

**Study of Responses to the Census 2000 Race Question Instruction: “Mark One or More Races” and Bridging to Single Race Distributions**

**David Hubble, James Poyer, and Michael Bentley, U.S. Census Bureau.**

**David Hubble, U.S. Census Bureau, Mail Stop 9200, Washington, DC 20233**

**Key Words: Census 2000, Bridging Methods, Race Distributions**

**Introduction**

The Office of Management and Budget (OMB) issued revised standards in 1997 on the collection and presentation of data on race and ethnicity (OMB, 1997). OMB designed these standards partly to measure the increased diversity of the American population. Census 2000 was the first Decennial Census to ask respondents to “mark one or more races”. Data users should understand how Census 2000 race data relates to race distributions from previous censuses, current surveys, and administrative records where respondents were instructed to report only one race. In response to this need, we conducted a study designed to produce a datafile from the Census Quality Survey (CQS) for users to create “bridging” parameters between the two race data collection methodologies. The main purpose of this paper is to serve as an introduction to the CQS and its purpose. After providing the background, this paper presents the CQS methodology and preliminary analyses. It concludes with a discussion on future work.

**Background**

Consistent with previous studies, including the 1995 Race and Ethnicity Supplement to the Current Population Survey (Tucker, et al., 1996), 1996 National Content Survey and the 1996 Race and Ethnic Targeted Test (U.S. Census Bureau, 1996 and 1997, respectively), Census 2000 estimated that 2.4% of the total population reported two or more races and that the proportion varied widely by race (see Table 1).

These results present some real challenges to users of Census 2000 data. For many Federal and non-Federal programs Census 2000 data are used to measure change over time. Without further guidance, these programs would be hard pressed to disentangle real changes in population, economic, social, and health conditions by race from methodological changes resulting from using the new standards for collecting data on race. So, to permit meaningful comparisons of data collected under the old standards with data collected under the new

standards, users may need procedures for bridging the two.

Table 1  
Percent Reporting Two or More Races  
by Race in Census 2000 (Jones and Smith, 2001)

Race	Alone or in Combination	Alone	In combination	Percent in combination
White	216,930,975	211,460,626	5,470,349	2.5%
Black	36,419,434	34,658,190	1,761,244	4.8%
AI/AN	4,119,301	2,475,956	1,643,345	39.9%
Asian	11,898,828	10,242,998	1,655,830	13.9%
NH/OPI	874,414	398,835	475,579	54.4%
SOR	18,521,486	15,359,073	3,162,413	17.1%

The following abbreviations for racial groups will be used in tables throughout this paper: Black for Black or African American, AI/AN for American Indian and Alaska Native, NH/OPI for Native Hawaiian and Other Pacific Islander, and SOR for Some other race.

This is especially true for those populations that are more sensitive to changes in the standards for collecting race data: American Indian/Alaska Native and Native Hawaiian/Other Pacific Islander.

In order to address this and other issues and to ensure that tabulation methodologies would be developed and coordinated among the Federal agencies, OMB assembled a group of statistical and policy analysts drawn from the Federal agencies that generate or use these data. From this effort, the OMB issued “Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity” (OMB, 2001). Specifically, this report provides guidelines for bridging race data collected

NOTE: This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review by the Census Bureau than its official publications. This report is released to inform interested parties and to encourage discussion.

using the new 1997 standards (mark one or more races) to the old 1977 standards (mark one race).

This OMB report uses data from the National Health Interview Survey and the Current Population Survey to demonstrate how several probabilistic whole and deterministic fractional assignment methods affect race distributions on several outcome measures.

However, a significant drawback to using these surveys is that there are several methodological differences in how they were conducted relative to how Census 2000 was conducted. These differences include race question wording, mode of data collection, processing and editing of the data, and time of data collection. It seems quite plausible that the computed fractions or probabilities could be seriously biased by these differences. So, in response to these concerns, we conducted a study in 2001 in which we collected race data using both the “mark one or more races” instruction and the “mark one race” instruction for the same persons emulating Census 2000 procedures, as much as possible.

## Methodology

### Sample Design and Selection

For the Census Quality Survey (CQS) study, we selected a stratified random sample of approximately 55,000 households in the spring of 2001. In order to make the sample design more efficient, we needed to select from a frame that would allow us to concentrate our sample more on persons who reported more than one race. So, households from the Census 2000 Hundred Percent Edited Detail File (HEDF) containing at least one person reporting more than one race were oversampled to allow us to produce more reliable bridging parameters. We also stratified at the state level to ensure sufficient sample size for producing sub-national geographic bridging parameters (U.S. Census Bureau, 2002).

