Claudette E. Bennett and Deborah H. Griffin, U. S. Census Bureau

Claudette E. Bennett, U.S. Census Bureau, Mail Stop 8800, Washington, DC 20233 Key Words: Census 2000, Census 2000 Supplementary Survey, Race, Hispanic Origin

Introduction¹

The American Community Survey (ACS) is designed to replace the decennial census "long form" (Alexander 2000). It is being promoted as a more accurate, timely, and reliable source of data on the social, economic, and housing characteristics of the United States population and housing stock. When fully implemented, the ACS will sample about 3 million addresses each year, the largest survey during the intercensal years. Results from the ACS will provide current information on the entire population down to census tracts, allowing policy makers, academic researchers, government agencies, private businesses, and the public access to information on the changing condition of the United States in a timely manner.

The ACS is in the testing and development phase. For the years 1999-2001 there were 31 sites included in the ACS. An operational feasibility test was conducted as part of Census 2000. This test - the Census 2000 Supplementary Survey (C2SS) - had a nationally representative sample of about 700,000 addresses and was conducted simultaneously with the 2000 decennial census. The C2SS was designed to be used in combination with the 31 ACS test sites, to produce estimates for the nation, states, and counties and places of 125,000 or greater populations.

It is important to undertake a comparison of the C2SS estimates with the 2000 decennial census data. The data from Census 2000 are seen as the "gold standard," providing the most accurate "snapshot" of the United States. Further, Census 2000 data and C2SS data both allow for a thorough examination of population characteristics at low levels of geography. Thus, comparing C2SS data on race and Hispanic origin to what was reported in Census 2000 provides a starting point for evaluating the validity of the C2SS estimates, as well as ultimately the estimates produced from the ACS. This type of initial comparison allows a judgment to be made about whether or not the C2SS can provide reliable data for racial and ethnic populations.

This paper compares the C2SS estimates with Census 2000 data at the national level, focusing specifically on race and Hispanic origin distributions. We examine the wording of each of the questions used to collect these data. In the final section, we discuss the observed trends and patterns, and provide possible explanations for differences found when making the above comparisons.

Background

In response to legislative, programmatic, and administrative requirements of the federal government, the Office of Management and Budget (OMB) in 1977 issued Statistical Policy Directive Number 15, *Race and Ethnic Standards for Federal Statistics and Administrative Reporting*, (OMB 1977). In these standards, four minimum race

categories were established: American Indian or Alaskan Native; Asian or Pacific Islander; Black; and White. Additionally, the standards established two ethnic categories: Hispanic origin and Not of Hispanic origin. Because Directive No. 15 defined race and Hispanic origin as two separate and distinct concepts, people of Hispanic origin may be of any race. These standards were used for over 20 years in the decennial censuses, and national surveys of the population. In 1993, the OMB began an extensive review of the 1977 standards and in October 1997, issued revisions to these standards. The new standards established five minimum race categories: American Indian or Alaska Native; Asian; Black or African American; Native Hawaiian or Other Pacific Islander; and White. Additionally, the revised standards allowed people to select or mark one or more races when selfidentification is the method of data collection.

The inclusion of more than one race reporting in the C2SS and in Census 2000, creates six single race and 57 two or more races combinations of the six single races. These 63 groups can be reduced into seven mutually exclusive and exhaustive racial categories - - six single race groups and a *Two or more races* category. Alternatively, the 57 combinations can be combined with each of the single races represented in the combinations to create six overlapping *race alone or in combination* categories, see Grieco (2001) for more information on these approaches.

Overview of Survey Methodology

The C2SS was part of the demonstration program for the American Community Survey (ACS). Its primary objective was to evaluate the feasibility of collecting long form data outside the decennial census. In addition, the data allow us to understand the impact of differences in ACS collection methods compared with the decennial long form methods. Both the C2SS and the Census 2000 long form are household surveys - an address is sampled and data are collected from a household member, usually providing data for all members of the household. Both surveys rely largely on mail as a primary mode of data collection with follow-ups conducted to collect data for nonrespondents. After data are captured, edit and consistent data. Both data sources use weighting techniques to correct for noninterviews and to control final counts to the census.

