Comparison of Data on Ancestry: Census 2000, C2SS, and 2005 ACS Angela Brittingham Population Division U.S. Census Bureau

January 7, 2008

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on the statistical and methodological issues in this report are those of the author and not necessarily those of the U.S. Census Bureau.

INTRODUCTION

This report is one in a series that compares data from the American Community Survey (ACS) with data from the 2000 Census of Population and Housing. This report compares national estimates of ancestry, including type of report, such as single, multiple, unclassified, or no ancestry reported. It also analyzes specific groups reported, such as German or Irish. The Census 2000 Supplementary Survey (C2SS) was a nationally representative sample within the ACS program. This analysis compares C2SS and Census 2000 distributions, highlights differences that are both statistically and substantively different, and suggests possible explanations. The paper will also compare the 2005 ACS ancestry data with the C2SS data on ancestry.

METHODOLOGY

The tables included in this report compare the size of the largest ancestry groups from the Census with those in the C2SS and the ACS. Comparisons consist primarily of percentage-point differences between the two distributions. The tables display the Census, C2SS, and ACS estimates, the margins of error from which 90-percent confidence intervals of the estimates can be derived, and the difference between the estimates. In the case of frequency distributions, the difference is calculated as the percent difference between the two estimates, while in the case of relative frequency distributions (percentage distributions), the difference is calculated as the percentage-point difference between the two estimates.

_

¹ Most reports in this series compare the ACS with the Current Population Survey (CPS). However, because the CPS does not ask a question on ancestry, this report compares ACS ancestry data with Census 2000 ancestry data.

At the national level, survey variances were small, resulting in many statistically significant differences between the Census, the C2SS, and the ACS distributions. This report focuses on statistically significant differences of 0.5 percentage points or more. This yardstick can vary based on the relative size of the category. For example, for groups constituting a relatively large percentage of the population (such as people reporting German ancestry), a 0.5 percentage-point difference in the estimates might be considered small, while for groups constituting a smaller percentage of the population (such as people reporting Swedish ancestry), a 0.5 percentage-point difference could be considered large. This decision is subjective, and users can apply their own standards to interpret the data presented in this report.

The remainder of this section examines differences in methodology between Census 2000, the C2SS, and the ACS.

Sample Frame

The long-form questionnaire used in Census 2000 was sent to a sample of approximately 1-in-6 households. This sample was designed to produce national, state, and substate estimates of many social and economic characteristics from questions that were not included on the Census 2000 short form that was sent to the entire population.²

_

² For a detailed explanation of the Census 2000 sampling frame and data collection procedures, see U.S. Census Bureau, 2002.

The C2SS was conducted to demonstrate the operational feasibility of ACS methods.

The C2SS distributions in this report come from information collected in the year 2000 from the original 36 ACS test counties plus another sample of 1,203 counties selected and surveyed using then-current ACS operational and data collection methods. The C2SS surveyed a national sample of housing units, both occupied and vacant.³

Unlike the C2SS, the sample for the 2005 ACS was drawn from all counties and county-equivalents in the United States.⁴ The 2005 ACS also included all municipios in Puerto Rico, although the Puerto Rico data were excluded from the estimates in this report.

One difference between the two universes is that Census 2000, unlike the C2SS and the 2005 ACS, included individuals living in group quarters (e.g. college dormitories, emergency and transitional shelters, workers' dormitories, and group homes). For the purposes of this report, all individuals who were living in group quarters were excluded from the Census 2000 estimates.

Sample Size and Mode of Data Collection

Census 2000 data were primarily collected via mail-out/mail-back questionnaires, and the remainder gathered through enumerator-assisted visits using paper forms. The sample

³ For a detailed explanation of the C2SS survey and comparisons with Census 2000 sample items, see U.S.