### CQS Initial Contact

To increase the utility of the study, two treatment panels were developed with half the sample in each panel. For both panels, an Initial Contact data collection occurs. The Initial Contact attempts to emulate Census 2000 data collection procedures by using a mailout/mailback questionnaire followed by a personal visit non-response follow up (NRFU) operation for those who did not mail the form back. One respondent was asked to provide information for all persons living in the household and the name of the person who had completed the Census 2000 form. The only difference in the panels is that Panel A

used the “mark one or more races” instruction, while Panel B used the “mark one race” instruction in the Initial Contact (see Table 2). The Initial Contact was conducted in the Summer of 2001 with a response rate of 97%.

### CQS Recontact

Both panels are then recontacted to collect additional information, but more importantly to be re-asked the race question with the opposite instruction they had received in the Initial Contact. In the Recontact, every effort is made to interview the same Initial Contact household respondent. The Recontact used telephone interviews when possible, and personal visits otherwise. Note, a proportion of the population will report two or more races even when instructed to report only one race. However, because the Recontact is collected by enumerators, we added a special follow up question for those respondents that reported two or more races in Panel A that inquires as to how they typically respond in situations where they are asked to provide a single race. The Recontact was conducted in the Fall of 2001 with a response rate of 92%.

### Census 2000 Matches

To complete the data collection, we computer matched sampled persons from both panels to the Census 2000 person-level file. The probabilistic matching algorithm used several variables as matching criteria, including self-reported name, sex, age, date of birth, and address (Berning, 2002). The matching operation matched approximately 85% of sampled initial contact persons to the Census 2000 file.

### Source of Bridging Data

This two panel design was used to allow the datafile users to create alternative sets of bridging parameters. The first set (Panel A) uses the CQS Initial Contact “mark one or more races” data and the Recontact “mark one race” data at the person level to estimate bridging parameters (in bold in Table 2). The second set (Panel B) of bridging parameters uses the Census 2000 “mark one or more races” data and the CQS Initial Contact “mark one race” data to estimate bridging parameters.

There are potential advantages to both panels in computing bridging parameters. For Panel A, the advantages include that the data is not constrained by the ability to match sample persons to Census 2000 and the greater likelihood that the same person was the household respondent for the “mark one or more” and “mark one”

Table 2  
 Census Quality Survey (CQS) Data Collection Sequence:  
 Race Instruction by Panel

	Data Collection Contact		
CQS Panel	Census 2000 Matches	CQS Initial Contact (Mailback & NRFU: June-Aug. 2001)	CQS Recontact (Phone & Visit: Aug. - Oct. 2001)
A	Mark one or more races	<b>Mark one or more races</b>	<b>Mark one race</b>
B	<b>Mark one or more races</b>	<b>Mark one race</b>	Mark one or more races

NOTE: Boldface type indicates the data intended to be used to estimate bridging parameters for each panel. Other combinations are possible though.

race questions. For Panel B, the advantages include that the “mark one or more” race distribution is the Census 2000 data itself and that the mode of data collection for “mark one” race question replicated Census 2000 procedures. Panel B may be best suited for bridging between the 1990 and 2000 Censuses, while Panel A may be better suited for bridging Census 2000 to current surveys and administrative records, especially when using the “choose only one” followup question results to compute parameters.

Note, bridging parameter estimates can be computed using combinations of data other than those in bold in Table 2. In fact, such analyses may be useful in more fully understanding the underlying nature of the CQS data.

**Analyses**

The implementation of the above described design has resulted in an enormously rich set of data. Though our exploration is still preliminary in nature, we see these data as being very important in understanding the reporting of race and ethnicity in Census 2000 and beyond. While the focus has been on serving the purpose of producing bridging parameters, these data will also be extremely useful in researching several aspects of the reporting of race and ethnicity. This paper summarizes a few topics analyzed by Census Bureau staff for this study with the hope of illustrating the study’s potential. These results and data processing are preliminary and should be not interpreted as completed work.