The Census 2000 Supplementary Survey

The C2SS, conducted as part of Census 2000, used the questionnaire and methods developed for the ACS to collect demographic, social, economic, and housing data from a national sample. Data collection for the C2SS began in January 2000 in 1,203 counties and ran through December 2000 (U.S. Census Bureau 2002). Approximately 58,000 addresses were sampled each month. Although the survey was conducted in only 1,203 counties, it is important to note that when added to data from the 2000 ACS comparison sites, this sample size is sufficient to produce data for every state in the Nation, as well as for most counties and metropolitan areas above 250,000 in population.

The C2SS collected data using three different data collection

¹NOTE: 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

modes: mail, telephone, and personal visit. During the initial phase, addresses were mailed a pre-notice letter to advise the occupant that they have been selected to participate in the survey. About one week later, the questionnaire arrived in the mail. Questionnaires were available only in English. Respondents were asked to return the completed questionnaire. Reminder cards were sent to all addresses and for those addresses that did not return the questionnaire, after about three weeks, a replacement questionnaire was mailed. If no questionnaire was returned after the replacement questionnaire, a telephone follow-up was attempted to obtain the information. Finally for nonrespondents, after mail and telephone attempts, a sub-sample was selected for personal visit interviewing. Permanent current survey field representatives conducted the C2SS nonresponse followup interviews. Both the telephone and personal visit interviews in the C2SS were conducted using computer-assisted technology.

Data from all modes were edited and imputation methods were used to provide missing responses. As is the case with the decennial census long form, the C2SS estimates are weighted for noninterviews to bring them into closer agreement with the census counts. This weighting is done primarily to correct for coverage error and takes race, Hispanic origin, age and sex into account. The weighting used in the C2SS did not attempt to control for all differences - but to be able to see differences in how race, Hispanic origin, sex, and age are measured and to keep from distorting other characteristics (which might have happened if agreement had been forced without taking into account differences in measurement.)

Census 2000

Census 2000 relied largely on mailout/mailback methods of data collection. Addresses received pre-notice letters a few days before questionnaires were delivered by the U. S. Postal Service or by census enumerators. Respondents were asked to complete the questionnaires and return them by mail. Reminder cards were sent to all households. Questionnaires were available, upon request, in Spanish, Chinese, Korean, Vietnamese, and Tagalog. After about a month, all nonresponding households were identified and followed up by a personal visit. Nearly 500,000 interviewers were hired and trained to collect census data using paper-and-pencil.

Mail and interviewer questionnaires were captured using Optical Mark and Optical Character Recognition methods. Edit and imputation methods similar to those used in the C2SS were used in the census to correct for missing and inconsistent data. Noninterview adjustments were also made.

Methods for Collecting Race and Hispanic Origin Data

Race

The wording and format of the race and Hispanic origin questions were tested extensively over the course of the decade. The wording for the question on race used on the C2SS mail questionnaire was similar to that used on the mail questionnaire in Census 2000 - both of which adhere to the revised OMB 1997 standards for collecting, tabulating, and presenting data on race. While wording for the response categories was identical on both instruments, there were slight differences in format.

The wording of the C2SS and the Census 2000 race questions used in telephone and personal visit interviewing differed from the wording used in the mail. Certain differences in the question wording are needed to accommodate the mode of data collection, but terminology changes also were used that were not strictly necessary. On the mail questionnaire, respondents were instructed to *mark one or more races to indicate what this person considers himself/herself to be.* In both the telephone and personal visit modes for the C2SS, the wording of the question on race requested that the respondent *choose one or more categories that best indicate his/her race or races.* The response categories were basically the same. However, in both telephone and personal visit interviews, examples were provided for the "Other Asian" and "Other Pacific Islander" response option. No examples were included on the C2SS or Census 2000 mail questionnaire.

The Census 2000 nonresponse followup questionnaire included a minor variation in the wording of the race question from how the question was posed on the census mail form. Response categories were identical to those on the census mail form and to those used in the C2SS.

