

OVERVIEW OF RESULTS OF NEW RACE AND HISPANIC ORIGIN QUESTIONS IN CENSUS 2000

Jorge del Pinal, Elizabeth Martin, Claudette Bennett, and Art Cresce
 U. S. Census Bureau, Washington DC, 20233

Keywords: Multiple race reporting, questionnaire effects

In October of 1997, the Office of Management and Budget (OMB) made significant revisions to the standards for federal data on race and ethnicity (OMB, 1997). These standards were implemented in the 2000 Census of Population and Housing. The most significant change to the OMB standards was to allow respondents to report one or more races. In addition, the “Asian and Pacific Islander” category was split into two: an “Asian” and a “Native Hawaiian and Other Pacific Islander” category. New standards suggest the question on Hispanic/Latino origin be asked ahead of race. In previous censuses the reverse was true. There were also some terminology changes. The Census Bureau sought and received permission to include a “Some other race” category to be used by respondents who do not identify with any of the other categories on the questionnaire (Grieco and Cassidy, 2001). The Census 2000 Hispanic origin and race items are as follows:

- NOTE: Please answer BOTH Questions 5 and 6.
5. Is Person 1 Spanish/Hispanic/Latino? Mark (X) the “No” box if not Spanish/Hispanic/Latino.
- No, not Spanish/Hispanic/Latino
 - Yes, Mexican, Mexican Am., Chicano
 - Yes, Puerto Rican
 - Yes, Cuban
 - Yes, other Spanish/Hispanic/Latino—*Print group*
6. What is Person 1's race? Mark (X) one or more races to indicate what this person considers himself/herself to be.
- White
 - Black, African Am., or Negro
 - American Indian or Alaska Native—*Print name of enrolled or principal tribe*
 - Asian Indian
 - Chinese
 - Filipino
 - Japanese
 - Korean
 - Vietnamese
 - Other Asian—*Print race*
 - Native Hawaiian
 - Guamanian or Chamorro
 - Samoan
 - Other Pacific Islander— *Print race*
 - Some other race— *Print race*

(See <http://www.census.gov/dmd/www/2000quest.html> for questionnaire facsimiles.)

This paper consists of two parts. Part I reports selected findings from the revised race and Hispanic origin questions from Census 2000. Part II documents the effects on reporting of the changes that were made in the race and Hispanic origin questions. An experiment was conducted during Census 2000 in which 2000-style and 1990-style questionnaires were mailed to randomly-selected panels of households. Results show that questionnaire differences affected reporting, and must be taken into account when assessing population changes from 1990 to 2000.

I. CENSUS 2000: RACE AND HISPANIC ORIGIN OVERVIEW

All write-in entries for the Census 2000 questions on Hispanic or Latino origin and on race were captured with Optical Character Recognition (OCR) technology during the data capture operation and subsequently coded automatically or by expert coders.

Based on the information collected for 100 percent of the population,¹ Census 2000 showed that about 12.5 percent of respondents were Hispanic or Latino, and 87.5 percent were not. As shown in Table 1, about 75.1 percent of respondents were classified by race as White; 12.3 percent as African-American (Black); 0.9 percent as American Indian or Alaskan Native (AIAN); 3.6 percent Asian; and 0.1 percent Native Hawaiian and other Pacific Islander (NHOPI)². About 5.5 percent of the population were classified as “Some other race” (SOR). However, these numbers represent the population that reported only one race. About 2.4 percent (6.8 million) of people reported having two or more races (Grieco and Cassidy, 2001, pp. 4-5).

When the distribution of race is crossed by Hispanic origin, we get two very different distributions, as shown in Table 1. Among Hispanics, 47.9 percent were classified as White, compared to 79.1 percent of non-Hispanics while 42.2 percent of Hispanics are classified as Some other race, compared to 0.2 percent of non-Hispanics. Most of the Hispanics classified in the SOR category also gave their Hispanic origin as their race. Hispanics, it seems, were

¹ The figures shown refer to actual responses provided by respondents as well as responses assigned during the editing and imputation processes unless otherwise stated.

