RESEARCH REPORT SERIES
(Statistics #2011-04)

Recursive Partitioning for Racial Classification Cells

Aaron Gilary

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: March 28, 2011

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Recursive Partitioning for Racial Classification Cells

Abstract:
From the time the Census Bureau introduced an option to identify with multiple races on its survey questionnaires, researchers within the Census Bureau have sought the best way to aggregate the possible responses into categories while preserving the information from an increasingly multiracial country. Classifying racial data into categories helps provide information to Census stakeholders so they can measure the Census Bureau's performance in identifying and correctly enumerating each population. As planning intensified for the 2010 Census Coverage Measurement study, research staff analyzed the Matching and Correct Enumeration rates of multiracial populations, in order to model the data.

The paper details the techniques used to build models for Census Coverage data, by applying stepwise regression to the concept of CART modeling to partition the data into cells, and adding information criteria as a method of cross-validation. The paper also addresses: the specific issues inherent in modeling Dual-System Estimation data for this topic, and how they were addressed; the patterns of racial identification that were discovered; and the recommendations that were ultimately proposed.

Introduction: Multiracial Modeling for the Census Bureau

After many decades of collecting respondent race as a single characteristic, the Census Bureau provided each respondent the option to identify with more than one race on the 2000 Census questionnaire. In order to link racial identification from the 2000 Census to the earlier classifications where each respondent could only identify with one race, a system of rules was developed to collapse multiple race responses into domains which approximated the traditional, single race categories (Farber 2001).

The Census Coverage Measurement (CCM) Estimation team investigated these classification rules in a 2009 meeting. The Census currently uses post-stratification to form cells for racial classification, but debated whether logistic regression should be used instead to formulate these cells and whether the current arrangement was the optimal one for a population increasingly reporting more than one race.

The Census asks about race using six categories: American Indian, Black, Native Hawaiian/Pacific Islander, Asian, White, and Some Other Race. As noted, each respondent can identify with one or more of these categories. These categories, together with Hispanic origin, offer 126 different options for race/origin classification. Should the Census Bureau use these options to expand its domain structure for 2010 to provide more information about populations identifying with more than one race?

To help answer this question, research began on an alternative method for forming race/origin subgroups. Recursive partitioning methods were applied to the CCM data sets as a way to develop new domains, as researchers strove for a balance between a minimal number of classification errors, and homogeneity in the rates used to evaluate Census Coverage.

Part I: Census Coverage, and our partitioning methods

The Census Coverage Measurement sample survey is a follow-up to the Census itself, conducted for an independent verification of the Bureau's performance capturing the nation's population. It is an area sample that contains two sources of data – the Post Enumeration Sample ("P-Sample") selected randomly from across the country as the verification sample, and the Enumeration Sample ("E-Sample") that comes from the corresponding Census records (Bell and Cohen 2009).

Census Coverage Measurement uses Dual-System Estimation as a means of evaluating the Census data. The standard dual-system estimator employed by Census Coverage holds that the estimate of the total population ($\widehat{N}_{++}$) can be derived from the Census and P-Sample totals as in the equation:

$$\widehat{N}_{++} = ((N_{+1} * N_{1+})/N_{11}),$$

Where
$\widehat{N}_{++}$ is the estimate of the total number of people;
N_{+1} is the number of people counted in the E-Sample;
N_{1+} is the number of people counted in the P-Sample;
N_{11} is the number of people counted in both the E-Sample and the P-Sample.

The counts in the formula are adjusted to eliminate duplicate, fictitious, or otherwise erroneous enumerations, and therefore reflect the number of *actual* people counted in each survey (Wright and Hogan 1999).

The Dual-System Estimate assumes that the P-Sample and E-Sample are independent and that the ratios of survey totals in the $\hat{N}_{++}$ equation are synthetic at an aggregate level, meaning that they can be applied to smaller geographic levels with no error other than classification error (Hogan 2003). This model for estimating the total population relies on several assumptions, including the Post-stratification Assumption, which holds that the total population may be calculated by producing estimates across any variable and then aggregating them for an estimate of the total (Wolter 1986).

The *Match Rate* (i.e., the proportion of persons in the P-Sample with a valid Census record) was used as the dependent variable for the P-Sample modeling, and the *Correct Enumeration (CE) Rate* (i.e., the proportion of Census records that are correct) was the dependent variable for the E-Sample modeling in this research study. The Match and CE rates are important because they measure the Census Bureau's success at population estimation, both overall and with respect to variables such as race, age, sex, and household tenure status (i.e., owners versus renters). The Census Bureau currently uses these covariates to separate the data into post-strata that are homogenous according to match and CE rates. There were 416 such post-strata in 2000 (Hogan 2003).

In order to expand the racial domain structure, a recursive partitioning method was chosen to partition respondent records from both samples into new race/ethnicity cells because there was a desire that the cells be mutually exclusive and because such a method would not exclude higher order interactions between covariates. The goal was to create a model to partition the records into cells through the following steps:

-- Start with a group of variables as potential covariates for the modeling (in this case, Hispanic Origin and the six racial categories, which are used to predict the CE and Match Rates).
-- Add the most significant variable to the model using stepwise logistic regression (stopping the procedure after one step).
-- Use that variable to split the dataset into two groups in a branching structure.
-- Then use stepwise regression again to find the most significant remaining variable for each of *those* groups. This procedure is repeated until the regression model (or "tree") has been expanded as far as possible.
-- Finally, prune back the tree using a model selection criterion. The Schwartz Bayesian Criterion (SBC) is used here, as it assesses a penalty for each parameter added to the model based on the log of the total number of records, and therefore favors more parsimonious models. This scaled back tree is the final model.

This procedure uses the basic concept of Classification and Regression Trees (CART). But unlike CART, which uses classification based on percentages, the procedure uses a likelihood-based selection. The likelihood mechanism used weighted observations in proportion to their sampling weight, adjusted to sum to the sample size. Any possible correlation between observations is ignored.

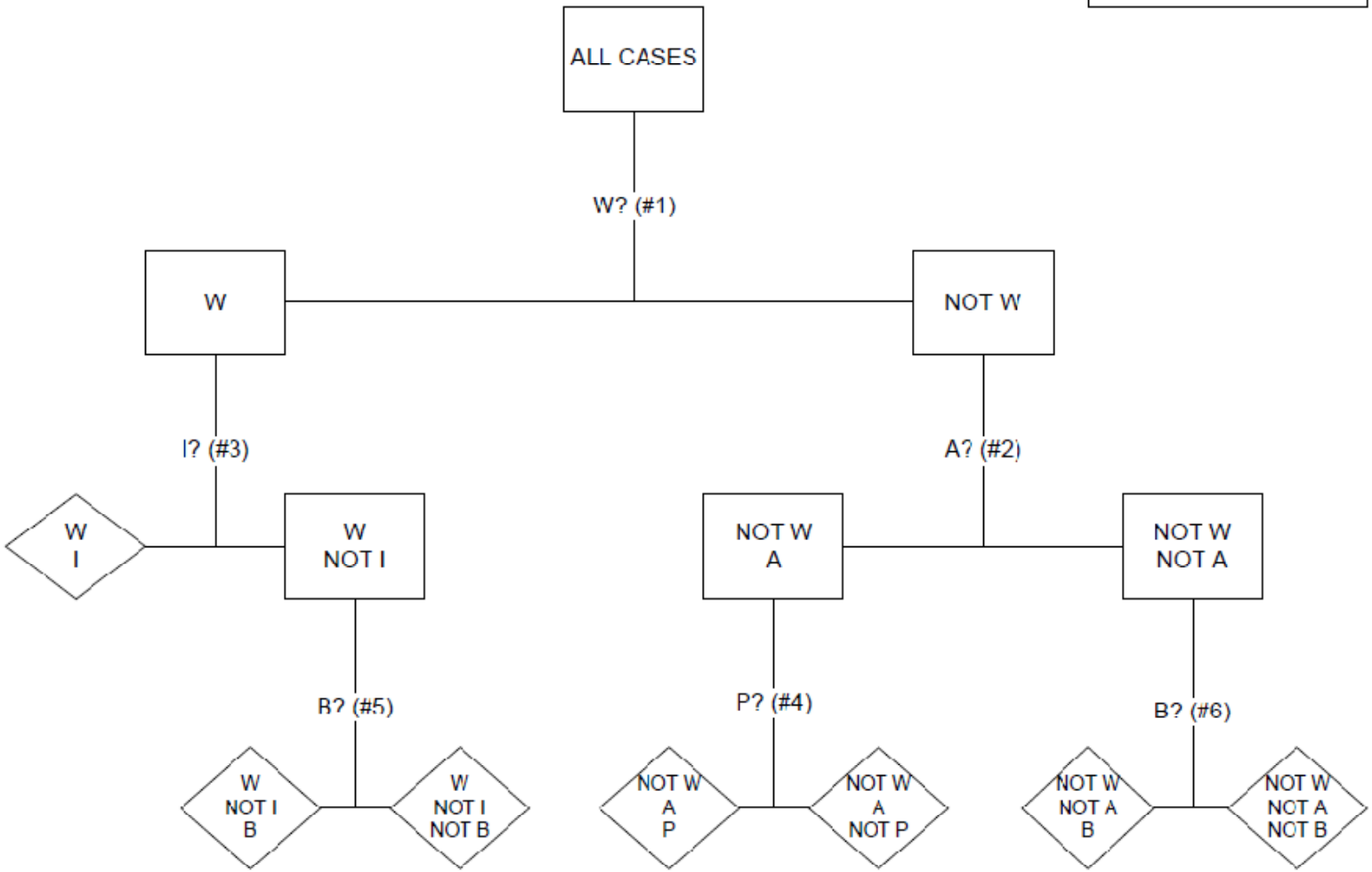
Ultimately, the procedure created four trees (Match Rates and CE Rates for both Non-Hispanics and Hispanics). The procedures detailed in the next three sections use the Non-Hispanic Match Rate results as a proxy for all four trees. The tree for the Non-Hispanic Match Rate is given on the next page as **Flowchart 1**.