Hispanic Origin

Both the wording and format on the mail questionnaires for the Hispanic origin question in the C2SS were similar to those used in Census 2000 and adhere to the 1997 revised OMB standards for collecting and presenting data on Hispanic origin. The question read, "Is this person Spanish/Hispanic/ Latino?" The response categories were also similar. The Census response categories were double-banked. Although the response categories were the same, the question version used in Census 2000 nonresponse followup was quite different. It asked, "Are any of the persons that I have listed Mexican, Puerto Rican, Cuban, or of another Hispanic or Latino group?"

The question on Hispanic origin used during telephone and personal visit follow-ups in the C2SS differed from both the Census 2000 mail and nonresponse followup forms. The question was presented in two parts. Part 1, asked "*Are you Spanish, Hispanic or Latino?*" Part two asked for specific detailed Hispanic origins, such as Mexican, Puerto Rican, and Cuban. A fifth category was, "Other Spanish/ Hispanic." While examples such as Argentinean, Columbian, Dominican, Nicaraguan, were provided in the C2SS, no such examples were provided during Census 2000 nonresponse follow-up interviews.

Analysis

Race Alone

The population that reported only one race category is referred to as the *race alone* population. Six major race categories are reflected - *White alone, Black or African American alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone,* and *Some other race alone.* Persons choosing more than one of these six race categories are referred to as the *Two or more races* population. The combination of the six *alone* categories and the one *Two or more races* category represents seven mutually exclusive and exhaustive categories. The distributions presented in Tables 1 through 3 are based on these seven race categories. These data are based on the data released and subsequent minor revisions have been made that could impact these findings.

The major findings from Census 2000 also hold for the C2SS - nearly 98 percent of all persons reported only one race and the largest group were White. About 12 percent of the population were *Black or African American alone*, about 4 percent were *Asian alone*, and less than one percent were *American Indian and Alaska Native alone* or *Native Hawaiian and Other Pacific Islander alone* (Grieco 2001).

Table 1: Comparison of Census 2000 and C2SS Race Distributions - Household Population Only

Race(s)	Census 2000	C2SS
Total household population 273,643,000	100.00	100.00
One race	97.55	97.88*
White	75.33	77.48*
Black or African American	12.04	11.75*
American Indian and Alaska Native	0.88	0.77*
Asian	3.67	3.81*
Native Hawaiian and Other Pacific Islander	0.14	0.16
Some other race	5.50	3.90*
Two or more races	2.45	2.12*

*C2SS significantly different from Census 2000 at $\alpha = 0.10$.

There are however, noteworthy differences in the distributions for the race *alone* categories observed in the C2SS compared with 100 percent Census 2000. Table 1 compares the race distributions for the C2SS to those for the household population in Census 2000. Because the C2SS is based on a sample, confidence intervals exist for the C2SS results. The percent distributions in the C2SS that were statistically significant from the percent distributions in Census 2000 are flagged (*). All other apparent differences could be explained by sampling error.

The area of greatest difference is in the percent of persons with a race of *White alone* (75.3 percent in Census 2000, 77.5 percent in the C2SS) and of *Some other race alone* (5.5 percent in Census 2000, 3.9 percent in the C2SS). Most of this difference is for Hispanic respondents. In addition, a significantly lower proportion of persons in the C2SS had Two or more races (2.4 percent and 2.1 percent, respectively). Small, but significant differences also exist for *Black or African American alone*, *American Indian or Alaska Native alone*, and *Asian alone*. Some of these differences may result, in part, from reporting differences in the race groups.

Race Alone for Non-Hispanics

As was seen in Census 2000, the C2SS found very low rates of *Some other race alone* for the non-Hispanic population. Only about five percent of the people reporting *Some other race alone*, were non-Hispanic. Both Census 2000 and C@SS found that about 19 percent of all non-Hispanics reported a race of *Black or African American alone*, *American Indian and Alaska Native alone*, *Asian alone*, or *Native Hawaiian and Other Pacific Islander alone* (Grieco 2001). But significant differences were observed for each of these groups except Native Hawaiian and Other Pacific Islander in the C2SS and Census 2000 for the non-Hispanic population. Table 2 summarizes these results. Higher proportions of the C2SS non-Hispanic population reported a race of *White alone*, *Asian alone*, and *Some other race alone*. Lower C2SS percent distributions were found for *Black or African American alone*, *American Indian and Alaska Native* alone, and for Two or more races.

