Bridging Between Two Standards for Collecting Information on Race and Ethnicity: An Application to Census 2000 and Vital Rates

JENNIFER D. PARKER, PHD[a]
NATHANIEL SCHENKER, PHD[b]
DEBORAH D. INGRAM, PHD[a]
JAMES A. WEED, PHD[c]
KATHERINE E. HECK, MPH[d]
JENNIFER H. MADANS, PHD[e]

SYNOPSIS

Objectives. The 2000 Census, which provides denominators used in calculating vital statistics and other rates, allowed multiple-race responses. Many other data systems that provide numerators used in calculating rates collect only single-race data. Bridging is needed to make the numerators and denominators comparable. This report describes and evaluates the method used by the National Center for Health Statistics to bridge multiple-race responses obtained from Census 2000 to single-race categories, creating single-race population estimates that are available to the public.

Methods. The authors fitted logistic regression models to multiple-race data from the National Health Interview Survey (NHIS) for 1997–2000. These fitted models, and two bridging methods previously suggested by the Office of Management and Budget, were applied to the public-use Census Modified Race Data Summary file to create single-race population estimates for the U.S. The authors also compared death rates for single-race groups calculated using these three approaches.

Results. Parameter estimates differed between the NHIS models for the multiple-race groups. For example, as the percentage of multiple-race respondents in a county increased, the likelihood of stating black as a primary race increased among black/white respondents but decreased among American Indian or Alaska Native/black respondents. The inclusion of county-level contextual variables in the regression models as well as the underlying demographic differences across states led to variation in allocation percentages; for example, the allocation of black/white respondents to single-race white ranged from nearly zero to more than 50% across states. Death rates calculated using bridging via the NHIS models were similar to those calculated using other methods, except for the American Indian/Alaska Native group, which included a large proportion of multiple-race reporters.

Conclusion. Many data systems do not currently allow multiple-race reporting. When such data systems are used with Census counts to produce race-specific rates, bridging methods that incorporate geographic and demographic factors may lead to better rates than methods that do not consider such factors.

[a]Office of Analysis, Epidemiology, and Health Promotion, National Center for Health Statistics, Hyattsville, MD

[b]Office of Research and Methodology, National Center for Health Statistics, Hyattsville, MD

[c]Division of Vital Statistics, National Center for Health Statistics, Hyattsville, MD

[d]Department of Human and Community Development, University of California at Davis, Davis, CA

[e]Office of the Director, National Center for Health Statistics, Hyattsville, MD

Address correspondence to: Jennifer D. Parker, PhD, Population Epidemiology Branch, NCHS, 3311 Toledo Rd., Rm. 6415, Hyattsville, MD 20782; tel. 301-458-4419; fax 301-458-4036; e-mail <jdparker@cdc.gov>.

In 1997, the Office of Management and Budget (OMB) issued revised standards for the collection and reporting of "race" and "ethnicity" data within the federal statistical system.[1] The 1997 standards replaced OMB Directive 15, the 1977 Directive enacted to increase the comparability of federal data.[2] OMB Directive 15 required federal agencies to report race-specific tabulations using four single-race categories: American Indian or Alaska Native (AIAN), Asian or Pacific Islander (API), black, and white. Hispanic ethnicity was to be collected and reported separately, when possible. Among other revisions, the 1997 standards allow the selection of more than one racial group to describe a person, and separate the API group into two groups (Asian, Native Hawaiian or other Pacific Islander); Hispanic origin is still collected separately. Full implementation of the 1997 standards was mandated for federal data systems by 2003, and some systems implemented the standards earlier. Census 2000, for example, collected information on race and ethnicity using the 1997 standards.[3] However, the timeline for updating other data systems varies. In addition, the standards apply only to federal data collection systems; thus, state-based systems, including vital statistics systems, are not required to implement the 1997 standards, although they may do so to maintain comparability with other data systems.

The differential adaptation of the 1997 standards by states poses several challenges for the calculation of vital statistics such as death and birth rates.[4–6] These annually reported statistics are calculated by dividing the number of events obtained from vital records by the population at risk obtained from the Census Bureau's resident population files.[5,6] While Census 2000 allowed respondents to select more than one racial category, vital records are still being collected in most states using single-race formats. As a result, numerators and denominators for race-specific rates are often incompatible. Additional challenges as states implement the revised standards at different times include difficulties in calculating trend statistics as well as difficulties in making inter-state comparisons. Furthermore, these issues exacerbate the existing limitations of race information.[7–9] The racial categories used by the federal government are social-political constructs that have changed historically in response to social attitudes and political concerns;[1] as such, their meaning, and consequently race-specific rates, must be considered within our social context.[10] Despite their problems, however, race-specific rates highlight disparities in health and access to care that are not yet fully understood using other social indicators.

To address the need to combine data collected under OMB Directive 15 and the 1997 standards during the transition period, the OMB issued *Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity*,[10] which includes a detailed discussion of "bridging" methods. These methods assign multiple-race responders to single-race categories by assigning some proportion of those reporting a specific combination of races to each of the races defining the group. The assignment is done so as to approximate in the aggregate how the individuals in this group would have responded had they been asked to report only a single-race. It is important to note that, for population estimates, the goal of bridging is to approximate the size of the single-race groups rather than to approximate how each individual would have responded to the traditional single-race question. Different bridging methods make different assumptions about the relationship between responses to multiple-race and single-race questions. One method that is often proposed, equal allocation, allocates equal proportions of each multiple-race group to its component single races.

