A Practical Approach to Using Multiple-Race Response Data:
A Bridging Method for Public-Use Microdata

Carolyn A. Liebler and Andrew Halpern-Manners
Department of Sociology and
Minnesota Population Center
University of Minnesota

September 2006

This project was begun while the first author was funded by "IPUMS-Redesign" (NIH GRANT R01-HD043392), Steven Ruggles, P.I. We thank John Robert Warren, Deborah D. Ingram, Elaine M. Hernandez, C. Matthew Snipp, and J. Trent Alexander for their helpful feedback and the Minnesota Population Center for its invaluable research support. Address comments to: Carolyn Liebler, Department of Sociology, 267 – 19th Ave. S., Minneapolis, MN 55455; email: liebler@soc.umn.edu.

ABSTRACT

Revised federal policies require that multiple-race responses be allowed in all federal data collection efforts, but many researchers find the multitude of race categories and variables very difficult to use. Important comparability issues also interfere with using multiple-race data in analyses of multiple datasets and/or multiple points in time. These difficulties have, in effect, discouraged the use of the more nuanced new data on race. We present a practical method for incorporating multiple-race respondents into analyses that use public-use Microdata. We extend prior work by the National Center for Health Statistics (NCHS) in which they use multiple-race respondents' preferred single race and other characteristics to develop a model predicting preferred single race (if forced to choose). In this paper, we apply the NCHS-generated regression coefficients to public-use Microdata with limited geographic information. We include documentation and dissemination tools for this practical and preferable method of including multiple-race respondents in analyses.

Race is a contextual, contingent, complicated, and life-directing social construct (c.f. Harris and Sim 2002; Cornell and Hartmann 1998; Root 1996). In the late 1990s, American federal policies for collecting data on race changed to better reflect this relatively recent understanding. The revised policies require that multiple-race responses be allowed in all federal data collection efforts and encourage data creators to provide as much detail as possible about their respondents' race reports (OMB 1997, 2000). The change in requirements came into effect in time for Census 2000, and many other data collection efforts have followed suit.[1] As witnesses to this great change in the way race is recorded, contemporary researchers have the opportunity to describe our complex social world more accurately and with more nuance.

In reality, however, many researchers find the multitude of race categories and variables unwieldy and very difficult to use, discuss, and/or interpret (Snipp 2003; Harrison 2002) for several reasons. Definitional changes and inconsistent classification schemes interfere with the calculation of statistics using multiple data sets which measure and record race differently. Data sets differ depending on who collected the data and when it was collected. For example, most race information gathered by state governments—including vital statistics such as births, deaths, and marriages—continues to be collected using the single-race response format. Similar challenges arise in studies that rely on time-series data with inconsistent classification schemes to obtain estimates of change over time in a racialized aspect of American society. Even the most straightforward research questions and widely used data sources are susceptible to complications (Warren and Halpern-Manners 2006).

[1] For example, complex race responses have been allowed in the Current Population Survey (CPS), the National Longitudinal Study of Adolescent Health (Add Health), the National Health Interview Survey (NHIS), and the American Community Survey (ACS).

In this paper, we offer a practical tool for dealing with these and related problems in a sophisticated and meaningful way which allows multiple-race respondents to be included in analyses.[2] The technique provides a "bridge" between the old race question format, where only one race response was coded, and the new "mark all that apply" system.[3] The bridging tool we present allows a researcher to recode complex race data in two alternative ways to identify (1) which single race a multiple-race respondent would have been most likely to report if he or she had been forced to choose only one; and (2) the predicted probability that the multiple-race respondent would have reported each single race, if forced to choose.[4] Unlike existing methods, which we describe in detail in the next section, the bridging method outlined here can be used reliably in conjunction with most public-use data sets to calculate changes across time in the same data series, to compare data sets with different race question formats, or simply to code "race" in a dataset with a very high number of race categories.

BACKGROUND

Three Problematic Approaches

In order to avoid the complexity inherent in data with multiple-race responses, researchers typically use one of three approaches to categorize multiple-race responses. Some focus their analyses only on sample members whose race responses are the least complex—for example, single-race whites and single-race blacks. Remaining groups are commonly excluded from the analysis (the first approach discussed below) or they are consigned to a miscellaneous "other"

[2] In situations where the data are available and the researcher is willing, each single-race group and specific multiple-race groups are better analyzed separately.
[3] While it would be preferable to bring the old forced-choice single-race data in line with the new, enhanced standard, this is not practical. Older data include only a single race response; there is little hope of knowing what other race(s) the respondent would have included if allowed.
[4] We use the term "race" to refer to the categories presented in a survey question and thus the number of "races" differs across survey instruments. We use "multiracial" and "multiple-race" to describe people who reported two or more "races" and we use "monoracial" and "single-race" to describe people who reported only one "race." Note that Latino/Hispanic is not usually included on "race" questions.

category that is not substantively meaningful (the second approach, below). A third and less common approach is to group multiple-race individuals with their single race counterparts such that an American Indian-black biracial person is included in the American Indian group and the black group.

Many scholars *exclude multiracial people* from their study in order to simplify their analyses. Using this "exclusionary method" (as we call it) assumes that multiracial people are, on average, the same as monoracial people, such that removing them from the sample does not bias the sample or any substantive conclusions that may be derived from it. In most cases, this is not true; multiracials and monoracials are known to differ on a number of important characteristics. Table 1 summarizes descriptive statistics for a variety of commonly used economic, geographic, and demographic characteristics, disaggregated according to racial group and multiple-race status. The first panel suggests that multiracial American Indians/Alaskan Natives earn roughly $4,000 more per year and are significantly less likely to live in a rural area, compared to their monoracial counterparts. Analogous differences of equal or greater magnitude exist among Pacific Islanders, Asians, and blacks. In other words, restricting the sample to monoracials introduces unnecessary bias by systematically underrepresenting distinctive portions of minority populations; a bias that may be exacerbated with time as the multiracial population continues to grow (Goldstein and Morning 2000; Waters 2000; Lee 2001).

TABLE 1 ABOUT HERE

The second problematic approach to using complex race data involves grouping all multiracial persons into a *single residual category*. Although it seems to be a reasonable simplification of an extremely complex situation, this approach creates its own problems. When diverse populations are collapsed into a single group, the results for this residual group cannot be interpreted and all cultural relevance is lost. As Burhansstipanov and Satter note, "'Other' data have the same effectiveness as having 'no' data" (2000:1721). It represents a loss of potential understanding of the smaller groups' experiences (NCVHS 2005). It is also frustrating for members of small minority groups such as American Indians, Pacific Islanders, and multiracial people (as well as for those who worked to recruit them into the study).

A third and less frequently used approach is to *include multiracials with each of their related monoracial groups*—that is, information about a black-white biracial person is included in the "black alone or in combination" category and in the "white alone or in combination" category (we term this the "all-inclusive method").[5] This approach is difficult to use because all-inclusive case totals do not sum to the number of cases in the study. Another major drawback to this approach is that it does not provide cross-time comparability. In Figure 1, for example, we present changes in personal income among American Indians/Alaskan Natives (1996 to 2002) using data collected in two ways. The Current Population Survey (CPS) used a forced-choice single-race approach throughout the period. In comparison, Census 2000 and its successor the American Community Survey (ACS) used the "mark all that apply" system beginning in 2000. As Figure 1 shows, using all-inclusive counts offers a distorted sense of recent trends in income, differing by as much as 14 percent from estimates produced using the data where only a single race response was allowed.

FIGURE 1 ABOUT HERE

Two Types of Bridging Methods

As alternatives to these problematic approaches, federal agencies, policy makers, and academics have developed and tested a variety of options for recoding or "bridging" multiple-race responses into single-race categories. The most simplistic of these options recodes each

[5] The term "race alone or in combination" is used by the U.S. Census Bureau to describe this classification scheme.

individual's multiracial response into a pre-determined single-race response.[6] This is called *whole assignment* because each multiple-race respondent's case is coded wholesale into a single-race category that is determined based on the race responses involved. Whole assignment methods are straightforward to apply and explain, but are not especially useful for analyzing change over time because the choice of allocation scheme can have a powerful effect on the results. For example, because about 40 percent of the American Indian population was multiracial in Census 2000, the choice of whole assignment methods can markedly affect the measured size of the American Indian population. In addition to yielding varying estimates for the size and composition of a group, deterministic whole assignment fails to represent all of the respondent's self-identified races – an attribute that further constrains the ability of survey data to capture the complexity and meaning of race. See Table 2 for an overview of key terms, including whole assignment.

TABLE 2 ABOUT HERE

A more nuanced technique is *fractional assignment*. Most fractional assignment bridging methods apply a pre-determined fractional weight to each multiple-race response. For example, the method of "equal fractional assignment" recodes a biracial Chinese-white response into two responses (Chinese and white), each with a weight of 0.5 so that the total number of cases remains constant.[7] When providing advice about how to use the newly available race information, the Federal Office of Management and Budget (OMB) proposed a variety of

[6] There are several strategies for pre-determining which single race that is used. The "smallest group" method, for instance, reallocates multiple-race responses to the category with the smallest single-race count. Other examples of fractional assignment discussed elsewhere include largest group other than white and largest group. These are discussed at length elsewhere (c.f. OMB 2000).

[7] Equal fractional assignment is just one of several fractional assignment strategies. Strategies vary in terms of the rules used to apportion multiple-race persons into single-race categories (see OMB 2000).

fractional assignment methods because of their simplicity and their potential to inform a wide variety of situations (OMB 2000, Appendix C).[8]

Evaluations of whole assignment and fractional assignment methods indicate that fractional assignment provides a closer approximation to past racial distributions (Lee 2001; Grieco 2002; Parker and Makuc 2002; Heck et al. 2003). However, fractional assignment algorithms which pre-determine fractions for each multiracial group fail to account for the substantial variation in single-race identification patterns found within each multiracial population (Parker et al. 2004). For instance, a Native Hawaiian-white in Hawaii is much more likely to report single-race Native Hawaiian, if forced to choose, than is a part-Hawaiian living on the mainland (Kana'iaupuni and Liebler 2005). Also, fractional allocation methods which use pre-determined fractions have been shown to provide substantially biased results either in favor of or against smaller groups, in part because they do not take into account the racial context in which the respondent lives (Lee 2001; Grieco 2002; Parker and Makuc 2002; Schenker and Parker 2003; Heck et al. 2003; Mays et al. 2003).

The NCHS Regression Method

In hopes of building a more reliable bridge federal researchers and policy makers have focused their efforts on data collected in the National Health Interview Survey (NHIS) (OMB 2000; Schenker and Parker 2003; Ingram et al. 2003; Parker et al. 2004).[9] Since 1982, the NHIS has allowed multiple race responses, asking each multiple-race respondent a follow-up question to

[8] Nine criteria were used to help identify the promising tabulation procedures to be used by federal agencies (OMB 2000:13-14) releasing data and reports directly to the general public. They are: (1) measure change over time; (2) minimize disruptions to the single-race distribution; (3) have include a range of applicability; (4) meet confidentiality and reliability standards; (5) be statistically defensible; (6) be easy to use; (7) need little statistical knowledge to use; (8) be easy to explain and understand; and (9) reflect the respondents choices as much as possible.

[9] Alternative bridging proportions were developed by Allen and Turner (2001) using data from the 1990 census race and ancestry questions.

identify a single race that best describes him or herself. With this information, the National Center for Health Statistics (NCHS) used multivariate methods to predict the single-race response preferred by each multiracial NHIS respondents (1997-2000). Their predictors were the respondent's age, sex, Hispanic origin, and racial context. They derived the racial context measures using county-level geographic information available in the private data.[10] Using the resulting regression coefficients and parallel measures of personal characteristic and racial context, researchers can apply these results to other data to predict a multiple-race respondent's most likely single-race report and/or to generate weights to be used for fractional assignment. The primary purpose of this paper is to aid researchers in making this application to public-use data with limited geographic information.

There are several benefits to the NCHS's method of attending to each person's characteristics and racial context when simplifying multiple-race data. First, studies of multiracial identification patterns have consistently found that *place* matters to racial identity (Allen and Turner 2001; Grieco 2002; Liebler 2004; Kana'iapuni and Liebler 2005). Age, sex, and Hispanic origin are also known to be related to racial identities and responses to questions about race (López 1996; Root 1996; Xie and Goyette 1998; Waters 1999; Rodríguez 2000; Schenker and Parker 2003; Liebler 2004; Kana'iaupuni and Liebler 2005). Second, this method is documented to have reduced bias and enhanced predictive power (Schenker and Parker 2003), and thus it can better capture real change over time in variables of interest while minimizing disruptions to the single-race distribution. And third, the method of fractional assignment better represents the identities of multiple-race respondents than whole assignment because of its

[10] The NCHS group collaborated with the Census Bureau to derive the county-level contextual variables for NHIS respondents.

emphasis on the non-zero probability of reporting each race.[11] In other words, fractional assignment reflects the respondent's choice—which was to be associated with multiple races rather than a single race.

Despite the high quality of the NCHS regression method, researchers attempting to apply their results face substantial challenges. As we discuss in the next section, without access to extremely detailed geographic identifiers—identifiers that are rarely if ever available without special consent—the NCHS method is impossible to fully implement. This paper represents an effort to resolve this problem. We provide computer code (in Appendix A), which researchers across disciplines can apply to public-use Microdata with state-level geographic identifiers. This code applies the NCHS regression results (with necessary adjustments for limited geographic information) to calculate the most likely single race (whole assignment) and the appropriate weights (fractional assignment) for each multiple-race respondent. In the remainder of the paper, we describe the specific compromises necessary for using data with limited geographic detail, introduce the usage of the allocated race data, and document some biases resulting from limitations on geographic information.