Table 2: Comparison of Census 2000 and C2SS Race Distributions - Not Hispanic/Latino Household Population Only

Race(s)	Census 2000	C2SS
Total household population reporting as Not Hispanic or Latino	239,051,000 100.00	239,309,000 100.00
One race	98.11	98.26*
White	79.30	79.58*
Black or African American	13.49	13.21*
American Indian and Alaska Native	0.84	0.76*
Asian	4.15	4.31*
Native Hawaiian and Other Pacific Islander	0.14	0.16
Some other race	0.19	0.25*
Two or more races	1.89	1.74*

*C2SS significantly different from Census 2000 at $\alpha = 0.10$.

Race Alone for Hispanics

In Census 2000, the overwhelming majority of people reporting *White alone, Black or African American alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, and Two or more races were not Hispanics.* However, over 95 percent of people reporting *Some other race alone* were of Hispanic origin, see Table 3. When C2SS data on race are compared with Census 2000 for the Hispanic population, the percent distribution of race in the C2SS shows that *White alone* was much higher in the C2SS (62.9 percent) than in Census 2000 (47.9 percent). This was counterbalanced by a significantly lower proportion of *Some other race alone* in the C2SS (29.4 percent) when compared to Census 2000 (42.2 percent).

Table 3 shows that most of the C2SS race distributions for the Hispanic population differed from Census 2000. However, the major conclusions drawn from Census 2000 on race of Hispanics are also true in the C2SS. Nine-out-of-ten Hispanics in Census 2000 reported *White alone* or *Some other race alone* (Grieco 2001). The C2SS found the same result, however, a greater contribution to that 90 percent came from *White alone* than from *Some other race alone*. Less than 4 percent of Hispanics or Latinos reported either *Black or African American, American Indian or Alaska Native, Asian, or Native Hawaiian or Other Pacific Islander* in Census 2000 (Grieco 2001). This was similar in the C2SS, but the rate was just under 3 percent.

The race distributions for Hispanics suggests conceptual problems that the race question presents for Hispanics. The high rate of *Some other race alone* indicates that many Hispanics do not consider themselves to be White, Black, and so forth. A greater understanding of these problems is needed to develop race concepts for the 2010 Census and the ACS.

Table 3: Comparison of Census 2000 and C2SS Race Distributions - Hispanic/Latino Household Population Only

Race(s)	Census 2000	C2SS
Total household population reporting as Hispanic or Latino	34,593,000 100.00	34,334,000 100.00
One race	93.69	95.21*
White	47.89	62.91*
Black or African American	1.97	1.57*
American Indian and Alaska Native	1.16	0.86*
Asian	0.34	0.31*
Native Hawaiian and Other Pacific Islander	0.13	0.16
Some other race	42.21	29.39*
Two or more races	6.31	4.79*
*C2SS significantly different from Census 2000 at $\alpha = 0.10$.		

Two or More Races

Table 4 details some of the differences in race distributions for the *Two or more races* population. Overall, significantly higher proportions of *Two or more races* were found in Census 2000, compared with the C2SS. Most of those differences seem to be explained by greater numbers of *Some other race* as a second race category reported in Census 2000. A large proportion of the *Two or more races* in Census 2000 resulted when one of the two races was *Some other race*. When the data are broken down by Hispanic or Latino, this difference is highlighted in the Hispanic or Latino population. About 5.1 percent of all persons reporting as Hispanic or Latino in Census 2000 report two races when *Some other race* is included. In the C2SS, that rate is only 3.4 percent for the Hispanic population.

Race Alone or In Combination

The *race alone or in combination* categories are based on the combination of persons who reported one race and persons who reported that same race in addition to one or more other races. Previous tables looked at six *race alone* categories. Table 5 presents data for six *race alone or in combination* categories. Unlike Tables 1 through 3, the *alone or in combination* categories are tallies of total responses to the race question, and thus are not mutually exclusive with respect to the population. Consequently, the sum of these six race categories equals the number of reported races, not the total population. Additionally, the table uses total household population as the base, so the rows will not sum to 100 percent.

