Bridging Estimates by Race for the

Tobacco Use Supplement to the Current Population Survey – (TUS-CPS)

William W. Davis, Anne M. Hartman, James T. Gibson

July 2007

Vers 071307

Abstract

Due to a directive from the Office of Management and Budget (OMB), the Current Population Survey (CPS) changed its race/ethnicity questions in January 2003. Respondents may now select more than one race when answering the survey. This paper provides a method to construct single race estimates using data from the post-2003 Tobacco Use Supplement to the CPS (TUS-CPS). The method is useful when trends over time are being examined for single race groups using both pre-2003 and post-2003. The method uses the post-2003 race/ethnicity responses to multiply impute the (unknown) pre-2003 race/ethnicity response. The method is particularly useful for racial groups whose respondents often report multiple races. This is the case for two races that are often underserved; the American Indian or Alaska Native (AIAN) and the Native Hawaiian or Other Pacific Islander (NHOPI).

1. Introduction

Due to a directive from the Office of Management and Budget (1997), the Current Population Survey (CPS) changed its race/ethnicity questions in January 2003. The old and new race-ethnicity questions for CPS are shown in Appendix 1 Table A-1. The major differences in the questions are the following:

- Respondents may now select more than one race when answering the survey.
- The Asian or Pacific Islander (API) category was split into two categories: Asian and Native Hawaiian or Other Pacific Islander (NHOPI).
- The ethnicity question was reworded to ask directly whether the respondent was Hispanic (an ethnicity rather than a race).

The change in wording impacts smoking estimates and trends made by race/ethnicity from the Tobacco Use Supplement to the CPS (TUS-CPS). To understand racial smoking trends, we need to be able to compare estimates made under the two different race-reporting systems. If we were just interested in trends at the National level by gender, we could ignore this change in the race/ethnicity questions. However, the NCI as part of the Department of Health and Human Services (DHHS) is committed to reducing health disparities. Since some of the smaller race/ethnicity groups are underserved, we need to understand trends by race/ethnicity – including smoking trends. The methodology that we develop for TUS-CPS is useful in this process.

When a change is made in question wording on a continuing survey, it is standard to ask the two different forms of the question to a random sample of survey respondents – the so-called overlap sample. Using this sample of responses, a model can be constructed that allows the prediction of responses from one set of questions given a particular response to the other question. The overlap sample and modeling process provides a bridge between the two sets of questions. We refer to this process for the race/ethnicity questions as race bridging.

In May 2002, the Bureau of Labor Statistics (BLS) sponsored a CPS supplement that asked the new race/ethnicity question to all sample people. The BLS used this sample to compare estimates of unemployment using the two race/ethnicity responses (Bowles *et al*, 2003, Tucker *et al*, 2002). Using the information from this supplement, we propose a race bridging approach that allows the user of post-2003 TUS-CPS data to calculate estimates that are comparable to the pre-2003 TUS-CPS system.

With the mandated change in race coding possibilities, OMB recognized that approaches to make data comparable would be needed and proposed the following six methods based on their simplicity and their wide applicability (OMB, 1997): largest group, smallest group, largest group other than white, plurality, equal fractions, and NHIS fractions. However, there were problems with all six methods so none has found widespread use.

The most sophisticated race bridging work has been done by National Center for Health Statistics (NCHS) researchers and applied to their vital statistics (Ingram *et al*, 2003; Parker et. al., 2004). Logistic regression models including person-level and county-level covariates were fit to 4 years of pooled National Health Interview Survey (NHIS) data (1997–2000). These regression equations were used to bridge multiple-race responses obtained from Census 2000 to single-race categories. In addition, multiple imputation (Rubin, 1987) was used to assess the variability in the estimates in Schenker and Parker (2003).

With the post-2003 race/ethnicity classifications, CPS data users can calculate means, percentages, cross tabulations and regressions using "only" or "any mention" race/ethnicity categories. For example, the user can calculate current smoking of Non-Hispanic Black males 18+, where we include those who list "NH Black Only" or we include everyone who lists NH Black (and other races) for the "any mention" estimate. The "only" classification gives too little emphasis to the mixed race respondents while the "any mention" classification gives too much emphasis to them. Neither of these estimates is exactly comparable to the pre-2003 CPS system, where multi-racial reporting was not allowed.

In Section 2, we provide summary information from the CPS May 2002 Supplement on Race. In Section 3, we describe the CPS overlap sample results, describe our method for race bridging using multiple imputation, and apply the method to the 2003 Tobacco Use Special Cessation Supplement to the CPS (a special topic tobacco use supplement in the general National Cancer Institute (NCI)/Centers for Disease Control and Prevention (CDC) co-sponsored TUS-CPS series) sample respondents. In Section 4, we compare the results from the May 2002 supplement with those obtained from CPS supplement results obtained in February 2002 and February 2003. In Section 5, we provide additional discussion of the race bridging method.

2. Data obtained from the CPS May 2002 Supplement

In this section we discuss the CPS May 2002 Supplement data used and provide descriptive statistics concerning the race/ethnicity responses. The U.S. Census Bureau carries out the CPS survey and some supplements for BLS, and additional supplements for various other government agencies. The CPS is a household survey which is the Nation's primary source of labor force statistics for the entire population; its sampling plan is documented in Current Population Survey (2002). The CPS has a 4-8-4 rotation scheme where households are in sample for four months, then out of sample for eight months, and then in sample again for four months. The TUS interview in each set of four interviews is typically carried out in the household while the other three are carried out by telephone.

The race/ethnicity questions are typically answered for all family members by a single respondent on the first interview. To obtain the most accurate relationship between responses between the two classification systems, it would have been ideal for the same person to have answered the old and new forms of the race/ethnicity questions in CPS May 2002 survey, but this was not always the case. In fact, the old race/ethnicity question was not repeated in May 2002 so both the old and the new questions were only given in May 2002 to those entering the survey. For those not entering the survey, the race/ethnicity initially established was used as the response using the old question. Thus, the following two possible important differences occurred in the race/ethnicity questions for the May 2002 overlap sample:

- The questions were answered at different times for most of the respondents.
- The questions may not have been answered by the same person due to the use of proxy responses.

These differences reduce the concordance in responses from those obtained from a person at a single time.

We could not directly utilize the May 2002 CPS supplement information. However, the Census Bureau provided us cross-tabulations of the number of responses obtained in specified cells, which we use to construct the race bridging methodology.