The research community stands to benefit from our efforts to document and disseminate this bridging method for public-use Microdata. Using this methodologically sound and substantively meaningful approach to generate simplified race variables avoids problems of bias and/or incomparability inherent in other methods. While the number of multiracial individuals may seem small now, this number is very likely to grow (Goldstein and Morning 2000; Waters 2000; Lee 2001); it is important to implement a good bridging method as early as possible during

[11] Multiracial individuals can have considerable leeway in how they report their races (Nobles 2000; Liebler 2001; Wallace 2001; Rockquemore and Brunsma 2002). Although fractional assignment may be an improvement, dividing identities into separate pieces does not (necessarily) well-represent how multiple-race individuals see themselves (Liebler 2001; Wallace 2001; Rockquemore and Brunsma 2002; Renn 2004).

the transition to new race data so that research done now is not undermined in the future by questions about how multiracial responses were used (or ignored). We also hope to enhance comparability and consistency between research projects from a variety of disciplines by explicitly describing a practical method for approximating the unknown preferred single race of a multiple-race survey respondent.

METHODS

Building a Bridge for Public-Use Microdata

For ease of discussion, we label the bridging method presented here the "modified regression method" because it is a modification of the NCHS regression method. When using the modified regression method, the researcher calculates the probability of each single-race response, based on the multiracial individual's specific combination of races, as well as their age, sex, Hispanic origin, region, urbanization level of residence, and the racial diversity of the local area. It can be used as a whole allocation method or a fractional allocation method, depending on the user's preference. The computer code given in Appendix A can be used with most public-use data sets, including the 2000 Census 1% and 5% samples, the American Community Survey (ACS), and the Current Population Survey (CPS). State-level geographic detail is required to implement our computer code.

Compressing multiple-race responses. In order to deal with a full set of possible multiple-race responses in the context of (potentially) hundreds of ways to mark multiple races, we follow the Census Bureau's approach of summarizing multiracial responses into a very limited number of groups.[12] In the "modified race data" format that we use[13] there are 11 multiple-race

[12] For an example of the hundreds of unique multiracial categories able to be generated using the Census 2000 public-use microdata sample, see the variable RACE at www.ipums.org/usa/.

categories, representing all of the possible combinations of (1) American Indian and/or Alaska Native (AIAN or American Indian); (2) Asian and/or Pacific Islander (API or Asian/PI); (3) black or African American (B or black); and (4) white (W).[14] The Census 2000 form allowed respondents to mark "some other race" (SOR) as one of their races, but the modified race data format ignores these responses, thereby dramatically reducing the number of possible multiple-race categories. A respondent reporting black and SOR in Census 2000 is thus considered to be single-race black in the modified race data format. In this format, individuals reporting only SOR were assigned a "valid" single- or multiple-race response using hot-deck allocation. Also, because Asians and Pacific Islanders were usually tallied together in the past, and because bridging is intended to mimic the past, the modified race data format combines Asians and Pacific Islanders. Thus individuals who mark several Asian groups, or who mark an Asian and a Pacific Islander group, are not considered multiracial in this categorization scheme. Note that the federal government does not consider Hispanic/Latino to be a race and so it is not included in the modified race data format.

Limited geographic information. Most public-use microdata contain detailed information that must be kept confidential. A common way of reducing the risk of a breach in confidentiality is to restrict the amount of geographic information available. In the public version of the Census 2000 data (5% sample), for example, the lowest level of geographic information available is the person's PUMA ("Public Use Microdata Area")—a census-defined area with a minimum of

[13] To remain compatible with the NCHS method, we further reduced the bureau's modified race data format by including Native Hawaiian and other Pacific Islanders with Asians.
[14] These 11 include 6 two-race categories (AIAN-API, AIAN-B, AIAN-W, API-B, API-W, B-W), 4 three-race categories (AIAN-API-B, AIAN-API-W, AIAN-B-W, API-B-W), and 1 four-race category (AIAN-API-B-W). It should be noted that this is slightly different from the Census Bureau's modified race data file, which contains 31 racial groups.

100,000 people.[15] In other data (e.g., the CPS and the ACS), the lowest level of available geography is the state. As analysts at NCHS and the Census Bureau, Ingram et al. had access to geographic information down to the county level and incorporated this into their analysis.[16] Therefore, the geographic aspects of the NCHS regression method cannot be replicated using publicly available data—compromises must be made.

Racial composition of the area. Using the modified race data format, the NCHS regression method measures the racial composition of local areas using four variables: percent American Indian in the county, percent Asian/PI in the county, percent black in the county, and percent multiple-race in the county. When replicating these variables using data with limited geographic information, it is important to code race responses in the modified race data format before calculating the percent of each racial group in the area. Complete replication of the modified race data format is not always possible, however, because this format requires that "some other race" responses be allocated to American Indian or Alaskan Native, Asian/Pacific Islander, black, white, or some combination of these four groups. To maximize comparability with the NCHS method, we used the state-level Census 2000 modified race data (provided in Ingram et al. 2003) to calculate state-level racial composition for the bridging program provided in Appendix A.

Urbanization level. More consequential compromises must be made when working to characterize the urbanization level of each respondent's local area, given limited geographic information. The NCHS regression method measures urbanization level using four categories: large urban, large suburban, medium/small metropolitan, and nonmetropolitan. In this context,

[15] The lowest level of geography available in the Census 2000 1% public-use microdata is the Super-PUMA, which is a census-defined area containing at least 400,000 people.
[16] They used this information in two ways for their regressions predicting the probability of each potential single-race response of a multiracial respondent: (1) they calculated the racial composition of the respondent's county using the internal Census Bureau files; and (2) they coded the urbanization level of each respondent's local area.

"large" is defined as a city of 1 million or more population (see Eberhardt et al. 2001:78-80). To create these categories for respondents in publicly available microdata, the researcher needs three pieces of information: whether the person lives in a "large" city; if so, whether he/she lives in the urban part or the suburban part of the city; and if not, whether he/she lives in a smaller city as opposed to an area not defined as a city. In most situations, this information is not available. Again, compromises must be made.

For multiracial respondents in the Census 2000 5% public-use microdata file, most of the needed information can be gathered and 79.1% of respondents can be directly coded into one of these four categories at the PUMA level.[17] However, for one of two reasons, only incomplete information is available for the remaining respondents. In the analysis, 13.9% of respondents live in an identified large metropolitan area, but their central city status is not given in the data. For each of these respondents, we assign a value ($0 < U < 1$) for their "large urban" indicator variable, which represents the proportion of people in that metropolitan area who live in the central city. We assign another value ($1 - U$) indicating the proportion of people in the metropolitan area who live outside of the central city. The remaining 7.0% of respondents in the Census 2000 5% sample live in a PUMA whose composition is mixed such that it would breach confidentiality to report whether the respondent lives in a nonmetropolitan or metropolitan area. For these respondents, we assigned them the state average for each of the four urbanization-level variables such that the four variables sum to 1.

In data with state as the lowest geographic identifier, further compromises are necessary. For the program included in Appendix A, we used the full-count data from Census 2000 (SF1, Table GCT-PH1) to calculate the proportion of individuals in the state who live in each of the

[17] Using other information provided by the Census Bureau (Census 2000 Summary File 1, Table P2 or Table GCT-PH1), the "large" cities can be identified (list is available from the authors on request). Residence in a "large urban area" was coded as inside of the area's "central city".

four types of places. Then we assigned each person in the state the same value for each of the four urbanization level indicators. For example, 29.59% of people in Minnesota in 2000 (1,456,119 of 4,919,479) lived in nonmetropolitan areas. Thus every resident of Minnesota is assigned a value of .2959 for their “nonmetropolitan” variable. This geographic restriction forces the questionable—but for our purposes necessary—assumption that all residents are equally likely to live in a nonmetropolitan area; in truth, multiracial individuals are geographically concentrated in complex ways (Jones and Symens Smith 2001; Farley 2002).[18]

Selecting a data set. We used data from Census 2000 in the imputations and calculations above, as well as in the bridging program provided in Appendix A. We use this data source, rather than a more recent one, for three reasons. First, Census 2000 is widely used and it is appropriate to build race bridges using contemporaneous data. Second, the Census 2000 full-count population data (SF1) provides detailed information that is not available elsewhere. Substituting more modern data for some of the imputations and calculations would muddle the situation further. Third, most non-census datasets to which one might apply this race bridge method are at most only a few years younger than Census 2000 and thus are served reasonably well by a bridge based on the year 2000. Because bridging proportions change over time (Schenker and Parker 2003), however, the numbers in the program in Appendix A will gradually become less accurate in predicting single-race responses will need to be updated using results from the 2010 Census.

APPLICATION OF THE METHOD

[18] To improve upon this assumption, a researcher would use information about which types of multiracial individuals are likely to live in which cities, including whether they are likely to live in the central city areas and how much their distribution is different from that of the single-race population. This refinement is beyond the scope of this paper.

Our bridging regression equations use individual-level information about multiple-race people in 11 multiple-race categories to assign each person four weights. Each weight represents the predicted probability that the person would have reported that single race (AIAN, API, B, or W) if forced to choose only one. These weights can be used for fractional assignment or, alternatively, the individuals can be assigned to their most heavily weighted single race. In this section, we provide a brief overview of our bridging program and give examples of how to use the variables it creates.

The data in need of bridging must be individual-level data. Each multiple-race individual's age, sex, Hispanic origin (yes/no), and state of residence must be included. Specific values and variable names used by the bridging program are listed at the top of the program in Appendix A. Only the 11 multiple-race responses found in the modified race data format can be bridged using this program; the codes and values of these are also listed at the top of the program. Five variables are created by the bridging program. The first four – AIPROB, APIPROB, BPROB, and WPROB – represent the probability that the individual would have reported each of the four single races. These can be interpreted as weights and can be used for fractional assignment. The fifth created variable – ONERACE – provides the single race which the person is most likely to have reported. ONERACE is the variable to use if whole assignment is preferred.

Practical Issues with Fractional Assignment. To incorporate other respondents, researchers using the four variables AIPROB, APIPROB, BPROB, and WPROB will need to assign single-race individuals a value of 1 on the relevant variable. Researchers may also need to create a variable indicating a single-race "some other race" response (included here as SORPROB) if the source data include this category. Using these variables in multivariate analyses simply requires that the researcher include this set of continuous variables in the model instead of the more familiar strategy of measuring race through dummy variables.

The advantage of using fractional assignment over whole assignment is best shown through example. The data given below represent a single-race white person (Person 1), as well as an AIAN/white person who is likely to have said white single race if forced to choose, but has a non-zero probability of choosing AIAN (Person 2). Person 3 is a different AIAN/white person whose characteristics and context (not shown) make it more probable that she or he would have reported American Indian in the past (AIPROB = .546). The final column (ONERACE) provides the single race that is assigned to each person using whole assignment.

	AIPROB	APIPROB	BPROB	SORPROB	WPROB	ONERACE
Person 1	0	0	0	0	1	white
Person 2	.101	0	0	0	.899	white
Person 3	.546	0	0	0	.454	AIAN

Note that unlike whole assignment, fractional assignment retains information about each of the respondents' self-reported races—an identification that included two or more groups on purpose. As we show in the next section, this sensitivity greatly increases the ability of the bridge to approximate the results that would have been produced using the old "mark one" classification system.

Practical Issues with Whole Assignment. Despite having desirable properties, some researchers will find the fractional assignment too cumbersome to use effectively. For this situation, we generate the ONERACE variable, which represents whole assignment – the assignment of the whole case to a single race, based on the multiracial respondent's personal characteristics and context. Whole assignment approximates fractional assignment in cases, such as Person 2, above, where the highest probability variable is close to 1. In cases such as Person 3, however, more error is introduced by the use of the whole assignment method. Groups like

AIAN/API whose mean probabilities are typically near 0.5 are especially affected by the decision to use whole assignment.

If ONERACE is used in a multivariate analysis, we recommend also including an indicator of whether the respondent reported multiple races. Because of increased precision and decreased bias, we favor the fractional allocation method represented in the probability variables. However, the whole assignment method provided here is preferable to all of the other whole assignment methods discussed previously because it incorporates other important information about the individual's into the prediction of his or her most likely single race. Both recoding methods given here allow for meaningful variation in the single-race assignments of multiracial people.

A Cautionary Note. At the individual level, bridged race should be treated with caution; it is a point estimate with high standard error. The independent variables in the bridging regression explain only a small part of the variance captured in the complex race question (Schenker and Parker 2003; Parker et al. 2004). Bridged estimates were developed with the intention that they be used at the aggregate level so that errors at the individual level would average out. One consequence of high error at the individual level is that bridged race is not appropriate for use as a dependent variable. This is especially true if predictors include age, sex, Hispanic origin, and/or racial context measures.

RESULTS

To evaluate our results, we compare our fractional assignment weights (calculated using state-level geographic information) to those calculated by the NCHS team (using county-level diversity data and non-imputed urbanization data). The comparisons are intended to reveal the biases inherently introduced when using state-level data instead of more detailed data.

In Table 3, we present the mean value for each of the fractional assignment variables (AIPROB, APIPROB, BPROB, and WPROB) within each of the 11 multiple-race categories. For example, when county racial composition and true urbanization information is used in the calculation, multiracial AIAN/API respondents have a mean probability of .404 of picking American Indian/Alaskan Native over Asian/Pacific Islander, if forced to choose just one. Using publicly available state-level detail about racial diversity and imputing urbanization information decreases this mean probability to .363. This artificial difference in means is a direct result of the compromises necessitated by state-level geographic information. The difference between the NCHS regression method and our modified regression method is relatively large for the AIAN/API group and for the API/W group, and is slight for the other nine multiracial groups. The reason for these differences is not surprising given the necessary compromises—unlike other racial groups, Asian/Pacific Islanders (including those who are multiracial) are especially likely to live in large urbanized areas, while American Indians (including those who are multiracial) are especially likely to live in rural areas. State-level data hides these variations in context.