Bridging methods that utilize available data on how individuals who report multiple races respond to a single-race question when given the opportunity are likely to be more accurate (i.e., to more closely approximate self-reported race/ethnicity) than those that a priori assign all multiple-race responders to one of the corresponding single-race groups or assign a set proportion, by equal allocation for example, to each of the single-race groups. Some of the bridging methods described in the OMB Guidance are based on data from the National Health Interview Survey (NHIS), a household survey conducted annually by the National Center for Health Statistics (NCHS), which has allowed multiple-race responses for all respondents since 1982. Prior to 1982, limited multiple-race reporting had been allowed since 1976. NHIS respondents who report more than one race are asked, in a follow-up question, to select the group that best represents their race. Until recently, this single-race response has been the only one retained for public-use files and publications.[11] Under the assumption that the response to the follow-up question is closely related to the response that would have been given if only a single-race question had been asked, NHIS data can be used to develop more realistic bridging strategies than those not based on existing data. An NHIS bridging method described in the OMB guidance, deterministic NHIS-fractions, uses the distribution of responses to the follow-up question for the total sample as allocation proportions.[10]

While the NHIS-fractions strategy is useful, it is possible to improve on the method by incorporating characteristics that are associated with the probability of reporting a particular single-race category. Schenker and Parker[12] extended the method by using logistic regression models to predict primary race (i.e., the single race group that multiple-race respondents reported when asked to select a "best" race) as a function of several covariates available in NHIS data and several contextual variables for the county of residence; their method will be referred to as the NHIS-regression method. They concluded that bridging methods that include individual and contextual predictors can lead to better predictions and better variance estimates. They also found that relationships between primary race and predictors differed among race groups and changed between time periods. Because vital statistics are calculated for groups defined by Hispanic origin, age, sex, and geographic unit (county, state, and national), the population estimates used as denominators can more closely approximate the underlying distributions if they are based on models that incorporate allocation proportions that vary according to individual and contextual predictors.

The objectives of this article are to announce the creation of a file of bridged population estimates from the 2000 Census and to summarize and evaluate the NHIS-regression method that was used by NCHS to create these estimates. The bridged file is based on the Census Bureau's in-house version of the Census 2000 Modified Race Data Summary file (MR file),[13] which is a county-level file with population counts by sex, Hispanic origin, age categories, and 31 single- and multiple-race groups. To create the MR file, the Bureau of the Census imputes respondents with missing or non-standard data into one of the standard categories. Using bridging proportions developed at NCHS, the Bureau created a bridged version of the MR file, the Bridged MR file,[14,15] by bridging the multiple-race counts in the Bureau's in-house version of the MR file to single-race groups. The Bridged MR counts have been and will be used as denominators for calculating rates in a number of data systems. In particular, they can be used to calculate vital rates in which the numerators are derived from records obtained under a single-race reporting system. Single-race population counts for 2001 and subsequent years will be estimated using this methodology until bridging is no longer necessary. The initial bridging models were tailored toward the specific task of estimating denominators for 2000. These models will be modified as new information becomes available. The Bridged MR file is available on the NCHS website.[14]

For the present study, to evaluate the method used to create the Bridged MR file and to demonstrate the use of bridged Census data, we used the NHIS-regression method to calculate single-race population estimates, bridging a public-use version of the MR file, and compared these estimates with those obtained by bridging the same file using NHIS-fractions and equal allocation. We could not use the Bridged MR file to generate these estimates because the file does not include the original multiple-race counts. The only difference between the Bureau's in-house MR file used to create the Bridged MR file available on the NCHS website and the public-use version of the MR file used here is that the former contains counts by single-year of age whereas the latter contains counts by broader age groups 0, 1–4, 5–9, 10–14, . . ., 80–84, and ≥85 for purposes of confidentiality. We used the three sets of single-race population estimates obtained by bridging the public-use version of the MR file as denominators to calculate death rates for the United States, using as numerators counts of deaths for 2000 collected under the single-race reporting system. We compared these death rates to determine whether the three bridging methods yielded very different rates.

METHODS

NHIS

The NHIS is a continuous household survey designed to measure the health status of residents of the United States.[16–18] Data from the 1997 to 2000 surveys were used in developing the logistic regression models in this study. Each year about 40,000 households are included in the sample, covering about 100,000 possible respondents. As in the decennial Census, information for some household adults and children younger than 18 years of age is provided by proxy in the NHIS. However, for ease of presentation, we refer to all individuals included in the survey as respondents. Since 1982, approximately 1% to 2% of survey respondents have reported more than one race each year.[11]

The NHIS allows respondents to choose races from a list of 15 race categories (white, black, American Indian, Aleut, Eskimo, Chinese, Filipino, Hawaiian, Korean, Vietnamese, Japanese, Asian Indian, Samoan, Guamanian, and other Asian or Pacific Islander) that can be collapsed into the four single-race categories specified in OMB Directive 15, i.e., AIAN, API, black, and white. Tabulations from the survey include an additional category, other race, for respondents who mention a race group not included on the list. Multiple-race responses are available on in-house versions of NCHS files, but are suppressed on public-use versions for confidentiality. In fitting the bridging models described below, if a multiple-race response included other race, we dropped the other race response, consistent with the Census Bureau's approach to the creation of the MR file. For example, respondents who reported black and other race were included in the single-race black group; respondents who reported AIAN, API, and other race were included in bridging models for the AIAN/API group.

From 1997 through 2000, 4,898 NHIS respondents reported more than one race. The four single-race groups in the 1977 standard imply 11 multiple-race groups: AIAN/API, AIAN/black, AIAN/white, API/black, API/white, black/white, AIAN/API/black, AIAN/API/white, AIAN/black/white, API/black/white, and AIAN/API/black/white. The likelihood of providing a primary-race response differed among race groups (Table 1). Since our task was to predict one of the four single-race categories, only the 3,956 multiple-race respondents who reported one of the single-race groups as their primary race in the follow-up question were included in fitting the bridging models described below; the results of a previous analysis that included "no primary race" as a possible outcome are reported in Schenker and Parker.[12]

Regression models

We fitted logistic regression models to multiple-race NHIS data for 1997–2000 to obtain estimated probabilities for each multiple-race group of reporting each of the possible primary-race categories (e.g., for the black/white group, the probability of reporting black as primary race and the probability of reporting white as primary race). These probabilities were used to calculate size estimates for the U.S. population of the four single-race groups defined by OMB Directive 15. A separate logistic regression model was developed for each two-race group with more than 100 respondents in the NHIS: black/white, AIAN/white, API/white, black/AIAN, and black/API. For the AIAN/black/white group, which also had more than 100 respondents, a multi-logit model, which allowed more than two outcomes, was fitted.