Issues

Initially, the project studied Hispanics and Non-Hispanics separately to address a question about the relationship between race and ethnicity identification for Hispanics. As a result, this analysis is divided into a Non-Hispanic and a Hispanic section. Hispanic ethnicity was coded on the same scale as the races for this project, even though race and ethnicity are different variables.

FLOWCHART 1: P-SAMPLE TREE (NON-HISPANICS)

W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER



It must be noted that this procedure does not study main effects beyond the first variable. But regression as classification is not in the scope of this work – the goal is to find classification cells to address multiracial populations.

There were also some issues of classification variability, in terms of respondents interpreting the question of their racial ethnicity in different ways. This variability within the modeling may be due to the different modes of the survey: the Census uses a mail return questionnaire which is self-administered, while Census Coverage Measurement uses Computer Assisted Personal Interviewing (CAPI) in which an interviewer administers the questionnaire to respondents. A limitation of this study is that with a subjective measure such as race, there is no objectively correct answer against which to compare.

Creation of the Regression Tree

Flowchart 1 illustrates the partitioning mechanism by which all the Non-Hispanic cases filter into one of the classification cells. It represents the tree after the SBC has been employed and the branches have been pruned. The process begins with all the cases in one bucket, at which point the model goes through all the covariates and picks out 'White' (W) as the most important covariate. The data is separated into those records where 'White' was marked, and those where it was not. At the next step, the process repeats for those two distinct groups (White/Not White). The American Indian covariate (I) is found to be the most important for the 'White' subgroup, but the Asian covariate (A) is found to be the most important for the non-White subgroup. Then, each of the new subgroups is tested and new covariates emerge. This process is extended for many stages, although it is ultimately pruned back through stepwise regression to what is seen in the flowchart.

The final cells are represented by the diamond shape. The key in the upper right shows the letters used to denote each racial category. The numbers reflect the ranked statistical significance of each split: the first split, on the 'White' covariate, is the most significant.

All cases will fit into one of these cells based on their combination of racial responses. Racial combinations with similar match rates will be grouped together into the same cell. Some of the variables in this chart correspond to a higher match rate than the baseline (White, Asian); others correspond to a lower match rate (American Indian, Black, and Pacific Islander).

It is worth pointing out that the Match Rate tree is different from the CE tree for the same Non-Hispanic population (the other trees are represented by Flowcharts 2- 4 in the Appendix) and that match and CE rates can be poorly correlated for the different subgroups. Pacific Islander identifiers have a high CE rate and a very low match rate, while respondents who marked both White and Other have a low CE rate relative to match rate. White identifiers have higher match rates and CE rates than any other race.

Application of SBC to the model

After the framework for the tree was established, each variable addition to the model was evaluated using the SBC. The SBC used here selects a smaller model than Akaike's Information Criterion (AIC), due to its penalty function. But while the criteria may vary, all of them will essentially measure the improvement in the model log-likelihood with an extra covariate added, and compare it to a penalty for that extra parameter.

Table 1 on the next page displays all the variables that were added to the Non-Hispanic Match Rate model and how those variables affected the weighted log-likelihood. This chart is meant to illustrate the effects of the variables. Each row of the chart represents a stage of the variable selection process. It shows a given subsection of the data set, defined by variables that were marked and/or unmarked and how each subsection changes when a new variable is added. The first two columns show the marked and unmarked variables (respectively) which conditionally define the cell, and the third

Table 1: P-Sample Racial Parameters to Add (Non-Hispanics)

Step	Marked	Not Marked	New variable	-2x Old Log Likelihood	-2x New Log Likelihood	Difference	SBC
1	-	-	W	297,037	293,672	3,365.71	Y
2	-	W	A	77,858	77,690	168.24	Y
3	W	-	I	215,813	215,785	28.52	Y
4	A	W	P	14,495	14,470	25.14	Y
5	W	I	B	213,139	213,116	23.50	Y
6	-	WA	B	63,195	63,179	15.68	Y
7	WI	-	P	2,645	2,633	12.37	N
8	WIP	-	A	53	42	11.29	N
9	BO	WAP	I	533	523	9.54	N
10	WB	I	O	1,176	1,168	7.99	N
11	O	WAB	P	5,894	5,888	6.48	N
12	W	IB	O	211,940	211,934	5.54	N
13	W	IBO	A	210,653	210,647	5.42	N
14	P	WABO	I	632	627	4.78	N
15	A	WP	I	14,236	14,232	3.94	N
16	-	WAB	O	9,114	9,110	3.83	N
17	W	IBOA	P	209,687	209,683	3.63	N
18	WIB	P	O	183	180	2.93	N
19	AIB	WP	O	17	14	2.66	N
20	AB	WPI	O	175	173	2.36	N
21	A	WPI	B	14,160	14,158	1.93	N
22	AP	W	O	234	233	1.56	N
23	AI	WP	B	72	71	1.37	N
24	WIO	PB	A	78	76	1.24	N
25	WB	IO	A	1,100	1,099	1.20	N
26	OP	WAB	I	46	45	1.16	N
27	AI	WPB	O	54	53	1.16	N
28	WBA	IO	P	30	29	1.02	N

W=WHITE B=BLACK A=ASIAN
P=PACIFIC ISLANDER I=AMERICAN INDIAN
O=OTHER

column shows the potential variable being evaluated for inclusion to the model for that group. (For example, the fourth row shows the variable selection process applied to those records that checked Asian but did not check White, and the new variable being evaluated is Pacific Islander.) The first row is for a cell with no conditions on what has been marked, i.e., the entire P-Sample.

The subsequent columns show:

-- the log-likelihood before the addition of the new variable, multiplied by (-2) – as log-likelihood is used within model selection criteria;

-- the log-likelihood before the addition of the new variable, multiplied by (-2);

--the difference between these scaled log-likelihoods (the chart is sorted by this difference);

--whether or not the difference was large enough for the SBC to include the variable in the model (SBC uses the log of N as the criterion, and so it sets a cutoff of about 13.4 for including a new variable).

The first row shows the inclusion of the initial covariate 'White' as a parameter generated a difference of almost 3,400, so the SBC dictated that it is used in the model. Once this variable was incorporated, the model split the original cell based on the presence or absence of the White covariate, as discussed in the previous section.

The second row is only for the group that DID NOT mark White. The new race variable suggested for this group is 'Asian'. Here the difference is about 170 – a much smaller effect than was seen in the previous step, but the SBC again dictated that we use it in our model. Note that the total for the old likelihood is different for this step because when each cell is split, a new test is performed.

As soon as this variable is accepted, the new model features one cell for those who marked White, another for those who marked Asian but not White, and a third for those who marked neither White nor Asian.

Ultimately, the model selected six variables for inclusion, as represented by the top six rows. Since each new variable included effectively adds one extra cell to the model, the final model for the Non-Hispanic Match Rate will have seven nodes (this is reflected in Table 1, and the observation numbers in that table correspond to the numbers marked on the Flowchart).

Match Rate/Correct Enumeration Rate Cells

Similar procedures were carried out for Hispanic match rates and Hispanic and Non-Hispanic CE rates, resulting in four final trees. **Table 2** shows the cells for those four trees. For the P-Sample, there are seven Non-Hispanic cells and seven Hispanic cells. For the E-Sample, there are five Non-Hispanic cells and two Hispanic cells. The unweighted sample sizes, weighted sample sizes, and match rates are given. (Note that all rates in this analysis use weighted data, and the total weights are in proportion to the sample size.)

