey undercoverage and unit nonresponse is that ACS estimates may be biased if survey data are more incomplete for particular subgroups. For example, if Black males between the ages of 18 and 24 are disproportionately missing from the sample, the ACS could report incorrectly on other characteristics of Black males. The similar effects of survey undercoverage and unit nonresponse and the difficulty in cleanly disaggregating the two errors pointed to the need for a combined measure—the sample completeness ratio.

To evaluate the ACS, sample completeness measures were calculated using the C2SS results relative to Census 2000. Specifically, initially weighted C2SS household population estimates without adjustments for nonresponse or coverage error were divided by the total household population counts from Census 2000. This measure of sample completeness indicates how representative the C2SS sample is overall, relative to the decennial counts. A similar ratio was calculated for the Census 2000 sample. The sample household population, weighted only by their probabilities of selection, were divided by the total household population counts from Census 2000.

Measurement and Processing Error

Measurement and processing errors are two additional components of accuracy affecting the quality of the ACS. Measurement error refers to the difference between the observed value of a variable gathered during data collection and the true, unobserved value of the variable. Response error will occur if a respondent does not understand the meaning of a question or fails to recall the information accurately. Deliberate misreporting is another example of response error. Interviewer error can be a source of systematic measurement error if interviewers are not properly trained or if they misinterpret their procedures.

Office of Management and Budget (2001) explains measurement error as follows.

Measurement error comes from four primary sources in survey data collection: the questionnaire, as the official presentation or request for information; the data collection method, as the way in which the request for information is made; the interviewer, as the deliverer of the questions; and the respondent, as the recipient of the request for information. These sources comprise the entirety of data collection, and each source can introduce error into the measurement process.

Processing error occurs during the series of operations that convert questionnaire entries to machine-readable information and published estimates. For example, a data entry clerk may continually key a particular item incorrectly during data capture, or there may be an error in the CATI or CAPI instrument or a response record transmission error. In addition, clerical coding is needed for some items, and coding errors are possible. In the ACS, a detailed set of edits is used. Errors introduced through these edits are another possible source of processing error.

Accurately assessing the extent and nature of measurement and processing errors and determining how to minimize them is a difficult task requiring an ongoing research and testing program. For this review, analysts studied the processes in place to ensure that measurement and processing errors are not introduced. This includes quality assurance and quality control procedures.

Results

<u>High survey response rates suggest minimal error introduced by unit nonresponse</u>

Survey response rates are calculated annually for the ACS. For this paper, a comparable survey response rate was calculated for the Census 2000 sample. The C2SS national weighted survey response rate was 95.1 percent[2]. The Census 2000 sample response rate was 91.2 percent. Like the decennial census long form, the ACS is conducted as a mandatory survey, which has led to high levels of mail response and cooperation in telephone and personal visit follow-ups.

To assess the potential variability in rates of unit nonresponse across different areas, this paper studied 21 of the ACS test sites. Figure 1 compares the 2000 ACS survey response rates to the Census 2000 sample response rates at the site level. The estimated levels of unit nonresponse for the 2000 ACS were consistently lower than the levels of unit nonresponse for the Census 2000 sample for all 21 sites.

<u>Levels of item nonresponse were lower in the ACS than in Census 2000</u>

Table 1 details the summary allocation rates across all population and housing items for occupied housing units common to Census 2000 and the C2SS. Fifty-four population items and 29 housing items are included in this summary. The data are shown separately by mode of data collection. The levels of allocation in the C2SS are consistently lower than those from the Census 2000 sample. The completeness of the population data collected by mail was fairly similar. The greatest differences are seen in the data collected in the follow-up operations.

Figure 2 displays summary allocation rates for occupied housing units in Census 2000 and in the 2000 ACS for each of the 21 ACS test sites. This comparison was undertaken to see if the national findings held at smaller geographic levels. These rates

[2]Note that this rate differs slightly from the rate cited in U.S. Census Bureau (2001). This paper uses data from the C2SS after changes were made in the weighting methods.

reflect the proportion of household population and occupied housing unit data that required allocation. For example, in Pima, AZ, over 6 percent of the population and housing data required for occupied housing units were the result of allocation in the 2000 ACS, while in Census 2000, the rate was over 9 percent. This chart shows that the national findings hold in each of these 21 test sites.

Completeness ratios were similar for the C2SS and the Census 2000 sample

The C2SS sample completeness ratio for the total population was comparable to the ratio for the Census 2000 sample - 0.902 versus 0.914. A ratio of 0.902 indicates that the C2SS represented about 90 percent of the total Census 2000 household population. The C2SS estimate of the household population based only on the probabilities of selection and before any adjustments for noninterview, fell 10 percentage points below the official Census 2000 count of the same population. The household population estimated by the Census 2000 sample, adjusted only by the probabilities of selection, was 91 percent of the full census count. The completeness ratios suggest similar levels of coverage and nonresponse in the C2SS and the Census sample.