The other multiple-race groups had too few respondents to support the fitting of separate models to the NHIS data. Predictions for these groups were derived from a combined multi-logit model fitted using all multiple-race respondents. The motivation behind this approach is that information about the associations between primary race (as reported on the follow-up question) and selected covariates for the smaller race groups can be approximated using the associations for the larger groups. Although previous evidence[12] and the results shown below suggest that a separate model would be preferable for each multiple-race group, the combined model was considered reasonable, given the data constraints. With the goal of

Table 1. Sample sizes and weighted percent distribution of single-race responses selected by multiple-race respondents, NHIS, 1997–2000

		Single-race response				
Multiple-race response	n	*AIAN*	*API*	*Black*	*White*	*None*
AIAN/API	27	—	—	—	—	—
AIAN/black	393	13.3	—	78.7	—	8.0
AIAN/white	1,593	21.2	—	—	74.0	4.8
API/black	130	—	33.8	51.0	—	15.2
API/white	1,147	—	39.6	—	41.2	19.2
Black/white	1,138	—	—	45.4	26.9	27.7
AIAN/API/black	12	—	—	—	—	—
AIAN/API/white	70	1.4	54.5	—	35.0	9.1
AIAN/black/white	346	6.9	—	27.6	8.5	57.0
API/black/white	38	—	—	—	—	—
AIAN/API/black/white	4	—	—	—	—	—

NOTES: Percentages were calculated using NHIS survey weights. The primary race distribution is not shown for multiple-race groups with fewer than 50 respondents.

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

NHIS = National Health Interview Survey

balancing race detail with an estimable model, we considered several ways of representing the multiple-race groups in the multi-logit model. We decided to include three indicator variables to describe the multiple-race groups: not black, not AIAN, and not API. For the multi-logit model, the coefficients for the indicator variables were constrained to zero for the corresponding primary race outcomes; for example, the parameter estimate for the variable "not black" was constrained to zero for the primary race outcome black. Although this representation of the multiple-race groups does not separately identify respondents who selected white as one of their race groups from those who did not, 87% of multiple-race respondents in the NHIS data selected white as one of their race groups. Hence, the bias associated with this lack of identifiability is likely to be small.

Covariates

Since the results of fitting the bridging models were to be applied to the MR file, the few individual factors available in both the MR file and the NHIS were included in the models: single year of age; Hispanic origin, yes or no; and sex.

County of residence is also available on in-house versions of NHIS data files and the MR file, enabling us to include geographic and contextual variables in the models. For each respondent, we added the region of the country (Northeast, Midwest, South, or West) and a county-specific index of urbanicity.[19] Using the data provided on the MR file, we also created four contextual variables for each county by calculating the percentages of the county's population who reported single-race AIAN, API, or black, and the percentage who reported more than one race. These variables were used as indicators of a county's racial composition.

The covariates were included in all models to make the models comparable across multiple-race groups. We included age in years in each model as a continuous variable. Single-race population percentage variables were included in models when appropriate. Different forms of these percentages were used as appropriate. Percent single-race black was included in the model for black/white respondents, for example, but not in the model for AIAN/white respondents. Percent multiple-race was included in each model, whereas percent single-race white was not included in any model. For AIAN/black and black/white respondents, the square of percent single-race black improved the model fit, indicating that the probability of a respondent's choosing black as a primary race (i.e., identifying as black in response to the follow-up question) increases relatively rapidly as the percentage of county residents who reported black as their only race increased. The logarithm of percent single-race AIAN improved the fit for the AIAN/black and AIAN/white respondents, indicating that the likelihood of reporting AIAN as a primary race increases slowly as the percentage of single-race AIAN reporters in the county increases. For the combined model, the percentage variables for black and API were included unaltered and the logarithm of percent single-race AIAN was used.

Stata software, incorporating the complex sample design and the survey weights of the NHIS, was used to fit the models.[16,20] Although we do not know whether use of the survey weights is appropriate for this application of the NHIS, standard analytic practice is to incorporate the survey weights and we had no evidence to suggest we should do otherwise.

Bridged population estimates

Single-race probabilities (i.e., the probability of identifying a particular race in response to the single-race question) were generated from the fitted NHIS models for each county, multiple-race group, and combination of individual variables. For ages ≥70 years, the corresponding probabilities for age 69 were assigned. Probabilities for the six largest multiple-race groups were obtained from the corresponding race-specific models; those for the smaller multiple-race groups were obtained from the combined model. Since the combined model produces primary-race probabilities for all four primary-race groups, the probabilities from the composite model were rescaled to sum to 1.0 after exclusion of the inapplicable primary-race categories; for example, for the AIAN/API/black group, the probability of white single race was set to 0 and the remaining probabilities were increased to sum to 1.0.

For this study, we applied the bridging models to the public-use MR file to estimate the size of single-race populations in the U.S. As the models were fitted using single year of age and the public-use MR file provides population counts for broader age intervals, predictions for the mid-point of each interval were used; age was set to 3 for the interval 1–4 years. We then compared single-race population counts calculated using three methods: first, the NHIS-regression bridging method described above; second, equal allocation, where multiple-race groups were equally divided among the corresponding single-race groups; and third, NHIS-fractions, where multiple-race groups were divided into the corresponding single-race groups using the fractions derived from the 1997–2000 NHIS (Table 1) after omission of those who did not provide a specific primary race and rescaling of the fractions to sum to 1.0; for groups with fewer than 50 respondents, equal allocation was used. For this study, no attempts were made to round population counts to whole numbers at any level of geography.

Death rates

Using mortality records from 2000[6] and the three sets of population estimates, we calculated death rates for single-race population groups. Because multiple-race groups, in general, are younger than single-race groups,[11] we also compared two cause-specific death rates: coronary heart disease (International Classification of Diseases, 10th Revision [ICD-10] codes I00–L09, I11, I13, I20-I51), a leading cause of death among older age groups, and homicide (X85–Y09, Y87.1), a leading cause of death among younger age groups.[21]

RESULTS

Bridging models

Generally, the few demographic variables available were not strongly associated with single-race responses among multiple-race respondents.

The coefficients from the regression models show that the strengths of the associations between selected covariates and primary race and the directions of these associations differed among race groups, as shown in Table 2. For example, increasing age was associated with a higher estimated likelihood of choosing API as primary race among the API/white respondents but associated with a lower estimated likelihood of choosing AIAN among the AIAN/white respondents. Similarly, as the percent of multiple-race people living in an area increased, the estimated likelihood of reporting primary race as black among the black/white respondents or API among the API/white respondents increased; a high proportion of multiple-race respondents decreased the estimated likelihood of AIAN as a primary-race response for the AIAN/white group.