TABLE 3 ABOUT HERE

Figure 2 provides a demonstration of the results of applying our modified regression method to census data in order to better represent real temporal changes in a population's characteristics, rather than changes induced by the racial classification method. Figure 2 is similar to Figure 1, discussed above, but also includes estimates of American Indian income calculated using the variables generated by using our modified regression method. The additional lines illustrate the face validity of our method to describe trends across time. The race question in the CPS did not change between 1997 and 2002, and so changes in the CPS estimates of American Indian income can be seen as real changes. Compared to estimates yielded using the all-inclusive method (of categorizing all part-American Indians as American Indian) and the exclusionary method (of including only single-race American Indians) for coding Census 2000 public use microdata, the modified regression method more closely approximates CPS estimates for each of the years in question. Whereas the estimates for the all-inclusive and exclusionary methods vary by as much as 5 percent from the CPS benchmark, at more than one time point the modified regression method results—particularly those that use fractional assignment—nearly replicate the CPS results exactly.

DISCUSSION

Researchers in sociology (Snipp 2003), public health (Mays et al. 2003; NCVHS 2005), and education policy (Renn and Lunceford 2004; Warren and Halpern-Manners 2006) have highlighted the need for new methods for incorporating newly complex race data into historical analyses that require time-series data with consistent measures of race. This paper represents an effort to provide useful tools for researchers working with modern survey data; tools that are at once sensitive to the complexities of race and the need for historical compatibility.

Unlike previous methods, researchers can apply the bridging technique that we have described to a wide variety of commonly used and publicly available microdata sets, thereby avoiding the pitfalls of folding all multiracial persons into a single residual category or dropping such cases altogether. In addition to reflecting respondents' choices as much as possible, the modified regression method allows researchers to carry out more accurate cross-time comparisons by retaining historically consistent and substantively meaningful groupings of people. In other words, this method is both preferable and practical.

We have presented two ways of applying the modified regression method: fractional assignment and whole assignment.   Both methods take into account multiracial respondents' key characteristics and contextual information in order to predict the person's likely single race answer to a forced-choice question. The whole assignment method presented here provides the single race most likely to be reported by each multiracial respondent; the fractional assignment method is slightly more cumbersome, but represents a more nuanced approach to a complex situation by providing non-zero predicted probabilities (or weights) of each race response provided by the multiracial respondent. Both methods, we must caution, represent no more than educated guesses about an unobserved situation.

Because of the fluidity and context-dependence of race, measuring it at all in a survey remains inherently challenging. However, the multiple-race population exists and is increasing. Emasculating multiracial responses through aggregation or exclusion introduces bias and misrepresents populations. By disseminating a sophisticated, practical, and well-documented approach to race bridging, we hope to advance researchers' ability to communicate with one another about the approach employed and allow them to retain much of the meaningful information that *is* able to be gathered through a survey question about race.

REFERENCES

Allen, James P. and Eugene Turner. 2001. “Bridging 1990 and 2000 Census Race Data: Fractional Assignment of Multiracial Populations.” *Population Research and Policy Review* 20:513-533.

Burhansstipanov, Linda and Delight E. Satter. 2000. “Office of Management and Budget Racial Categories and Implications for American Indians and Alaska Natives.” *American Journal of Public Health* 90(11):1720-1723.

Cornell, Stephen and Douglas Hartmann. 1998. *Ethnicity and Race: Making Identities in a Changing World*. Thousand Oaks, CA: Pine Forge Press.

Eberhardt, Mark S., Deborah D. Ingram, Diane M. Makuc et al. 2001. “Urban and Rural Health Chartbook.” Health United States, 2001. Hyattsville, MD: National Center for Health Statistics.

Farley, Reynolds. 2002. “Racial Identities in 2000: The Response to the Multiple-Race Response Option.” Pp. 33-61 in *The New Race Question: How the Census Counts Multiracial Individuals,* Joel Perlmann and Mary C. Waters, eds. New York: Russell Sage Foundation.

Goldstein, Joshua R. and Ann J. Morning. 2000. “The Multiple-Race Population of the United States: Issues and Estimates.” *Proceedings of the National Academy of Sciences* 97(11):6230-6235.

Grieco, Elizabeth M. 2002. “An Evaluation of Bridging Methods Using Race Data from Census 2000.” *Population Research and Policy Review* 21:91-107.

Harris, David R. and Jeremiah Joseph Sim. 2002. “Who is Multiracial? Assessing the Complexity of Lived Race.” *American Sociological Review* 67:614-627.

Harrison, Roderick J. 2002. “Inadequacies of Multiple-Response Race Data in the Federal Statistical System” Pp. 137-160 in *The New Race Question: How the Census Counts Multiracial Individuals,* Joel Perlmann and Mary C. Waters, eds. New York: Russell Sage Foundation.

Heck, Katherine E., Jennifer D. Parker, C. Jane McKendry, and Gilberto F. Chávez. 2003. “Mind the Gap: Bridge Methods to Allocate Multiple-Race Mothers in Trend Analyses of Birth Certificate Data.” *Maternal and Child Health Journal* 7(1):65-70.

Ingram, Deborah D., Jennifer D. Parker, Nathaniel Schenker, James A. Weed, Brady Hamilton, Elizabeth Arias, and Jennifer H. Madans. 2003. “United States Census 2000 Population with Bridged Race Categories.” *Vital Health Statistics* 2(135). National Center for Health Statistics.

Jones, Nicholas A. and Amy Symens Smith. 2001. “The Two or More Races Population: 2000.” Census 2000 Brief C2KBR/01-6.

Kana’iaupuni, Shawn M. and Carolyn A. Liebler. 2005. “Pondering Poi Dog: Place and Racial Identification of Multiracial Native Hawaiians.” *Ethnic and Racial Studies* 28(4):687-721.

LaVeist, Thomas A. 1994. “Beyond Dummy Variables and Sample Selection: What Health Services Researchers Ought to Know about Race as a Variable.” *Health Services Research* 29(1):1-16.

Lee, Sharon M. 2001. “Using the New Racial Categories in the 2000 Census.” Washington: The Annie E. Casey Foundation and The Population Reference Bureau.

Liebler, Carolyn A. 2001. *The Fringes of American Indian Identity.* Dissertation. University of Wisconsin-Madison.

Liebler, Carolyn A. 2004. “Ties on the Fringes of Identity.” *Social Science Research* 33:702-723.

Lopez, Ian F. Haney. 1996. *White by Law: The Legal Construction of Race*. NY: New York University Press

Mays, Vickie M., Ninez A. Ponce, Donna L. Washington, and Susan D. Cochran. 2003. “Classification of Race and Ethnicity: Implications for Public Health.” *Annual Review of Public Health* 24:83-110.

National Committee on Vital and Health Statistics (NCVHS). 2005. *Eliminating Health Disparities: Strengthening Data on Race, Ethnicity, and Primary Language in the United States.* U.S. Department of Health and Human Services. Available at: http://www.cdc.gov/nchs/data/misc/EliHealthDisp.pdf

Nobles, Melissa. 2000. Shades of Citizenship: Race and Census in Modern Politics. Stanford, CA: Stanford Univ. Press.

Office of Management and Budget (OMB) 1997. “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.” Federal Register 62FR58781-58790. Available at http://www.whitehouse.gov/omb/fedreg/1997standards.html.

Office of Management and Budget (OMB). 2000. *Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity*. Available at http://www.whitehouse.gov/omb/inforeg/re_guidance2000update.pdf.

Parker, Jennifer D. and Diane M. Makuc. 2002. “Methodologic Implications of Allocating Multiple-Race Data to Single-Race Categories.” *Health Services Research* 37(1):201-213.

Parker, Jennifer D. Nathaniel Schenker, Deborah D. Ingram, James A. Weed, Katherine E. Heck, and Jennifer H. Madans. 2004. “Bridging Between Two Standards for Collecting Information on Race and Ethnicity: An Application to Census 2000 and Vital Rates.” *Public Health Reports* 119:192-205.

Renn, Kristen A. 2004. *Mixed Race Students in College: the Ecology of Race, Identity, and Community on Campus*. Albany, NY: SUNY Press.

Renn, Kristen A. and Christina J. Lunceford. 2004. “Because the Numbers Matter: Transforming Postsecondary Education Data on Student Race and Ethnicity to Meet the Challenges of a Changing Nation.” *Educational Policy* 18(5):752-783.

Rockquemore, Kerry Ann and David L. Brunsma. 2002. *Beyond Black: Biracial Identity in America*. Thousand Oaks, CA: Sage.

Rodríguez, Clara E. 2000. *Changing Race : Latinos, the Census, and the History of Ethnicity in the United States*. NY: New York University Press

Root, Maria P.P. (ed.) 1996. *The Multiracial Experience: Racial Borders as the New Frontier*. Newbury Park, CA: Sage.

Schenker, Nathaniel and Jennifer D. Parker. 2003. “From single-race reporting to multiple-race reporting: using imputation methods to bridge the transition.” *Statistics in Medicine* 22: 1571–87.

Snipp, C. Matthew. 2003. “Racial Measurement in the American Census: Past Practices and Implications for the Future.” *Annual Review of Sociology*. 29:563-588.

Wallace, Kendra R. 2001. *Relative/Outsider: The Art and Politics of Identity among Mixed Heritage Students*. Westport, CT: Ablex.

Warren, John Robert and Andrew Halpern-Manners. 2006. “Is the Glass Emptying or Filling Up? Reconciling Divergent Trends in High School Completion and Dropout.” Working paper.

Waters, Mary C. 1999. *Black Identities: West Indian Immigrant Dreams and American Realities*. Cambridge, MA: Harvard University Press.

Waters, Mary C. 2000. “Immigration, Intermarriage, and the Challenges of Measuring Racial/Ethnic Identities.” *American Journal of Public Health* 90(11):1735-1737.

Xie, Yu and Kimberly Goyette. 1998. "The Racial Identification of Biracial Children with One Asian Parent: Evidence from the 1990 Census." *Social Forces* 76:547-570.

Zuckerman, Marvin. 1990. “Some Dubious Premises in Research and Theory on Racial Differences: Scientific, Social, and Ethical Issues.” *American Psychologist* 45(12):1297-1303.

Table 1. Race group differences on selected economic, demographic, and geographic characteristics in 2000

	American Indian/ Alaskan Native		Pacific Islander		Asian		Black	
	Multiracial[a]	Monoracial[b]	Multiracial	Monoracial	Multiracial	Monoracial	Multiracial	Monoracial
Mean personal income (ages 15+)	21,522†	17,879	21,121†	20,338	22,882†	27,449	20,851†	28,722
Percent living in a rural area	18.0†	33.7						
Percent living in Hawaii			34.1†	29.7				
Percent under age 18	31.6	33.1	39.0†	31.3	44.7†	23.9	36.3†	24.4

Source : Integrated Public-Use Microdata Series (IPUMS), Census 2000, 5% sample

† Indicates a statistically significant difference ($p < .001$) between multiracial and monoracial subgroups.

[a] Includes persons who reported, for example, American Indian/Alaskan Native as well as other racial group(s).

[b] Includes persons who only reported one race.

Table 2. Definitions of Key Terms

Term	***Definition***
Whole assignment	Each multiple-race respondent is assigned to only one single-race category
Fractional assignment	Each multiple-race respondent is assigned a weight for each of their reported races. Weights sum to 1
AIAN	American Indian and/or Alaskan Native
API	Asian and/or Pacific Islander
B	Black
W	White
SOR	Some other race
PUMA	Public Use Microdata Area
Modified regression method	A practical method (presented in this paper) for simplifying complex race codes while retaining substantive meaning.
AIPROB	Fractional assignment weight variable for American Indian/Alaska Native. Calculated using the modified regression method.
APIPROB	Fractional assignment weight variable for Asian/Pacific Islander. Calculated using the modified regression method.
BPROB	Fractional assignment weight variable for black. Calculated using the modified regression method.
WPROB	Fractional assignment weight variable for white. Calculated using the modified regression method.
ONERACE	Race variable representing whole assignment. Calculated using the modified regression method.