2.1 <u>CPS May 2002 response summary using the post-2003 classification system</u>

As shown in Table A-1 beginning in January 2003, the respondent may select one or more of the following five races: White, Black or African American, Asian, American Indian or Alaska Native (AIAN), and Native Hawaiian or Other Pacific Islander (NHOPI). Since multiple race reporting is allowed, one might expect that 31 ($=2^5$-1) possible race groups would be reported. However, only 21 race groups are reported on the CPS post-2003 public-use files -- due to confidentiality considerations. We add ethnicity as a 22nd group by combining all Hispanics together – independent of race. For example, a respondent to the post-2003 race/ethnicity questions, who responded "No" to the "Are you Spanish, Hispanic or Latino" question and listed only "White" to the race question, would be in the "non-Hispanic (NH) White Only" group. Table 1 shows the number and percentage of race/ethnicity responses (weighted and un-weighted) of adults in the CPS May 2002 Supplement.

Table 2 shows the number of "any mention" and single race respondents for the five races allowed in the new reporting system for non-Hispanics. For example, the "NH Black only" estimates are based on those in category 2 in table 1 while the NH Black "any mention" estimates would also include those in categories 6, 9, 13-14, and 17-19. In the table the races are ordered by the ratio of "single races" to "any mentions". The table shows that the AIAN group has the smallest fraction of single race respondents (44.4%) followed by the NHOPI (66.9%). The other three races have percentages over 90%; for these races the "single race" and the "any mention" estimates should be similar in most cases while this would not necessarily be the case for AIAN and NHOPI. Thus, race bridging has the most potential for offering results that differ from the single race and the "any mention" results for two of the most underserved races, the AIAN and the NHOPI.

Table 1. Race/ethnicity totals and percentages from the CPS May 2002 Supplement using the post-2003 classification system

		Unweighted		Weighted	
No.	CPS Race/ethnicity group	Total	Percentage	Total	Percentage
0	Hispanic	10,490	9.436%	26,483,420	11.973%
1	NH White only	83,877	75.446%	157,215,671	71.074%
2	NH Black Only	9,857	8.866%	24,982,634	11.294%
3	NH AIAN Only	1,065	0.958%	1,033,250	0.467%
4	NH Asian Only	3,712	3.339%	8,557,714	3.869%
5	NH NHOPI Only	349	0.314%	428,815	0.194%
6	NH White-Black	121	0.109%	213,999	0.097%
7	NH White-Asian	167	0.150%	229,920	0.104%
8	NH White-AIAN	1,138	1.024%	1,554,817	0.703%
9	NH Black-AIAN	130	0.117%	215,646	0.097%
10	NH Asian-NHOPI	66	0.059%	50,865	0.023%
11	NH White-NHOPI	59	0.053%	62,130	0.028%
12	NH AIAN-Asian	6	0.005%	13,512	0.006%
13	NH Black-Asian	9	0.008%	12,828	0.006%
14	NH Black-NHOPI	9	0.008%	12,766	0.006%
15	NH White-Asian-NHOPI	39	0.035%	29,913	0.014%
16	NH White-Asian-AIAN	12	0.011%	9,617	0.004%
17	NH White-Black-AIAN	52	0.047%	78,327	0.035%
18	NH White-Black-Asian	0	0.000%	0	0%
19	NH White-Black-AIAN-Asian	2	0.002%	4,563	0.002%
20	NH 2 or 3 Races	10	0.009%	7,546	0.003%
21	NH 4 or 5 Races	5	0.004%	2,428	0.001%
	Total	**111,175**	**100.000%**	**221,200,435**	**100.000%**

Table 2. Number of "single race" and "any mention" for non-Hispanics (NH) from the CPS May 2002 Supplement using the post-2003 classification system

	Single race	Any mention	Ratio "single race" to "any mention"
NH White	83,877	85,467	98.1%
NH Black	9,857	10,179	96.8%
NH Asian	3,712	4,013	92.5%
NH NHOPI	349	522	66.9%
NH AIAN	1,065	2,399	44.4%

Table 3 shows the weighted and unweighted totals and percentages of the single race or ethnicity (No. 0-5 in table 1), specified multiple race (No. 6-19 in table 1), and unspecified multiple race (No. 20-21 in table 1). Since the weighted percentages reflect the overall National population, we emphasize these. The table shows that the single race category is almost 99% (98.87%) of the population. Also, the "specified multiple race" category composes only 1.1% of the population while the "unspecified multiple race" category is only 0.004% of the population. The unspecified multiple race individuals are difficult to include in "any mention" analyses but can be included in bridged race analyses using imputation.

Table 3. Race/ethnicity groups totals and percentages (weighted and unweighted) from the CPS May 2002 Supplement using the post-2003 classification system

		Unweighted		Weighted	
No.	CPS Race group	Total	Percentage	Total	Percentage
0-5	Single race or ethnicity	109,365	98.372%	218,701,504	98.870%
6-19	Specified Multiple race	1,810	1.628%	2,488,957	1.125%
20-21	Unspecified Multiple race	15	0.015%	9,974	0.004%
	Total	**111,175**	**100.000%**	**221,200,435**	**100.000%**

3.2 CPS May 2002 response summary using the post- and pre-2003 classification systems

Table 4 shows the number of responses under the pre-2003 classification system for the single race or ethnicity categories (numbered 0 to 5) in the post-2003 classification system. The table also shows the percentage agreement in each row. For example, the first row shows that 9,246 of the 10,490 classified Hispanic in the post-2003 classification were also classified as Hispanic in the pre-2003 classification, which is an agreement of 88.1% (=9,246/10,490). Since the pre-2003 Asian and Pacific Islander category was split into two post-2003 categories (Asian and NHOPI), the agreement percentage of both NH Asian and NH NHOPI are computed with numerator obtained from NH Asian or Pacific Islander (NH API) in table 4. For the six post-2003 classifications, the percentage agreement varies between 84.5% for NH AIAN only and 98.0% for NH White only. The overall agreement percentage for these six categories was 96.1% (=105,085/109,350). We attribute the failure to achieve closer to the expected 100% agreement to limitations in the overlap sample responses -- the use of proxy responses and the use of responses made at different times.