² For brevity, we use the abbreviations in parentheses to refer to race groups with long names.

much more likely to be classified as SOR and to report two or more races than were non-Hispanics.

Another question of great interest is how many of the possible 63 categories (Grieco and Cassidy, 2001) of the major race groupings would respondents select? Similar to findings from the American Community Survey (del Pinal et al., 2001) more than nine of every ten (93.3 percent) respondents selected combinations that included only two races in Census 2000. Another 6.0 percent gave combinations of three races and 0.7 percent gave combinations of four or more races. This distribution also holds by Hispanic origin. Among Hispanics, 94.9 percent selected combinations of two races, 4.5 percent three races, 0.6 percent four or more races. Many of the combinations that Hispanic selected included "SOR" as one of the combinations as will be shown later. Among non-Hispanics, 92.5 percent selected combinations of two races, 6.8 percent 3 races, and 0.7 percent 4 or more races.

The other question of interest is what were the largest combinations of two or more races? About 32.3 percent of those who selected two or more races were reported as "White and SOR." The next largest combination was "White and AIAN" with about 15.9 percent, followed by "White and Asian" with 12.7 percent, "White and Black" with 11.5 percent, "Black and SOR" with 6.1 percent, "Asian and SOR" with 3.6 percent, and "Black and AIAN" with 2.7 percent. Considering only the combinations of two or more races which exclude "SOR" pairs, the largest combinations were "White and AIAN" with 15.9 percent, "White and Asian" with 12.7 percent, "White and Black" with 11.5 percent, "Black and AIAN" with 2.7 percent, "Asian and NHOPI" with 2.0 percent, "White and NHOPI" with 1.7 percent and "White and Black and AIAN" with 1.6 percent.

Almost half of the 6.8 million two or more race responses reported in Census 2000 involved combinations with SOR, many of which were essentially ethnicities. When we eliminate from consideration two-race combinations involving SOR, we find that about 3.7 million respondents remain in the two or more race category. About 6.4 million reported exactly two races, and that number drops to 3.4 million if we exclude pairs that include SOR. Less than one-half a million respondents (458,000) reported three or more races, and that drops to about 279,000 excluding SOR combinations. The effect is even more dramatic for Hispanics. About 2.2 million Hispanics reported two or more races, but that drops to about 365,000 when SOR combinations are excluded. On the other hand, there were about 4.6 million respondents of two or more races among non-Hispanics, but that drops to 3.3 million excluding SOR combinations of two races. Thus it is clear that the inclusion of SOR responses tends to dramatically increase

the apparent reporting of two or more races.

Some Other Race Reporting

About 15.4 million respondents were classified in the SOR alone category. That number increases to 18.5 million when we add in about 3.2 million SOR responses that were given along with one or more other races (Grieco and Cassidy, 2001, p.10). About 97.0 percent of the SOR alone responses were Hispanic or Latino. Similarly, 90.4 percent of the SOR responses, either alone or in combination with other races, were Hispanic. Thus it is clear that reporting of SOR is highly related to how Hispanics report in race.

Many responses to race are "ethnic" terms. The Census Bureau developed a method which was called the "90 Percent Rule" to reclassify ethnic responses in the race question into an OMB race category. The method is empirically based using 1990 Census sample data as reported and not imputed. Single ancestry responses (which are primarily ethnic responses) were cross-tabulated by race responses. If 90 percent or more of respondents of a specific ancestry group selected a particular race, then that race was assigned to respondents who reported that particular ethnic response in the race question in Census 2000. If less than 90 percent of respondents in 1990 selected any particular race category, then SOR was assigned. This operation was performed during the Census 2000 coding operation. For example, if "German" was entered as a response in race, it was coded as "White," "Jamaican" was coded as "Black or African American," and "Lebanese" as "White." Examples of ethnicities that did not meet the "90 Percent Rule" were "Mexican," "Bermudan," and "Guyanan," which were classified as SOR.