There are numerous possible explanations for this 9-10 percent under-representation. C2SS and Census unit nonresponse, which occur when efforts to collect the survey data are unsuccessful, contributes to this shortfall. Sample completeness levels in the C2SS can also be affected by differences in the address frame used for C2SS sampling and the final Census 2000 housing unit inventory. Since the sample completeness ratios are expressed relative to the decennial census, coverage errors in the census can also affect both the estimated ratios and their interpretation.

Some processes exist to control measurement and processing errors but additional analysis is needed

ACS methods currently include some procedures to help control measurement and processing errors. These procedures are either inherent to the use of ACS methods or are applied as part of the quality assurance activities. The CATI and CAPI operations benefitted from several quality assurance activities.

Because both CATI and CAPI use computer-assisted instruments to conduct interviews, the interviews conducted in these survey phases benefit from numerous checks and edits built into the software to ensure consistency and accuracy. For example, the software prevents most errors such as out-of-range responses or skipped questions. Additionally, in CATI, supervisors monitor calls to check for other interviewer errors such as asking the questions incorrectly or entering answers that differ from ones provided by respondents. A formal quality control recheck program is built into the CAPI operation. The work of field interviewers is sampled and respondents are recontacted to determine if there is evidence of falsification or other substandard performance.

To help ensure that processing errors are not introduced during keying, a quality assurance program designed to keep work unit total error rate below 1.5 percent has been implemented, preventing keying from being a serious source of error. Subject matter experts review unedited as well as edited data to minimize errors being introduced by the edit and imputation procedures.

As seen from these examples, ACS methods and quality assurance are helping to control measurement and processing errors. However, it is critical that the collection and processing operations continue to be monitored and, when appropriate, procedures changed to reduce measurement and processing errors.

Conclusions

Implementing the ACS is integral to the Census Bureau continuing to successfully achieve its demographic and socioeconomic data collection mission. The Census Bureau has a constitutional and statutory mandate to enumerate the population and housing of the U.S. as well as to collect detailed demographic and socioeconomic data needed by policymakers at all levels of government. Given rapid demographic change and the ever-increasing demand for more timely and relevant information, the Census Bureau must move away from the massive and nearly overwhelming effort to collect the detailed data from a one-in-six sample of households in the country over a 6 month period once every ten years. Hence, the rationale for designing, developing, and implementing the ACS–to provide more current demographic and socioeconomic data throughout the decade. Implementing the ACS will not only meet detailed demographic and socioeconomic data needs but will also simplify, streamline, and improve the 2010 census enumeration.

It was essential to demonstrate that the ACS could be scaled up from 36 counties to a national survey, which was the rationale for the C2SS conducted in 2000. Although the simultaneous conduct of Census 2000 had some negative effects on the C2SS, the C2SS data collection and processing operations were successfully completed and response rates remained high. Adjustments are being made to further optimize operational activities. Such ongoing improvement is integral to the power of the ACS. A review of 2001 rates shows continued improvements in these areas. Although the outcomes discussed in this paper do not convey the whole story, they are important performance measures indicating that the C2SS was a well-managed and executed nationwide survey.

This paper documents the quality of the ACS by discussing timeliness briefly and focusing largely on the quality dimension of accuracy. Overall, accuracy was high and compared well to the Census 2000 long form sample. The shift from 36 counties to a national sample in over 1,200 counties did not appear to have any major impact on the level of nonsampling error. It is known that sampling error in the ACS is larger than that of the census sample. Decennial samples have always been larger than the planned 5 year aggregated ACS samples slated to replace the census sample in 2010. It has also been expected--and this paper confirms--that nonsampling error in the ACS will be smaller than that experienced with the census long form samples.

Bibliography

AAPOR (2002). "Standard Definitions - Final Dispositions of Case Codes and Outcome Rates for Surveys."

Groves, Robert M. and Mick P. Couper (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.

Lessler, Judith T. and William D. Kalsbeck (1992). *Nonsampling Error In Surveys*. New York: Wiley.

Navarro, A. "Coverage in the 2001 American Community Survey and 2001 Supplementary Survey," Memorandum for Alan Tupek.

Salvo, Joseph J. and Arun Peter Lobo (2002). "The American Community Survey: Quality of Response by Mode of Data Collection in the Bronx Test Site." Paper presented at the American Statistical Association Meetings.

Schindler, Eric, Richard Griffin and Carolyn Swan (1992). "Weighting the 1990 Census Sample." Paper presented at the American Statistical Association Meetings.

Office of Management and Budget (2001). "Statistical Policy Working Paper 31: Measuring and Reporting Sources of Errors in Survey," Statistical Policy Office, Office of Information and Regulatory Affairs.

U.S. Census Bureau (2001). "Demonstrating Operational Feasibility." American Community Survey Report Series, Report 1.

U.S. Census Bureau (2002). "Demonstrating Survey Quality." American Community Survey Report Series, Report 2.