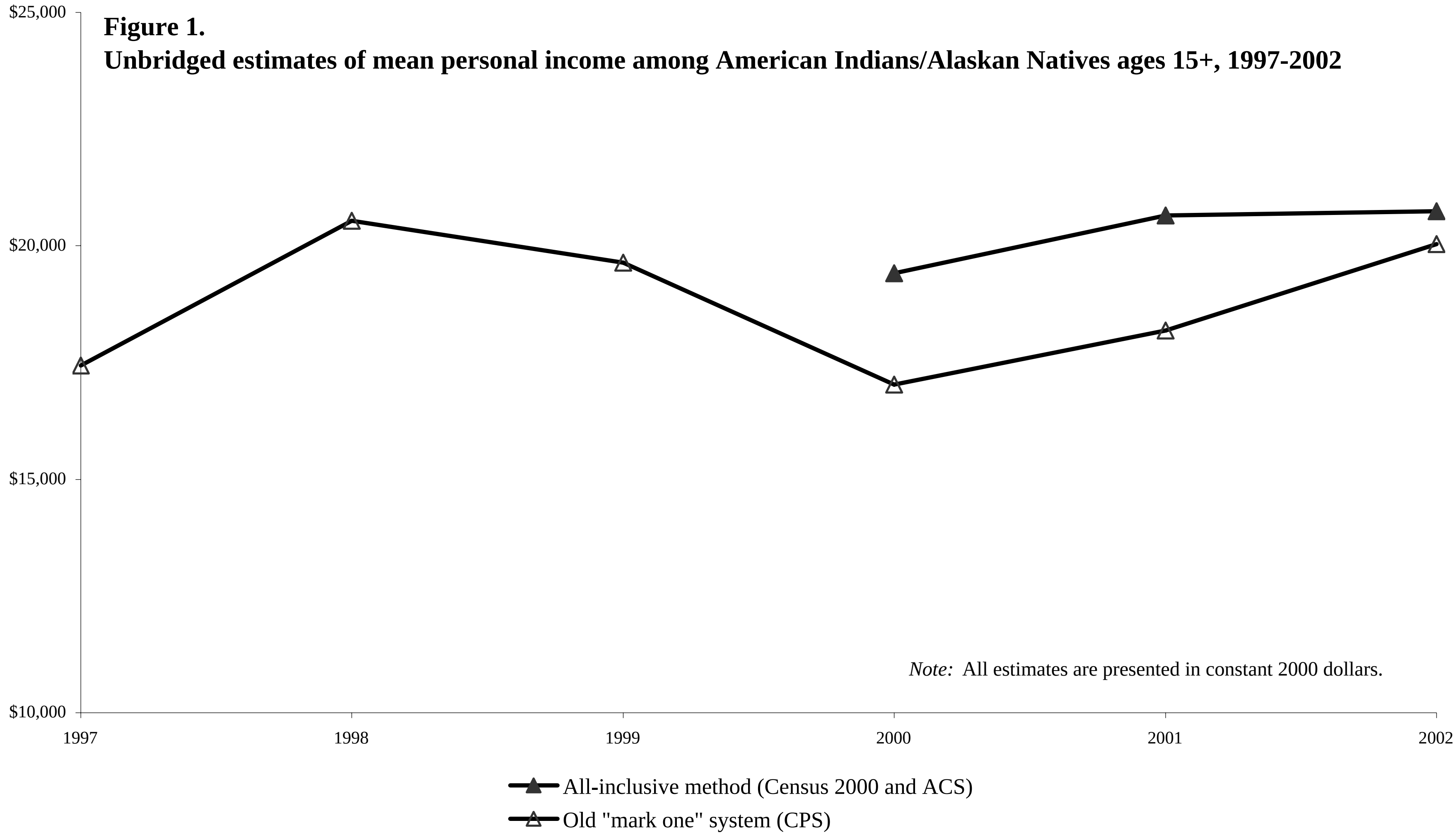
Table 3. Mean Probability of Assignment to Each Single-Race Group, by Level of Geographic Information

Multiple-Race Response	AIAN		API		Black		White	
	County†	State^	County†	State^	County†	State^	County†	State^
AIAN & API	0.404	0.363	0.596	0.637				
AIAN & black	0.186	0.163			0.814	0.838		
AIAN & white	0.205	0.221					0.795	0.779
API & black			0.370	0.350	0.630	0.650		
API & white			0.327	0.401			0.673	0.599
Black & white					0.621	0.607	0.379	0.394
AIAN & API & black	0.286	0.327	0.253	0.255	0.461	0.418		
AIAN & API & white	0.024	0.023	0.043	0.084			0.933	0.893
AIAN & black & white	0.195	0.192			0.572	0.626	0.233	0.182
API & black & white			0.104	0.103	0.113	0.098	0.782	0.800
AIAN & API & black & white	0.010	0.010	0.009	0.009	0.020	0.013	0.960	0.967

† NCHS regression estimates using restricted-use county-level data are from Ingram et al., 2003, Table 9.

^ State columns use the Census 2000 5% microdata and represent the probability variables generated by the modified regression method presented here and the program provided in Appendix A.

Figure 1.
Unbridged estimates of mean personal income among American Indians/Alaskan Natives ages 15+, 1997-2002



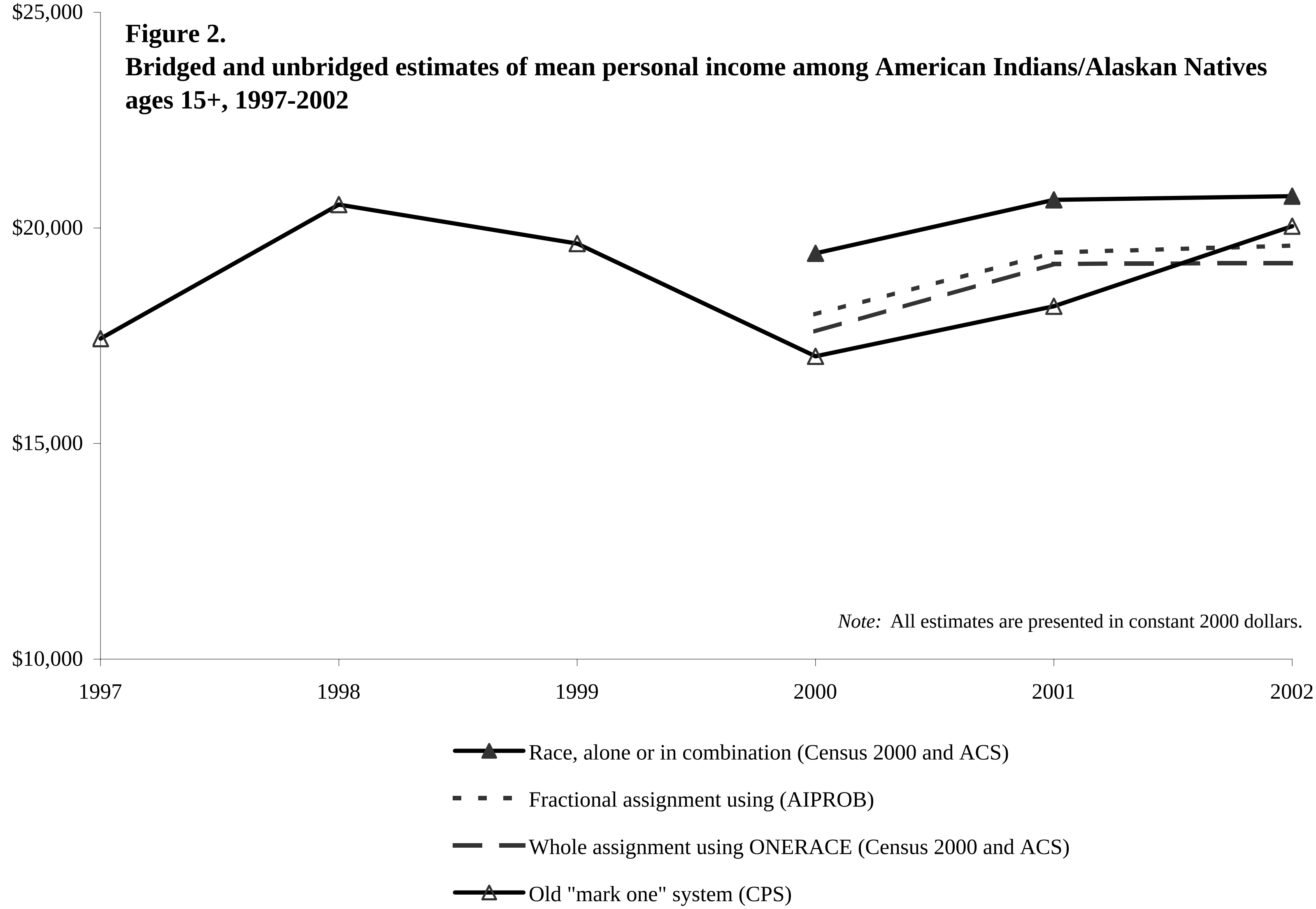


Figure 2.
Bridged and unbridged estimates of mean personal income among American Indians/Alaskan Natives ages 15+, 1997-2002