⁴ For a detailed explanation of the 2005 ACS sampling frame and data collection procedures, see U.S. Census Bureau, 2006a.

size for the long form was around 43 million people, and the overall response rate was 91.2 percent.⁵

The C2SS interviewed a total of 329,589 households.⁶ Data were collected continuously throughout the year using a combination of mail-out/mail-back questionnaires, Computer-Assisted Telephone Interviewing (CATI), and Computer-Assisted Personal Interviewing (CAPI). Each month a unique national sample of addresses received an ACS questionnaire. Individuals at addresses that did not respond were telephoned during the second month of collection if a phone number for the address was available, and personal visits were conducted during the third and the last month of data collection for a subsample of the remaining nonresponding units. The C2SS achieved an overall survey response rate, calculated as the initially weighted estimate of interviews divided by the initially weighted estimate of cases eligible to be interviewed, of 95.1 percent.

The 2005 ACS used the same methods as the C2SS, although with a larger sample size. The 2005 ACS interviewed a total of 1,924,527 households. ⁷ The overall response rate for the 2005 ACS was 97.3 percent.

The Census 2000 field workers were mainly temporary employees, while the ACS employs experienced permanent interviewers for CATI and CAPI data collection. Differences in mode of data collection and experience of field workers could have an effect on how people report their ancestry.

⁵ Griffin, Love, and Obenski, 2003.

⁶ http://www.census.gov/acs/www/SBasics/SSizes/SSizes2.htm.

^{7 &}lt;a href="http://www.census.gov/acs/www/acs-php/quality">http://www.census.gov/acs/www/acs-php/quality measures sample 2005.php>.

Residence Rules

Census 2000 employed different residence rules than did the C2SS and the ACS to determine which individuals in a household were eligible for interview; the ACS uses the concept of current residence while Census 2000 used the concept of usual residence.

This difference may contribute to variation in the universes for which social characteristics are described.

The C2SS and the ACS interviewed everyone in the sample housing unit on the day of interview who was living or staying there for more than two months, regardless of whether or not they maintained a usual residence elsewhere, or if they did not have a usual residence. If a person who usually lived in the housing unit was away for more than two months at the time of the survey contact, he or she was not considered to be a current resident of that unit. This rule recognizes that people can have more than one place where they live or stay over the course of a year, and these people may affect the estimates of the characteristics of the population for some areas.

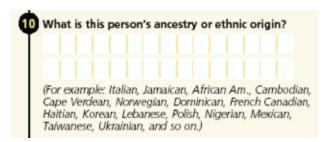
For Census 2000, each person was enumerated as an inhabitant of his or her "usual residence." Usual residence is the place where the person lives and sleeps most of the time. If a person had no usual residence, the person was to be counted where he or she was staying on Census Day (April 1, 2000). College students were counted as residents of the area in which they were living while attending college. Children in boarding schools below the college level were counted at their parental home. For Census 2000,

people living in group quarters were included in the enumeration, but they were excluded in the analysis in this paper.

Question Wording

The C2SS and the ACS ancestry question and examples were worded exactly the same as the ancestry question and examples on Census 2000 (See Figure 1). The only difference was where the question was placed in relation to the nativity questions. In the C2SS and the ACS, ancestry was asked after the nativity questions; namely, place of birth, citizenship status, and year of entry. In Census 2000, ancestry was asked before the nativity questions. This difference in question order could contribute to differences in reporting.

Figure 1. Reproduction of the Question on Ancestry From Census 2000.



Source: U.S. Census Bureau, Census 2000 questionnaire.

Data Capture

In Census 2000, the C2SS, and the ACS, up to two ancestries were captured for each person. Specific responses took precedence over more general responses, for example, if

a respondent reported European and French, only French would be shown for that person.

Also, American was captured only if the person provided it and no other response.

Religious responses were coded as "Other," and not tabulated in any of the data products.

All data on ethnic groups in this report come from the data captured in response to the ancestry question. There are some groups included in this report that have not been shown on most Census 2000 and ACS tabulations of ancestry data; these include African American, Mexican, and American Indian. Typically, these groups are shown in tabulations of race and Hispanic origin. However, to examine the most common ethnic groups reported in response to the ancestry question, these groups were included in this report regardless of their status as race and Hispanic groups.

Item Nonresponse

Item nonresponse occurs when a respondent or a household does not provide complete and usable information for a data item. Item nonresponse for the ancestry question was 18.1 percent of the household population in Census 2000, compared with 10.3 percent in the C2SS, and 7.8 percent in the 2005 ACS. Unlike most other questions on the Census, the C2SS, and the ACS, missing responses for ancestry were never allocated from other household members or imputed from neighbors; data missing for an individual was left blank.