Table 4. Number of pre-2003 responses for the single race or ethnicity post-2003 classifications using CPS May 2002

	Pre-2003 CPS Classification						
Post-2003 CPS classification	Hispanic	NH White	NH Black	NH AIAN	NH API	Total	Agreement
Hispanic	9,246	952	118	57	117	10,490	88.1%
NH White only	561	82,220	532	227	337	83,877	98.0%
NH Black only	97	349	9,340	18	53	9,857	94.8%
NH AIAN only	6	121	34	900	4	1,065	84.5%
NH Asian only	37	190	34	17	3,434	3,712	92.5%
NH NHOPI only	3	37	6	4	299	349	85.7%
Total	9,950	83,869	10,064	1,223	4,244	109,350	96.1%

3. Race bridging Method

The CPS May 2002 overlap sample contains information that could be used to bridge the race/ethnicity responses forward or backward as follows:

- Bridge forward: A method is provided to impute pre-2003 race/ethnicity responses given the subject's post-2003 race/ethnicity response. Then, the imputation method is applied to a post-2003 dataset, which allows race/ethnicity estimates on the post-2003 dataset that are comparable to those that would have been obtained if the pre-2003 race/ethnicity questions had been used.
- Bridge backward: A method is provided to impute post-2003 race/ethnicity responses given the subject's pre-2003 race/ethnicity response. Then, the imputation method is applied to a pre-2003 dataset, which allows race/ethnicity estimates on the pre-2003 dataset that are comparable to those that would have been obtained if the post-2003 race/ethnicity questions had been used.

Both of the bridges are potentially useful, but we show only how to carry out forward bridging. With forward bridging we can compare post-2003 TUS-CPS smoking estimates by race/ethnicity with those previously reported using pre-2003 data and definitions. Furthermore, we can bridge forward by using only the post-2003 TUS-CPS datasets. Had we chosen to bridge backward, we would apply the model to older pre-2003 TUS-CPS datasets.

The most common application of imputation occurs when there is item non-response in survey data. That is the case when a survey respondent does not answer a particular question; for example, a question specifying family income. When a single imputation is used for missing values, the data are usually analyzed as if the imputed values were equivalent to observed values. The uncertainty in the process used to generate the imputed values is usually ignored. If this uncertainty is incorporated correctly, the inference typically yields wider confidence limits and fewer statistically significant relationships. In contrast to item non-response, the change in the race/ethnicity questions in 2003 results in an "artificial" missing data problem since it was not caused by the respondent's failure to answer a question. However, viewing the problem as a missing data problem allows the large amount of research on this subject to be applied (Schenker, 2004).

For “forward” race bridging using a TUS-CPS post-2003 dataset, the “new” race/ethnicity is observed and the “old” race/ethnicity is missing for all respondents. However, for single race/ethnicity post-2003 respondents (No. 0 to 5 in Table 1), we have chosen to assume that we know the missing pre-2003 race/ethnicity response, where the assumed values are shown in Table 5.

Table 5. Assumed relationship between post- and pre-2003 single race/ethnicity categories

Observed Post-2003 category	Assumed Pre-2003 category
Hispanic	Hispanic
NH White only	NH White
NH Black only	NH Black
NH AIAN only	NH AIAN
NH Asian only	NH API
NH NHOPI only	NH API

For example, table 5 shows that we assume all those who chose “NH White only” using the post-2003 questions would have responded “NH White” using the pre-2003 questions. Table 4 shows that this assumption was true for over 96% of the May 2003 CPS respondents in the six categories number 0-5 in Table 1. We expect that this percentage would have been substantially higher if proxy responses were eliminated from the process. With this assumption, we only need to impute the pre-2003 race/ethnicities for those who reported the following using the post-2003 classification system:

- Those who were non-Hispanic
- Those who provided multiple race responses

In summary to bridge forward, we impute values only for those categories listed as 6-21 in Table 1 while for categories numbered 0-5 in table 1 the pre-2003 race/ethnicity is specified in Table 5. There are only the four possible imputed responses using the pre-2003 classification system: White, Black, AIAN, and API (and all are non-Hispanic so we are imputing race rather than race/ethnicity). To carry out the imputation, we used a statistical model with two predictors: gender and age (considered as a categorical variable with two levels). Now, we describe the imputation process more completely.

3.1 Multiple Imputation

Multiple imputation (MI) is a technique for reflecting the uncertainty due to the imputation (Rubin, 1987). In multiple imputation several sets, say *M*, of imputations are produced for all missing data resulting in set of *M* complete data sets. The analysis of each of these *M* complete data sets is carried out using the same method that would be used for a single data set. Then, the multiple analyses are combined to reflect an inference that contains the proper variability. Usually the imputations are obtained through a statistical model. If a

Bayesian approach is adopted, the *M* imputations are obtained from the posterior predictive distribution. To show the steps necessary to carry out the MI for the CPS race/ethnicities, we define the following:

- Categories i=1,…,I formed by post-2003 (multiple) race/ethnicity, age, and gender.
- Label the pre-2003 race/ethnicity categories j=1,…, 4 (White, Black, AIAN, API).
- Define n_{ij} as the number of May 2002 CPS subjects who were in the (i,j) cell and define a vector $\underline{n}_i=(n_{i1},n_{i2},n_{i3},n_{i4})$ for i=1,…,I.
- Define p_{ij} as the probability of the (i,j) cell and a vector $\underline{p}_i=(p_{i1},p_{i2},p_{i3},p_{i4})$ for i=1,…,I..

If we ignore the CPS survey design, the sample likelihood, *L*, is proportional to

$$L \propto \prod_{i=1}^{I}\prod_{j=1}^{4} p_{ij}^{n_{ij}} \qquad (1)$$

Equation (1) shows that the vectors $\underline{n}_1$, $\underline{n}_2$, …$\underline{n}_I$ are independent multinomial distributions given the parameter vector $\underline{p}_i$. If we adopt a Bayesian approach with non-informative prior information about the parameters, the posterior distribution of $\underline{p}_i$ is the Dirichlet with parameter vector $\underline{n}_i$ (e.g., Gelman *et al.*, 2004, p. 83). For each *i*, the Dirichlet distribution can be obtained by generating independent samples from the gamma distribution (with different shape parameters) and normalizing these (Gelman *et al.*, 2004, p. 582). That is, for i=1,…,I, we can generate x_{ij}~G(n_{ij},1) for j=1,…,4 and then determine the probabilities from $p_{ij} = x_{ij} / \sum_{j=1}^{4} x_{ij}$, where G(a,b) denotes the gamma distribution with shape parameter *a* and scale parameter *b*.