Table 4: Comparison of Census 2000 and C2SS Race Distributions of Two or More Races by Hispanic/Latino -Household Population Only

Race(s)	Census 2000	C2SS
Total household population 273,643,000	100.00	100.00
Two or more races	2.45	2.12*
Two races (including Some other race)	1.08	0.55*
Two races (excluding Some other race) and three or more races	1.37	1.57*
Total household population reporting as Not Hispanic or Latino	239,051,000	239,309,000
Two or more races	1.89	1.74*
Two races (including Some other race	0.50	0.15*
Two races (excluding Some other race) and three or more races	1.39	1.59*
Total household population reporting as Hispanic or Latino	34,593,000	34,334,000
Two or more races	6.31	4.79*
Two races (including Some other race)	5.09	3.35*
Two races (excluding Some other race) and three or more races	1.22	1.45*

*C2SS significantly different from Census 2000 at $\alpha = 0.10$.

Census 2000 and the C2SS found that *White alone or in combination* with at least one other race, was the largest of all *alone or in combination* categories, representing over three fourths of the total population. The next two largest categories in both the census and the C2SS were *Black or African American alone or in combination* and *Some other race alone or in combination*. Overall, significantly higher proportions of *White alone or in combination* were found in the C2SS, compared with Census 2000. Most of the these differences can be explained by the lower reporting of *Some other race alone or in combination* in the C2SS.

A lower proportion of *Black alone or in combination* was found in the C2SS compared with Census 2000. Additional research is needed to explain why this is the case. Whereas significant differences were found in the alone categories for *American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander* populations, they were not significantly different when assessed as alone or in combination. This suggests that some of the differences are due to differences in patterns of reporting *Two or more races*.

Hispanic Origin

Table 6 compares the percent distribution of total Hispanic or Latino in Census 2000 and the C2SS. The proportion of the population identified as Hispanic or Latino was not significantly different in the Census and the C2SS. Some differences however, were found in the detailed Hispanic origin categories. A higher proportion of the total household population were *Mexican* in the C2SS, while a lower proportion were classified as *Other Hispanic* may be partially explained by the use of examples in the telephone and personal visit data collection phases of the C2SS.

 Table 5: Comparisons of Census 2000 and C2SS Race

 Distributions - Household Population Only

Race (s)	Census 2000	C2SS
Total household population 273,643,000	100.00	100.00
White alone or in combination	77.3	79.2*
Black or African American alone or in combination	12.7	12.4*
American Indian and Alaska Native alone or in combination	1.5	1.5
Asian alone or in combination	4.3	4.3
Native Hawaiian and Other Pacific Islander alone or in combination	0.3	0.3
Some other race alone or in combination	6.6	4.5*

*C2SS significantly different from Census 2000 at $\alpha = 0.10$.

 Table 6: Comparisons of Census 2000 and C2SS Hispanic

 Origin Distributions - Household Population Only

100.00	100.00
87.36	87.45
12.64	12.55
7.4	7.9*
1.2	1.3
0.4	0.5
3.6	3.0*
	87.36 12.64 7.4 1.2 0.4

*C2SS significantly different from Census 2000 at $\alpha = 0.10$.

Limitations

These comparisons are based on aggregate distributions after all data processing. For this reason, any differences noted in this paper may not strictly represent reporting differences. Most tables compare final data from Census 2000 100 percent data and the C2SS after all editing and weighting. Additional analysis is underway to compare the C2SS and Census 2000 responses to the race questions at the both the aggregate and response record levels. This paper may not strictly represent

Conclusions

The American Community Survey is an important instrument for the future collection of demographic, social, and economic information for the nation. Understanding the results of this survey is critical during the testing phases in order to evaluate the estimates that are produced. Of particular importance is investigating data from the ACS for racial and ethnic populations, since this is a growing segment of the U.S. population.