A major difference in the data collection follow-up programs may also have contributed to differences in the item nonresponse rates. The C2SS and the ACS employed a nonresponse follow-up system, which allowed phone interviewers to re-contact people

who had left certain items on the questionnaire missing. In this way, they were able to fill in missing data on ancestry, for example, that someone may have left blank when they completed their paper questionnaire. In contrast, the Census 2000 content-edit follow-up program was cancelled because of time and budget constraints, and therefore respondents were never re-asked ancestry questions that were left blank initially.

Data Editing Procedures

The Census 2000, the C2SS, and the ACS edit rules for ancestry were designed to ensure that the final edited data are as usable as possible. These rules were used to identify and account for ambiguous or unclear responses. In each case where a problem was detected, pre-established edit rules governed its resolution; for example, in cases where "Indian" was the reported ancestry, information about race and place of birth was evaluated to try to determine if the person was of American Indian or Asian Indian origin.

The edits for Census 2000, the C2SS and the ACS were the same. Less than 1 percent of the data were edited in any of the data collection efforts.

Controls and Weighting

Differences in the selection of controls and the calculation of weights between the two surveys may lead to differences in estimates. The Census 2000 long form, the C2SS, and the ACS are all weighted to account for the probability of selection and housing-unit nonresponse.

After the initial weighting, data from the C2SS and the ACS were controlled to be consistent with independent population estimates. Data from the C2SS were controlled at the county level to the household population and number of housing units as of Census Day, April 1, 2000. Data from the 2005 ACS were controlled to the corresponding independent estimates for July 1, 2005.⁸

Estimates from the Census 2000 sample were obtained from an iterative ratio-estimation procedure that assigned a weight to each sample person. The estimation procedure used to assign the weights was performed in geographically defined weighting areas that were usually formed of contiguous geographic units within counties.

Because the C2SS and ACS are controlled to both the total population and the total number of housing units, the C2SS and ACS files, like the Census 2000 files, contain both person weights and housing unit weights to account for different coverage rates between people and housing units.

-

⁸ The C2SS and the ACS do not control to the county level for small counties. For small counties, ACS groups the counties into weighting areas with a population of approximately 30,000. Data are then controlled at the weighting area level. For more information regarding the application of population controls in the ACS, please refer to U.S. Census Bureau, 2006.

RESULTS

Comparison of ACS and Census 2000 Data

Type of Response and Item Nonresponse

Table 1 shows the number and the percentage of the population who specified a single ancestry or multiple ancestries, as well as those who did not specify an ancestry, either by providing an unclassifiable response or not responding to the question at all. The proportion of people who specified an ancestry differed in the C2SS and Census 2000. In the C2SS, 88.3 percent of people specified an ancestry, compared with 81.0 percent in Census 2000. More people reported both single and multiple ancestries in the C2SS. In the C2SS, 62.6 percent reported a single ancestry compared with 58.7 percent in Census 2000. In the C2SS, 25.8 percent reported multiple ancestries compared with 22.3 percent in Census 2000. The percentage reporting an unclassifiable ancestry was also higher in the C2SS than in Census 2000 (1.4 percent compared with 0.9 percent). The percentage who did not report any answer to the ancestry question was lower in the C2SS than it was in the Census (10.3 percent compared with 18.1 percent).

Ancestries Reported

To examine the differences in the ancestries that were reported, Table 2 shows the 15 largest ancestry groups in Census 2000 and their corresponding numbers and percentages in the C2SS. In order to account for the larger nonresponse to the ancestry question in Census 2000, the percentage for each ancestry was calculated as the number who

reported the particular ancestry divided by the total number of people who specified any ancestry.

Eight of the 15 ancestry groups had higher percentages in the C2SS than in Census 2000, although most of these were differences of less than 0.5 percentage-points and therefore not considered substantive. Only English and American Indian had percentages in the C2SS that were substantively higher than those in Census 2000. English ancestry was reported by 10.9 percent of those who specified an ancestry in Census 2000, compared with 11.7 percent in the C2SS. Likewise, American Indian was reported by 3.5 percent in Census 2000 and 4.3 percent in the C2SS.