The two trees are constructed differently, as the different racial variables do not have the same degree of importance for the two samples. They do share their first three Non-Hispanic nodes in common:

--White;

-- Asian but not White;

-- Neither White nor Asian.

There was a brief attempt to combine the E-Sample and P-Sample trees into one unified tree, but this idea was rejected because it involved creating too many slices from different combinations of data. For example, if one sample split Hispanic cases based upon White identification and the other split those cases upon Asian identification, the unified tree would split the cases based on both variables and the nodes would be too specific to be of use.

The conclusions were mixed for this segment. The procedure did create a group of heterogeneous cells that were informative and easy to interpret, but there were concerns about

inconsistency between 2000 and 2010, as well as concerns about how to address the effects of imputation on this data, if this procedure were to be implemented.

Part I Conclusion

While the initial results provided a new framework for grouping the data, they were not entirely conclusive. They did not address whether this procedure could work in a localized way within the domains, or how consistent racial identification is between the two samples.

Part II: Expanding Domains into Sub-Domains

After further discussion, it was decided to apply the partitioning methods to split each of the original seven domains further, so that the domains could retain historic consistency with the previous census. At the same time, the team wanted to measure the stability of racial partitions with "matched records" contained in both the census and the P-Sample. These two investigations are the focus of this section.

Cell Modeling Within the Domain Structure

The definitions of the current domains are given in **Table 3**. These definitions come from a DSSD Memorandum, but they are reorganized here as an informal guide to each scenario (Haines 2001).

There are two things to note here about this classification structure. First, the domain definitions affected the branching and kept some models very small. For example, the Non-Hispanic Asian domain did not split at all, because there was nothing to split: an Asian respondent who marked any additional category in her response would have been automatically placed into a different domain. Second, there are geographical factors at play here in addition to race/ethnicity. Residence in Hawaii or on an American Indian Reservation contributed to the classification as well.

Table 4 shows the results of recursive partitioning when applied to the E-Sample and the P-Sample. For each of those, the top graph shows the domains' totals, weighted totals, and rates; and the bottom graph shows this information partitioned into the cell model. All totals are proportionally weighted to the sample size, and there may be some rounding errors.

Although there were some similarities, the samples' splits within the domains are based on different racial indicators. And the splits dictated by the recursive partitioning model, especially for the P-Sample, can be affected by either size or rate differential. In Domain 2, Off-Reservation American Indians were split by the 'Asian' covariate, based on a small population of Asian identifiers in that domain with a much lower match rate. In Domain 3, Hispanics were split by the 'White' covariate based on a very large population of White identifiers with a match rate that was only slightly smaller.

Examining Consistency of Domain Classification

After the domain classification was finished, the project focused on measures of stability to gauge the consistency of racial classification for the two models. The study involves a different data set, comprised of the 578,300 P-Sample cases that matched to Census enumerations. The samples need to be linked to study consistency so that two race/origin responses are present on each record.

Table 5 shows the overall stability of the domains by illustrating how E-Sample cases map back to the same P-Sample domain and vice versa. The data here is a little bit mixed. Over 96 percent of cases

have consistent race domains for the two samples, but that number is somewhat inflated by the mostly White Domain 7; both Off-Reservation American Indians and Native Hawaiian/Pacific Islanders are under 76 percent for the two samples. This finding is consistent with the results in Farber (2001). It appears that the most stable ones are the most clear-cut, such as American Indians on a Reservation or Non-Hispanic Asians, but overall most domains are around the 90 percent level.

Table 6 uses all of the domain subgroups to examine the stability of the model. The two different models were used to split up the data into every possible combination. Different combinations are marked by the letters (A, B, C, D) for the P-Sample cell and (X, Y) for the E-Sample cell. (For example, in the Hispanic Domain, the P-Sample tree uses 'White' as a covariate and the E-Sample tree uses 'NHPI', so each combination among those variables was examined.) The data is given in the attachment, and the segments listed at the bottom show the domains and the relevant variables for each of the cells. If a racial variable is not listed, it is not relevant to the process of forming the cells.

Ultimately, these models are not very stable, whether evaluated together (as in this table) or separately. The splits being created do not lead to a clear-cut cell model. This illustrated the problem with creating one model from the Census and P-Sample – merging boundaries yields an unwieldy model which is too sparse to signify anything.

Explanations for Inconsistency

Data analysis of the inconsistent cases helped to uncover one source of the inconsistency problem. It was found that splitting the matched data into cases where the race/origin answer combinations do and do not match shows that 89 percent of respondents match their identification exactly between the E-Sample and the P-Sample. But the results of splitting that data into single-race and multi-race identifiers present a different picture:

88.4 percent of matched respondents marked one race on the Census, and matched that in the CCM follow-up;
.55 percent of matched respondents marked multiple races on the Census, and matched those in the CCM follow-up;
9.3 percent of matched respondents marked one race on the Census, but changed at least one answer in the CCM follow-up;
1.7 percent of matched respondents marked multiple races on the Census, but changed at least one answer in the CCM follow-up.

Table 7 has the breakdown for all four possibilities sorted by the likeliest race/origin responses. (The number of racial responses listed will not sum to the total, which includes other combinations that are not listed.)

To put it another way, the data shows that *less than 24 percent of respondent records self-identifying as more than one race on the Census later used those exact same categories to self-identify in the Census Coverage Measurement follow-up*. Even taking into account the aforementioned issue with 'Some Other Race' identifiers, this fact captures the problem in creating a single model – for those records that are multiracial identifiers, there is a limited effectiveness to these racial data. (Note: Hispanic Ethnicity does not count as a race here when assessing whether a Census respondent identifies with more than one race, but it does count as a category when assessing consistency versus inconsistency in the two samples.)

Issues

It should be noted that there are about 500 cases nationwide for each sample case in Census 2000, and that ratio is increasing for Census 2010 as the sample is being cut. The more weighting is used, the more the analysis is reliant on modeling assumptions.

There was also an issue in the way Some Other Race cases were coded for Coverage data. The ACS Implementation Report explains that "detailed review of the edits used in Census 2000 led to the discovery of a difference on enumerator returns" (Griffin et al. 2004). As a result, the number of Census respondents identifying as Some Other Race was incorrectly inflated. But this error was limited in scope and should not affect broader conclusions.

This report is meant to highlight important issues in Multiracial/Multiethnic classification. More definitive results, using tests and hypotheses, could be obtained in a follow-up analysis.

Part II Conclusion

Although the domain subgroups offered a new method for classifying the data, the inconsistency in race reporting between the samples largely undermined these findings. At this point there was not enough evidence to seriously consider overturning the current domain assignments. However, there was still interest in using the subgroups found with this procedure to evaluate if the domains needed any adjustments.

Part III: Following Up and Assembling Information

Studying target populations from Parts I and II

After those findings were presented and discussed, the estimation team proposed four race/origin subgroups to investigate more closely, based on unexpected results they had shown in the earlier research. Those groups were:

(A) *'Hispanic' and 'NHPI' identifiers in the HISPANIC domain (Domain 3)*. This group is possibly a specific ethnic population. All Hispanic identifiers will be put into Domain 3 unless there are geographic concerns, but the group's match and CE rates will be compared to Hispanic and Native Hawaiian/Pacific Islander (NHPI) baselines (from Domains 3 and 5, respectively).

(B) *'Black' and 'White' identifiers in the NON-HISPANIC BLACK domain (Domain 4)*. These cases are generally put into the Non-Hispanic Black domain. They are compared to the baselines from the Non-Hispanic White and Some Other Race domain (Domain 7).

(C) *'Asian' and 'NHPI' identifiers in the NATIVE HAWAIIAN/PACIFIC ISLANDER domain (Domain 5)*.If someone identifies with both of these races, they are generally placed with NHPIs. This cohort is compared to NHPIs (Domain 5) and Non-Hispanic Asians (Domain 6).

(D) *'Some Other Race' identifiers in the NON-HISPANIC WHITE OR "SOME OTHER RACE" domain (Domain 7)*. The scheme for Domain 7 is fairly complex, but it basically includes 'White' identifiers and 'Some Other Race' identifiers (along with a few heterogeneous examples such as Black & Asian & NHPI). The 'Some Other Race' group is split from the rest of Domain 7, and studied separately.

The results are given in **Table 8**. The first chart goes through how the domains are actually arranged, including the unweighted and weighted totals, correct enumeration rates and match rates (which are computed separately), and consistency rates. The consistency percentage given on the right hand side denotes the weighted percentage of records in the Census that were classified in the same domain in the P-Sample.