Census 2000 Modified Race

The Census Bureau developed procedures for modifying race data to eliminate SOR responses because SOR is not used by other Federal Agencies in their data collections. The resulting data which do not contain SOR responses are used for population estimates, projections, and survey controls. Other agencies also need these data for denominators for vital rates. The basic procedure used the race as reported if SOR is not involved. If SOR is the only race reported, that response was blanked, and a new one was imputed from other household members if possible, or from nearby households of the same Hispanic origin if not. If SOR appeared in combination with one or more additional races, SOR was blanked and the other race or races were kept. For example, "White and SOR" became "White," and "Black and AIAN and SOR" became "Black and AIAN."

Table 1 shows the effect of the modified race procedure. For the total population SOR disappears while all other

groups increase proportionately except two or more races, which declines. Proportionately more of the responses went to the “White” category. This is not surprising given the previous discussion of the role that SOR plays in two or more races reporting. The effect on the race distribution for the Hispanic population is also not surprising given that most of the SOR reporting is by Hispanics. The “White” category went from about 47.9 percent to 92.1 percent and the two or more races decreased from 6.3 percent to 1.4 percent. All other race categories increased somewhat as well. The impact on the non-Hispanic race distribution is very minor but that is not surprising as very little of the SOR reporting involved non-Hispanics.

Discussion and Conclusion

Census 2000 showed proportionately little reporting of two or more races. But Goldstein and Morning (2000) point out that not all people of mixed racial background are aware of that heritage. And if they are, they still may not always identify with that heritage. Jones and Smith (2002) have shown that a substantial proportion of children who could have reported as two or more races in Census 2000 based on their parents’ race, did not do so. Goldstein and Morning (2000) also point out that identification may vary with questionnaire design, public awareness of the option to report more than one race, and the desirability of reporting more than one race at any given time or place. It also depends on how the single-race groups are defined. It would increase, for example, by splitting one category into two (as was done in the case of the former “Asian and Pacific Islander” category). And as presented above, whether or not the “Some other race” category is used, substantially increased the number of the two or more races population.

In Census 2000 over 90 percent of the combinations reported involved pairs of races rather than the higher order combinations of three, four or more races. The proportion is even smaller if SOR responses are excluded. Virtually all of the SOR responses involve “ethnicities” reported in race that we were not able to assign into an OMB race category. Again virtually all of these involved a Hispanic identifier and most of the SOR respondents were Hispanic. Procedures were developed to eliminate SOR responses for the purposes of estimates and survey controls. The Hispanic population was most affected by the modified race procedure.

Table 1. Distribution of Race and Modified Race by Hispanic or Latino Origin

	Race	Modified Race
Total	100.0	100.0
White	75.1	81.1
Black	12.3	12.7
AIAN	0.9	1.0
Asian	3.6	3.8
NHOPI	0.1	0.2
SOR	5.5	–
Two or more	2.4	1.4
Hispanic or Latino	100.0	100.0
White	47.9	92.1
Black	2.0	3.9
AIAN	1.2	1.6
Asian	0.3	0.7
NHOPI	0.1	0.3
SOR	42.2	–
Two or more	6.3	1.4
Not Hispanic	100.0	100.0
White	79.1	79.5
Black	13.8	13.9
AIAN	0.8	0.9
Asian	4.1	4.2
NHOPI	0.1	0.2
SOR	0.2	–
Two or more	1.9	1.4

II. QUESTIONNAIRE EFFECTS ON REPORTING OF RACE AND HISPANIC ORIGIN

The Census 2000 mail short form questionnaire was different in important ways from the mail questionnaire used in the 1990 Census. Race and Hispanic reporting may be affected by a number of the design changes, including allowing respondents to report more than one race, questionnaire formatting changes, reversing the sequence of race and Hispanic items, wording changes, and elimination of examples in both items.