American was the only ancestry group that was substantively lower in the C2SS than in Census 2000. Of those who specified an ancestry, American was reported by 9.2 percent in Census 2000 and 8.3 percent in the C2SS.

Comparison of C2SS and 2005 ACS Data

Type of Response and Item Nonresponse

The percentage of the household population who specified an ancestry increased from 88.3 percent in the C2SS to 91.1 percent in the 2005 ACS (Table 1). The overall increase was driven by the increase in the percentage who reported a single ancestry, up from 62.6 percent in the C2SS to 65.3 percent in the 2005 ACS. There was no increase in the percentage who reported multiple ancestries (25.8 percent in both 2000 and 2005). The

^

⁹ The apparent difference in German was not statistically significant.

percentage reporting an unclassifiable ancestry decreased over the five years, from 1.4 percent in the C2SS to 1.0 percent in the 2005 ACS. The percentage of the population who did not report any answer to the ancestry question decreased from 10.3 percent in the C2SS to 7.8 percent in the 2005 ACS.

Ancestries Reported

Ten of the 15 selected ancestry groups decreased from 2000 to 2005 in terms of their proportion of the population (see Table 2). Four of those groups decreased by 0.5 percentage points or more, namely, English, German, Irish, and American. English decreased from 11.7 percent in the C2SS to 10.6 percent in 2005 ACS, German decreased from 19.2 percent to 18.7 percent, Irish decreased from 13.7 percent to 13.2 percent, and American decreased from 8.3 percent to 7.8 percent.

The number of people with Mexican ancestry increased in proportion over the time period, from 8.2 percent of the population in the C2SS (and in Census 2000) to 9.4 percent in the 2005 ACS. This change reflects an increase over the five years between the C2SS and 2005 ACS of about five million people who reported Mexican as their ancestry.

SUMMARY

This paper compared ancestry data from the Census 2000 Supplementary Survey (C2SS) with data from the Census 2000 long form, and then compared data from the C2SS with data from the 2005 American Community Survey (ACS). Because Census 2000 and the

C2SS were conducted during the same year, differences between the results of the two surveys should mainly be methodological. On the other hand, differences between the C2SS and the 2005 ACS are more likely to reflect actual population change, with methodological differences likely playing a lesser role.

Ancestry data from the C2SS are comparable to Census 2000, but with cautions due to a higher response rate for the question in the C2SS. The larger number of people reporting an ancestry in the C2SS resulted in larger numbers in nearly all of the largest ancestries examined; however, as a proportion of those who reported an ancestry, the C2SS was substantively higher only for English and American Indian. The general response of American as a person's only ancestry was reported less often in the C2SS than in Census 2000, suggesting that in the C2SS respondents may have provided more specific responses about where their ancestors were originally from.

The lower proportion of people in the C2SS who responded to the ancestry question by saying they were American may have been caused by methodological differences between Census 2000 and C2SS. People may have reported American as their ancestry for several reasons, including multiple ethnic backgrounds, an unknown ethnic background, national pride, citizenship status, or ancestors that have been here for a long time. American is an acceptable and widely used ancestry; in fact, in all three data collection efforts, about 20 million people reported American as their only ancestry. Some researchers, however, may be interested in more specific ethnic background information, and may prefer to see fewer responses of American and a higher proportion

of responses that reflect specific countries or ethnic groups, as found in the C2SS (and 2005 ACS) data.

One notable difference in methodology could be contributing to the higher response rates in the C2SS than in Census 2000, as well as the lower proportion of the population that reported American as their only ancestry. A full 41 percent of respondents in the C2SS reported through an interviewer-assisted interview, compared with 29 percent in Census 2000. 10 Differences in interviewer training could lead to higher response rates to the ancestry question when an interviewer is present. In fact, interviewer-assisted cases in the C2SS were more likely to include an ancestry than enumerator-assisted forms in the Census (87.4 percent of CATI forms and 90.4 percent of the CAPI forms in the C2SS compared with 81.0 percent of enumerator-assisted forms in Census 2000)¹¹. Another possible cause of differences in reporting could be question placement, as ancestry was asked after the nativity questions in the C2SS, and before the nativity questions in Census 2000.