The second chart shows the racial subgroups. The left column shows that A, B, C, and D have been removed from the existing domains and incorporated into new ones. The rates for these new domains are different from the initial domains in many instances, but the consistency percentages are much smaller.

Each row is defined as a separate group for these charts, and consistency is defined as matching to this exact same row. Per this definition, adding more structure for each domain will decrease the consistency rate, as it would add more requirements for consistency.

Note the unweighted numbers highlighted here. These totals may be affected, to varying degrees, by cuts to the coverage sample in 2010.

Table 9 expands on these results to show the rates and consistency percentages for any possible domain that could be constructed, with the target groups added and removed from domains where they might be classified. Note that the rates stay fairly consistent for these domains for all possible adjustments, except for a slight decrease in the consistency rate.

On one hand, there is no real effect to the rates of the parent domains when the subpopulation is removed because of the large number of monoracial identifiers keeping the rates stable. (The Native Hawaiian/Pacific Islander domain is the lone exception, since it is much smaller than all the others which might be affected.) However, it should be noted that the subpopulations may have a rate substantially different from the parents. So if the group is of interest by itself, there is a reason to split it out systematically; but if not, splitting does not really matter to the larger domain.

This domain separation has different advantages and disadvantages for the separating of each racial population:

(A) *'Hispanic' and 'NHPI' identifiers*. Under the current classification system Hispanics and NHPIs are placed into the Hispanic domain. The match rates are about the same as the larger homogenous groups, but the CE rates are drastically lower than each of those groups (and slightly closer to NHPIs). This may be worth studying, even though the sample is smaller and less consistent than other target groups.

(B) *'Black' and 'White' identifiers*. This group is placed into the Non-Hispanic Black domain but performs better than solely Black identifiers in both match and CE rate (although worse than solely White identifiers in both of these categories). Also, they seem to be fairly consistent relative to other multiracial populations, and the subgroup is relatively large.

(C) *'Asian' and 'NHPI' identifiers*. For some reason this group has match rates that are unusually low, and CE rates that are unusually high. Such unconventional data bears watching. The group is fairly consistent and has large raw totals, but the weighted totals are substantially smaller.

(D) *'Some Other Race' identifiers*. When 'Some Other Race' identifiers are separated from the others within this domain, their rates are much lower for both Match and Correct Enumeration. The population is relatively large for both the E-Sample and the P-Sample, although this group is also relatively inconsistent. There are other issues to consider for this subgroup, such as the fact that 'Other' identifiers are by definition difficult to place into any particular domain.

Studying Hispanic Ethnicity Identification

An estimation team member suggested further investigation into Census Coverage data with respect to Hispanic ethnicity information he had used (variables which labeled each Hispanic identifier by their ethnic ancestry, such as Cuban, Mexican, or Puerto Rican). His interests were: the effect of geographic boundaries on match/CE rate; the racial identification, and consistency, of different subpopulations; and how well these subgroups compare to Census domains.

How does geography affect Hispanic subgroups?

Table 10 lists the effects of geography on both CE and Match Rates for Hispanic subgroups. For each rate, the first chart shows these rates by Census Division, with Hispanic and Non-Hispanic subtotals below and then the grand total featured below. Census Divisions are groupings of states nested within the four Census Regions, and Divisions are used here to avoid confusion between different Hispanic populations. (For example, the use of Division distinguishes between Florida and Texas, which are both in the South Region but have different Hispanic demographics.)

The second chart standardizes the rate by subtracting the Hispanic baseline rate. This causes the Hispanic line to go to zero and the other lines are adjusted accordingly. According to the P-Sample charts, there are several match rates that stand out with a better or worse than the overall Hispanic rates from that division. Cuban-Americans have a very high match rate, even higher than Whites. The match rate is especially high in the South Atlantic division (Division 5), which includes the state of Florida where the bulk of the Cuban-American population is concentrated. Not many Hispanic populations stand out in their CE rate; Cubans have a higher CE rate in Division Five, but only slightly.

The third chart shows the relevant sample sizes for these populations. This indicates that Cuban-Americans and Mexican-Americans in the South Atlantic division have a critical mass, but some of the other Hispanic groups with large discrepancies are too small for those differences to be statistically significant. The number of Cuban-Americans in the South Atlantic division dwarfs the Cuban-American population anywhere else.

Are Hispanic subgroups consistent, as groups?

The first question to pursue with this Hispanic subgroup data was how consistent the codes are. **Table 11** modifies some previous work with all domains, displaying totals and rates for each racial cohort on the Census and P-Sample and how they match back to each other, as seen in Table 5. Consistency is in the 80-85 percent range for the three major groups. It is a little lower for 'Other Hispanic,' but that is an aggregated category.

Note the difference between the 11,600 cases changing from Non-Hispanic in the Census to Other Hispanic in the P-Sample, and the 1,700 changing from Other Hispanic to Non-Hispanic. The error in Some Other Race processing mentioned earlier may be a factor here, or possibly the error could be due to differences in how these studies were conducted led to differences in racial identification. Also note that the grand total for this table is lower than for other tables with merged data (549,668 versus 578,300). There were stricter guidelines for this data set, and only those cases with detailed race information in both samples were used.

Are Hispanic subgroups consistent in racial identification?

The study expanded to examine the race of each subgroup. **Table 12** also modifies previous work, illustrating the aggregate E-Sample and P-Sample racial responses as variables called 'ecode' and 'pcode', as seen in Table 7. Table 12 shows the number of E-Sample and P-Sample cases for each combination (labeled 'ecases' and 'pcases'), and then all cases with a specific Census response and a specific P-Sample response on the merged file ('mcases').

Cuban-Americans are relatively white and homogenous – identifying as Hispanic and White almost exclusively. The other groups identified less consistently, and with no general pattern. Because of the lack of Black identifiers, the 'Other' response may be an unintended proxy for race, perhaps implying Black Hispanic as opposed to White Hispanic.

How do Hispanic subgroups compare to the larger domains?

Table 13 shows the statistics for all domains and breaks out subgroups of interest from the Hispanic domain, showing the rates for all subgroups and how they compare to the relevant domains. Unweighted totals are given in the first column, and weighted totals are used thereafter. The rates seem a little irregular across different groups, but Cubans do have a good match rate, as seen in Table 10.

The standard errors for these subgroup rates are given in **Table 14**. The natural log coefficient and the standard error (given in the first two columns) produce a point estimate and a confidence interval. Although there is a high standard error for the Cuban-American group, they still have a significantly better match rate than any other Hispanic cohort. This finding, combined with a consistent of White identification, suggests that Cuban-Americans may be a population worth further examination.

Effects of Imputed Cases on these Data

The study was expanded further to examine cases with imputed Race or Origin data. There were some small changes to the data patterns when we separated these cases from the ones with respondent reporting.

The CE rates for reported cases are slightly higher across the board than the rates for imputed cases, which seems logical. Hispanic Match Rates are generally about the same or slightly higher.

Among Hispanics, records identifying as Hispanic and 'Some Other Race' contain imputed data at a slightly higher rate than what we see in the reported data. The same goes for cases in the Other Hispanic subgroup.

Effects of Tenure on these Data

The study was also expanded to look at the differences between owners and renters (denoted by the 'tenure' variable), as they related to Hispanic populations. Tenure has a traditionally strong effect on the coverage rates, so it seems important to evaluate the Hispanic data in that context.

The CE rates are much higher for owners than for renters (the totals are .945 for Hispanic Owners and .948 for Non-Hispanic Owners, versus .909 for Hispanic Renters and .905 for Non-Hispanic Renters). This pattern is true across different divisions or subgroups. The South Atlantic Cuban Renters have a CE rate of .926, which compares favorably to any division or Hispanic cohort total among renters.

The Match rates are *also* much higher for owners than for renters (.913 for Hispanic Owners and .946 for Non-Hispanic Owners, versus .842 for Hispanic Renters and .875 for Non-Hispanic Renters). This pattern is true across different divisions or subgroups. In this context, the .912 match rate for South

Atlantic Cuban Renters is relatively strong. Even though it lags behind the South Atlantic Cuban Owners rate of .962, it compares favorably to any cohort of renters.

In terms of Match Rates, Hispanics trail Non-Hispanics by far for both owners and renters across all segments of the data. For CE rates, the picture is murkier. Puerto Rican owners are about on par with Non-Hispanic owners, and Other Hispanic renters are exceeding Non-Hispanic renters.