The combined model shows many similarities in the relationships between primary-race responses and the demographic covariates (Table 3). This is not surprising given that the largest multiple-race groups have the greatest influence on the combined-model estimates.

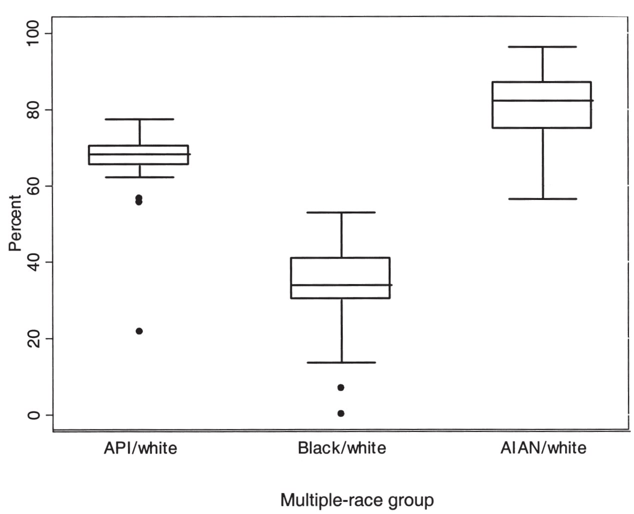
Population estimates

As did the distribution of primary-race responses for multiple-race groups shown in Table 1, the allocation under the NHIS-regression method to single-race groups differed across the multiple-race groups (Table 4). As expected, for the largest groups, the overall NHIS-regression allocations were reasonably consistent with the primary-race distribution in the NHIS, excluding the "no primary race" responses, and differed from what would be obtained by equal allocation.

Figure 1 shows box-plots depicting the percentage of the state multiple-race population bridged to single-race white under NHIS-regression allocation for the three largest multiple-race groups. With this method, unlike with the NHIS-fractions and equal allocation methods, the differences between states in demographic characteristics of the multiple-race reporters, and the corresponding individual and contextual variables, led to geographic variation in the single-race allocations.

Figure 2 shows box-plots depicting the percentage of the state single-race count that is attributable to multiple-race bridging after the original single-race counts are combined with the bridged multiple-race counts, for example, the part of the bridged single-race black count that was originally a multiple-race group. These distributions are influenced both by the relative counts of the single-race and multiple-race populations in each state as well as by the differing demographic characteristics of the multiple-race groups. As a result, these percentages also vary across states for the NHIS-fractions and equal allocation

Figure 1. Percentage of multiple-race population bridged to single-race white under NHIS-regression allocation, 50 states and the District of Columbia



NOTES: In each box-plot, the middle horizontal line represents the median of the distribution and the outer edges of the box are the 25th and 75th percentiles. Data points beyond the outside lines are considered outliers.

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

Table 2. Model coefficients from logistic regression models for selected multiple-race groups

Covariate	*Multiple-race group (single-race response predicted)*						
	AIAN/black (black)	*AIAN/white (AIAN)*	*API/white (API)*	*API/black (black)*	*Black/white (black)*	*AIAN/black/white (black)*	*AIAN/black/white (AIAN)*
Age	–0.05	–0.09[a]	0.10	0.06	0.06	0.36[a]	0.26
Hispanic origin, yes	–1.93[a]	0.89[a]	0.19	–0.10	–0.52	–0.83	0.36
Sex, male	–0.12	0.01	0.01	0.34	0.12	0.51	–0.44
Region							
Northeast	–0.88	0.21	–0.06	–0.46	–0.25	–3.46[b]	–4.54[a]
Midwest	–1.70[a]	0.09	–0.06	–3.92[a]	0.17	–3.79[b]	–3.82[b]
South	–0.98	–0.28	0.13	–1.48	–0.64	–2.27	–5.73[a]
West	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Urban/rural							
Large urban	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Large suburban	–0.44	–0.22	0.51	1.47	–0.08	2.31[a]	2.79[b]
Medium–small urban	0.88	–0.44[b]	0.07	1.68	0.29	0.75	2.27[a]
Nonmetropolitan	–0.38	–0.14	–0.63	0.13	0.58	1.65	4.18[a]
County population, 2000[c]							
Percent AIAN	–0.43[a]	0.51[a]	—	—	—	0.39	0.55
Percent API	—	—	0.007	–0.13	—	—	—
Percent black	0.00003	—	—	0.02	0.0008[a]	0.05[b]	0.11[a]
Percent multiple race	–0.17[a]	–0.08[b]	0.10[a]	0.31	0.32[b]	–0.03	–0.24[b]

[a]Coefficient differs from zero, $p<0.05$.

[b]Coefficient differs from zero, $p<0.1$.

[c]Logarithm of percent AIAN used in AIAN/white and AIAN/black models; square of percent black used in black/white and AIAN/black models.

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

NHIS = National Health Interview Survey

Table 3. Model coefficients for combined multi-logit model

Covariate	Single-race response predicted AIAN	API	Black
Race			
Not AIAN	—[a]	2.79[b]	2.20[b]
Not API	2.83[b]	—[a]	3.06[b]
Not black	0.97[b]	1.62[b]	—[a]
Age	–0.04	0.02	–0.02
Hispanic origin	0.84[b]	0.22	–0.59[b]
Male sex	0.02	0.01	–0.08
Region			
Northeast	0.60[b]	–0.13	0.40
Midwest	0.43	–0.15	0.20
South	–0.22	–0.25	–0.29
West	Reference	Reference	Reference
Urban/rural			
Large urban	Reference	Reference	Reference
Large suburban	0.16	0.46	0.12
Medium-small urban	–0.17	–0.09	–0.11
Nonmetropolitan	0.25	–0.15	–0.12
County population, 2000			
Log of percent AIAN	0.57[b]	0.07	–0.003
Percent API	0.04[b]	0.04[b]	0.05[b]
Percent black	0.04[b]	0.04[b]	0.06[b]
Percent multiple race	–0.10[b]	0.06[b]	–0.04

[a]Constrained to 0.

[b]Statistically significant, $p<0.05$.