Compared with the C2SS, the 2005 ACS had an even higher response rate to the ancestry question. Of the largest ancestries examined, the proportions English, German, Irish, and American all decreased at least 0.5 percentage points, while the proportion Mexican increased.

Stern, 2003.U.S. Census Bureau, 2004.

In summary, the proportion responding to the ancestry question was higher in the C2SS than in Census 2000, and in the 2005 ACS the response rate was even higher. The proportion who reported both single and multiple ancestries was higher in C2SS than in Census 2000. Between C2SS and the 2005 ACS, the proportion reporting a single ancestry increased, while the proportion reporting multiple ancestries remained even. Differences between the C2SS and Census 2000 may be due to interviewer or questionnaire differences, while differences between C2SS and the 2005 ACS may be more reflective of change in the population.

References

Griffin, Deborah, Susan Love, and Sally Obenski. 2003. "Can the American Community Survey Replace the Census Long Form?" Presentation at the Annual Meeting of the American Association for Public Opinion Research, Nashville, TN, August 16.

Stern, Sharon. 2003. "Counting People with Disabilities, A Comparison of Estimates in the Census 2000 and Census 2000 Supplementary Survey," Presentation at the Annual Meeting of the American Statistical Association.

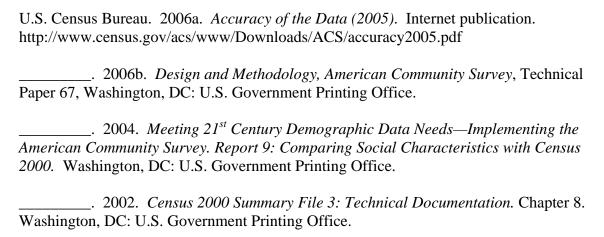


Table 1. Ancestry Reporting in the United States: Census 2000, C2SS, and ACS 2005

	Census 2000		C2SS		ACS 2005			
Characteristic	Estimate	Margin of error ¹	Estimate	Margin of error ¹	Estimate	Margin of error ¹	Difference between Census 2000 and C2SS ²	Difference between C2SS and ACS 2005 ³
NUMBER (in thousands)								
Total household population	273,637	(X)	273,643	(X)	288,378	(X)	_	5.4
Ancestry specified	221,642	36	241,750		262,797			8.7*
Single ancestry	160,589		171,178		188,321	193		10.0*
Multiple ancestry	61,053		70,572		74,476			5.5*
Ancestry not specified	51,995	36	31,894		25,581	129		-19.8*
Unclassified	2,343	8	3,786	128	3,016	41	61.6*	-20.3 *
Not reported	49,652	35	28,107	407	22,565	131	-43.4*	-19.7 *
PERCENT								
Total household								
population	100.0	(X)	100.0	` '	100.0	(X)		(X)
Ancestry specified	81.0	-	88.3	0.1	91.1	-	7.3*	2.8*
Single ancestry	58.7	-	62.6	-	65.3	_	3.9*	2.7 *
Multiple ancestry	22.3	-	25.8		25.8	0.1	3.5 *	-
Ancestry not specified	19.0	-	11.7	0.1	8.9	-	-7.3*	-2.8*
Unclassified	0.9	-	1.4		1.0	-	0.5 *	-0.3*
Not reported	18.1	-	10.3	0.1	7.8	-	-7.9*	-2.4*

(X) Not applicable.

Source: Census 2000 SF4 and special tabulation, C2SS Table PCT027 and special tabulation, and 2005 ACS Detailed Table B04007 and special tabulation.

^{*} Difference is statistically significant at the 90-percent confidence level.

⁻ Less than 0.05.

^{1.} This number added to and subtracted from the estimate yields the 90-percent confidence interval around the estimate.

^{2.} For the numbers, the difference is the percent difference and is calculated as {C2SS-Census}/Census}*100. For the percentages, the difference is the percentage-point difference and is calculated as C2SS-Census. All tests of significance are done on unrounded estimates and standard errors.

^{3.} For the numbers, the difference is the percent difference and is calculated as {(ACS 2005-C2SS)/C2SS}*100. For the percentages, the difference is the percentage-point difference and is calculated as ACS 2005-C2SS. All tests of significance are done on unrounded estimates and standard errors.