It is important to understand how the questionnaire differences affect reporting, because they must be taken into account when comparing 1990 and 2000 census data. For example, there was an increase from 3.9% who reported Some other race in 1990 to 5.5% who reported Some other race (alone) in 2000 (U.S. Census Bureau, 2001). What appears to be population change may result from questionnaire differences. By administering both 1990 and 2000-style mail questionnaires during Census 2000, we may attribute reporting differences solely to the effects of the questionnaires; the effect of population change is eliminated.

In order to assess questionnaire effects on reporting, data quality, and comparability, an experiment (the Alternative Questionnaire Experiment, or AQE) was conducted during Census 2000. 1990-style short forms were mailed to a random sample of 10,500 households. The form replicated 1990 question wordings, categories, matrix format, and sequencing. A control panel of about 25,000 households received 2000-style mail short form questionnaires.

About 72% in each panel responded by mail. Results exclude mail nonrespondents enumerated in nonresponse followup and segments of the population (e.g. American Indians on reservations, Alaska Natives) enumerated in other operations.

Data were edited by applying a simplified version of standard Census 2000 pre-edits and coding procedures to data from both forms. Missing data were not imputed or allocated, as they would be in fully edited census data. A content edit followup operation conducted in 1990 to obtain more complete responses from households providing insufficient data was not done in Census 2000 or the experiment. Differences in editing and processing may result in differences between results reported here and 1990 or 2000 census data. The results can support conclusions about questionnaire differences in the *quality and content of response data* they produce, but not about differences in *final data quality*.

Results are weighted to reflect sampling probabilities and are nationally representative of mail back areas. Standard

errors (in parentheses in the tables) and t-statistics are computed using stratified jackknife replication methods (Fay, 1998) that account for sample design and clustering of people within households. Differences denoted (*) are statistically significant at $p < .05$.

For a complete description of study methods and findings, see Martin, 2002.

Results: Hispanic Reporting Table 2 shows that nearly identical fractions of people were reported as Hispanic in 2000-style and 1990-style forms. The fraction reported as not Hispanic is larger in 2000-style forms, and the fraction with missing data is smaller. In past censuses, most people for whom origin was missing were non-Hispanic (McKenney et al., 1993). On this assumption, the Table 2 results suggest that questionnaire changes reduced item nonresponse but did not otherwise affect reporting as Hispanic. The actual effect would depend on how missing data were edited and imputed.

Table 2. Percent of people reported as Hispanic in mail questionnaires in Census 2000 AQE, by form type

Form type	2000-style	1990-style	t
Total persons	100.0%	100.0%	
Hispanic	11.2% (.29)	11.1% (.45)	.05
Not Hispanic	85.5% (.32)	74.4% (.62)	15.8*
Hispanic item blank, uncodable	3.3% (.14)	14.5% (.49)	21.9*
N	40,723	16,616	

It has been hypothesized that dropping examples from the Hispanic origin question in the Census 2000 mail form may have resulted in a loss of detail in Hispanic reporting. To examine this possibility, Hispanic origins were classified into four categories:

1. groups with *check boxes* (Mexican, Puerto Rican, Cuban) in both 1990 and 2000 forms;
2. groups listed as *examples* in the 1990 but not the 2000 form (Argentinian, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard);
3. *all other specific groups* with no check boxes and not listed as examples in either form; and
4. *general descriptors*: “Hispanic,” “Latino,” or “Spanish” were written in, rather than a specific group.

As shown in Table 3, the fraction of Hispanics who checked Mexican, Puerto Rican, or Cuban (or who wrote in one of

these groups) does not differ significantly by form (70% in 2000-style forms compared to 73% in 1990-style forms). Significantly more Hispanics reported in one of the “example groups,” or in one of the remaining non-checkbox, non-example groups, in 1990-style forms, while significantly fewer wrote in a general descriptor.