Finally, tenure was used to examine the Hispanic demographics. The Hispanic population has about a 1:1 ratio between owners and renters. For Non-Hispanics, the ratio is approximately 3:1. There are slightly more owners than renters among Cubans, but for other Hispanic groups the reverse is true. White Hispanics make up a sizable percentage of Hispanic owners, but a smaller percentage of Hispanic renters.

Part III Conclusion

The Cuban subgroup may be worth studying. This cohort has very high match rates (on a par with Whites), and consistent White identification. Other variables had some effect on the data but did not exhibit the strong patterns that Cuban Americans did.

Future Research to Complete

Look at other covariates to introduce into the race model. One suggestion arising from this project involved using inconsistent racial identification as a dependent variable, and determining if it can be explained using other covariates, such as age/sex/tenure. Perhaps some of these characteristics could enhance the model. This research may be conducted as part of a future project.

Conclusion/Recommendations

Self-reported race is a complex and thorny topic to study, and it requires a balance between quantitative conclusions and qualitative knowledge of the subject, while also taking into account historical definitions and peculiarities that exist within the topic.

For both Non-Hispanics and Hispanics, the dual criteria of forming cells with similar Match and CE rates produced different domains that were difficult to reconcile. An approach based on obtaining a single tree may be needed – but this would require additional time for formulating the domains.

Unfortunately the scope of the project and the strength of its recommendation are restricted by the nature of the Dual-System Estimation. There are limitations inherent in the data when examining racial consistency in reporting because only the matched sample is available; the racial consistency of reporting for the unmatched population is unknown. There is also a large amount of inconsistency between the E-Sample and P-Sample: often they will contain different records, or the same record may report race differently.

The groups created by the recursive partitioning models are homogenous for Correct Enumeration and Matching, but they are not very consistent, particularly when it comes to people who identify with multiple races. As a result, it might be difficult to expand the race domains as the group may have envisioned. The issue with poor racial consistency may be a matter of understanding what is being asked, or it may be a matter of weak identification with different groups. Either way, it should be a concern that so many groups in this racial framework are difficult to identify.

At a minimum, the singling out of five groups for evaluation and sensitivity analysis is recommended. Those groups, all of which were covered earlier, are:

(A) Hispanic/NHPI identifiers in the HISPANIC domain (3).

(B) Black/White identifiers in the NON-HISPANIC BLACK domain (4).
(C) Asian/NHPI identifiers in the NATIVE HAWAIIAN/PACIFIC ISLANDER domain (5).
(D) "Some Other Race" identifiers in the NON-HISPANIC WHITE OR "SOME OTHER RACE" domain (7).
(E) "Cuban" identifiers among Hispanics.

As the Census 2010 data begins to arrive, there is likely to be a noticeably larger population reporting more than one race, and researchers should be well equipped to measure it. Studying the data of multiracial populations doesn't provide any clear-cut answers, but it does illustrate how the national racial composition changes, and where any related research should focus.

References

Bell, Robert M., and Cohen, Michael L., eds. (2009). *Coverage Measurement in the 2010 Census*. Washington, DC: National Academies Press.

Farber, James (2001). *DSSD Census 2000 Procedures and Operations Memorandum Series B-10.*

Griffin, Deborah H., Broadwater, Joan K., Leslie, Theresa F., Love, Susan P., Obenski, Sally M., and Raglin, David A. (2004). *Meeting 21st Century Demographic Data Needs – Implementing the American Community Survey, Report 4: Comparing General Demographic and Housing Characteristics with Census 2000*.

Haines, Dawn (2001). *DSSD Census 2000 Procedures and Operations Memorandum Series #Q-48.*

Hogan, Howard (2003). *The Accuracy and Coverage Evaluation: Theory and Design*. Statistics Canada: Volume 29, Number 2, pp.129-138.

Wolter, Kirk M. (1986). *Some Coverage Error Models for Census Data*. Journal of the American Statistical Association: Volume 81, Number 394, pp.338-346.

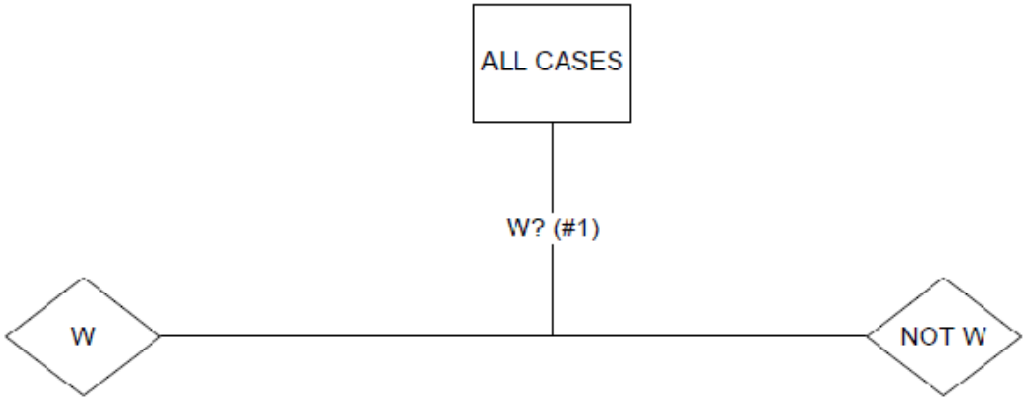
Wright, Tommy, and Hogan, Howard (1999). *Census 2000: Evolution of the Revised Plan*. Chance: Volume 12, Number 4, pp.11-19.

Appendix:

Flowchart 2:	P-Sample Cell Assignment (Hispanics)	19
Flowchart 3:	E-Sample Cell Assignment (Non-Hispanics)	20
Flowchart 4:	E-Sample Cell Assignment (Hispanics)	21
Table 2.a:	E-Sample Correct Enumeration Cells (Non-Hispanics)	22
Table 2.b:	E-Sample Correct Enumeration Cells (Hispanics)	22
Table 2.c:	P-Sample Match Rate Cells (Non-Hispanics)	23
Table 2.d:	P-Sample Match Rate Cells (Hispanics)	23
Table 3:	Summary of Census Domain Conditions	24
Table 4.a:	Recursive Partitioning Model (E-Sample)	25
Table 4.b:	Recursive Partitioning Model (P-Sample)	26
Table 5:	General Consistency of Race/Origin Domains	27
Table 6:	All Domain/Cell Racial Combinations for the E-Sample and P-Sample	28
Table 7.a:	Monoracial Census cases with Consistent Race/Origin Classifications	29
Table 7.b:	Multiracial Census cases with Consistent Race/Origin Classifications	30
Table 7.c:	Monoracial Census cases with Inconsistent Race /Origin Classifications	31
Table 7.d:	Multiracial Census cases with Inconsistent Race/Origin Classifications	32
Table 8:	Correct Enumeration and Match Rates for Domains and Key Groups	33
Table 9:	Alternative Race/Origin Domain Integration for Key Groups	34
Table 10.a:	Effect of Geography on Hispanic Subgroups (E-Sample)	35
Table 10.b:	Effect of Geography on Hispanic Subgroups (P-Sample)	36
Table 11:	Consistency of Hispanic subgroups	37
Table 12.a:	Common Race/Origin Codes for CUBAN subgroup	38
Table 12.b:	Common Race/Origin Codes for MEXICAN subgroup	38
Table 12.c:	Common Race/Origin Codes for PUERTO RICAN subgroup	39
Table 12.d:	Common Race/Origin Codes for OTHER HISPANIC subgroup	39
Table 13:	Correct Enumeration and Match Rates for Hispanic subgroups	40
Table 14:	Standard Errors for Hispanic groups	41