A problem arises in the above when there are no subjects in a cell, that is, when n_{ij}=0 for some pair (i,j). In this case, the Dirichlet distribution result stated above is not valid, and we cannot generate a sample from the gamma distribution with shape parameter equal to zero. To circumvent this problem, we collapse race/ethnicity, age, gender categories that have a small number of responses (those less than 20). However, we still have some cells with n_{ij}=0. A standard solution to this problem is to assume a proper (informative) prior distribution of the form $f(\underline{p}_1,...,\underline{p}_I) \propto \prod_{i=1}^{I}\prod_{j=1}^{4} p_{ij}^{a-1}$. When a=0, this is the non-informative prior distribution. For this application we suggest using a small positive value such as a=0.01 to minimize the problem with cells with no respondents. The analyst could check for sensitivity of the analysis to this informative prior distribution by comparing the results using the multiply imputed values with different small values of *a* to insure that this choice does not impact the conclusions.

With this modification to the prior distribution, the posterior distribution of the i^{th} parameter vector, $\underline{p}_i$, is the Dirichlet with parameter vector (n_{i1}+a,n_{i2}+a,…,n_{i4}+a). We generate this vector in the same fashion as when a=0; that is, we generate x_{ij}~G(n_{ij}+a,1) for j=1,…,4 and then normalize these values to obtain $p_{ij} = x_{ij} / \sum_{j=1}^{4} x_{ij}$.

In summary, each of the *M* sets of imputations can be created as follows;

1. For each i=1,…,I, sample from the Dirichlet distribution and determine the vector $\underline{p}_i$ of cell probabilities
2. For each person with multiple race/ethnicity
   a. Determine the appropriate cell i=1,…,I based on age, gender, and race/ethnicity and the associated probability vector $\underline{p}_i$ from step 1.
   b. Draw one of the four primary race/ethnicities using the probabilities determined in 2a.

3.2. Using the May 2002 overlap sample to multiply impute race

Table 6 shows the cross classification of race responses from the CPS May 2002 supplement for the multiple race categories. For each of the "new" race categories that are provided with the public use dataset the number of people n_{ij}, who selected the possible responses under the "old" race classification system are shown in the table. The table includes two age categories (15-44, 45+) and two gender categories. The table shows I=23 distinct categories.

Table 6. Total number of multiple race responses for the CPS May 2002 supplemental sample by age and sex for non-Hispanics

Cat	New race	age	sex	White	Black	AIAN	API	Total
1	White-Black	15-44	All	58	38	2	2	100
2	White-Black	45+	All	9	8	2	0	19
3	White-Asian	15-44	All	71	0	0	70	141
4	White-Asian	45+	All	10	0	0	15	25
5	White-AIAN	15-44	Female	266	1	61	1	329
6	White-AIAN	45+	Female	229	2	35	0	266
7	White-AIAN	15-44	Male	255	1	57	2	315
8	White-AIAN	45+	Male	201	2	20	1	224
9	Black-AIAN	15-44	All	8	62	1	1	72
10	Black-AIAN	45+	All	1	54	1	1	57
11	Asian-NHOPI	15-44	All	2	0	0	36	38
12	Asian-NHOPI	45+	All	1	0	0	27	28
13	White-NHOPI	All	All	25	0	0	34	59
14	AIAN-Asian	All	All	0	0	3	3	6
15	Black-Asian	All	All	1	5	0	3	9
16	Black-NHOPI	All	All	0	6	0	3	9
17	White-Asian-NHOPI	All	All	10	0	0	29	39
18	White-AIAN-Asian	All	All	9	0	1	2	12
19	White-Black-AIAN	15-44	All	13	18	1	0	32
20	White-Black-AIAN	45+	All	9	10	1	0	20
21	White-Black-AIAN-Asian	All	All	0	2	0	0	2
22	2 or 3 Races	All	All	4	1	1	4	10
23	4 or 5 Races	All	All	1	0	0	4	5

We collapsed over gender and age in table 6 if there were not sufficient row totals. The goal of the collapsing process was to obtain more accurate row entries and to eliminate zero cell entries by using only categories with at least 20 total respondents. Clearly, this was not possible for the race/ethnicity groups that had fewer than 20 total respondents. Even with the collapsing, some rows have a small number of respondents.

A single category in the CPS May 2002 had no responses (NH White-Black-Asian). However, in future TUS-CPS surveys we may obtain respondents who identify with this class. Since this is a three race classification, we suggest grouping with the "2 or 3 races" category in table 6 for imputation purposes.

Inspection of the cross-tabulations suggests that age is a more important determinant of the CPS pre-2003 race response than gender. Thus, we collapsed cells in these tables over gender before (possibly) collapsing over age.

Other researchers may want to use the methodology developed here to carry out TUS-CPS analyses. To assist the interested user, in Appendix 2 we provide generic SAS code to carry out the MI process.

We ignored the survey design in the MI process, but we assume that the survey design would be incorporate in any analysis which used the MI race/ethnicities. In particular, the MI race/ethnicity values can be used in SUDAAN (Research Triangle Institute, 2004), which incorporates the design of the Current Population Survey in the analysis through replicate weights.[1]

Without the assumption made in table 5, we would have imputed pre-2003 responses for *all* post-2003 respondents – not just the multiple race responders. The method given above could have been applied to the single race responders, but the sample code (and data input) given in the Appendix would have to be modified to implement this change. However, since table 4 shows an agreement percentage over 96% in the single race/ethnicity categories, the vast majority of the imputations made for single race/ethnicity categories would satisfy the assumptions of table 5. This was one reason that we chose not to impute these categories. But the most important reason was that we believed the agreement percentage would have been even higher than 96% if proxy responses were eliminated.

3.3. Multiple imputation for multiple race responders in the TUS-CPS 2003 sample

The process described above was used to carry out MI of the pre-2003 race for the TUS-CPS surveys conducted in February, June and November of 2003. There were 330,234 respondents in the 2003 TUS-CPS, and, of those, 4,488 (1.4%) listed more than one race. We implemented the MI to these 4,488 TUS-CPS respondents using the categories and cell counts shown in Table 6. Also, we used M=5, which is the most common choice in multiple imputation, so there were 22,440 (=5 x 4,448) multiply imputed races.