Table 3. Detailed Hispanic reporting in Census 2000 AQE, by form type

Form type	2000-style	1990-style	t
All Hispanics	100.0%	100.0%	
<i>Check box groups</i>	70.2% (1.25)	73.2% (1.77)	1.4
<i>Example groups</i>	6.4% (.63)	11.2% (1.17)	3.6*
<i>All other detailed Hispanic origins</i>	4.2 (.50)	8.7% (1.23)	3.4*
<i>General descriptor</i>	11.9% (.88)	1.9% (.42)	10.3*
Blank, uncodable write-in	7.2% (.66)	5.0% (.79)	2.2*
N	5,163	3,091	

Thus, the 1990-style form elicited more detailed reports of Hispanic origin than the 2000-style questionnaire for all three categories of Hispanic groups: those with separate check boxes, those listed as examples in 1990 but not 2000, and the remaining groups, with the differences significant for the latter two categories. Overall, about 93% of Hispanics reported a specific group in 1990-style forms, compared with 81% who filled out 2000-style forms. In the latter, Hispanics tended to describe their ethnicity in general rather than specific terms. About 12% gave Hispanic, Latino, or Spanish as their “group,” compared with 2% in 1990-style forms. There were also more uncodable write-in entries in 2000-style questionnaires.

Results: Race Reporting by Hispanics

Questionnaire effects on race reporting were also observed, especially for Hispanics. Table 4 shows that race item nonresponse is much lower in 2000-style forms than in 1990-style forms. (Race is missing if no box is checked and no codable write-in entry is present.)

Table 4. Percentage with missing data on race in Census 2000 AQE, by form type and Hispanic origin

Form type	2000-style	1990-style	t
Total population	3.3% (.16)	6.0% (.33)	7.3*
Hispanics	20.8% (1.14)	30.5% (1.89)	4.4*
Non-Hispanics	.6% (.06)	1.5% (.18)	5.0*
Hispanic origin missing	13.2% (1.39)	9.7% (1.05)	2.0*

Even with reduced race item nonresponse for Hispanics and non-Hispanics in the 2000 form, nonresponse remains very high for Hispanics, who are far more likely to leave race blank than non-Hispanics. Race nonresponse is higher in Census 2000 forms for people who were also missing information on Hispanic origin. (There are many fewer such people in 2000-style forms, as seen in Table 2.)

Table 5 shows form differences in Hispanics’ race reporting. Missing or uncodable responses are excluded, so the distribution approximates the distribution that would be obtained were missing data imputed.

Table 5. Race distributions for Hispanics, by form type

Form type	2000-style	1990-style	t
Total Hispanics	100.0	100.0	
White	49.0 (1.57)	39.9 (2.35)	3.2*
Black	2.1 (.37)	2.3 (.60)	.3
AIAN	1.5 (.38)	.7 (.29)	1.6
Asian	.6 (.22)	.9 (.43)	.6
NHOPI	<.1 (.01)	.2 (.12)	1.1
SOR	39.0 (1.56)	51.5 (2.42)	4.3*
Two or more races	7.8 (.73)	4.6 (.86)	2.9*

Not surprisingly, there is more reporting of two or more races in 2000-style forms, which explicitly allowed this option, than in 1990-style forms, which asked respondents to “fill ONE circle.” In addition, Hispanics’ reporting as

White was higher by about 10 percentage points, and reporting as Some other race was lower by the same amount, in 2000-style forms compared with 1990-style census forms. Prior research (Bates et al., 1995) suggests that these large differences are probably due to the effects of reversing the order of Hispanic and race items, as well as the new “one or more” option.