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

NHIS = National Health Interview Survey

bridge methods. For the black population counts, the multiple-race allocation varies more across states under the NHIS-regression method than under either NHIS-fractions or equal allocation. For the AIAN counts, equal allocation leads both to greater variation across states and higher percentages overall, likely indicative of the large size of the multiple-race AIAN/white group. For the API counts, the box-plots for the NHIS-fractions and equal allocation methods are close, due to the nearly even split of primary-race responses within the NHIS sample of API/white respondents who selected a primary race; the smaller variation and lower proportions for the NHIS-regression method may be attributed to the model parameters capturing state-specific characteristics.

The overall single-race percentages in the population (single-race reporters and bridged multiple-race reporters combined) under the NHIS-regression allocation, NHIS-fractions allocation, and equal allocation methods are shown in Table 5 under "Denominator distribution." Under both the NHIS-regression and NHIS-fraction allocations, the percentage of the population in the AIAN group was more than 10% lower than under equal allocation.

Death rates

National death rates for the API, black, and white groups were similar under NHIS-regression, NHIS-fractions, and equal allocation bridging (Table 5). Death rates for the AIAN group were much lower under equal allocation than under the NHIS-regression or NHIS-fractions bridging due to the larger percentage of the multiple-race AIAN/white population assigned to the single-race AIAN denominator under equal allocation. For heart disease and homicide, the pattern was similar to that observed for all-cause mor-

Table 4. Percent distribution of single-race assignment after application of the NHIS-regression method to bridge multiple-race counts to single-race categories: public-use Census Modified Race Summary file, United States, 2000

	Single-race assignment			
Multiple-race response	*AIAN*	*API*	*Black*	*White*
AIAN/API	63.3	36.7	—	—
AIAN/black	15.9	—	84.1	—
AIAN/white	22.4	—	—	77.6
API/black	—	41.4	58.6	—
API/white	—	40.9	—	59.1
Black/white	—	—	62.9	37.1
AIAN/API/black	26.8	25.4	47.8	—
AIAN/API/white	2.2	8.7	—	89.1
AIAN/black/white	18.7	—	57.4	23.9
API/black/white	—	12.0	11.9	76.1
AIAN/API/black/white	0.9	1.0	2.1	95.9

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

NHIS = National Health Interview Survey

tality; death rates were closer across the three methods for the white, black, and API groups than for the AIAN group. Within the white population, findings were similar for Hispanics and non-Hispanics (not shown); the relatively few Hispanics of other race groups did not allow for stable death rate calculations.

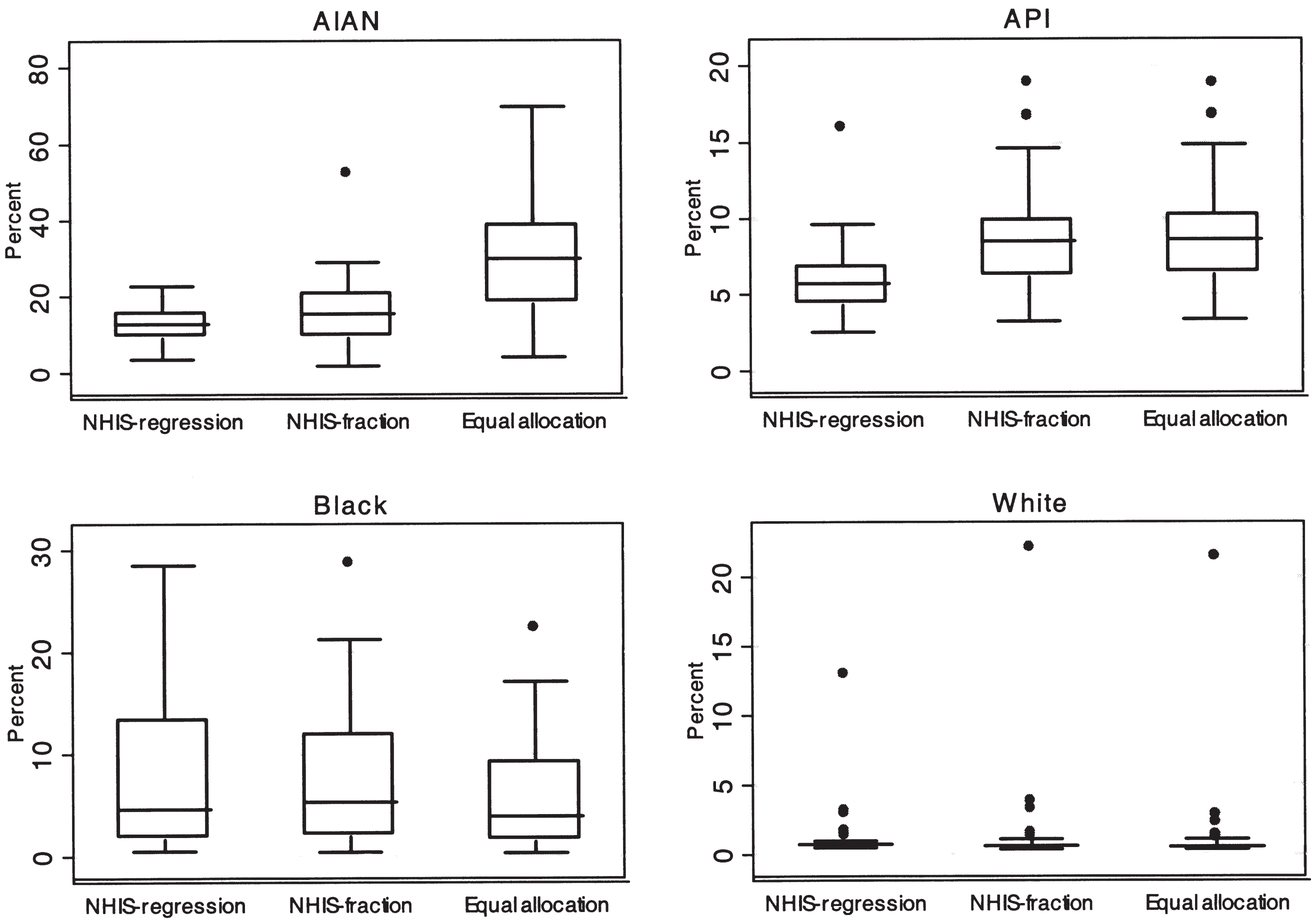
In the API, black, and white groups, relative differences in age-specific death rates under NHIS-regression allocation and equal allocation tended to decrease with age (Figure 3), likely due to larger numbers of multiple-race respondents in the younger age groups than in the older age groups. In contrast, for the AIAN group, the absolute differences were considerably larger, and increased with age. Age-specific differences between death rates calculated using the NHIS-fractions and NHIS-regression denominators were generally smaller than the differences shown in Figure 3 for equal allocation but followed a similar pattern (not shown); for the API group, the differences were similar to those shown in Figure 3 due to the similarity of NHIS-fractions and equal allocation for the API/white group.