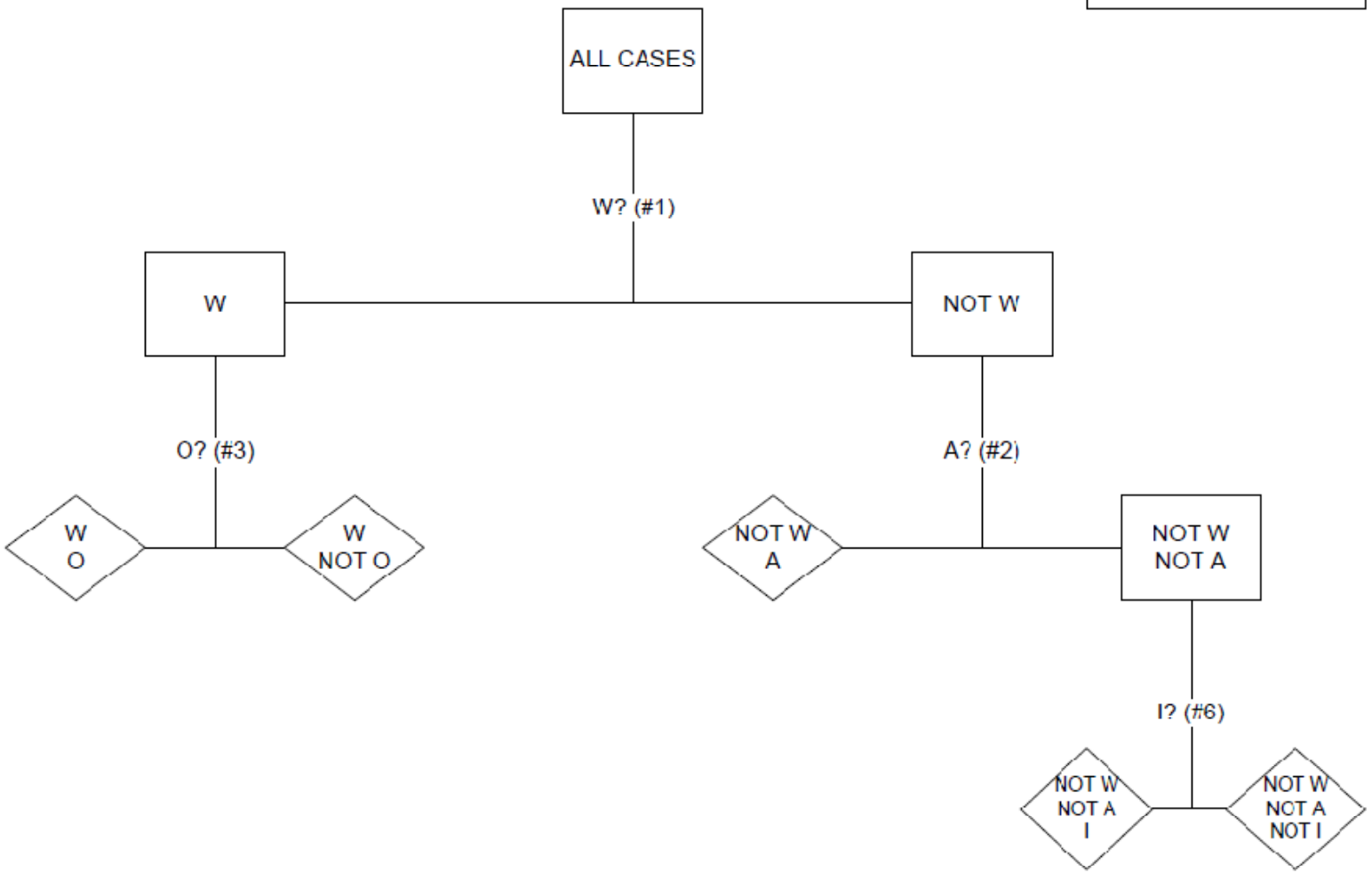
FLOWCHART 2: P-SAMPLE TREE (HISPANICS)

W=WHITE
B=BLACK
A–ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER



FLOWCHART 3: E-SAMPLE TREE (NON-HISPANICS)

W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER



FLOWCHART 4: E-SAMPLE TREE (HISPANICS)

W=WHITE
B=BLACK
A–ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER

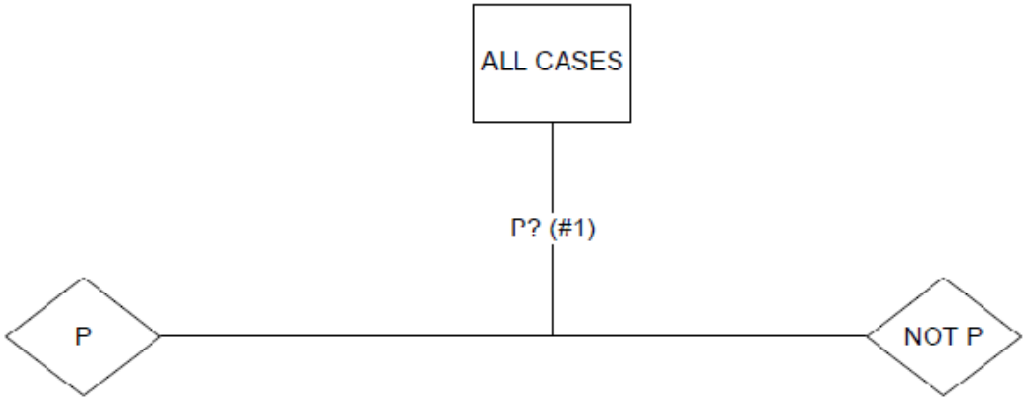


Table 2(a): E-Sample Correct Enumeration Cells (Non-Hispanics)

CELL	MARKED	NOT MARKED	N	WGTD_N	RATE
1	WO	--	2,115	1,901	0.898
2	W	O	461,360	507,147	0.942
3	A	W	31,200	27,129	0.927
4	I	WA	18,688	4,937	0.919
5	--	WAI	98,185	83,719	0.904
TOTAL NON-HISPANIC			*611,548*	*624,833*	*0.936*

Table 2(b): E-Sample Correct Enumeration Cells (Hispanics)

CELL	MARKED	NOT MARKED	N	WGTD_N	RATE
1	P	--	503	291	0.840
2	--	P	100,849	87,776	0.926
TOTAL HISPANIC			*101,352*	*88,067*	*0.926*
TOTAL E-SAMPLE			***712,900***	***712,900***	***0.935***

W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER

Table 2(c): P-Sample Match Rate Cells (Non-Hispanics)

CELL	MARKED	NOT MARKED	N	WGTD_N	RATE
1	WI	--	5,033	4,587	0.916
2	WB	I	1,931	1,908	0.908
3	W	IB	408,903	447,600	0.936
4	AP	W	503	232	0.797
5	A	WP	25,937	22,888	0.906
6	B	WA	82,244	70,967	0.873
7	--	WAB	25,882	12,794	0.885
TOTAL NON-HISPANIC			*550,433*	*560,976*	*0.926*

Table 2(d): P-Sample Match Rate Cells (Hispanics)

CELL	MARKED	NOT MARKED	N	WGTD_N	RATE
1	W	--	33,225	31,613	0.887
2	--	W	56,919	47,989	0.874
TOTAL HISPANIC			*90,144*	*79,602*	*0.879*

TOTAL P-SAMPLE	***640,577***	***640,577***	***0.920***

W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER

Table 3: Summary of Census Domain Conditions

DOMAIN 1: AMERICAN INDIAN OR ALASKA NATIVE ON RESERVATIONS
*Marked AMERICAN INDIAN/ALASKA NATIVE
*Lives on a reservation

DOMAIN 2: OFF-RESERVATION AMERICAN INDIAN OR ALASKA NATIVE
*Marked AMERICAN INDIAN/ALASKA NATIVE
*Does not live on a reservation
*Either lives in Indian Country, OR did not mark any of these: HISPANIC, BLACK, ASIAN, WHITE, or OTHER

DOMAIN 3: HISPANIC
*Marked HISPANIC
*Does not live in Indian Country, OR did not mark AMERICAN INDIAN/ALASKA NATIVE
*Does not live in Hawaii, OR did not mark NATIVE HAWAIIAN/PACIFIC ISLANDER

DOMAIN 4: NON-HISPANIC BLACK
*Marked BLACK
*Does not live in Indian Country, OR did not mark AMERICAN INDIAN/ALASKA NATIVE
*Did not mark HISPANIC
*Does not live in Hawaii, OR did not mark NATIVE HAWAIIAN/PACIFIC ISLANDER
*Marked no more than 1 of the following races: NATIVE HAWAIIAN/PACIFIC ISLANDER, ASIAN, WHITE, or OTHER

DOMAIN 5: NATIVE HAWAIIAN OR PACIFIC ISLANDER
*Marked NATIVE HAWAIIAN/PACIFIC ISLANDER
*Does not live in Indian Country, OR did not mark AMERICAN INDIAN/ALASKA NATIVE
*Either lives in Hawaii, OR did not mark any of these: HISPANIC, BLACK, WHITE, or OTHER

DOMAIN 6: NON-HISPANIC ASIAN
*Marked ASIAN
*Does not live in Indian Country, OR did not mark AMERICAN INDIAN/ALASKA NATIVE
*Did not mark any of these: HISPANIC, BLACK, NATIVE HAWAIIAN/PACIFIC ISLANDER, WHITE, or OTHER

DOMAIN 7: NON-HISPANIC WHITE OR "SOME OTHER RACE"
*Either marked 3 or more races, OR marked either WHITE or OTHER and did not mark BLACK.
*Does not live in Indian Country, OR did not mark AMERICAN INDIAN/ALASKA NATIVE
*Did not mark HISPANIC
*Does not live in Hawaii, OR did not mark NATIVE HAWAIIAN/PACIFIC ISLANDER