All variances were estimated using the Variance Estimation for Complex Samples (VPLX) software developed by the Census Bureau (Fay, 1998). Mailout stratification and household level reporting were incorporated into variance estimates by using the mailout strata and household level clusters. Mailout sampling intervals were used to create sample weights, and were not adjusted (as of this date) to compensate for differential nonresponse. All estimates used initial mailout weights.

In this paper, we restrict our analysis to the non-Hispanic population, since our background analysis indicates that the bridging parameters differ between Hispanics and non-Hispanics. This is supported by previous Censuses, which indicate that about half of Hispanics report their ethnic origin as their race, which is recoded to “some other race”. This reporting pattern also occurred among Hispanics reporting more than one race. Also, Hispanics were more likely to report two or more races than non-Hispanics (6.3% vs. 1.9%). Consequently, we expect that the bridging parameters for the Hispanic population will be quite different, relative to the non-Hispanic population.

As initial analysis steps, we assess the representativeness of our sample in a couple of ways. First, we compare the race distribution of CQS respondents that are matched to Census 2000 persons by panel to each panel’s entire CQS sample. Next, we compare CQS data by panel using the “mark one more races” instruction to Census 2000 responses for CQS respondents “matched” to Census 2000.

Finally, for illustrative purposes, we discuss the potential use of bridging parameters to measure the change in population counts from the 1990 Census to Census 2000 by race.

Representativeness of Matched Persons

Each CQS Panel’s race distribution for persons matched to Census 2000 appears to be representative of the race reporting for each panel. Table 3 reveals that the “matched person” subpopulation does not differ from the corresponding panel’s reporting for all respondents. In particular, the percentage reporting two or more races is unaffected when restricted to “matched” persons (2.0% for Panel A and 1.0% for Panel B). Although non-Hispanic Blacks, Asians, and American Indians and Alaska Natives are less likely to be matched than person reporting other races, this difference is not significant at the p<.10 level. Thus, with the representativeness of matched persons established, we restrict further analysis in this paper to CQS matched persons.

Table 3  
Percent Distribution of All Census Quality Survey (CQS) Respondents and Respondents Matched to Census 2000 by Race - CQS Initial Contact by CQS Panel Non-Hispanics For the U.S.

Race	CQS Panel A “mark one or more races” instruction		CQS Panel B “mark one race” instruction	
	All Persons (std. error)	Persons Matched to Census 2000 (std. error)	All Persons (std. error)	Persons Matched to Census 2000 (std. error)
White Alone	80.6% (1.18%)	81.9% (1.16%)	80.1% (1.23%)	81.3% (1.26%)
Black Alone	11.2% (0.99%)	10.4% (0.96%)	12.7% (1.13%)	12.1% (1.16%)
AI/AN Alone	0.7% (0.16%)	0.6% (0.14%)	0.9% (0.18%)	0.8% (0.17%)
Asian Alone	4.4% (0.63%)	3.8% (0.61%)	3.7% (0.49%)	3.4% (0.49%)
NH/OPI Alone	<0.1% (0.02%)	<0.1% (0.02%)	0.1% (0.03%)	0.1% (0.03%)
SOR Alone	0.4% (0.15%)	0.4% (0.14%)	0.9% (0.27%)	0.8% (0.26%)
Two or More Races	2.0% (0.30%)	2.0% (0.33%)	1.0% (0.18%)	1.0% (0.20%)
Missing or not able to code	0.7% (0.18%)	0.7% (0.19%)	0.5% (0.14%)	0.5% (0.14%)

Representativeness of the CQS to Census 2000

Each CQS panel appears to be representative of Census 2000. Table 4 indicates that aggregated race reporting among CQS respondents to the “mark one or more races” instruction closely resembles Census 2000 race reporting for each panel. Note, the “mark one or more races” data collection in the Initial Contact for Panel A and the Recontact for Panel B. No race group appears to be significantly different than Census 2000 (p<.10 level) in either panel, including the Two or More Races category.

A secondary, but equally interesting interpretation of this result is that, apparently, the differences in the race question sequence and the data collection modes between the two CQS panels (as reflected in Table 2) did not have a significant impact on the reported race distribution.

Table 4  
Percent Distribution of Census Quality Survey (CQS) Respondents Matched to Census 2000 by Race - Census 2000 and CQS “Mark One or More Races” Instruction Data Collection Contact by CQS Panel - Non-Hispanics For the U.S.