In table 7, we show the MI results aggregated over all individuals in the race, gender and age cells and also aggregated over the M=5 imputations. The total number of TUS-CPS 2003 respondents in each of the categories is the total in Table 7 divided by 5. For example, there were 331 (=1655/5) TUS-CPS 2003 respondents in category 1 (new race = white/black,

[1] The TUS-CPS replicate weights are not contained in the public use file but can be obtained from the National Cancer Institute upon request.

both genders, and age = 15-44). In table 7, we also show the p-value of the standard chi-square hypothesis test, which tests the homogeneity of the simulated values to those shown in table 6.[2] None of the p-values are statistically significant at the 0.05 significance level.

The use of a proper prior distribution with a>0 allows the possibility of a value to be imputed in a cell (in table 6) that had no respondents in the overlap sample (table 7). However, comparison of tables 6 and 7 shows that this did not occur in the TUS-CPS 2003 multiple imputations of race.

Table 7. Multiply imputed race responses for the TUS-CPS 2003 sample: total number of respondents by age and sex for non-Hispanics and test of homogeneity with the frequencies of the same category from table 6

Cat	New race/ethnicity	age	sex	White	Black	AIAN	API	Total	p-value
1	White-Black	15-44	All	892	717	27	19	1,655	0.67
2	White-Black	45+	All	234	151	45	0	430	0.81
3	White-Asian	15-44	All	825	0	0	870	1,695	0.70
4	White-Asian	45+	All	182	0	0	253	435	0.86
5	White-AIAN	15-44	Female	2,756	6	688	5	3,455	0.78
6	White-AIAN	45+	Female	2,746	22	387	0	3,155	0.91
7	White-AIAN	15-44	Male	2,697	15	729	24	3,465	0.64
8	White-AIAN	45+	Male	2,349	49	283	9	2,690	0.63
9	Black-AIAN	15-44	All	87	545	12	21	665	0.78
10	Black-AIAN	45+	All	6	566	12	11	595	0.96
11	Asian-NHOPI	15-44	All	50	0	0	575	625	0.54
12	Asian-NHOPI	45+	All	23	0	0	407	430	0.68
13	White-NHOPI	All	All	460	0	0	615	1,075	0.95
14	AIAN-Asian	All	All	0	0	23	17	40	0.73
15	Black-Asian	All	All	20	91	0	84	195	0.84
16	Black-NHOPI	All	All	0	22	0	18	40	0.52
17	White-Asian-NHOPI	All	All	22	0	0	53	75	0.68
18	White-AIAN-Asian	All	All	26	0	1	8	35	0.67
19	White-Black-AIAN	15-44	All	178	165	17	0	360	0.52
20	White-Black-AIAN	45+	All	118	100	7	0	225	0.77
21	White-Black-AIAN-Asian	All	All	0	10	0	0	10	NA*
22	2 or 3 Races	All	All	367	106	130	397	1,000	0.99
23	4 or 5 Races	All	All	27	0	0	63	90	0.63
	Total	**All**	**All**	**14,065**	**2565**	**2361**	**3449**	**22,440**	

* Chi-square test not applicable since all responses are in a single category.

[2] The assumptions of the chi-square test are not satisfied since the multiply imputed races are generated from five independent Dirichlet distributions – not the multinomial distribution. However, the test is asymptotically valid as the Dirichlet converges to the multinomial as the sample size goes to infinity.

4. Stability over time of the TUS-CPS responses

The results of this paper are based on the data obtained from the CPS May 2002 Supplement. If the nature of the responses changes over time, it could impact the validity of the method proposed here. Thus, we used responses from CPS supplements to check the stability over time of the race/ethnicity response frequencies. In particular, we use results obtained from the February 2002 Tobacco Use Supplement (TUS) and the February 2003 Tobacco Use Special Cessation Supplement (TUSCS). Since the time period from the May 2002 CPS supplement and the February 2003 TUSCS is less than one year, one would not expect a large change in the race/ethnicity response patterns.

A unique feature of the CPS is its panel design where each household in the sample is surveyed for four consecutive months and then for four more consecutive months nine months later (see, Current Population Survey, 2002). Due to this sampling strategy persons who were in their 1st, 2nd, or 3rd month in sample in February 2002 when the TUS data was collected were potentially also in the February 2003 sample for panel months #5, 6 and 7 when the TUSCS-CPS was fielded. We refer to those who responded to both the February 2002 TUS and the February 2003 TUSCS as the overlap sample. The overlap sample respondents answered the new race/ethnicity questions in February 2003 and the old race/ethnicity questions in February 2002. We have restricted attention to the 68,954 February 2003 survey respondents who were age 15 and older -- due to our interest in smoking outcomes. Of these, 22,598 (=32.8%) were in the overlap sample; that is, they also responded to the February 2002 TUS. This is slightly below the 37.5% (=3/8) that would be achieved if all in the three eligible (of the eight total) panels from February 2002 TUS-CPS also participated in the February 2003 TUSCS-CPS.

Tables 8a and 8b compares the number and percentage of respondents in the February 2003 overlap sample with the May 2002 supplement sample based on their ethnicity categorization (Hispanic or not) and on whether they responded with a single or multiple race. Table 8a shows the values unweighted while table 8b shows the weighted values. One would not expect the weighted percentages to change much in this short time span and that is the case -- as the largest weighted percentage difference is only 0.4% (for single race Hispanics).

Table 8a. Race ethnicity responses of May 2002 and February 2003 overlap unweighted

ethnicity	Single/multiple race?	Feb 2003 overlap	May 2002
Hispanic	Single	1,741 (7.7%)	10,177 (9.2%)
Hispanic	Multiple	30 (0.1%)	313 (0.3%)
Non-Hispanic	Single	20,576 (91.1%)	98,860 (88.9%)
Non-Hispanic	Multiple	249 (1.1%)	1,825 (1.6%)
Total	All	22,598 (100.0%)	111,175 (100.0%)

Table 8b. Race ethnicity responses of May 2002 and February 2003 overlap weighted

	Single/multiple race?	Feb 2003 overlap		May 2002	
Hispanic	Single	27,335,760	(12.2%)	25,999,023	(11.8%)
Hispanic	Multiple	476,392	(0.2%)	484,397	(0.2%)
Non-Hispanic	Single	193,978,598	(86.6%)	192,218,084	(86.9%)
Non-Hispanic	Multiple	2,297,890	(1.0%)	2,498,931	(1.1%)
Total	All	224,088,640	(100.0%)	221,200,435	(100.0%)

Now, we do further comparisons on the multiple race respondents. Since there were only 30 Hispanic multiple race respondents in the overlap sample (table 8a), we restrict the comparisons to the non-Hispanics where we compare the 249 February 2003 overlap respondents with the 1,825 May 2002 supplement respondents.