DISCUSSION

We have demonstrated a method for bridging the multiple-race categories in the MR file to single-race groups using models fitted to NHIS data. Building on earlier work,[12] we used individual and contextual information available for both NHIS data and the MR file to improve upon the initial approaches outlined by the OMB. Although not the intention of the 1997 standards or of the respondents providing answers to Census 2000, exclusive and exhaustive single-race population estimates are required to calculate rates and make comparisons using data collected under the 1977 Directive and the 1997 revised standards. The problem of non-comparability due to changes in classification systems is not unique to race reporting and has also arisen, for example, with the recent adaptation of the 10th revision of the International Classification of Diseases,[22] and with the changes in Census industry and occupation classification schemes from 1970 to 1980.[23–25]

Furthermore, we have demonstrated the use of bridged population data in calculating death rates. With the exception of statistics calculated for the single-race AIAN group, overall death rates calculated using the NHIS-regression allocation did not differ appreciably from those calculated using equal or NHIS-fraction allocation. Nevertheless, the variation shown across states in Figures 1 and 2 demonstrates that the regression models, by incorporating geographic and demographic variation, may produce death rates for sub-national areas, such as states or metropolitan areas, that differ from those calculated using the other bridging methods. The larger differences for the AIAN group can be attributed to the large number of re-

Figure 2. Percentage of single-race count attributed to multiple-race allocation, by single-race group, 50 states and the District of Columbia



NOTES: In each box-plot, the middle horizontal line represents the median of the distribution and the outer edges of the box are the 25th and 75th percentiles. Data points beyond the outside lines are considered outliers. The ranges of the Y-axes differ among race groups.

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

spondents who chose AIAN and another race as well as the relatively small proportion of multiple-race AIAN responders who chose AIAN as a primary race (Table 1); the choice of bridging method has previously been shown to have the largest effect on this group.[26,27] Although national mortality statistics for the AIAN group were prone to inaccuracy prior to multiple-race reporting,[9,28] incorporation of the bridged multiple-race responses will make interpreting the AIAN data even more difficult. On the other hand, bridging the denominator counts to obtain single-race death rates is likely more informative than calculating death rates for multiple-race groups, which are likely to be unrealistic for some time.[29]

The population estimates and death rates under NHIS-regression allocation reported here are based on bridging the public-use MR file rather than using the Bridged MR file, the file created by the Census Bureau using the predictions from the models shown in this article. Use of the public-use MR file allowed us to calculate the single-race allocations for multiple-race respondents; only the final single-race counts are available on the Bridged MR file. More important for our evaluation, use of the public-use MR file allowed us to create bridged single-race counts based on NHIS-fractions and equal allocation for comparison. However, the results from bridging the public-use MR file and from the Bridged MR file differ somewhat due to

Table 5. Single-race percent distribution and corresponding crude and age-adjusted death rates for all causes and for selected causes, by race after multiple-race counts were bridged to single-race categories by use of three bridging methods, United States, 2000

			Death rates					
		Denominator distribution	*All causes*		*Coronary heart disease*[a]		*Homicide*[b]	
Single race category	*Bridging method*	*Percent*	*Crude*	*Age-adjusted*	*Crude*	*Age-adjusted*	*Crude*	*Age-adjusted*
AIAN	NHIS-regression	1.06	379.3	697.9	80.7	175.0	6.8	6.7
	NHIS-fractions	1.07	378.8	673.1	80.6	167.9	6.8	6.7
	Equal allocation	1.21	333.0	569.3	70.8	141.0	5.9	5.9
API	NHIS-regression	4.18	296.2	505.5	77.3	145.7	3.0	3.0
	NHIS-fractions	4.22	293.4	504.8	76.6	145.6	3.0	2.9
	Equal allocation	4.23	293.0	504.3	76.5	145.5	3.0	2.9
Black	NHIS-regression	13.00	781.3	1,121.7	211.9	324.9	21.5	20.6
	NHIS-fractions	13.01	780.8	1,121.9	211.8	324.9	21.5	20.5
	Equal allocation	12.92	785.8	1,126.1	213.1	326.1	21.6	20.7
White	NHIS-regression	81.75	900.3	849.9	270.2	253.4	3.6	3.6
	NHIS-fractions	81.70	900.8	850.0	270.4	253.5	3.6	3.6
	Equal allocation	81.63	901.6	850.7	270.6	253.7	3.6	3.6

NOTES: NHIS-regression bridges multiple-race counts based on model predictions. NHIS-fractions bridges multiple-race counts based on national estimates of single race distributions. Equal allocation allocates multiple-race counts equally among corresponding single-race groups.

[a]ICD-10 codes I00–I09, I11, I13, I20–I51.

[b]ICD-10 codes X85–Y09, Y87.1.