Race	CQS Panel A		CQS Panel B	
	Census 2000 (std. error)	CQS Initial Contact (std. error)	Census 2000 (std. error)	CQS Recont. (std. error)
White Alone	83.1 % (1.13%)	81.9% (1.16%)	81.8% (1.25%)	80.7% (1.35%)
Black Alone	10.4% (0.98%)	10.4% (0.96%)	12.1% (1.17%)	12.0% (1.21%)
AI/AN Alone	0.6% (0.20%)	0.6% (0.14%)	0.7% (0.17%)	0.7% (0.17%)
Asian Alone	3.7% (0.61%)	3.8% (0.61%)	3.4% (0.52%)	4.0% (0.69%)
NH/OPI Alone	<0.1% (0.01%)	<0.1% (0.02%)	<0.1% (0.03%)	<0.1% (0.03%)
SOR Alone	0.3% (0.11%)	0.4% (0.14%)	0.2% (0.14%)	0.3% (0.09%)
Two or More Races	1.8% (0.18%)	2.0% (0.33%)	1.6% (0.15%)	1.7% (0.24%)
Missing or not able to code	-0-	0.7% (0.19%)	-0-	0.5% (0.15%)

Illustration of Potential Use of Bridging Parameters

Recall, the most important criteria for choosing a bridging method is its ability to match how the respondent would have responded under the old standards had that been possible. A corollary to this is that it may also be equally important to choose a data processing method that matches how the data were processed under the old standards. We will use the 1990 Census to illustrate.

Even though the respondents were instructed to “mark one” race in the 1990 Census, a small percent marked more than one race. This same phenomenon can be seen in Table 3, where 1.0% of the respondents in Panel B marked two or more races, even though instructed to “mark one” race. In 1990, this issue was handled in the processing of the data by taking the first race marked on the form. That is, a response of marking both White and Black was processed as White alone.

So, in the case of computing appropriate Census 2000 bridging parameters to the 1990 Census the first step could be to process the two or more races responses from Panel B using a 1990 Census style race edit. Taking these results, along with those multiple race respondents who did provide a single race response in Panel B, we could compute parameters specific to the task of bridging the Census 2000 population counts to 1990 Census population counts by race.

Without bridging parameters, we can only guess at the percent change for a specific race between 1990 and 2000 (see Table 5). The Race Alone column is the low end estimate and Race Alone or in Combination is the high end estimate (U.S. Census Bureau, 2001). Both of these change estimates are flawed in their ability to measure real change because they reflect both real change and methodological changes. By applying appropriate bridging parameters to the population that reported two or more races in Census 2000, we can produce percent change from 1990 population estimates by race that minimize the Census 2000 methodological differences and more truly reflect the real change.

Table 5  
Percent Change of Population from 1990 to 2000  
for Census 2000 Race Alone and Race Alone or in  
Combination -  
Non-Hispanics For the U.S.

Race	1990 Census	2000 Census Race Alone	2000 Census Race Alone or in Combination
	Number	Number & % Change From 1990	Number & % Change From 1990
White	188,128,296	194,552,774 3.4%	198,177,900 5.3%
Black	29,216,293	33,947,837 16.2%	35,383,751 21.1%
AI/AN	1,793,773	2,068,883 15.3%	3,444,700 92.0%
Asian	6,642,481	10,123,169 52.4%	11,579,494 74.3%
NH/OPI	325,878	353,509 8.5%	748,149 129.6%
SOR	249,093	467,770 87.8%	1,770,645 610.8%

### Future Work

As stated earlier, the CQS study has provided an extremely rich set of data. Much work has yet to be done to fully exploit its potential.

#### Developing Bridging Parameters

In issuing the revised standards in 1997, the OMB Federal Register Notice provided general guidance on the tabulating of race data, but beyond that it outlined four areas where further research was needed on how to tabulate race data (OMB, 1997). The following four questions were raised:

1. How should the data be used to evaluate conformance with program objectives in the area of equal employment opportunity and other anti-discrimination programs?
2. How should the decennial census data for many small population groups with multiple heritages be used to develop sample designs and survey controls for major demographic surveys?
3. How should the 1997 standards be introduced in the vital statistics program which obtains the number of births and deaths from administrative records, but uses intercensal population estimates in determining the rates of births and deaths?
4. And more generally, how can meaningful comparisons be made of data collected under the previous standards and data that will be collected under the 1997 standards?