First, we compare the total number of responses in the 23 categories used in tables 6 and 7 for non-Hispanics in table 9. The May 2002 column coincides with the total column of table 6 (with the addition of the 8 respondents who were classified as non-Hispanic using the new questions and Hispanic using the old questions) while the February 2003 overlap column shows the total number of respondents from the overlap sample. For example, there were 8 respondents in the overlap who reported their race as "white-black" and were in the age range 15-44. The last column of the table shows the ratio of the number of respondents in the overlap to the total number of respondents in the row (May 2002 and Feb. 2003 overlap). For example, there were 1817 non-Hispanic) multiple race respondents in the May 2002 sample and 249 (non-Hispanic) multiple race respondents in the February 2003 overlap sample. The table shows that 12.0% (=249/(249+1817) of the respondents were in the overlap sample. If there is stability in responses over time, one would expect a roughly constant ratio for all the categories. Of course, when the sample size is small in both categories, the ratio estimate has a large variance and significant deviations from 12% could be expected. Two categories with large differences from 12% (and substantial sample sizes) are the following:

- Large decrease in the use of cat 17 "white-Asian-NHOPI" with 39 in the May 2002 and none in the overlap sample for a ratio of 0%.
- Large increase in the use of cat 22 "2 or 3 races" with 10 in the May 2002 and 15 in the overlap sample for a ratio of 60%

For any category in table 9 with at least 60 responses for both surveys combined, we performed a Pearson chi-square test of the homogeneity of the response patterns using the old single-response category system. There were four possible responses (white, black, AIAN, or API) so the chi-square test had 3 degrees of freedom (or fewer if some of the cells had no observations). The chi-square statistics and p-values of these tests are shown in Table 10 (raw data that the tests are based upon are not shown here). Table 10 shows that 9 categories met the sample size threshold, and none of these tests indicate significant changes in the response patterns within categories.

Table 9. Total responses by category for the May 2002 and February 2003 overlap samples

Cat	New race	age	sex	May 2002	Feb 2003 overlap	Ratio of Feb 2003 overlap to total
1	White-Black	15-44	All	100	8	7.3%
2	White-Black	45+	All	19	6	23.1%
3	White-Asian	15-44	All	141	12	7.8%
4	White-Asian	45+	All	25	3	10.7%
5	White-AIAN	15-44	Female	329	38	10.4%
6	White-AIAN	45+	Female	266	48	15.2%
7	White-AIAN	15-44	Male	315	20	5.9%
8	White-AIAN	45+	Male	224	38	14.4%
9	Black-AIAN	15-44	All	72	10	12.0%
10	Black-AIAN	45+	All	57	9	13.6%
11	Asian-NHOPI	15-44	All	38	8	17.4%
12	Asian-NHOPI	45+	All	28	4	12.5%
13	White-NHOPI	All	All	59	18	23.4%
14	AIAN-Asian	All	All	6	0	0.0%
15	Black-Asian	All	All	9	2	18.2%
16	Black-NHOPI	All	All	9	0	0.0%
17	White-Asian-NHOPI	All	All	39	0	0.0%
18	White-AIAN-Asian	All	All	12	1	7.7%
19	White-Black-AIAN	15-44	All	32	5	13.5%
20	White-Black-AIAN	45+	All	20	3	13.0%
21	White-Black-AIAN-Asian	All	All	2	1	33.3%
22	2 or 3 Races	All	All	10	15	60.0%
23	4 or 5 Races	All	All	5	0	0.0%
	Total	All	All	1817	249	12.0%

Table 10. Chi-square tests of homogeneity of responses within categories for the May 2002 and February 2003 overlap samples

Cat	New race	age	sex	p-value	chi-square	Total sample size
1	White-Black	15-44	All	0.88	0.7	108
3	White-Asian	15-44	All	0.28	1.2	153
5	White-AIAN	15-44	Female	0.47	2.5	367
6	White-AIAN	45+	Female	0.83	0.4	314
7	White-AIAN	15-44	Male	0.78	1.1	335
8	White-AIAN	45+	Male	0.77	1.1	262
9	Black-AIAN	15-44	All	0.66	1.6	82
10	Black-AIAN	45+	All	0.92	0.5	66
13	White-NHOPI	All	All	0.79	0.1	77

5. Discussion

We chose to implement only forward race bridging in this paper. This allows us to compare post-2003 TUS-CPS smoking estimates by race/ethnicity with those previously reported using pre-2003 data and definitions. Backward bridging could be implemented using the same methodology and would allow us to compare pre-2003 smoking estimates by race/ethnicity with those previously reported using post-2003 data and definitions.

The race bridging method is not necessary for certain TUS-CPS analyses. For example, those interested in smoking trends by gender at the National level do not need this method. However, the method is useful when trends over time are being examined for single race groups using both pre-2003 and post-2003 data. The method is particularly useful for racial groups whose respondents often report multiple races. This is the case for two races that are often underserved in health care: the American Indian or Alaska Native (AIAN) and the Native Hawaiian or Other Pacific Islander (NHOPI).

We implemented the race bridging methodology using multiple imputation (MI) of the CPS pre-2003 race/ethnicities. We chose to impute only the multiple race/ethnicities. This is an important assumption and slightly different results would be obtained if all pre-2003 race/ethnicities were imputed. We felt that this assumption was justified based on the high correlation of responses in the single race/ethnicity categories and also based on the assumption that the correlation would have been even higher if proxy responses were eliminated in the overlap sample.

There are a number of limitations in the race/bridging methodology. First, the race bridging method depends on the accuracy of the probabilities as estimated using the cross-tabulation given in table 6. Some of the probabilities are inaccurately estimated due to small sample size. Thus, the imputation would have been improved if there had been a larger overlap sample.

Another limitation of the methodology is the dependence on the race/bridging results from the CPS survey conducted in May 2002. Since the race-ethnicity composition of the United States is changing rapidly, the relationship between the post-2003 and the pre-2003 race/ethnicities as described in table 6 will eventually become inaccurate.

A third limitation was the lack of information available to develop accurate predictions of the pre-2003 race/ethnicity given the post-2003 race/ethnicity. We could not directly utilize the May 2002 CPS supplement information -- due to confidentiality considerations especially concerning geographical identification of respondents. Thus, we did not use logistic regression models to estimate probabilities using an adaptation of NCHS's logistic regression race bridging method. Based on NCHS analysis, in addition to age, gender, and Hispanic origin the race responses should vary by geographic factors. Use of additional geographic information would have improved the accuracy of the race/ethnicity multiple imputation process.