Table 4(a): Recursive Partitioning Model (E-Sample)

#	DOMAIN	TOTAL	WEIGHTED TOTAL	WEIGHTED CE	WEIGHTED RATE
1	American Indian Alaska Native on Reservation	14,438	1,126	1,025	0.910
2	Off Reservation American Indian Alaska Native	4,664	3,701	3,403	0.919
3	Hispanic	100,786	87,934	81,444	0.926
4	Non-Hispanic Black	97,767	84,001	76,025	0.905
5	Native Hawaiian Pacific Islander	2,890	1,305	1,188	0.910
6	Non-Hispanic Asian	29,731	26,067	24,179	0.928
7	Non-Hispanic White or Some Other Race	462,624	508,766	479,346	0.942
		712,900	712,900	666,610	0.935

#	DOMAIN	DOMAIN MODEL	TOTAL	WEIGHTED TOTAL	WEIGHTED CE	WEIGHTED RATE
1	AIAN on Res.	(1) ALL DOMAIN 1	14,438	1,126	1,025	0.910
2	Off Res AIAN.	(2) ALL DOMAIN 2	4,664	3,701	3,403	0.919
3	Hispanic	(3A) Pacific Islander	286	230	187	0.813
		(3B) Not Pacific Islander	100,500	87,704	81,257	0.926
4	N-H Black	(4A) White	1,677	1,642	1,536	0.935
		(4B) Not White	96,090	82,359	74,489	0.904
5	NHPI	(5A) Asian	1,151	457	439	0.961
		(5B) Not Asian	1,739	848	749	0.883
6	N-H Asian	(6) ALL DOMAIN 6	29,731	26,067	24,179	0.928
7	N-H White or SOR	(7A) Other	4,039	3,629	3,268	0.901
		(7B) Not Other	458,585	505,137	476,078	0.942
			712,900	712,900	666,610	0.935

Table 4(b): Recursive Partitioning Model (P-Sample)

#	DOMAIN	TOTAL	WEIGHTED TOTAL	WEIGHTED MATCH	WEIGHTED RATE
1	American Indian Alaska Native on Reservation	12,956	985	855	0.868
2	Off Reservation American Indian Alaska Native	3,833	3,059	2,708	0.885
3	Hispanic	89,586	79,477	69,841	0.879
4	Non-Hispanic Black	84,127	72,922	63,717	0.874
5	Native Hawaiian Pacific Islander	2,624	1,322	1,134	0.858
6	Non-Hispanic Asian	25,153	22,140	20,084	0.907
7	Non-Hispanic White or Some Other Race	422,298	460,673	430,875	0.935
		640,577	640,577	589,214	0.920

#	DOMAIN	DOMAIN MODEL	TOTAL	WEIGHTED TOTAL	WEIGHTED MATCH	WEIGHTED RATE
1	AIAN on Res.	(1) ALL DOMAIN 1	12,956	985	855	0.868
2	Off Res AIAN.	(2A) Asian	13	11	6	0.506
		(2B) Not Asian	3,820	3,047	2,702	0.887
3	Hispanic	(3A) White	33,050	31,554	27,977	0.887
		(3B) Not White	56,536	47,923	41,864	0.874
4	N-H Black	(4A) White	1,813	1,802	1,644	0.912
		(4B) Not White	82,314	71,120	62,073	0.873
5	NHPI	(5) ALL DOMAIN 5	2,624	1,322	1,134	0.858
6	N-H Asian	(6) ALL DOMAIN 6	25,153	22,140	20,084	0.907
7	N-H White or SOR	(7A) White & AIAN	4,051	4,150	3,805	0.917
		(7B) White & Black, Not AIAN	111	103	84	0.816
		(7C) White, Not AIAN or Black	408,375	447,432	418,984	0.936
		(7D) Not White	9,761	8,988	8,002	0.890
			640,577	640,577	589,214	0.920

Table 5: General Consistency of Race/Origin Domains

Domains for E-Sample and P-Sample Matched Cases (Weighted)

		P							
		1	2	3	4	5	6	7	
E	1	820	0	4	1	0	0	11	836
	2	0	1,711	127	93	4	30	940	2,905
	3	2	56	62,507	485	21	88	5,179	68,338
	4	2	77	550	59,810	28	90	2,659	63,216
	5	0	1	33	6	772	93	122	1,027
	6	0	35	205	111	163	18,117	2,057	20,688
	7	12	759	4,305	1,483	140	1,052	413,541	421,292
		836	2,639	67,731	61,989	1,128	19,470	424,509	578,300

(557,278 weighted cases -- 96.4% -- have consistent race domains for the two samples)

E-Sample Domains onto P-Sample Domains (Read ACROSS)

		P							
		1	2	3	4	5	6	7	
E	1	98.09%	0.00%	0.48%	0.12%	0.00%	0.00%	1.32%	1
	2	0.00%	58.90%	4.37%	3.20%	0.14%	1.03%	32.36%	1
	3	0.00%	0.08%	91.47%	0.71%	0.03%	0.13%	7.58%	1
	4	0.00%	0.12%	0.87%	94.61%	0.04%	0.14%	4.21%	1
	5	0.00%	0.10%	3.21%	0.58%	75.17%	9.06%	11.88%	1
	6	0.00%	0.17%	0.99%	0.54%	0.79%	87.57%	9.94%	1
	7	0.00%	0.18%	1.02%	0.35%	0.03%	0.25%	98.16%	1

P-Sample Domains onto E-Sample Domains (Read DOWN)

		P						
		1	2	3	4	5	6	7
E	1	98.09%	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%
	2	0.00%	64.84%	0.19%	0.15%	0.35%	0.15%	0.22%
	3	0.24%	2.12%	92.29%	0.78%	1.86%	0.45%	1.22%
	4	0.24%	2.92%	0.81%	96.48%	2.48%	0.46%	0.63%
	5	0.00%	0.04%	0.05%	0.01%	68.44%	0.48%	0.03%
	6	0.00%	1.33%	0.30%	0.18%	14.45%	93.05%	0.48%
	7	1.44%	28.76%	6.36%	2.39%	12.41%	5.40%	97.42%
		1	1	1	1	1	1	1

Table 6: All Domain and Cell Race/Origin Combinations for the E-Sample and P-Sample

E \ P	1	2A	2B	3AX	3AY	3BX	3BY	4AX	4BY	5X	5Y	6	7AX	7AY	7BX	7BY	7CX	7CY	7DX	7DY	
1	820	0	0	0	1	0	2	0	2	0	0	0	0	0	0	0	0	11	1	0	837
2A	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	5
2B	0	3	1,706	0	17	0	39	0	77	0	1	33	0	155	0	0	3	582	16	0	2,632
3AX	0	0	0	3	5	0	5	0	0	0	1	0	0	0	0	0	0	4	0	0	18
3AY	0	0	63	9	16,486	19	7,251	8	79	1	7	66	0	25	5	1	73	2,871	45	0	27,009
3BX	0	0	0	5	24	17	42	0	4	2	9	4	0	0	0	0	1	5	0	0	113
3BY	4	1	63	12	16,960	49	21,622	17	444	4	9	135	0	14	0	2	52	1,092	116	0	40,596
4AX	0	0	4	0	7	0	16	743	444	0	0	0	0	29	21	2	6	223	77	0	1,572
4BY	1	0	89	0	63	2	397	206	58,417	3	3	111	3	81	5	20	19	862	127	8	60,417
5X	0	0	0	0	0	2	2	0	0	177	59	56	0	0	0	0	0	10	1	0	307
5Y	0	0	4	0	8	3	6	4	23	88	447	108	0	4	0	0	2	99	25	0	821
6	0	0	30	0	40	2	46	1	89	71	22	18,117	0	20	0	0	36	712	283	0	19,469
7AX	0	0	5	0	15	0	2	5	10	0	0	0	0	20	0	0	2	80	1	1	141
7AY	0	0	385	0	20	0	28	30	103	0	0	8	6	764	0	0	13	2,253	26	0	3,636
7BX	0	0	1	0	2	0	7	10	21	0	0	0	0	2	2	0	1	7	0	0	53
7BY	0	0	0	0	0	0	0	1	16	0	0	0	0	0	0	6	0	6	0	0	29
7CX	0	0	3	0	248	0	66	9	15	1	3	42	0	4	0	0	91	1,653	20	0	2,155
7CY	9	1	486	6	2,710	6	605	181	1,333	14	55	756	6	866	5	5	729	402,638	249	0	410,660
7DX	1	1	58	4	735	2	723	87	829	23	26	1,251	2	69	6	4	338	3,337	320	0	7,816
7DY	0	0	0	0	0	0	0	0	8	0	0	0	0	1	0	0	0	0	0	4	13
	835	7	2,898	39	37,341	102	30,859	1,302	61,914	384	642	20,688	17	2,054	44	40	1,366	416,447	1,307	13	578,299