Appendix A

```
#delimit ;
*********************************************************************

This program is designed to bridge multiple-race responses to single-
race codes using limited geographic information.

It is based on the method and results reported in Ingram DD, Parker JD,
Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. "United States
Census 2000 population with bridged race categories." National Center
for Health Statistics. Vital Health Stat 2(135). 2003.

The following abbreviations are used below:
 AIAN = American Indian and/or Alaska Native
 API = Asian and/or Pacific Islander
 B = black or African American
 W = white
 SOR = "some other race"
Note that "other race" responses are ignored in this bridging method.

*********************************************************************

The following variables are used in the program and must be provided
for each individual in the input file. Note that variable names are
case-sensitive in STATA.

MRDRACE
1  = AIAN and API
2  = AIAN and B
3  = AIAN and W
4  = API and B
5  = API and W
6  = B and W
7  = AIAN and API and B
8  = AIAN and API and W
9  = AIAN and B and W
10 = API and B and W
11 = AIAN and API and B and W

20 = AIAN only
30 = API only
40 = B only
50 = W only
60 = SOR only

AGE10
continuous variable -- respondent's age divided by 10

HISP
1 = of Spanish, Hispanic, or Latino origin
0 = not Spanish, Hispanic, or Latino

MALE
1 = male
0 = female

STATEFIP
FIPS coding is used for all states and Washington, DC
01 = Alabama
02 = Alaska
04 = Arizona
05 = Arkansas
06 = California
08 = Colorado
09 = Connecticut
10 = Delaware
11 = District of Columbia
12 = Florida
13 = Georgia
15 = Hawaii
```

```
16 = Idaho
17 = Illinois
18 = Indiana
19 = Iowa
20 = Kansas
21 = Kentucky
22 = Louisiana
23 = Maine
24 = Maryland
25 = Massachusetts
26 = Michigan
27 = Minnesota
28 = Mississippi
29 = Missouri
30 = Montana
31 = Nebraska
32 = Nevada
33 = New Hampshire
34 = New Jersey
35 = New Mexico
36 = New York
37 = North Carolina
38 = North Dakota
39 = Ohio
40 = Oklahoma
41 = Oregon
42 = Pennsylvania
44 = Rhode Island
45 = South Carolina
46 = South Dakota
47 = Tennessee
48 = Texas
49 = Utah
51 = Virginia
50 = Vermont
53 = Washington
54 = West Virginia
55 = Wisconsin
56 = Wyoming

*********************************************************************

The following variables are created by the program. The "PROB" variables
range from 0 to 1 and sum to 1. ONERACE is the single race most likely
to have been reported by that individual if he/she was forced to choose.

AIPROB  = probability of reporting American Indian and/or Alaskan Native
APIPROB = probability of reporting Asian and/or Pacific Islander
BPROB = probability of reporting black/African American
WPROB = probability of reporting white

ONERACE = historically compatible bridged single race
1 = American Indian and/or Alaska Native
2 = Asian and/or Pacific Islander
3 = Black/African American
4 = White

*********************************************************************
*********************************************************************
*********************************************************************
;
* Source: The racial composition of states as of April 1, 2000
are from Ingram et al. 2003, Table 2, which was derived from
the Census 2000 Modified Race Data Summary File;

gen lnpctAI=0;
 label variable lnpctAI "natural log of %AI in state";
replace lnpctAI=        -0.693147181    if STATEFIP==   1       ;
replace lnpctAI=        2.753660712     if STATEFIP==   2       ;
```

```
replace lnpctAI=        1.648658626     if STATEFIP==   4       ;
replace lnpctAI=        -0.356674944    if STATEFIP==   5       ;
replace lnpctAI=        0.182321557     if STATEFIP==   6       ;
replace lnpctAI=        0.182321557     if STATEFIP==   8       ;
replace lnpctAI=        -1.203972804    if STATEFIP==   9       ;
replace lnpctAI=        -0.916290732    if STATEFIP==   10      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   11      ;
replace lnpctAI=        -0.916290732    if STATEFIP==   12      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   13      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   15      ;
replace lnpctAI=        0.336472237     if STATEFIP==   16      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   17      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   18      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   19      ;
replace lnpctAI=        0               if STATEFIP==   20      ;
replace lnpctAI=        -1.609437912    if STATEFIP==   21      ;
replace lnpctAI=        -0.510825624    if STATEFIP==   22      ;
replace lnpctAI=        -0.510825624    if STATEFIP==   23      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   24      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   25      ;
replace lnpctAI=        -0.510825624    if STATEFIP==   26      ;
replace lnpctAI=        0.09531018      if STATEFIP==   27      ;
replace lnpctAI=        -0.916290732    if STATEFIP==   28      ;
replace lnpctAI=        -0.693147181    if STATEFIP==   29      ;
replace lnpctAI=        1.824549292     if STATEFIP==   30      ;
replace lnpctAI=        -0.105360516    if STATEFIP==   31      ;
replace lnpctAI=        0.336472237     if STATEFIP==   32      ;
replace lnpctAI=        -1.609437912    if STATEFIP==   33      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   34      ;
replace lnpctAI=        2.282382386     if STATEFIP==   35      ;
replace lnpctAI=        -0.693147181    if STATEFIP==   36      ;
replace lnpctAI=        0.262364264     if STATEFIP==   37      ;
replace lnpctAI=        1.589235205     if STATEFIP==   38      ;
replace lnpctAI=        -1.609437912    if STATEFIP==   39      ;
replace lnpctAI=        2.079441542     if STATEFIP==   40      ;
replace lnpctAI=        0.336472237     if STATEFIP==   41      ;
replace lnpctAI=        -1.609437912    if STATEFIP==   42      ;
replace lnpctAI=        -0.510825624    if STATEFIP==   44      ;
replace lnpctAI=        -0.916290732    if STATEFIP==   45      ;
replace lnpctAI=        2.116255515     if STATEFIP==   46      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   47      ;
replace lnpctAI=        -0.356674944    if STATEFIP==   48      ;
replace lnpctAI=        0.336472237     if STATEFIP==   49      ;
replace lnpctAI=        -0.916290732    if STATEFIP==   50      ;
replace lnpctAI=        -1.203972804    if STATEFIP==   51      ;
replace lnpctAI=        0.470003629     if STATEFIP==   53      ;
replace lnpctAI=        -1.609437912    if STATEFIP==   54      ;
replace lnpctAI=        -0.105360516    if STATEFIP==   55      ;
replace lnpctAI=        0.832909123     if STATEFIP==   56      ;

gen pctAI=0;
 label variable pctAI "% Am.Ind. in state";
replace pctAI=  0.5     if STATEFIP==   1       ;
replace pctAI=  15.7    if STATEFIP==   2       ;
replace pctAI=  5.2     if STATEFIP==   4       ;
replace pctAI=  0.7     if STATEFIP==   5       ;
replace pctAI=  1.2     if STATEFIP==   6       ;
replace pctAI=  1.2     if STATEFIP==   8       ;
replace pctAI=  0.3     if STATEFIP==   9       ;
replace pctAI=  0.4     if STATEFIP==   10      ;
replace pctAI=  0.3     if STATEFIP==   11      ;
replace pctAI=  0.4     if STATEFIP==   12      ;
replace pctAI=  0.3     if STATEFIP==   13      ;
replace pctAI=  0.3     if STATEFIP==   15      ;
replace pctAI=  1.4     if STATEFIP==   16      ;
replace pctAI=  0.3     if STATEFIP==   17      ;
replace pctAI=  0.3     if STATEFIP==   18      ;
replace pctAI=  0.3     if STATEFIP==   19      ;
replace pctAI=  1       if STATEFIP==   20      ;
```

```
replace pctAI=  0.2     if STATEFIP==   21      ;
replace pctAI=  0.6     if STATEFIP==   22      ;
replace pctAI=  0.6     if STATEFIP==   23      ;
replace pctAI=  0.3     if STATEFIP==   24      ;
replace pctAI=  0.3     if STATEFIP==   25      ;
replace pctAI=  0.6     if STATEFIP==   26      ;
replace pctAI=  1.1     if STATEFIP==   27      ;
replace pctAI=  0.4     if STATEFIP==   28      ;
replace pctAI=  0.5     if STATEFIP==   29      ;
replace pctAI=  6.2     if STATEFIP==   30      ;
replace pctAI=  0.9     if STATEFIP==   31      ;
replace pctAI=  1.4     if STATEFIP==   32      ;
replace pctAI=  0.2     if STATEFIP==   33      ;
replace pctAI=  0.3     if STATEFIP==   34      ;
replace pctAI=  9.8     if STATEFIP==   35      ;
replace pctAI=  0.5     if STATEFIP==   36      ;
replace pctAI=  1.3     if STATEFIP==   37      ;
replace pctAI=  4.9     if STATEFIP==   38      ;
replace pctAI=  0.2     if STATEFIP==   39      ;
replace pctAI=  8       if STATEFIP==   40      ;
replace pctAI=  1.4     if STATEFIP==   41      ;
replace pctAI=  0.2     if STATEFIP==   42      ;
replace pctAI=  0.6     if STATEFIP==   44      ;
replace pctAI=  0.4     if STATEFIP==   45      ;
replace pctAI=  8.3     if STATEFIP==   46      ;
replace pctAI=  0.3     if STATEFIP==   47      ;
replace pctAI=  0.7     if STATEFIP==   48      ;
replace pctAI=  1.4     if STATEFIP==   49      ;
replace pctAI=  0.4     if STATEFIP==   50      ;
replace pctAI=  0.3     if STATEFIP==   51      ;
replace pctAI=  1.6     if STATEFIP==   53      ;
replace pctAI=  0.2     if STATEFIP==   54      ;
replace pctAI=  0.9     if STATEFIP==   55      ;
replace pctAI=  2.3     if STATEFIP==   56      ;


gen pctAPI=0;
 label variable pctAPI "% Asian or PI in state";
replace pctAPI= 0.8     if STATEFIP==   1       ;
replace pctAPI= 4.7     if STATEFIP==   2       ;
replace pctAPI= 2.1     if STATEFIP==   4       ;
replace pctAPI= 0.9     if STATEFIP==   5       ;
replace pctAPI= 11.8    if STATEFIP==   6       ;
replace pctAPI= 2.5     if STATEFIP==   8       ;
replace pctAPI= 2.6     if STATEFIP==   9       ;
replace pctAPI= 2.2     if STATEFIP==   10      ;
replace pctAPI= 2.8     if STATEFIP==   11      ;
replace pctAPI= 1.8     if STATEFIP==   12      ;
replace pctAPI= 2.3     if STATEFIP==   13      ;
replace pctAPI= 56.7    if STATEFIP==   15      ;
replace pctAPI= 1.1     if STATEFIP==   16      ;
replace pctAPI= 3.6     if STATEFIP==   17      ;
replace pctAPI= 1       if STATEFIP==   18      ;
replace pctAPI= 1.3     if STATEFIP==   19      ;
replace pctAPI= 1.9     if STATEFIP==   20      ;
replace pctAPI= 0.8     if STATEFIP==   21      ;
replace pctAPI= 1.3     if STATEFIP==   22      ;
replace pctAPI= 0.8     if STATEFIP==   23      ;
replace pctAPI= 4.1     if STATEFIP==   24      ;
replace pctAPI= 4       if STATEFIP==   25      ;
replace pctAPI= 1.9     if STATEFIP==   26      ;
replace pctAPI= 3       if STATEFIP==   27      ;
replace pctAPI= 0.7     if STATEFIP==   28      ;
replace pctAPI= 1.2     if STATEFIP==   29      ;
replace pctAPI= 0.6     if STATEFIP==   30      ;
replace pctAPI= 1.4     if STATEFIP==   31      ;
replace pctAPI= 5.3     if STATEFIP==   32      ;
replace pctAPI= 1.4     if STATEFIP==   33      ;
replace pctAPI= 6       if STATEFIP==   34      ;
```

```
replace pctAPI= 1.3     if STATEFIP==   35      ;
replace pctAPI= 5.9     if STATEFIP==   36      ;
replace pctAPI= 1.5     if STATEFIP==   37      ;
replace pctAPI= 0.6     if STATEFIP==   38      ;
replace pctAPI= 1.2     if STATEFIP==   39      ;
replace pctAPI= 1.5     if STATEFIP==   40      ;
replace pctAPI= 3.4     if STATEFIP==   41      ;
replace pctAPI= 1.9     if STATEFIP==   42      ;
replace pctAPI= 2.5     if STATEFIP==   44      ;
replace pctAPI= 1       if STATEFIP==   45      ;
replace pctAPI= 0.6     if STATEFIP==   46      ;
replace pctAPI= 1.1     if STATEFIP==   47      ;
replace pctAPI= 2.9     if STATEFIP==   48      ;
replace pctAPI= 2.5     if STATEFIP==   49      ;
replace pctAPI= 0.9     if STATEFIP==   50      ;
replace pctAPI= 3.9     if STATEFIP==   51      ;
replace pctAPI= 6.1     if STATEFIP==   53      ;
replace pctAPI= 0.6     if STATEFIP==   54      ;
replace pctAPI= 1.8     if STATEFIP==   55      ;
replace pctAPI= 0.7     if STATEFIP==   56      ;

gen pctB=0;
 label variable pctB "% Black in state";
replace pctB=   26.1    if STATEFIP==   1       ;
replace pctB=   3.7     if STATEFIP==   2       ;
replace pctB=   3.3     if STATEFIP==   4       ;
replace pctB=   15.7    if STATEFIP==   5       ;
replace pctB=   7       if STATEFIP==   6       ;
replace pctB=   4       if STATEFIP==   8       ;
replace pctB=   9.8     if STATEFIP==   9       ;
replace pctB=   19.6    if STATEFIP==   10      ;
replace pctB=   61.1    if STATEFIP==   11      ;
replace pctB=   15.2    if STATEFIP==   12      ;
replace pctB=   29      if STATEFIP==   13      ;
replace pctB=   1.9     if STATEFIP==   15      ;
replace pctB=   0.5     if STATEFIP==   16      ;
replace pctB=   15.3    if STATEFIP==   17      ;
replace pctB=   8.5     if STATEFIP==   18      ;
replace pctB=   2.2     if STATEFIP==   19      ;
replace pctB=   5.9     if STATEFIP==   20      ;
replace pctB=   7.4     if STATEFIP==   21      ;
replace pctB=   32.6    if STATEFIP==   22      ;
replace pctB=   0.6     if STATEFIP==   23      ;
replace pctB=   28.3    if STATEFIP==   24      ;
replace pctB=   6.4     if STATEFIP==   25      ;
replace pctB=   14.4    if STATEFIP==   26      ;
replace pctB=   3.7     if STATEFIP==   27      ;
replace pctB=   36.4    if STATEFIP==   28      ;
replace pctB=   11.3    if STATEFIP==   29      ;
replace pctB=   0.3     if STATEFIP==   30      ;
replace pctB=   4.1     if STATEFIP==   31      ;
replace pctB=   7       if STATEFIP==   32      ;
replace pctB=   0.8     if STATEFIP==   33      ;
replace pctB=   14.4    if STATEFIP==   34      ;
replace pctB=   2.1     if STATEFIP==   35      ;
replace pctB=   17.6    if STATEFIP==   36      ;
replace pctB=   21.8    if STATEFIP==   37      ;
replace pctB=   0.6     if STATEFIP==   38      ;
replace pctB=   11.6    if STATEFIP==   39      ;
replace pctB=   7.7     if STATEFIP==   40      ;
replace pctB=   1.7     if STATEFIP==   41      ;
replace pctB=   10.3    if STATEFIP==   42      ;
replace pctB=   5.7     if STATEFIP==   44      ;
replace pctB=   29.7    if STATEFIP==   45      ;
replace pctB=   0.7     if STATEFIP==   46      ;
replace pctB=   16.5    if STATEFIP==   47      ;
replace pctB=   11.8    if STATEFIP==   48      ;
replace pctB=   0.9     if STATEFIP==   49      ;
replace pctB=   0.5     if STATEFIP==   50      ;
```

```
replace pctB=   19.9    if STATEFIP==   51      ;
replace pctB=   3.4     if STATEFIP==   53      ;
replace pctB=   3.2     if STATEFIP==   54      ;
replace pctB=   5.8     if STATEFIP==   55      ;
replace pctB=   0.8     if STATEFIP==   56      ;

gen pctBsq=0;
  label variable pctBsq "% Black in state, squared";
replace pctBsq= 681.21  if STATEFIP==   1       ;
replace pctBsq= 13.69   if STATEFIP==   2       ;
replace pctBsq= 10.89   if STATEFIP==   4       ;
replace pctBsq= 246.49  if STATEFIP==   5       ;
replace pctBsq= 49      if STATEFIP==   6       ;
replace pctBsq= 16      if STATEFIP==   8       ;
replace pctBsq= 96.04   if STATEFIP==   9       ;
replace pctBsq= 384.16  if STATEFIP==   10      ;
replace pctBsq= 3733.21 if STATEFIP==   11      ;
replace pctBsq= 231.04  if STATEFIP==   12      ;
replace pctBsq= 841     if STATEFIP==   13      ;
replace pctBsq= 3.61    if STATEFIP==   15      ;
replace pctBsq= 0.25    if STATEFIP==   16      ;
replace pctBsq= 234.09  if STATEFIP==   17      ;
replace pctBsq= 72.25   if STATEFIP==   18      ;
replace pctBsq= 4.84    if STATEFIP==   19      ;
replace pctBsq= 34.81   if STATEFIP==   20      ;
replace pctBsq= 54.76   if STATEFIP==   21      ;
replace pctBsq= 1062.76 if STATEFIP==   22      ;
replace pctBsq= 0.36    if STATEFIP==   23      ;
replace pctBsq= 800.89  if STATEFIP==   24      ;
replace pctBsq= 40.96   if STATEFIP==   25      ;
replace pctBsq= 207.36  if STATEFIP==   26      ;
replace pctBsq= 13.69   if STATEFIP==   27      ;
replace pctBsq= 1324.96 if STATEFIP==   28      ;
replace pctBsq= 127.69  if STATEFIP==   29      ;
replace pctBsq= 0.09    if STATEFIP==   30      ;
replace pctBsq= 16.81   if STATEFIP==   31      ;
replace pctBsq= 49      if STATEFIP==   32      ;
replace pctBsq= 0.64    if STATEFIP==   33      ;
replace pctBsq= 207.36  if STATEFIP==   34      ;
replace pctBsq= 4.41    if STATEFIP==   35      ;
replace pctBsq= 309.76  if STATEFIP==   36      ;
replace pctBsq= 475.24  if STATEFIP==   37      ;
replace pctBsq= 0.36    if STATEFIP==   38      ;
replace pctBsq= 134.56  if STATEFIP==   39      ;
replace pctBsq= 59.29   if STATEFIP==   40      ;
replace pctBsq= 2.89    if STATEFIP==   41      ;
replace pctBsq= 106.09  if STATEFIP==   42      ;
replace pctBsq= 32.49   if STATEFIP==   44      ;
replace pctBsq= 882.09  if STATEFIP==   45      ;
replace pctBsq= 0.49    if STATEFIP==   46      ;