Even with the limitations cited above, we feel that the race bridging methodology presented here will be useful in the documentation of smoking trends for racial groups using TUS-CPS.

Acknowledgement

The authors wish to thank Dr. Nathaniel Schenker and Dr. Jennifer Parker of the National Center for Health Statistics, Dr. Ram Tiwari of the National Cancer Institute, and Joe Zou of IMS for comments, which have significantly improved the manuscript. Also, we wish to thank Greg Weyland of the U. S. Census Bureau for providing the CPS May 2002 overlap cross tabulations for race/ethnicity that formed the basis of our approach to race bridging.

6. References

Bowles M, Ilg RE, Miller S, Robison E, Polivka A. Revisions to the Current Population Survey effective in January 2003. *Employment and Earnings*, January 2003.

Current Population Survey "Design and Methodology. Technical Paper 63RV, May 2002, U.S. Department of Labor, Bureau of Labor Statistics.

Gelman A, Carlin JB, Stern HS, and Rubin DB. *Bayesian Data Analysis* (2^{nd} ed.). Chapman & Hall/CRC, Boca Raton, 2004.

Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. United States Census 2000 population with bridged race categories. *Vital Health Stat* 2003; 135:1-63.

Office of Management and Budget. Revisions to the standards for the classification of Federal data on race and ethnicity. *Federal Register* 62FR58781-58790, 1997.

Parker JD, Schenker N, Ingram DD, Weed JA, Heck KE, Madans JH. Bridging between two standards for collecting information on race and ethnicity: an application to Census 2000 and vital rates. *Public Health Reports* 2004; 119: 192-205.

Research Triangle Institute. SUDAAN Example Manual, Release 9.0. Research Triangle Park, NC; Research Triangle Institute, 2004.

Rubin DB. *Multiple imputation for nonresponse in Surveys*. John Wiley, New York, 1987.

Schenker N and Parker JD. From single-race reporting to multiple-race reporting: using imputation methods to bridge the transition. *Stat in Medicine* 2003; 22: 1571-1587.

Schenker N. Bridging across changes in classification systems. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.* A. Gelman and X-L. Meng (eds). John Wiley & Sons, New York, 2004: pp 117-128.

Tucker C, Miller S, Parker J. Comparing census race data under the old and new standards. Chapter 19 in *The New Race Question: How the Census counts multiracial individuals.* J. Perlmann and M Waters (eds). Russell Sage Publications, 2002.

Appendix 1: Old and New CPS Race/Ethnicity Questions

Table A-1. Comparison of CPS questions on race and ethnicity[1]

Prior to January 2003		Starting in January 2003
What is your race?		Are you Spanish, Hispanic, or Latino?
Respondents are shown a flash card with the following:		Yes
RACE		No
1. White		
2. Black		
3. American Indian, Eskimo, or Aleut		
4. Asian or Pacific Islander		
(Chinese, Filipino, Japanese, Asian Indian, Korean,		
Vietnamese, Laotian, Thai, Other Asian, Hawaiian,		
Samoan, other Pacific Islander)		
What is your origin or descent? [2]		Please choose one or more races that you consider yourself to be
Respondents are shown a flash card with the following:		*Respondents are shown a flash card with the following:*
ORIGIN OR DESCENT		**CHOOSE ONE OR MORE**
	12 Mexican	
01 German	**14** Puerto Rican	White
02 Italian	**15** Cuban	Black or African American
03 Irish	**16** Central or South American	American Indian or Alaska Native
04 French (Hispanic Countries)		Asian
05 Polish	**17** Other Hispanic	Native Hawaiian or Other Pacific Islander
06 Russian	**20** Afro-American	
07 English	**26** Dutch	
08 Scottish	**27** Swedish	
10 Mexican-American	**28** Hungarian	
11 Chicano	**30** Another group not listed	

[1] The question wording is slightly different when the questions are asked during interviews by telephone.

[2] Individuals whose answers were coded in categories 10 through 17 were classified as Hispanics.