Figure 1: Comparison of Survey and Sample Response Rates - Census 2000 Sample and 2000 ACS (21 Test Sites)

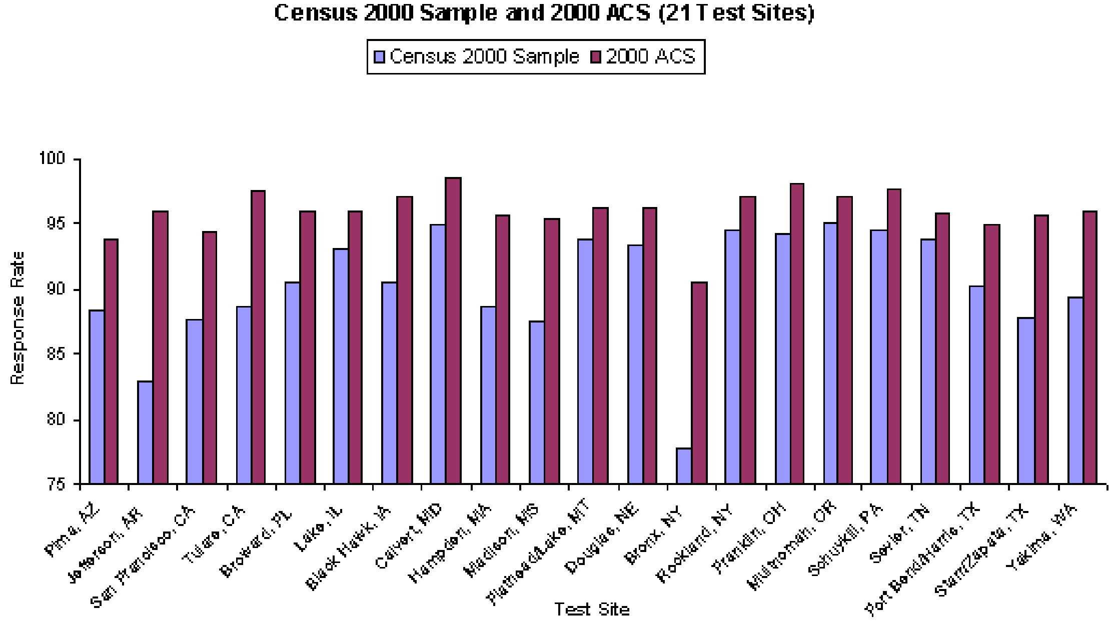

Figure 2: Comparison of Summary Allocation Rates for Occupied Housing Units in the Census 2000 Sample and 2000 ACS (21 Test Sites)

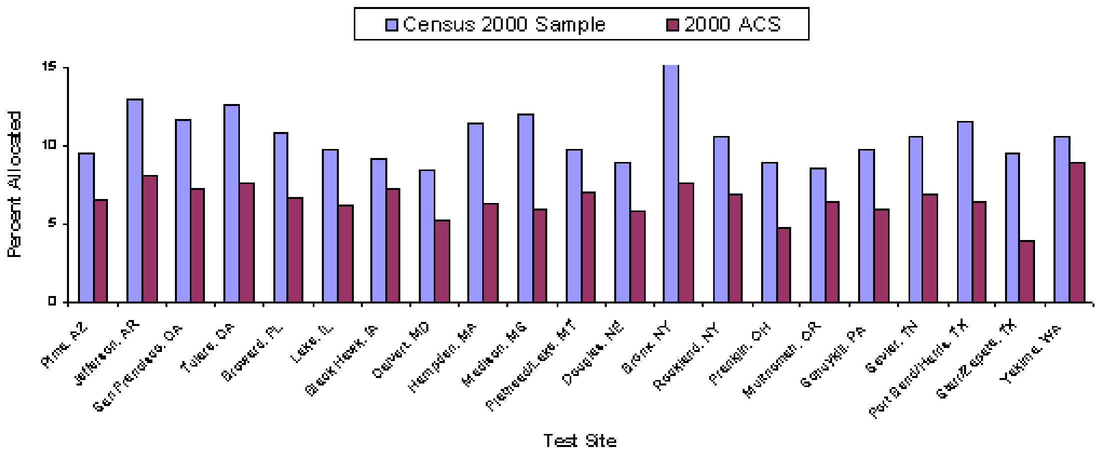

Table 1: Comparison of Summary Allocation Rates for Occupied Housing Units in the Census 2000 Sample and the C2SS

Universe	Census 2000 Sample (percent allocated)	C2SS (percent allocated)
All modes - All Items	10.4	6.6
* Population Items	9.9	6.6
* Housing Items	12.5	6.3
Mail Interviews - All Items	9.2	7.7
* Population Items	8.7	8.1
* Housing Items	11.5	5.9
Nonresponse Follow-up Interviews - All Items	13.3	5.0
* Population Items	12.9	4.5
* Housing Items	14.9	7.1