In thinking about the possible answers to the above questions and in developing methods to compute the appropriate bridging parameters for measuring 1990 to 2000 population growth rates by race, as illustrated in the analyses section above, it has become clear that there may be no single best bridging technique. The best method for computing parameters from the CQS study dataset to bridge from Census 2000 multiple race data to a given single race dataset is directly related to how the single race data was collected and processed.

#### Research the Reporting of Race

The CQS study with its split-panel design, Initial and Recontact interviewing stages, mixed-mode data collections, and matching back to Census 2000 responses, provides us with the capability of researching many aspects of the process by which respondents answer questions about their race or races.

One area of particular interest is the relatively high level of inconsistency in the reporting of race for matched people from Census 2000 to the CQS. The inconsistency was especially pronounced in the reporting of two or more races. Table 6 shows that less than half of two or more races matched people in Census 2000 reported two or more races in the CQS. Even though the total race distributions are comparable, this reporting inconsistency reduced the CQS sample size of two or more races people significantly. Further analyses are needed to understand the source of this inconsistency.

Table 6  
Number of Reported Races for Census 2000 and  
Panel A Initial Contact Matched Persons -  
Non-Hispanics For the U.S.

Census 2000	CQS Panel A		
	One Race	2+ Races	Total
One Race	*96.9% (n=35,320)	1.3% (n=1,995)	98.2% (n=37,315)
2+ Races	1.1% (n=9,228)	0.7% (n=8,050)	1.8% (n=17,278)
Total	98.0% (n=44,548)	2.0% (n=10,045)	

\*Percent estimates reflect application of initial weights

**Limitations**

Complete weighting procedures have not been developed or applied yet. The two main issues here relate to the need to adjust for nonresponse and the possible need to trim large weights. (Note, the large weights were brought about by the need to use highly differential sampling rates in order to oversample the relatively rare population of households containing at least one person reporting two or more races.)

**References**

Berning, Mike. (2002) "Matching Process for the Census Quality survey to Evaluate Responses to the Census 2000 Questions on Race". U.S. Census bureau internal document.

Fay, Robert E. (1998) "VPLX: Program Documentation". U.S. Bureau of the Census internet site.

Harrison, Roderick J. (2000) "Inadequacies of Multiple Response Race Data in the Federal Statistical System". Paper Presented at the Levy Institute Conference on

Multiraciality: How Will the New Census Data Be Used?

Harris, David R. and Justin L. Thomas. (2001) "The Educational Costs of Being Multiracial: Evidence from a National Survey of Adolescents." Presented at the Annual Meetings of the Population Association of America.

Jones, Nicholas and Amy Symens Smith. (2001) "The Two or More Races Population: 2000 - Census 2000 Brief". U.S. Census Bureau.

Office of Management and Budget. (1997) "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity". OMB Revised Statistical Policy Directive No. 15.

Office of Management and Budget. (2001) "Federal Register Notice: Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity". Federal Register Vol. 66, No. 10, pp. 3829-31 (with three appendices).

Tucker, Clyde, Ruth McKay, Brian Kojetin, Roderick Harrison, Manuel de la Puente, Linda Stinson, and Ed Robison. (1996) "Testing Methods of Collecting Racial and Ethnic Information: Results of the Current Population Supplement on Race and Ethnicity". Bureau of Labor Statistics, BLS Statistical Notes No. 40.

U.S. Census Bureau. (1996) "Findings on Questions on Race and Hispanic Origin Tested in the 1996 National Content Survey". Population Division Working Paper No. 16.

U.S. Census Bureau. (1997) "Results of the 1996 Race and Ethnic Targeted Test". Population Division Working Paper No. 18.

U.S. Census Bureau. (2001a) "Study Plan - The Census Quality - Evaluation of Responses to the Race Question". U.S. Census Bureau internal document.

U.S. Census Bureau. (2001b) "Table 4. Difference in Population by Race and Hispanic or Latino Origin, for the United states: 1990 to 2000." [www.census.gov/population/cen2000/phc-t1/tab04.pdf](http://www.census.gov/population/cen2000/phc-t1/tab04.pdf). Internet release date: April 2, 2001.

U.S. Census Bureau. (2002) "Program Master Plan - Census Quality Survey to Evaluate Responses to the Census 2000 Question on Race". Internal document.