replace pctBsq= 272.25  if STATEFIP==   47      ;
replace pctBsq= 139.24  if STATEFIP==   48      ;
replace pctBsq= 0.81    if STATEFIP==   49      ;
replace pctBsq= 0.25    if STATEFIP==   50      ;
replace pctBsq= 396.01  if STATEFIP==   51      ;
replace pctBsq= 11.56   if STATEFIP==   53      ;
replace pctBsq= 10.24   if STATEFIP==   54      ;
replace pctBsq= 33.64   if STATEFIP==   55      ;
replace pctBsq= 0.64    if STATEFIP==   56      ;

gen pct2race=0;
 label variable pct2race "% 2+ races or non-Hisp 'other' in state";
replace pct2race=       0.8     if STATEFIP==   1       ;
replace pct2race=       4.6     if STATEFIP==   2       ;
replace pct2race=       1.3     if STATEFIP==   4       ;
replace pct2race=       1.1     if STATEFIP==   5       ;
replace pct2race=       2.2     if STATEFIP==   6       ;
replace pct2race=       1.6     if STATEFIP==   8       ;
replace pct2race=       1.1     if STATEFIP==   9       ;
```

```
replace pct2race=       1.1     if STATEFIP==   10      ;
replace pct2race=       1.3     if STATEFIP==   11      ;
replace pct2race=       1       if STATEFIP==   12      ;
replace pct2race=       0.9     if STATEFIP==   13      ;
replace pct2race=       15.2    if STATEFIP==   15      ;
replace pct2race=       1.3     if STATEFIP==   16      ;
replace pct2race=       0.9     if STATEFIP==   17      ;
replace pct2race=       0.9     if STATEFIP==   18      ;
replace pct2race=       0.8     if STATEFIP==   19      ;
replace pct2race=       1.5     if STATEFIP==   20      ;
replace pct2race=       0.8     if STATEFIP==   21      ;
replace pct2race=       0.7     if STATEFIP==   22      ;
replace pct2race=       0.8     if STATEFIP==   23      ;
replace pct2race=       1.3     if STATEFIP==   24      ;
replace pct2race=       1.1     if STATEFIP==   25      ;
replace pct2race=       1.3     if STATEFIP==   26      ;
replace pct2race=       1.2     if STATEFIP==   27      ;
replace pct2race=       0.6     if STATEFIP==   28      ;
replace pct2race=       1.1     if STATEFIP==   29      ;
replace pct2race=       1.5     if STATEFIP==   30      ;
replace pct2race=       1       if STATEFIP==   31      ;
replace pct2race=       2.1     if STATEFIP==   32      ;
replace pct2race=       0.8     if STATEFIP==   33      ;
replace pct2race=       1.1     if STATEFIP==   34      ;
replace pct2race=       1.4     if STATEFIP==   35      ;
replace pct2race=       1.3     if STATEFIP==   36      ;
replace pct2race=       0.9     if STATEFIP==   37      ;
replace pct2race=       0.9     if STATEFIP==   38      ;
replace pct2race=       1.1     if STATEFIP==   39      ;
replace pct2race=       4       if STATEFIP==   40      ;
replace pct2race=       2.1     if STATEFIP==   41      ;
replace pct2race=       0.8     if STATEFIP==   42      ;
replace pct2race=       1.3     if STATEFIP==   44      ;
replace pct2race=       0.7     if STATEFIP==   45      ;
replace pct2race=       1.1     if STATEFIP==   46      ;
replace pct2race=       0.8     if STATEFIP==   47      ;
replace pct2race=       0.9     if STATEFIP==   48      ;
replace pct2race=       1.2     if STATEFIP==   49      ;
replace pct2race=       1       if STATEFIP==   50      ;
replace pct2race=       1.3     if STATEFIP==   51      ;
replace pct2race=       2.6     if STATEFIP==   53      ;
replace pct2race=       0.8     if STATEFIP==   54      ;
replace pct2race=       0.9     if STATEFIP==   55      ;
replace pct2race=       1.2     if STATEFIP==   56      ;

* 14 states have no cities of more than 1 million population;
* source for total state pop & # not metro: Census table GCT-PH1: state;

gen medmetro= 0;
 label variable medmetro "MSA less than 1 million";
replace medmetro=       0.6991   if STATEFIP==  1       ;
replace medmetro=       0.4152   if STATEFIP==  2       ;
replace medmetro=       0.8001   if STATEFIP==  10      ;
replace medmetro=       0.7232   if STATEFIP==  15      ;
replace medmetro=       0.3925   if STATEFIP==  16      ;
replace medmetro=       0.4532   if STATEFIP==  19      ;
replace medmetro=       0.3660   if STATEFIP==  23      ;
replace medmetro=       0.3386   if STATEFIP==  30      ;
replace medmetro=       0.5258   if STATEFIP==  31      ;
replace medmetro=       0.5690   if STATEFIP==  35      ;
replace medmetro=       0.4422   if STATEFIP==  38      ;
replace medmetro=       0.3457   if STATEFIP==  46      ;
replace medmetro=       0.2782   if STATEFIP==  50      ;
replace medmetro=       0.3000   if STATEFIP==  56      ;

gen nonmetro= 0;
 label variable nonmetro "not in metro area";
replace nonmetro=       0.3009   if STATEFIP==  1       ;
replace nonmetro=       0.5848   if STATEFIP==  2       ;
```

```
replace nonmetro=       0.1999   if STATEFIP==  10      ;
replace nonmetro=       0.2768   if STATEFIP==  15      ;
replace nonmetro=       0.6075   if STATEFIP==  16      ;
replace nonmetro=       0.5468   if STATEFIP==  19      ;
replace nonmetro=       0.6340   if STATEFIP==  23      ;
replace nonmetro=       0.6614   if STATEFIP==  30      ;
replace nonmetro=       0.4742   if STATEFIP==  31      ;
replace nonmetro=       0.4310   if STATEFIP==  35      ;
replace nonmetro=       0.5578   if STATEFIP==  38      ;
replace nonmetro=       0.6543   if STATEFIP==  46      ;
replace nonmetro=       0.7218   if STATEFIP==  50      ;
replace nonmetro=       0.7000   if STATEFIP==  56      ;

* 11 states and DC have 1+ 'big' city and no big cities cross
  its state lines;
* source for central city population is: Census table GCT-PH1:
  Population, Housing Units, Area, and Density: 2000 (SF1)
  US and PR -- Metropolitan Area, in Central City, Not in
  Central City, County, and (in selected States) County Subdivision;

gen bigurban= 0;
 label variable bigurban "in central city, MSA >= 1 million";
replace bigurban=       0.30194982       if STATEFIP==  6       ;
replace bigurban=       0.128947302      if STATEFIP==  8       ;
replace bigurban=       0.048375233      if STATEFIP==  9       ;
replace bigurban=       1                if STATEFIP==  11      ;
replace bigurban=       0.146517683      if STATEFIP==  12      ;
replace bigurban=       0.050873559      if STATEFIP==  13      ;
replace bigurban=       0.114202672      if STATEFIP==  22      ;
replace bigurban=       0.142947226      if STATEFIP==  26      ;
replace bigurban=       0.454740787      if STATEFIP==  36      ;
replace bigurban=       0.182724202      if STATEFIP==  40      ;
replace bigurban=       0.300103396      if STATEFIP==  48      ;
replace bigurban=       0.127595807      if STATEFIP==  49      ;

gen bigsubrb= 0;
 label variable bigsubrb "outside central city, MSA >= 1 million";
replace bigsubrb=       0.461780454      if STATEFIP==  6       ;
replace bigsubrb=       0.361439587      if STATEFIP==  8       ;
replace bigsubrb=       0.299029676      if STATEFIP==  9       ;
replace bigsubrb=       0                if STATEFIP==  11      ;
replace bigsubrb=       0.488469989      if STATEFIP==  12      ;
replace bigsubrb=       0.451443867      if STATEFIP==  13      ;
replace bigsubrb=       0.185133462      if STATEFIP==  22      ;
replace bigsubrb=       0.413484445      if STATEFIP==  26      ;
replace bigsubrb=       0.300745866      if STATEFIP==  36      ;
replace bigsubrb=       0.131229616      if STATEFIP==  40      ;
replace bigsubrb=       0.28697212       if STATEFIP==  48      ;
replace bigsubrb=       0.469723071      if STATEFIP==  49      ;

replace medmetro=       0.203166702      if STATEFIP==  6       ;
replace medmetro=       0.348356912      if STATEFIP==  8       ;
replace medmetro=       0.608941541      if STATEFIP==  9       ;
replace medmetro=       0                if STATEFIP==  11      ;
replace medmetro=       0.29337837       if STATEFIP==  12      ;
replace medmetro=       0.189882725      if STATEFIP==  13      ;
replace medmetro=       0.454798594      if STATEFIP==  22      ;
replace medmetro=       0.265574873      if STATEFIP==  26      ;
replace medmetro=       0.165288916      if STATEFIP==  36      ;
replace medmetro=       0.294151775      if STATEFIP==  40      ;
replace medmetro=       0.261381836      if STATEFIP==  48      ;
replace medmetro=       0.167735626      if STATEFIP==  49      ;

replace nonmetro=       0.033103025      if STATEFIP==  6       ;
replace nonmetro=       0.161256199      if STATEFIP==  8       ;
replace nonmetro=       0.043653549      if STATEFIP==  9       ;
replace nonmetro=       0                if STATEFIP==  11      ;
replace nonmetro=       0.071633958      if STATEFIP==  12      ;
replace nonmetro=       0.307799849      if STATEFIP==  13      ;
```

```
replace nonmetro=       0.245865272      if STATEFIP==  22      ;
replace nonmetro=       0.177993457      if STATEFIP==  26      ;
replace nonmetro=       0.079224431      if STATEFIP==  36      ;
replace nonmetro=       0.391894406      if STATEFIP==  40      ;
replace nonmetro=       0.151542647      if STATEFIP==  48      ;
replace nonmetro=       0.234945497      if STATEFIP==  49      ;

* 25 states have at least one city that crosses state lines;
* Source for proportions of city populations in each state: Census
  table GCT-PH1: Population, Housing Units Area, and Density: 2000 (SF1)
;
replace bigurban=       0.414412241      if STATEFIP==  4       ;
replace bigurban=       0.011355668      if STATEFIP==  5       ;
replace bigurban=       0.282950925      if STATEFIP==  17      ;
replace bigurban=       0.150714862      if STATEFIP==  18      ;
replace bigurban=       0.10598629       if STATEFIP==  20      ;
replace bigurban=       0.075349134      if STATEFIP==  21      ;
replace bigurban=       0.153141576      if STATEFIP==  24      ;
replace bigurban=       0.152262797      if STATEFIP==  25      ;
replace bigurban=       0.131562313      if STATEFIP==  27      ;
replace bigurban=       0.022491049      if STATEFIP==  28      ;
replace bigurban=       0.151907607      if STATEFIP==  29      ;
replace bigurban=       0.215681614      if STATEFIP==  32      ;
replace bigurban=       0.001811942      if STATEFIP==  33      ;
replace bigurban=       0.089741393      if STATEFIP==  34      ;
replace bigurban=       0.215302513      if STATEFIP==  37      ;
replace bigurban=       0.144795427      if STATEFIP==  39      ;
replace bigurban=       0.161220509      if STATEFIP==  41      ;
replace bigurban=       0.125408903      if STATEFIP==  42      ;
replace bigurban=       0.390805241      if STATEFIP==  44      ;
replace bigurban=       0.020518167      if STATEFIP==  45      ;
replace bigurban=       0.210530644      if STATEFIP==  47      ;
replace bigurban=       0.179054064      if STATEFIP==  51      ;
replace bigurban=       0.150236534      if STATEFIP==  53      ;
replace bigurban=       0.003924841      if STATEFIP==  54      ;
replace bigurban=       0.127589752      if STATEFIP==  55      ;

replace bigsubrb=       0.249620572      if STATEFIP==  4       ;
replace bigsubrb=       0.00767104       if STATEFIP==  5       ;
replace bigsubrb=       0.431470822      if STATEFIP==  17      ;
replace bigsubrb=       0.159095622      if STATEFIP==  18      ;
replace bigsubrb=       0.156625402      if STATEFIP==  20      ;
replace bigsubrb=       0.214556845      if STATEFIP==  21      ;
replace bigsubrb=       0.718801822      if STATEFIP==  24      ;
replace bigsubrb=       0.419647665      if STATEFIP==  25      ;
replace bigsubrb=       0.451598424      if STATEFIP==  27      ;
replace bigsubrb=       0.01519327       if STATEFIP==  28      ;
replace bigsubrb=       0.397457556      if STATEFIP==  29      ;
replace bigsubrb=       0.489057566      if STATEFIP==  32      ;
replace bigsubrb=       0.00529123       if STATEFIP==  33      ;
replace bigsubrb=       0.736620477      if STATEFIP==  34      ;
replace bigsubrb=       0.255833222      if STATEFIP==  37      ;
replace bigsubrb=       0.296930475      if STATEFIP==  39      ;
replace bigsubrb=       0.298466011      if STATEFIP==  41      ;
replace bigsubrb=       0.380113017      if STATEFIP==  42      ;
replace bigsubrb=       0.520700703      if STATEFIP==  44      ;
replace bigsubrb=       0.020512119      if STATEFIP==  45      ;
replace bigsubrb=       0.177718631      if STATEFIP==  47      ;
replace bigsubrb=       0.346353879      if STATEFIP==  51      ;
replace bigsubrb=       0.318001902      if STATEFIP==  53      ;
replace bigsubrb=       0.061380765      if STATEFIP==  54      ;
replace bigsubrb=       0.170843691      if STATEFIP==  55      ;

replace medmetro=       0.218314625      if STATEFIP==  4       ;
replace medmetro=       0.475107728      if STATEFIP==  5       ;
replace medmetro=       0.134395332      if STATEFIP==  17      ;
replace medmetro=       0.412155445      if STATEFIP==  18      ;
replace medmetro=       0.303171977      if STATEFIP==  20      ;
replace medmetro=       0.198271846      if STATEFIP==  21      ;
```

```
replace medmetro=       0.055282691      if STATEFIP==  24       ;
replace medmetro=       0.389080526      if STATEFIP==  25       ;
replace medmetro=       0.120848773      if STATEFIP==  27       ;
replace medmetro=       0.322169836      if STATEFIP==  28       ;
replace medmetro=       0.128857875      if STATEFIP==  29       ;
replace medmetro=       0.16989106       if STATEFIP==  32       ;
replace medmetro=       0.591462438      if STATEFIP==  33       ;
replace medmetro=       0.17363813       if STATEFIP==  34       ;
replace medmetro=       0.204332593      if STATEFIP==  37       ;
replace medmetro=       0.36983601       if STATEFIP==  39       ;
replace medmetro=       0.271700261      if STATEFIP==  41       ;
replace medmetro=       0.340621171      if STATEFIP==  42       ;
replace medmetro=       0.029382278      if STATEFIP==  44       ;
replace medmetro=       0.658609197      if STATEFIP==  45       ;
replace medmetro=       0.290596196      if STATEFIP==  47       ;
replace medmetro=       0.255556427      if STATEFIP==  51       ;
replace medmetro=       0.362954883      if STATEFIP==  53       ;
replace medmetro=       0.358047473      if STATEFIP==  54       ;
replace medmetro=       0.380263159      if STATEFIP==  55       ;

replace nonmetro=       0.117652562      if STATEFIP==  4        ;
replace nonmetro=       0.505865564      if STATEFIP==  5        ;
replace nonmetro=       0.151182922      if STATEFIP==  17       ;
replace nonmetro=       0.278034071      if STATEFIP==  18       ;
replace nonmetro=       0.434216331      if STATEFIP==  20       ;
replace nonmetro=       0.511822175      if STATEFIP==  21       ;
replace nonmetro=       0.072773911      if STATEFIP==  24       ;
replace nonmetro=       0.039009012      if STATEFIP==  25       ;
replace nonmetro=       0.29599049       if STATEFIP==  27       ;
replace nonmetro=       0.640145845      if STATEFIP==  28       ;
replace nonmetro=       0.321776962      if STATEFIP==  29       ;
replace nonmetro=       0.12536976       if STATEFIP==  32       ;
replace nonmetro=       0.401434391      if STATEFIP==  33       ;
replace nonmetro=       0                if STATEFIP==  34       ;
replace nonmetro=       0.324531671      if STATEFIP==  37       ;
replace nonmetro=       0.188438088      if STATEFIP==  39       ;
replace nonmetro=       0.268613219      if STATEFIP==  41       ;
replace nonmetro=       0.153856908      if STATEFIP==  42       ;
replace nonmetro=       0.059111778      if STATEFIP==  44       ;
replace nonmetro=       0.300360517      if STATEFIP==  45       ;
replace nonmetro=       0.321154529      if STATEFIP==  47       ;
replace nonmetro=       0.219035631      if STATEFIP==  51       ;
replace nonmetro=       0.16880668       if STATEFIP==  53       ;
replace nonmetro=       0.576646921      if STATEFIP==  54       ;
replace nonmetro=       0.321303397      if STATEFIP==  55       ;

gen AIPROB=0;
label variable AIPROB  "probability of reporting Am. Ind. and/or Alaskan
Native";
replace AIPROB=1 if MRDRACE==20;
gen APIPROB=0;
label variable APIPROB "probability of reporting Asian and/or PI";
replace APIPROB=1 if MRDRACE==30;
gen BPROB=0;
label variable BPROB "probability of reporting black";
replace BPROB=1 if MRDRACE==40;
gen WPROB=0;
label variable WPROB "probability of reporting white";
replace WPROB=1 if MRDRACE==50;

gen constant=1;
gen notAI=0 ;   replace notAI=1 if MRDRACE==10;
gen notAPI=0;
gen notB=0  ;    replace notB=1 if MRDRACE==1 | MRDRACE==8;

gen northest= 0; replace northest=1 if STATEFIP==09 | STATEFIP==23 |
STATEFIP==25 | STATEFIP==33 | STATEFIP==34 | STATEFIP==36 |
STATEFIP==42 | STATEFIP==44 | STATEFIP==50;
 label variable northest "northeast region";
```