Conclusions: The experimental evidence shows that, compared to the 1990-style questionnaire, the 2000-style questionnaire resulted in more complete reporting in both the Hispanic origin and race items, indicating better response quality. However, Hispanics who filled out 2000-style mail questionnaires were more likely to report a general descriptor (such as Hispanic, Latino, or Spanish) than those who filled out 1990-style questionnaires. This indicates that the Census 2000 questionnaire design resulted in some loss of detail in Hispanic reporting, probably due in part to elimination of examples. (The design of the experiment does not permit estimation of separate effects of specific design features.) Race reporting by Hispanics is also influenced by the design of the questionnaire, with more Hispanics reporting as White and fewer as Some other race in 2000-style forms.

These questionnaire effects may confound comparisons of 1990 and 2000 census data. The degree of confounding cannot be inferred directly from the analysis reported here, which is restricted to mail short forms and does not employ fully edited data. However, we can infer from the experimental evidence that the differences in the design of 1990 and 2000 mail short forms would have resulted in an increase from the 1990 to the 2000 census in Hispanics’ reporting of White race, and a decline in their reporting of detailed Hispanic groups, *in the absence of true change in the racial or ethnic composition or identifications of the population*. The percentage of Hispanics who reported as White (alone) was 51.7 in 1990 and 47.9 in 2000 (U. S. Census Bureau, 2001). The questionnaire effect would have led more Hispanics to report as White. Therefore, we infer that the decline in White reporting would have been even larger had the 2000-style questionnaire not increased Hispanics’ reporting as White, compared to a 1990-style questionnaire. We can also infer that any measured decline from the 1990 to 2000 census in reporting of detailed Hispanic origins is overstated; the decline would be less if the 2000-style questionnaire had not resulted in less detailed reporting. These confounding effects of questionnaire differences must be taken into account when comparing 1990 and 2000 census data. We plan in future analyses to assess more thoroughly the magnitude of the effects.

References

- Bates, N., Martin, E. A., DeMaio, T. J., and de la Puente, M. (1995) "Questionnaire effects on measurements of race and Hispanic origin," Journal of Official Statistics 11:433-459.
- del Pinal, J., Taguba, L., Cresce, A. and Morning, A. (2001) "Reporting of Two or More Races in the 1999 American Community Survey," Levy Institute Working Paper # 329, <http://www.levy.org/publications/workp3.html>
- Fay, R. E. (1998) VPLX Program Documentation, Vol. 1. Census Bureau.
- Goldstein, J. and Morning, A. (2000) "The Multiple-Race Population of the United States: Issues and Estimates." Proceedings of the National Academy of Sciences 97(11):6230-35.
- Grieco, E., and Cassidy, R. (2001) "Overview of Race and Hispanic Origin," Census 2000 Brief, C2KBR/01-1, Washington, DC: U.S. Census Bureau.
- Jones, N. and Symens Smith, A. (2002) 'Who Is "Multiracial?" Exploring the Complexities and Challenges Associated With Identifying "the" Two or More Races Population in Census 2000.' Paper presented at the annual meetings of the Population Association of America, May 9–11, 2002, Atlanta, GA.
- Martin, E. (2002) Questionnaire Effects on Reporting of Race and Hispanic Origin: Results of a Replication of the 1990 Mail Short Form in Census 2000. Final report of Census 2000 Alternative Questionnaire Experiment. U.S. Census Bureau.
- McKenney, N., Bennett, C, Harrison, R., and del Pinal, J. (1993) "Evaluating Racial and Ethnic Reporting in the 1990 Census." Proceedings of the Am. Stat. Assoc.
- Office of Management and Budget, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity," Federal Register Vol. 62., No. 210, October 30, 1997.
- U.S. Census Bureau (2001) "Population by Race and Hispanic or Latino Origin for the United States: 1990 and 2000." Census 2000 PHC-T-1. Internet release date

NOTES

1. 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. We thank Marie Stetser and Mark Sands for helpful comments.
2. Contact: Jorge del Pinal, Population Division, U.S. Census Bureau, Washington, DC 20233-8800, (301) 763-4875, (301) 457-2644 (fax), email: jorge.h.delpinal@census.gov.