There are several key factors that must be kept in mind when comparing results of the C2SS and Census 2000, namely the purpose of the instrument, differences in data collection methodologies, differences in question wording and format, and the use of the estimates by the general population. In general, survey results indicate that small differences in how questions are asked can result in substantial differences in how people respond. For even the same question can elicit different responses from the same person at different points in time. The combination of these two factors means there is no simple straightforward answer to the question of why the results of the C2SS differ from those of Census 2000 by race and Hispanic origin.

Another important factor is the mission for the data. The goal of Census 2000 was to count everyone in the population and to get a complete count, including specific groups within the *Asian, Native Hawaiian and Other Pacific Islander*, and *Hispanic* populations. Its questionnaire was designed to do this and was highly successful. The C2SS, like the census long form, was designed to collect detailed characteristics of the population. It too was highly successful. It is possible that because of these different goals and small differences in methods the that respondents provided different results.

When comparisons are made of the race data from the C2SS and Census 2000, several differences are apparent. For example, Hispanics in the C2SS have a different pattern of responses to the race question than Hispanics in Census 2000; the biggest difference is that many fewer Hispanics in the C2SS report *Some other race* and many more report *White*. There are also differences for some race groups in the estimates for *alone* compared to the estimates of *alone or in combination*; in some cases, the differences are substantial.

There also are significant differences in the races reported by Hispanic respondents, and some differences in the two-or-more/one-race results, between the C2SS and Census 2000. Compared with Census 2000, C2SS tends to estimate a larger number of *Whites alone, Asians alone, and Native Hawaiians* and Other Pacific Islanders alone; and estimate a smaller number of Blacks or African Américans alone, American Indians and Alaska Natives alone, Some other race, and Two or more races population. These results are related to the reporting of race by Hispanics. The pattern for the race alone or in combination populations was similar to the race alone populations with two notable exceptions. First, the C2SS estimate for *the American Indian and Alaska Native* alone or in combination population is larger. Second, the Native Hawaiian and Other Pacific Islander alone or in combination population is smaller. When C2SS data on race are compared with Census 2000 data, the percent distribution of C2SS shows that respondents identified themselves about 1.8 percent more often as *White alone not Hispanic*, counterbalanced by a drop of about 1.6 percent in those who self-identified as Hispanic, *"Some other race" alone*. There were no other significant differences in the C2SS race distributions by race and Hispanic origin when compared with comparable Census 2000 race distributions. These results illustrate that the concept of race is complex and it appears that very minor differences in how data are collected and processed can affect the responses to a far greater degree than previously understood.

reporting across Census 2000 and the C2SS, the Census Bureau has established a working group to investigate how comparable and replicable data across different measurement modes and procedures can be obtained. Results of these investigations are expected early in 2005. Factors likely to be considered include revisions to the instructions to the questions, format changes, changes in procedures by mode, and improved question wording. Results reported in this paper provide further confirmation of the measurement problems associated with collecting data on race and Hispanic origin. They also demonstrate the vulnerability of using different measurement procedures to collect these data.

Finally, our findings suggest that caution should be used when interpreting data on race and Hispanic origin when different questionnaire design and modes of data collection are used.

References

Alexander, C. H. (2000) "American Community Survey" pp.26-28 in Anderson, M.J., (ed.). Encyclopedia of the U.S. Census. Washington, D.C.: Congressional Quarterly Press.

Grieco, E., and Cassidy. R. (2001) "Overview of Race and Hispanic Origin: Census 2000 Brief." March 2001.U.S. Census Bureau.

U.S. Government, Office of Management and Budget. (1977) "Race and Ethnic Standards for the Federal Statistics and Administrative Reporting." Adopted on May 12, 1977.

U.S. Government, Office of Management and Budget. (1997) "*Revisions to the Standards for the Classifications of Federal Data on Race and Ethnicity*," Federal Register, Vol. 62, No. 210, Thursday, October 30, 1997, p 58782-58790.

U. S. Census Bureau. (2002) Meeting 21st Century demographic Data Needs - Implementing the American Community Survey" Report 2 - Demonstrating Survey Quality.