Appendix A

```
gen midwest= 0; replace midwest=1 if STATEFIP==17 | STATEFIP==18 |
STATEFIP==19 | STATEFIP==20 | STATEFIP==26 | STATEFIP==27 |
STATEFIP==29 | STATEFIP==31 | STATEFIP==38 | STATEFIP==39 |
STATEFIP==46 | STATEFIP==55;
 label variable midwest "midwest region";

gen south= 0; replace south=1 if STATEFIP==01 | STATEFIP==05 |
STATEFIP==10 | STATEFIP==11 | STATEFIP==12 | STATEFIP==13 |
STATEFIP==21 | STATEFIP==22 | STATEFIP==24 | STATEFIP==28 |
STATEFIP==37 | STATEFIP==40 | STATEFIP==45 | STATEFIP==47 |
STATEFIP==48 | STATEFIP==51 | STATEFIP==54;
 label variable south "south region";

gen west=0; replace west=1 if STATEFIP==02 | STATEFIP==04 |
STATEFIP==06 | STATEFIP==08 | STATEFIP==15 | STATEFIP==16 |
STATEFIP==30 | STATEFIP==32 | STATEFIP==35 | STATEFIP==41 |
STATEFIP==49 | STATEFIP==53 | STATEFIP==56;
 label variable west "west region";

* MRDRACE=1 IS AI-API    ==> predict AI and API, then rescale;
gen mrd1AI =0; label variable mrd1AI "regression answer: AI, MRD group 1";
replace mrd1AI =
 (notAI    *   0) +
 (notAPI   *   2.83058) +
 (notB     *   0.97010) +
 (AGE10    *  -0.03967) +
 (HISP     *   0.84013) +
 (MALE     *   0.01914) +
 (northest *   0.59649) +
 (midwest  *   0.43237) +
 (south    *  -0.22255) +
 (bigsubrb *   0.15744) +
 (medmetro *  -0.17318) +
 (nonmetro *   0.25013) +
 (lnpctAI  *   0.56512) +
 (pctAPI   *   0.04203) +
 (pctB     *   0.03921) +
 (pct2race *  -0.09723) +
 (constant *  -5.29417)
if MRDRACE==1;
gen AItemp1= exp(mrd1AI)/(1+exp(mrd1AI)) if MRDRACE==1;

gen mrd1API =0; label variable mrd1API "regression answer: API, MRD group 1";
replace mrd1API =
 (notAI    *   2.78725) +
 (notAPI   *   0) +
 (notB     *   1.61570) +
 (AGE10    *   0.01946) +
 (HISP     *   0.21507) +
 (MALE     *   0.01283) +
 (northest *  -0.13221) +
 (midwest  *  -0.15172) +
 (south    *  -0.24854) +
 (bigsubrb *   0.46028) +
 (medmetro *  -0.09493) +
 (nonmetro *  -0.15342) +
 (lnpctAI  *   0.06996) +
 (pctAPI   *   0.03741) +
 (pctB     *   0.03590) +
 (pct2race *   0.06402) +
 (constant *  -5.73987)
if MRDRACE==1;
gen APItemp1= exp(mrd1API)/(1+exp(mrd1API)) if MRDRACE==1;

gen rescale1=1/(AItemp1+APItemp1);
replace AIPROB = AItemp1 * rescale1 if MRDRACE==1;
replace APIPROB = APItemp1 * rescale1 if MRDRACE==1;
drop  mrd1AI AItemp1 mrd1API APItemp1 rescale1;
```