AIAN = American Indian or Alaska Native

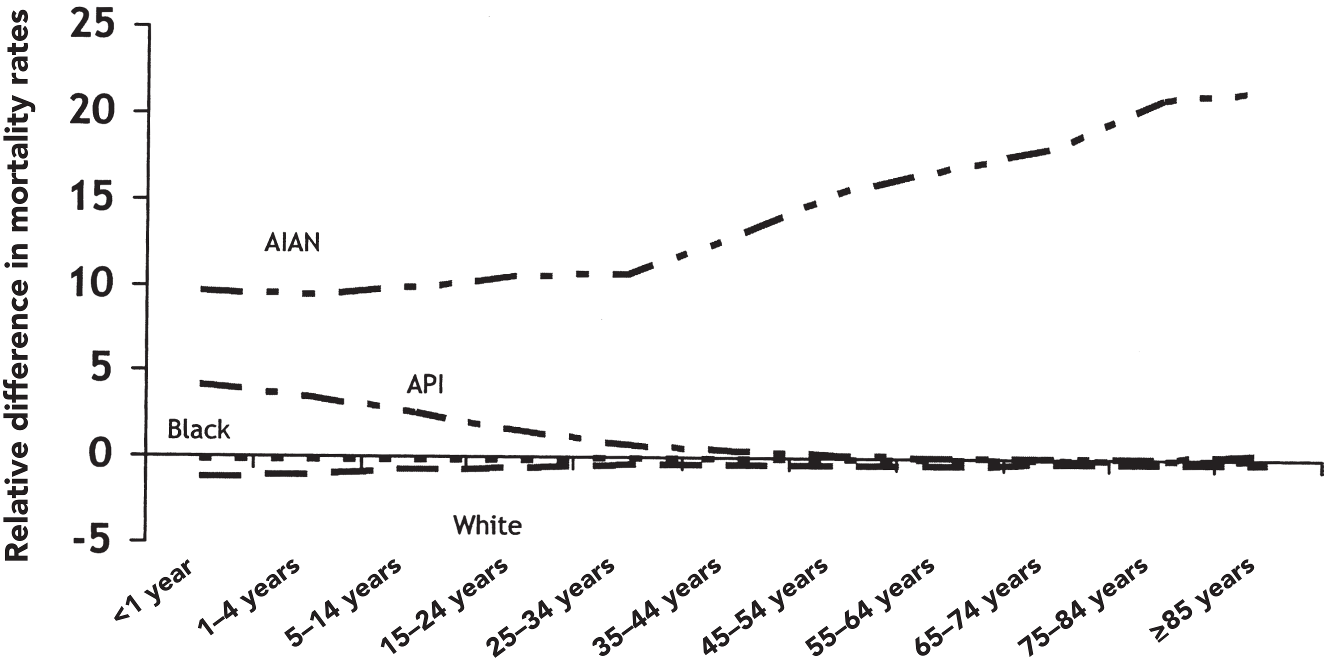
API = Asian or Pacific Islander

NHIS = National Health Interview Survey

rounding by the Census Bureau on the Bridged MR file and our application of the models to the midpoint of age intervals on the public-use file that were broader than single years of age. Overall national population totals used here differ by less than 0.5% for all four single-race groups; however, the AIAN counts for those ≥65 years of age are 2% to 3% higher in the bridged public-use MR file than in the Bridged MR file. As a result, death rates reported here differ slightly from the official death rates shown in the published reports that were calculated using the population counts from the Bridged MR file.[6,14]

The methods of collecting race information in vital statistics systems, interview surveys, and the decennial Census vary greatly. Primary race in the NHIS, which our bridging models are predicting, is collected using an interview survey and thus is not entirely comparable with the single race reported in the Census under either the 1977 OMB Directive or the 1997 standards; in turn, the use of bridged single-race estimates in creating population denominators could exacerbate the biases that would already exist even if there were no multiple-race reporting. Many problems existed in the use of race data before the 1997 standards, which were made more complicated with their adoption.[7] For example, although race information collected in interview surveys such as the NHIS and in the Census is considered self-reported, as recommended by the OMB, often a single member of a household provides data for all its members. Death records depend on informants or observer identification for race information. Both self-identification and observer-report may vary, depending on circumstances. The inherent imprecision in both self-reported and observer-identified race suggests that incompatibility between numerators and denominators existed in vital statistics data before the addition of multiple-race categories, as did imprecision in the racial categories, particularly for the AIAN group.[9] These reporting issues may affect multiple-race reporting more

Figure 3. Relative difference in age-specific death rates estimated using NHIS-regression allocation and equal allocation to generate denominators, by age and race



NOTES: Relative differences are calculated as the rate using the NHIS-regression allocation minus the rate using equal allocation, divided by the rate using the NHIS-regression allocation, multiplied by 100.

AIAN = American Indian or Alaska Native

API = Asian or Pacific Islander

than single-race reporting; however, the relatively small number of multiple-race responses likely leads to small additional biases in single-race statistics for all but the AIAN group.

Similarly, rate calculations that use the MR file typically treat the denominators as fixed and not subject to random variation. However, since the Census Bureau uses methods to modify these MR counts to adjust for non-response and to produce realistic counts across levels of geography, these counts are in fact subject to random variation. This random variation has typically been ignored in calculating vital rates, under the assumption that the error associated with the denominator is small relative to that associated with the numerator. That many calculations derived from population estimates are ratios, including death rates, increases the complexity of the error estimation and introduces some "ratio bias." The use of bridged single-race counts adds to the variability associated with population estimates, and the corresponding bias and variability in the vital rates. The extra uncertainty could be calculated using multiple imputation,[12,23–25] replication methods such as the jackknife or bootstrap, or possibly analytic approximations to multiple imputation.[30] We hope to assess the level of extra uncertainty in estimated rates due to the application of the NHIS-regression predictions in future evaluations.

Since our NHIS-regression models were fitted at the national level, we expect that predictions for local areas will not be as good as national predictions, despite our use of individual and contextual variables. On the other hand, the regression-based bridging method may do better for some local areas, though not all, than NHIS-fractions or equal allocation, which do not consider local variation. In addition, the predictions for the smaller multiple-race groups obtained from the combined model are not expected to be as good as those for the larger groups obtained from the individual models. These limitations are unavoidable due to our lack of data to fit separate bridging models at the local-area level and for smaller multiple-race groups.

Furthermore, our predictions are likely to be limited by the small number of covariates for which data were available. This too is unavoidable since we were limited to using the few individual covariates on the MR file for our models; additionally, our selection of contextual variables was limited by the urgency of producing the bridged file and the fact that our model-fitting occurred when data on many contextual variables were not yet available from the 2000 Census. We plan to consider adding variables from the 2000 Census as they become available.

Given the evidence that other variables are predictive of primary-race identification,[12,31] and the likelihood that these variables are associated with variables that are likely to be analyzed together with the Bridged MR file (e.g., deaths, births, diseases in calculating rates), it follows that relationships observed in such analyses could be biased due to attenuation.[32] The addition of contextual variables to our models could diminish such attenuation.