Table 2. Selected Ancestries in the United States: Census 2000, C2SS, and ACS 2005

Table 2. Selected Ancestries	Selected Ancestries in the United S				Census 2000, C2SS, and ACS 2005						
	Census 2000		C2	SS	ACS	2005	Difference				
Characteristic	Estimate	Margin of error ¹	Estimate	Margin of error ¹	Estimate	Margin of error ¹	between Census 2000 and C2SS ²	Difference between C2SS and ACS 2005 ³			
Total household population	273,637	(X)	273,643	(X)	288,378	(X)	0.0	5.4			
Total household population who specified an ancestry	221,642	36	241,750	370	262,797	129	9.1*	8.7*			
NUMBER (in thousands)											
German	42,172	32	46,424	434	49,179	125	10.1 *	5.9*			
Irish	30,008	28	33,049	247	34,669	114	10.1 *	4.9*			
African American	24,356	26	27,369	132	28,772	78	12.4*	5.1 *			
English	24,191	26	28,221	247	27,762	119	16.7*	-1.6*			
American	20,425	24	20,093	430	20,536	102	-1.6	2.2			
Mexican	18,220	23	19,761	189	24,742	107	8.5 *	25.2*			
Italian	15,452	21	15,907	264	17,235	87	2.9 *	8.4 *			
Polish	8,827	16	9,029	182	9,771	71	2.3 *	8.2*			
French	8,180			227	9,530		19.7 *	-2.7*			
American Indian	7,736			179	11,916		34.3 *	14.7*			
Scottish	4,805		5,406		5,859		12.5 *	8.4 *			
Dutch	4,472		5,203		5,079		16.4 *	-2.4			
Norwegian	4,399		4,525		4,601	37	2.9	1.7			
Scotch-Irish Swedish	4,250 3,932		5,205 4,342		5,289 4,260		22.5 * 10.4 *	1.6 -1.9			
PERCENT OF HOUSEHOLD POPULATION WHO SPECIFIED AN ANCESTRY											
German	19.0	-	19.2	0.2	18.7	-	0.2	-0.5 *			
Irish	13.5	-	13.7	0.1	13.2	-	0.1 *	-0.5 *			
African American	11.0	-	11.3	0.1	10.9	-	0.3 *	-0.4 *			
English	10.9	-	11.7	0.1	10.6	-	0.8 *	-1.1*			
American	9.2	-	8.3	0.2	7.8	-	-0.9*	-0.5 *			
Mexican	8.2		8.2		9.4		-	1.2*			
Italian	7.0	-	6.6	0.1	6.6	-	-0.4*	-			
Polish	4.0		3.7	0.1	3.7	-	-0.2*	-			
French	3.7	-	4.1	0.1	3.6	-	0.4*	-0.4*			
American Indian	3.5		4.3		4.5		0.8*	0.2*			
Scottish	2.2		2.2		2.2		0.1 *	-			
Dutch	2.0		2.2		1.9		0.1 *	-0.2 *			
Norwegian	2.0		1.9		1.8		-0.1 *	-0.1 *			
Scotch-Irish	1.9		2.2		2.0		0.2*	-0.1 *			
Swedish	1.8	-	1.8	0.1	1.6	-	-	-0.2*			

⁽X) Not applicable.

Source: Census 2000 special tabulation, C2SS special tabulation, and 2005 ACS Detailed Table B04006 and special tabulation.

^{*} Difference is statistically significant at the 90-percent confidence level.

⁻ Less than 0.05

^{1.} This number added to and subtracted from the estimate yields the 90-percent confidence interval around the estimate.

^{2.} For the numbers, the difference is the percent difference and is calculated as [(C2SS-Census)/Census*100]. For the percentages, the difference is the percentage-point difference and is calculated as C2SS-Census. All tests of significance are done on unrounded estimates and standard errors.

^{3.} For the numbers, the difference is the percent difference and is calculated as [(ACS 2005-C2SS)/C2SS*100]. For the percentages, the difference is the percentage-point difference and is calculated as ACS 2005-C2SS. All tests of significance are done on unrounded estimates and standard errors.