Appendix A

```
* MRDRACE=2 IS AI-B   ==> predict Black;
gen mrd2B =0; label variable mrd2B "regression answer: Black, MRD group 2";
replace mrd2B =
 (AGE10    *  -0.05461) +
 (HISP     *  -1.92602) +
 (MALE     *  -0.12359) +
 (northest *  -0.88349) +
 (midwest  *  -1.70126) +
 (south    *  -0.97935) +
 (bigsubrb *  -0.44211) +
 (medmetro *   0.88281) +
 (nonmetro *  -0.38427) +
 (lnpctAI  *  -0.43045) +
 (pctBsq   * 0.0000258) +
 (pct2race *  -0.16934) +
 (constant *   3.08086)
if MRDRACE==2;
replace BPROB= exp(mrd2B)/(1+exp(mrd2B)) if MRDRACE==2;
replace AIPROB= 1-BPROB if MRDRACE==2;
drop mrd2B;

* MRDRACE=3 IS AI-W   ==> predict AI ;
gen mrd3AI =0; label variable mrd3AI "regression answer: AI, MRD group 3";
replace mrd3AI =
 (AGE10    *  -0.08968) +
 (HISP     *   0.88834) +
 (MALE     *   0.00972) +
 (northest *   0.21233) +
 (midwest  *   0.09144) +
 (south    *  -0.28494) +
 (bigsubrb *  -0.22069) +
 (medmetro *  -0.44238) +
 (nonmetro *  -0.13978) +
 (lnpctAI  *   0.51235) +
 (pct2race *  -0.07906) +
 (constant *  -0.70527)
if MRDRACE==3;
replace AIPROB= exp(mrd3AI)/(1+exp(mrd3AI)) if MRDRACE==3;
replace WPROB= 1-AIPROB if MRDRACE==3;
drop mrd3AI;

* MRDRACE=4 IS API-B  ==> predict Black;
gen mrd4B =0; label variable mrd4B "regression answer: Black, MRD group 4";
replace mrd4B =
 (AGE10    *  0.05669) +
 (HISP     * -0.10458) +
 (MALE     *  0.33642) +
 (northest * -0.45997) +
 (midwest  * -3.92403) +
 (south    * -1.48264) +
 (bigsubrb *  1.46590) +
 (medmetro *  1.67953) +
 (nonmetro *  0.13301) +
 (pctAPI   * -0.13245) +
 (pctB     *  0.02078) +
 (pct2race *  0.31250) +
 (constant *  0.45883)
if MRDRACE==4;
replace BPROB= exp(mrd4B)/(1+exp(mrd4B)) if MRDRACE==4;
replace APIPROB= 1-BPROB if MRDRACE==4;
drop mrd4B;

* MRDRACE=5 IS API-W  ==> predict API ;
gen mrd5API =0; label variable mrd5API "regression answer: API, MRD group 5";
replace mrd5API =
 (AGE10    *  0.09568) +
 (HISP     *  0.19303) +
 (MALE     *  0.01393) +
```

```
 (northest * -0.05520) +
 (midwest  * -0.06453) +
 (south    *  0.12694) +
 (bigsubrb *  0.50556) +
 (medmetro *  0.07443) +
 (nonmetro * -0.62956) +
 (pctAPI   *  0.00735) +
 (pct2race *  0.09791) +
 (constant * -1.18887)
if MRDRACE==5;
replace APIPROB= exp(mrd5API)/(1+exp(mrd5API)) if MRDRACE==5;
replace WPROB= 1-APIPROB if MRDRACE==5;
drop mrd5API;

* MRDRACE=6 IS B-W    ==> predict Black;
gen mrd6B =0; label variable mrd6B "regression answer: Black, MRD group 6";
replace mrd6B =
 (AGE10    *  0.05532) +
 (HISP     * -0.52253) +
 (MALE     *  0.11948) +
 (northest * -0.25363) +
 (midwest  *  0.17140) +
 (south    * -0.64386) +
 (bigsubrb * -0.07649) +
 (medmetro *  0.28938) +
 (nonmetro *  0.57636) +
 (pctBsq   *  0.00079) +
 (pct2race *  0.31679) +
 (constant * -0.17533)
if MRDRACE==6;
replace BPROB= exp(mrd6B)/(1+exp(mrd6B)) if MRDRACE==6;
replace WPROB= 1-BPROB if MRDRACE==6;
drop mrd6B;

* MRDRACE=7 IS AI-API-B  ==> predict AI, API, and B, then rescale;
gen mrd7AI =0; label variable mrd7AI "regression answer: AI, MRD group 7";
replace mrd7AI =
 (notAI    *   0) +
 (notAPI   *   2.83058) +
 (notB     *   0.97010) +
 (AGE10    *  -0.03967) +
 (HISP     *   0.84013) +
 (MALE     *   0.01914) +
 (northest *   0.59649) +
 (midwest  *   0.43237) +
 (south    *  -0.22255) +
 (bigsubrb *   0.15744) +
 (medmetro *  -0.17318) +
 (nonmetro *   0.25013) +
 (lnpctAI  *   0.56512) +
 (pctAPI   *   0.04203) +
 (pctB     *   0.03921) +
 (pct2race *  -0.09723) +
 (constant *  -5.29417)
if MRDRACE==7;

gen mrd7API =0; label variable mrd7API "regression answer: API, MRD group 7";
replace mrd7API =
 (notAI    *   2.78725) +
 (notAPI   *   0) +
 (notB     *   1.61570) +
 (AGE10    *   0.01946) +
 (HISP     *   0.21507) +
 (MALE     *   0.01283) +
 (northest *  -0.13221) +
 (midwest  *  -0.15172) +
 (south    *  -0.24854) +
 (bigsubrb *   0.46028) +
 (medmetro *  -0.09493) +
```

```
 (nonmetro *  -0.15342) +
 (lnpctAI  *   0.06996) +
 (pctAPI   *   0.03741) +
 (pctB     *   0.03590) +
 (pct2race *   0.06402) +
 (constant *  -5.73987)
if MRDRACE==7;

gen mrd7B =0; label variable mrd7B "regression answer: B, MRD group 7";
replace mrd7B =
 (notAI    *   2.19772) +
 (notAPI   *   3.06153) +
 (notB     *   0) +
 (AGE10    *  -0.01691) +
 (HISP     *  -0.58721) +
 (MALE     *  -0.08093) +
 (northest *   0.40115) +
 (midwest  *   0.20136) +
 (south    *  -0.29365) +
 (bigsubrb *   0.12070) +
 (medmetro *  -0.11129) +
 (nonmetro *  -0.12077) +
 (lnpctAI  *  -0.00347) +
 (pctAPI   *   0.05396) +
 (pctB     *   0.05893) +
 (pct2race *  -0.03953) +
 (constant *  -5.21431)
if MRDRACE==7;

gen AItemp7= exp(mrd7AI)/(1+exp(mrd7AI)+exp(mrd7API)+exp(mrd7B)) if MRDRACE==
7;
gen APItemp7= exp(mrd7API)/(1+exp(mrd7API)+exp(mrd7API)+exp(mrd7B)) if
MRDRACE==7;
gen Btemp7= exp(mrd7B)/(1+exp(mrd7API)+exp(mrd7API)+exp(mrd7B)) if MRDRACE==7;
gen resc7=AItemp7+APItemp7+Btemp7 if MRDRACE==7;
gen rescale7=1/resc7;
replace AIPROB  = AItemp7  * rescale7 if MRDRACE==7;
replace APIPROB = APItemp7 * rescale7 if MRDRACE==7;
replace BPROB = Btemp7   * rescale7 if MRDRACE==7;
drop  mrd7AI mrd7API mrd7B APItemp7 AItemp7 Btemp7 resc7 rescale7;

* MRDRACE=8 IS AI-API-W  ==> predict AI and API, white is remainder;
gen mrd8AI =0; label variable mrd8AI "regression answer: AI, MRD group 8";
replace mrd8AI =
 (notAI    *   0) +
 (notAPI   *   2.83058) +
 (notB     *   0.97010) +
 (AGE10    *  -0.03967) +
 (HISP     *   0.84013) +
 (MALE     *   0.01914) +
 (northest *   0.59649) +
 (midwest  *   0.43237) +
 (south    *  -0.22255) +
 (bigsubrb *   0.15744) +
 (medmetro *  -0.17318) +
 (nonmetro *   0.25013) +
 (lnpctAI  *   0.56512) +
 (pctAPI   *   0.04203) +
 (pctB     *   0.03921) +
 (pct2race *  -0.09723) +
 (constant *  -5.29417)
if MRDRACE==8;

gen mrd8API =0; label variable mrd8API "regression answer: API, MRD group 8";
replace mrd8API =
 (notAI    *   2.78725) +
 (notAPI   *   0) +
 (notB     *   1.61570) +
 (AGE10    *   0.01946) +
```

```
 (HISP     *   0.21507) +
 (MALE     *   0.01283) +
 (northest *  -0.13221) +
 (midwest  *  -0.15172) +
 (south    *  -0.24854) +
 (bigsubrb *   0.46028) +
 (medmetro *  -0.09493) +
 (nonmetro *  -0.15342) +
 (lnpctAI  *   0.06996) +
 (pctAPI   *   0.03741) +
 (pctB     *   0.03590) +
 (pct2race *   0.06402) +
 (constant *  -5.73987)
if MRDRACE==8;

replace AIPROB = exp(mrd8AI) /(1+exp(mrd8AI)+exp(mrd8API)) if MRDRACE==8;
replace APIPROB= exp(mrd8API)/(1+exp(mrd8AI)+exp(mrd8API)) if MRDRACE==8;
replace WPROB= 1-AIPROB-APIPROB if MRDRACE==8;
drop mrd8AI mrd8API;

* MRDRACE=9 IS AI-B-W ==> predict AI and B, white is remainder;

gen mrd9AI =0; label variable mrd9AI "regression answer: AI, MRD group 9";
replace mrd9AI =
 (AGE10    *  0.26212) +
 (HISP     *  0.35986) +
 (MALE     * -0.43898) +
 (northest * -4.53976) +
 (midwest  * -3.82328) +
 (south    * -5.73385) +
 (bigsubrb *  2.78910) +
 (medmetro *  2.27176) +
 (nonmetro *  4.17804) +
 (pctAI    *  0.54579) +
 (pctB     *  0.11100) +
 (pct2race * -0.23972) +
 (constant * -0.64594)
if MRDRACE==9;

gen mrd9B =0; label variable mrd9B "regression answer: B, MRD group 9";
replace mrd9B =
 (AGE10    *  0.36140) +
 (HISP     * -0.83526) +
 (MALE     *  0.50777) +
 (northest * -3.45593) +
 (midwest  * -3.79144) +
 (south    * -2.27313) +
 (bigsubrb *  2.31011) +
 (medmetro *  0.75477) +
 (nonmetro *  1.64725) +
 (pctAI    *  0.39101) +
 (pctB     *  0.04985) +
 (pct2race * -0.02919) +
 (constant *  0.77004)
if MRDRACE==9;

replace AIPROB= exp(mrd9AI)/(1+exp(mrd9AI)+exp(mrd9B)) if MRDRACE==9;
replace BPROB= exp(mrd9B)/(1+exp(mrd9AI)+exp(mrd9B)) if MRDRACE==9;
replace WPROB=1-BPROB-AIPROB if MRDRACE==9;
drop mrd9AI mrd9B;

* MARS=10 IS API-B-W  ==> predict API and B, white is remainder;
gen mrd10AP =0; label variable mrd10AP "regression answer: API, MRD group 10";
replace mrd10AP =
 (notAI    *   2.78725) +
 (notAPI   *   0) +
 (notB     *   1.61570) +
 (AGE10    *   0.01946) +
 (HISP     *   0.21507) +
```

```
 (MALE     *   0.01283) +
 (northest *  -0.13221) +
 (midwest  *  -0.15172) +
 (south    *  -0.24854) +
 (bigsubrb *   0.46028) +
 (medmetro *  -0.09493) +
 (nonmetro *  -0.15342) +
 (lnpctAI  *   0.06996) +
 (pctAPI   *   0.03741) +
 (pctB     *   0.03590) +
 (pct2race *   0.06402) +
 (constant *  -5.73987)
if MRDRACE==10;

gen mrd10B =0; label variable mrd10B "regression answer: B, MRD group 10";
replace mrd10B =
 (notAI    *   2.19772) +
 (notAPI   *   3.06153) +
 (notB     *   0) +
 (AGE10    *  -0.01691) +
 (HISP     *  -0.58721) +
 (MALE     *  -0.08093) +
 (northest *   0.40115) +
 (midwest  *   0.20136) +
 (south    *  -0.29365) +
 (bigsubrb *   0.12070) +
 (medmetro *  -0.11129) +
 (nonmetro *  -0.12077) +
 (lnpctAI  *  -0.00347) +
 (pctAPI   *   0.05396) +
 (pctB     *   0.05893) +
 (pct2race *  -0.03953) +
 (constant *  -5.21431)
if MRDRACE==10;

replace APIPROB= exp(mrd10AP)/(1+exp(mrd10AP)+exp(mrd10B)) if MRDRACE==10;
replace BPROB= exp(mrd10B)/(1+exp(mrd10AP)+exp(mrd10B)) if MRDRACE==10;
replace WPROB=1-BPROB-APIPROB if MRDRACE==10;
drop mrd10AP mrd10B;

* MARS=11 IS AI-API-B-W ==> predict AI, API, and B, white is remainder;
gen mrd11AI =0; label variable mrd11AI "regression answer: AI, MRD group 11";
replace mrd11AI =
 (notAI    *   0) +
 (notAPI   *   2.83058) +
 (notB     *   0.97010) +
 (AGE10    *  -0.03967) +
 (HISP     *   0.84013) +
 (MALE     *   0.01914) +
 (northest *   0.59649) +
 (midwest  *   0.43237) +
 (south    *  -0.22255) +
 (bigsubrb *   0.15744) +
 (medmetro *  -0.17318) +
 (nonmetro *   0.25013) +
 (lnpctAI  *   0.56512) +
 (pctAPI   *   0.04203) +
 (pctB     *   0.03921) +
 (pct2race *  -0.09723) +
 (constant *  -5.29417)
if MRDRACE==11;

gen mrd11AP =0; label variable mrd11AP "regression answer: API, MRD group 11";
replace mrd11AP =
 (notAI    *   2.78725) +
 (notAPI   *   0) +
 (notB     *   1.61570) +
 (AGE10    *   0.01946) +
 (HISP     *   0.21507) +
```

```
 (MALE     *   0.01283) +
 (northest *  -0.13221) +
 (midwest  *  -0.15172) +
 (south    *  -0.24854) +
 (bigsubrb *   0.46028) +
 (medmetro *  -0.09493) +
 (nonmetro *  -0.15342) +
 (lnpctAI  *   0.06996) +
 (pctAPI   *   0.03741) +
 (pctB     *   0.03590) +
 (pct2race *   0.06402) +
 (constant *  -5.73987)
if MRDRACE==11;

gen mrd11B =0; label variable mrd11B "regression answer: B, MRD group 11";
replace mrd11B =
 (notAI    *   2.19772) +
 (notAPI   *   3.06153) +
 (notB     *   0) +
 (AGE10    *  -0.01691) +
 (HISP     *  -0.58721) +
 (MALE     *  -0.08093) +
 (northest *   0.40115) +
 (midwest  *   0.20136) +
 (south    *  -0.29365) +
 (bigsubrb *   0.12070) +
 (medmetro *  -0.11129) +
 (nonmetro *  -0.12077) +
 (lnpctAI  *  -0.00347) +
 (pctAPI   *   0.05396) +
 (pctB     *   0.05893) +
 (pct2race *  -0.03953) +
 (constant *  -5.21431)
if MRDRACE==11;

replace AIPROB= exp(mrd11AI)/(1+exp(mrd11AI)+exp(mrd11AP)+exp(mrd11B)) if
MRDRACE==11;
replace APIPROB= exp(mrd11AP)/(1+exp(mrd11AI)+exp(mrd11AP)+exp(mrd11B)) if
MRDRACE==11;
replace BPROB= exp(mrd11B)/(1+exp(mrd11AI)+exp(mrd11AP)+exp(mrd11B)) if
MRDRACE==11;
replace WPROB=1-AIPROB-APIPROB-BPROB if MRDRACE==11;
drop mrd11AI mrd11AP mrd11B;

drop  lnpctAI pctAI pctAPI pctB pctBsq pct2race medmetro nonmetro
bigurban bigsubrb constant notAI notAPI notB northest midwest south west;

* Note: When converting probabilities to ONERACE, equal probabilities
are decided in favor of the single-race group that has a higher average
probability listed in Table 9 of Ingram et al. 2003.;

gen ONERACE=0;
replace ONERACE=1 if MRDRACE==20;
replace ONERACE=2 if MRDRACE==30;
replace ONERACE=3 if MRDRACE==40;
replace ONERACE=4 if MRDRACE==50;
replace ONERACE=5 if MRDRACE==60;

replace ONERACE=1 if MRDRACE==1 & AIPROB>0.5;
replace ONERACE=2 if MRDRACE==1 & APIPROB>=0.5;

replace ONERACE=1 if MRDRACE==2 & AIPROB>0.5;
replace ONERACE=3 if MRDRACE==2 & BPROB>=0.5;

replace ONERACE=1 if MRDRACE==3 & AIPROB>0.5;
replace ONERACE=4 if MRDRACE==3 & WPROB>=0.5;

replace ONERACE=2 if MRDRACE==4 & APIPROB>0.5;
replace ONERACE=3 if MRDRACE==4 & BPROB>=0.5;
```

```
replace ONERACE=2 if MRDRACE==5 & APIPROB>0.5;
replace ONERACE=4 if MRDRACE==5 & WPROB>=0.5;

replace ONERACE=4 if MRDRACE==6 & WPROB>0.5;
replace ONERACE=3 if MRDRACE==6 & BPROB>=0.5;

replace ONERACE=2 if MRDRACE==7 & APIPROB>BPROB   & APIPROB>AIPROB;
replace ONERACE=1 if MRDRACE==7 & AIPROB>=APIPROB & AIPROB>BPROB;
replace ONERACE=3 if MRDRACE==7 & BPROB>=APIPROB  & BPROB>=AIPROB;

replace ONERACE=1 if MRDRACE==8 & AIPROB>APIPROB  & AIPROB>WPROB;
replace ONERACE=2 if MRDRACE==8 & APIPROB>=AIPROB & APIPROB>WPROB;
replace ONERACE=4 if MRDRACE==8 & WPROB>=AIPROB   & WPROB>=APIPROB;

replace ONERACE=1 if MRDRACE==9 & AIPROB>BPROB    & AIPROB>WPROB;
replace ONERACE=4 if MRDRACE==9 & WPROB>=AIPROB   & WPROB>BPROB;
replace ONERACE=3 if MRDRACE==9 & BPROB>=AIPROB   & BPROB>=WPROB;

replace ONERACE=2 if MRDRACE==10 & APIPROB>BPROB  & APIPROB>WPROB;
replace ONERACE=3 if MRDRACE==10 & BPROB>=APIPROB & BPROB>WPROB;
replace ONERACE=4 if MRDRACE==10 & WPROB>=APIPROB & WPROB>=BPROB;

replace ONERACE=2 if MRDRACE==11 & APIPROB>AIPROB & APIPROB>BPROB & APIPROB>
WPROB;
replace ONERACE=1 if MRDRACE==11 & AIPROB>=APIPROB & AIPROB>BPROB & AIPROB>
WPROB;
replace ONERACE=3 if MRDRACE==11 & BPROB>=APIPROB & BPROB>=AIPROB & BPROB>
WPROB;
replace ONERACE=4 if MRDRACE==11 & WPROB>=APIPROB & WPROB>=AIPROB & WPROB>=
BPROB;

label variable ONERACE "bridged single race";
label define oneracelbl 1 "Am.Ind." 2 "Asian/PI" 3 "black" 4 "white" 5 "SOR";
label values ONERACE oneracelbl;
```