The strength of the limited covariate set and the similar models for each race group is the consistency across areas and groups, reducing the bias associated with different methods of assigning single races to the multiple-race respondents.[26,27] The consistency across areas and groups is congruent with the principal goal of this project: improving data comparability under the 1977 Directive and 1997 standard.

The authors thank colleagues in the Population Division at the Census Bureau, and Elizabeth Arias, PhD, Brady E. Hamilton, PhD, and Paul D. Sutton, PhD, of the National Center for Health Statistics.

REFERENCES

1. Office of Management and Budget (US). Revisions to the standards for the classification of federal data on race and ethnicity. Fed Regist 1997;62;58781-90. Also available from: URL: http://www.whitehouse.gov/omb/fedreg/ombdir15.html
2. Office of Management and Budget (US). Appendix: Directive No. 15. Race and ethnic standards for federal statistics and administrative reporting (as adopted on May 12, 1977). In: Office of Management and Budget (US). Standards for the classification of federal data on race and ethnicity. August 1995 [cited 2002 Nov 15]. Available from: URL: http://www.whitehouse.gov/omb/fedreg/race-ethnicity.html
3. Census Bureau (US). Overview of race and Hispanic origin. Census 2000 Brief. Suitland (MD): Census Bureau; 2001 Mar.
4. Durch JS, Madans JH. Methodological issues for vital rates and population estimates: 1997 OMB standards for data on race and ethnicity. Vital Health Stat 4 2001;31:1-30.
5. Ventura SJ, Hamilton BE, Sutton PD. Revised birth and fertility rates for the United States, 2000 and 2001. Natl Vital Stat Rep 2003;51:1-18.
6. Arias E, Smith BL. Deaths: preliminary data for 2001. Natl Vital Stat Rep 2003;51:1-44.
7. Boehmer U, Kressin NR, Berlowitz DR, Christiansen CL, Kazis LE, Jones JA. Self-reported vs. administrative race/ethnicity data and study results. Am J Public Health 2002;92:1471-3.
8. Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. JAMA 2003;289:2709-16.
9. Rosenberg HM, Maurer JD, Sorlie PD, Johnson NJ, MacDorman MF, Hoyert DL, et al. Quality of death rates by race and Hispanic origin: a summary of current research, 1999. Vital Health Stat 2 1999;128:1-13.
10. Office of Management and Budget (US). December 15, 2000. Provisional guidance on the implementation of the 1997 standards for the collection of federal data on race and ethnicity [cited 2003 Aug 1]. Available from: URL: http://www.whitehouse.gov/omb/inforeg/re_guidance2000update.pdf
11. Sondik EJ, Lucas JW, Madans JH, Smith SS. Race/ethnicity and the 2000 Census: implications for public health. Am J Public Health 2000;11:1709-13.
12. Schenker N, Parker J. From single-race reporting to multiple-race reporting: using imputation methods to bridge the transition. Stat Med 2003;22:1571-87.
13. Census Bureau (US). Estimates data. National. Modified race data summary file [cited 2003 Sep 22]. Available from: URL: http://eire.census.gov/popest/data/national.php
14. National Center for Health Statistics (US). National Vital Statistics System. U.S. Census populations with bridged race categories. Bridged-race population estimates for July 1, 2000–July 1, 2002 [cited 2003 Sep 22]. Available from: URL: http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm
15. Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. United States Census 2000 population with bridged race categories. Vital Health Stat 2 2003;135:1-63.
16. Botman SL, Moore TF, Moriarty CL, Parsons VL. Design and estimation for the National Health Interview Survey, 1995–2004. Vital Health Stat 2 2003;1-41.
17. Blackwell DL, Tonthat L. Summary health statistics for the U.S. population: National Health Interview Survey, 1999. Vital Health Stat 10 2003:1-103.
18. National Center for Health Statistics (US). National Health Interview Survey (NHIS). Micro-data [cited 2003 Sep 22]. Available from: URL: http://www.cdc.gov/nchs/nhis.htm#micro-data
19. Eberhardt MS, Ingram DD, Makuc DM, Pamuk ER, Fried VM, Harper SB, et al. Urban and rural health chartbook. Health United States, 2001. Hyattsville (MD): National Center for Health Statistics; 2001.

20. Stata Corporation. Stata statistical software: release 6.0 College Station (TX): Stata Corporation; 1999.
21. Anderson RN. Deaths: leading causes for 1999. Natl Vital Stat Rep 2001;49:1-87.
22. Anderson, RN, Rosenberg, HM. Disease classification: measuring the effect of the tenth revision of the International Classification of Diseases on cause-of-death data in the United States. Stat Med 2003;22:1551-70.
23. Rubin DB, Schenker N. Interval estimation from multiply-imputed data: a case study using Census agriculture industry codes. J Official Stat 1987;3:375-87.
24. Clogg CC, Rubin DB, Schenker N, Schultz B, Weidman L. Multiple imputation of industry and occupation codes in Census public-use samples using Bayesian logistic regression. J Am Stat Assoc 1991;86:68-78.
25. Schenker N, Treiman DJ, Weidman L. Analyses of public-use decennial Census data with multiply-imputed industry and occupation codes. Appl Stat 1993;42:545-56.
26. Heck KE, Parker JD, McKendry CJ, Chavez GF. Mind the gap: bridge methods to allocate multiple-race mothers in trend analyses of birth certificate data. Matern Child Health J 2003;7:65-70.
27. Parker JD, Makuc DM. Methodologic implications of allocating multiple-race data to single race categories. Heath Serv Res 2002;37:203-15.
28. Burhansstipanov L, Satter DE. Office of Management and Budget racial categories and implications for American Indians and Alaska Natives. Am J Public Health 2000;90:1720-3.
29. Heck K, Parker JD, McKendry CJ. Multiple-race mortality data for California, 2000–2001. Pub Health Rep. *In press 2004.*
30. Schafer J, Schenker N. Inference with imputed conditional means. J Am Stat Assoc 2000;95:144-54.
31. Heck KE, Parker JD. Predicting first race reported on California birth certificates. Ethn Dis. *In press 2003.*
32. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91:473-89.