Appendix 2: Sample SAS code for merging probabilities with CPS data and generating five random race/ethnicity replicates

```
/* The following macro simulates a sample from the Dirichlet distribution with        */
/* parameter (X1+a,..,X4+a) for a=0.01.  For each row of RaceCnts with race counts    */
/* X1-X4 for the four single race ethnicities, generates four gamma random variables  */
/* N1-N4 and normalizes them so that their sum is 1. Outputs four probabilities P1-P4.*/

%Macro GenProb(Seed,P1,P2,P3,P4);
  Data RaceCnts;
    Set RaceCnts;
    A=.01;
    N1=RanGam(&Seed,X1+A);
    N2=RanGam(&Seed,X2+A);
    N3=RanGam(&Seed,X3+A);
    N4=RanGam(&Seed,X4+A);
    Sum=N1+N2+N3+N4;
    &P1=N1/Sum;
    &P2=N2/Sum;
    &P3=N3/Sum;
    &P4=N4/Sum;
    Drop N1-N4 Sum A;
%Mend;

/* The following datastep reads the 2003 CPS main survey data file.  Only variables    */
/* needed for matching with the probabilities are listed here.  The variables are as   */
/* follows:                                                                            */
/*                                                                                     */
/*    PrtAge: Persons age as of the end of survey week                                 */
/*     PESex: Sex (1=Male, 2=Female)                                                   */
/*  PRDTRace: Race                                                                     */
/*  PEHSPNON: Hispanic or Non-Hispanic (1=Hispanic, 2=Non-Hispanic)                    */
/*  PRPerTyp: Type of person record recode (2=Adult civilian household member)         */
Data CPS03;
  Infile CPS03 Missover;
  Input @0122 PrtAge   2.
        @0129 PESex    2.
        @0139 PRDTRace 2.
        @0157 PEHSPNON 2.
        @0161 PRPerTyp 2.;
  If (PrPerTyp=2) & (PrtAge=>15);
  If (15<=PrtAge<=44) Then AgeGrp=1;
  Else If (PrtAge>44) Then AgeGrp=2;
Run;

/* The following datastep reads the data from Table 6.: Race/ethnicity responses for   */
/* the CPS May 2002 supplemental sample: total number of respondents by age and sex    */
/* for non-Hispanics.                                                                  */
/*                                                                                     */
/* There weren't any respondents for the new race/ethnicity group White-Black-Asian    */
/* The numbers from 2 or 3 races will be used.  The variables are as follows:          */
/*                                                                                     */
/*    AgeGrp: Age Group (0=All, 1=15-44, 2=45+)                                        */
/*     PESex: Sex (0=Both sexes, 1=Male, 2=Female)                                     */
/*  PRDTRace: Race (2003 CPS)                                                          */
/*        X1: Count for White (2002 May CPS)                                           */
/*        X2: Count for Black (2002 May CPS)                                           */
/*        X3: Count for American Indian/Alaska Native (2002 May CPS)                   */
/*        X4: Count for Asian/Pacific Islander (2002 May CPS)                          */
Data RaceCnts;
  Infile Cards;
  Input @001 PRDTRace 2.
        @004 AgeGrp   2.
        @007 PESex    2.
        @010 X1       4.
        @015 X2       4.
        @020 X3       4.
```

```
        @025 X4       4.
        @030 Total    5.;
  PEHSPNon=2;
Cards;
 6  1  0   58   38    2    2   100   /* White-Black, 15-44, Both Genders               */
 6  2  0    9    8    2    0    19   /* White-Black, 45+, Both Genders                 */
 8  1  0   71    0    0   70   141   /* White-Asian, 15-44, Both Genders               */
 8  2  0   10    0    0   15    25   /* White-Asian, 45+, Both Genders                 */
 7  1  2  266    1   61    1   329   /* White-AIAN, 15-44, Female                      */
 7  2  2  229    2   35    0   266   /* White-AIAN, 45+ , Female                       */
 7  1  1  255    1   57    2   315   /* White-AIAN, 15-44, Male                        */
 7  2  1  201    2   20    1   224   /* White-AIAN, 45+ , Male                         */
10  1  0    8   62    1    1    72   /* Black-AIAN, 15-44, Both Genders                */
10  2  0    1   54    1    1    57   /* Black-AIAN, 45+, Both Genders                  */
14  1  0    2    0    0   36    38   /* Asian-NHOPI, 15-44, Both Genders               */
14  2  0    1    0    0   27    28   /* Asian-NHOPI, 45+, Both Genders                 */
 9  0  0   25    0    0   34    59   /* White-NHOPI, All Ages, Both Genders            */
13  0  0    0    0    3    3     6   /* AIAN-Asian, All Ages, Both Genders             */
11  0  0    1    5    0    3     9   /* Black-Asian, All Ages, Both Genders            */
12  0  0    0    6    0    3     9   /* Black-NHOPI, All Ages, Both Genders            */
18  0  0   10    0    0   29    39   /* White-Asian-NHOPI, All Ages, Both Genders      */
17  0  0    9    0    1    2    12   /* White-AIAN-Asian, All Ages, Both Genders       */
15  1  0   13   18    1    0    32   /* White-Black-AIAN, 15-44, Both Genders          */
15  2  0    9   10    1    0    20   /* White-Black-AIAN, 45+, Both Genders            */
19  0  0    0    2    0    0     2   /* White-Black-AIAN-Asian, All Ages, Both Genders */
20  0  0    4    1    1    4    10   /* 2 or 3 races, All Ages, Both Genders           */
21  0  0    1    0    0    4     5   /* 4 or 5 races, All Ages, Both Genders           */
16  0  0    4    1    1    4    10   /* White-Black-Asian, All Ages, Both Genders      */
;
Run;

%GenProb(631280,P11,P12,P13,P14); Run;
%GenProb(129857306,P21,P22,P23,P24); Run;
%GenProb(4478015,P31,P32,P33,P34); Run;
%GenProb(11148899,P41,P42,P43,P44); Run;
%GenProb(284959446,P51,P52,P53,P54); Run;

/* Output records for "All" Ages once for 15-44 and once for 45+                   */
Data RaceCnts;
  Set RaceCnts;
  If AgeGrp=0 Then Do;
    AgeGrp=1; Output;
    AgeGrp=2; Output;
    End;
  Else Output;
Run;

/* Output records for "All" Gender once for males and once for females             */
Data RaceCnts;
  Set RaceCnts;
  If PESex=0 Then Do;
    PESex=1; Output;
    PESex=2; Output;
    End;
  Else Output;
Run;

/* Sort the CPS dataset                                                           */
Proc Sort Data=CPS03;
  By PEHSPNon PRDTRace PESex AgeGrp;
Run;

/* Sort the race counts and probablities dataset                                  */
Proc Sort Data=RaceCnts;
  By PEHSPNon PRDTRace PESex AgeGrp;
Run;
```

```
/* Merge the CPS 2003 data with the probabilities dataset and create the 5           */
/* multiply imputed races.  For Hispanic and the race alone (i.e. White alone,       */
/* Black alone, ...) races in 2003 all 5 race variables get set to the appropriate   */
/* race.  The codes for Race1-Race5 are as follows                                   */
/*     1 = Non-Hispanic White                                                        */
/*     2 = Non-Hispanic Black                                                        */
/*     3 = Non-Hispanic American Indian/Alaska Native                                */
/*     4 = Non-Hispanic Asian/Pacific Islander                                       */
/*     5 = Hispanic                                                                  */
Data Together;
  Merge CPS03(In=In1)
        RaceCnts;
  By PEHSPNon PRDTRace PESex AgeGrp;
  If In1;
  If PEHSPNon=1 Then Do;
    Race1=5; Race2=5; Race3=5; Race4=5; Race5=5;
    End;
  Else If PEHSPNon=2 & PRDTRace In (1,2,3) Then Do;
    Race1=PRDTRace; Race2=PRDtRace; Race3=PRDtRace; Race4=PRDtRace; Race5=PRDtRace;
    End;
  Else If PEHSPNon=2 & PRDTRace In (4,5) Then Do;
    Race1=4; Race2=4; Race3=4; Race4=4; Race5=4;
    End;
  Else Do;
    Race1=RanTbl(7945631,P11,P12,P13,P14);
    Race2=RanTbl(9764325,P21,P22,P23,P24);
    Race3=RanTbl(9933511,P31,P32,P33,P34);
    Race4=RanTbl(21785664,P41,P42,P43,P44);
    Race5=RanTbl(1613219064,P51,P52,P53,P54);
    End;
  Drop P11-P14 P21-P14 P31-P34 P41-P44 P51-P54 X1-X4 Total;
Run;
```