1: AI/AN ON RESERVATION
2A: OFF-RESERVATION AI/AN (Asian)
2B: OFF-RESERVATION AI/AN (NOT Asian)
3AX: HISPANIC (White, NH/PI)
3AY: HISPANIC (White, NOT NH/PI)
3BX: HISPANIC (NH/PI, NOT White)
3BY: HISPANIC (NOT White, NOT NH/PI)
4AX: NON-HISP BLACK (White)
4BY: NON-HISP BLACK (NOT White)
5X: NH/PI (Asian)
5Y: NH/PI (NOT Asian)
6: NON-HISP ASIAN
7AX: NON-HISP WHITE/SOR (White, AI/AN, Other)
7AY: NON-HISP WHITE/SOR (White, AI/AN, NOT Other)
7BX: NON-HISP WHITE/SOR (White, Black, Other, NOT AI/AN)
7BY: NON-HISP WHITE/SOR (White, Black, NOT AI/AN, NOT Other)
7CX: NON-HISP WHITE/SOR (White, Other, NOT AI/AN, NOT Black)
7CY: NON-HISP WHITE/SOR (White, NOT AI/AN, NOT Black, NOT Other)
7DX: NON-HISP WHITE/SOR (NOT White, Other)
7DY: NON-HISP WHITE/SOR (NOT White, NOT Other)

(cell matches are given in YELLOW, other domain matches are given in BLUE)

Table 7(a): Monoracial Census cases with consistent Race/Origin (min .02%)

ecode	pcode	cases	percent
-W-----	-W-----	399,804	69.13
--B----	--B----	57,112	9.88
H-----O	H-----O	19,731	3.41
---A---	---A---	18,123	3.13
HW-----	HW-----	13,590	2.35
-----I-	-----I-	2,232	0.39
----P--	----P--	353	0.06
------O	------O	238	0.04
H-B----	H-B----	195	0.03
TOT		511,429	88.44

H=HISPANIC
W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER

Table 7(b): Multiracial Census cases with consistent Race/Origin (min .02%)

ecode	pcode	cases	percent
-WB----	-WB----	744	0.13
-W---I-	-W---I-	721	0.12
-W-A---	-W-A---	711	0.12
HW----O	HW----O	218	0.04
--BA---	--BA---	95	0.02
--B--I-	--B--I-	93	0.02
-W----O	-W----O	90	0.02
TOT		3,152	0.55

H=HISPANIC
W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER

Table 7(c): Monoracial Census cases with inconsistent Race/Origin classifications (min .02%)

ecode	pcode	cases	percent
HW-----	H-----O	14,995	2.59
H-----O	HW-----	6,019	1.04
-W-----	------O	3,096	0.54
HW-----	-W-----	2,477	0.43
-W-----	-W---I-	2,369	0.41
-W-----	HW-----	2,360	0.41
-W-----	-W----O	1,591	0.28
HW-----	HW----O	1,582	0.27
--B----	-W-----	1,345	0.23
-W-----	H-----O	1,046	0.18
---A---	------O	899	0.16
-W-----	--B----	877	0.15
H-----O	HW----O	724	0.13
HW-----	------O	658	0.11
--B----	------O	654	0.11
H-----O	------O	622	0.11
H-----O	-W-----	551	0.10
H-B----	H-----O	495	0.09
-----I-	-W---I-	482	0.08
-W-----	---A---	478	0.08
---A---	-W-----	475	0.08
-W-----	HW----O	471	0.08
-----I-	-W-----	470	0.08
--B----	-WB----	437	0.08
-W-----	-----I-	399	0.07
-W-----	-W-A---	397	0.07
--B----	--B---O	375	0.06
--B----	--B--I-	366	0.06
H----I-	H-----O	354	0.06
---A---	---A--O	328	0.06
---A---	-W-A---	285	0.05
H-B----	--B----	228	0.04
-W-----	-WB----	227	0.04
HW-----	-W----O	204	0.04
------O	-W-----	188	0.03
--B----	H-B----	179	0.03
H-B----	HW-----	150	0.03
--B----	H-----O	143	0.02
H----I-	HW-----	108	0.02
---A---	----P--	102	0.02
H-B----	H-B---O	100	0.02
H-----O	H-B----	100	0.02
------O	--B----	97	0.02
------O	H-----O	93	0.02
---A---	--B----	90	0.02
TOT		53,732	9.29

Table 7(d): Multiracial Census cases with inconsistent Race/Origin classifications (min .02%)

ecode	pcode	cases	percent
HW----O	H-----O	1,638	0.28
-W---I-	-W-----	849	0.15
HW----O	HW-----	842	0.15
-W----O	-W-----	705	0.12
-W----O	------O	329	0.06
-W-A---	-W-----	306	0.05
-W-A---	---A---	259	0.04
--B---O	--B----	242	0.04
---A--O	---A---	217	0.04
HW----O	-W-----	213	0.04
-WB----	--B----	202	0.03
-W---I-	-----I-	195	0.03
-WB----	-W-----	182	0.03
-W-A---	------O	132	0.02
--B--I-	--B----	127	0.02
H-B---O	H-----O	115	0.02
--B---O	------O	106	0.02
-WB----	------O	87	0.02
TOT		9,987	1.73

H=HISPANIC
W=WHITE
B=BLACK
A=ASIAN
P=PACIFIC ISLANDER
I=AMERICAN INDIAN
O=OTHER

Table 8: Correct Enumeration and Match Rates for Domains and Key Groups

How Domains Are Arranged Now:

#	DOMAIN	TOT	W_TOT	W_CE	W_RATE	TOT	W_TOT	W_MATCH	W_RATE	CONS %
1	Am. Indian/AK Native ON reservation	14,438	1,126	1,025	0.910	12,956	985	855	0.868	98
2	OFF reservation Am. Indian/AK Native	4,664	3,701	3,403	0.919	3,833	3,059	2,708	0.885	59
3	Hispanic	100,786	87,934	81,443	0.926	89,586	79,477	69,842	0.879	91
4	Non-Hispanic Black	97,767	84,001	76,025	0.905	84,127	72,922	63,717	0.874	95
5	Native Hawaiian or Pacific Islander	2,890	1,305	1,188	0.910	2,624	1,322	1,134	0.858	75
6	Non-Hispanic Asian	29,731	26,067	24,179	0.928	25,153	22,140	20,084	0.907	88
7	Non-Hispanic White or "Some Other Race"	462,624	508,766	479,346	0.942	422,298	460,674	430,875	0.935	98
		712,900	**712,900**	**666,610**	**0.935**	**640,577**	**640,577**	**589,215**	**0.920**	

Breaking Out Groups Of Interest:

#	DOMAIN	TOT	W_TOT	W_CE	W_RATE	TOT	W_TOT	W_MATCH	W_RATE	CONS %
1	Am. Indian/AK Native ON reservation	14,438	1,126	1,025	0.910	12,956	985	855	0.868	98
2	OFF reservation Am. Indian/AK Native	4,664	3,701	3,403	0.919	3,833	3,059	2,708	0.885	59
3-A	Hispanic	100,500	87,704	81,257	0.926	89,412	79,328	69,711	0.879	91
4-B	Non-Hispanic Black	96,090	82,359	74,489	0.904	82,314	71,120	62,073	0.873	94
5-C	Native Hawaiian or Pacific Islander	1,739	848	749	0.883	1,665	949	827	0.871	69
6	Non-Hispanic Asian	29,731	26,067	24,179	0.928	25,153	22,140	20,084	0.907	88
7-D	Non-Hispanic White or "Some Other Race"	458,585	505,137	476,078	0.942	409,936	449,098	420,488	0.936	97
A	HISPANIC + NHPI (But NOT Hawaii)	286	230	186	0.809	174	149	131	0.879	17
B	BLACK + WHITE	1,677	1,642	1,536	0.935	1,813	1,802	1,644	0.912	57
C	ASIAN + NHPI	1,151	457	439	0.961	959	373	307	0.823	46
D	SOME OTHER RACE	4,039	3,629	3,268	0.901	12,362	11,576	10,387	0.897	28
		712,900	**712,900**	**666,610**	**0.935**	**640,577**	**640,577**	**589,214**	**0.920**	

Table 9: Alternative Race/Origin Domain Integration for Key Groups

All Pairing Combinations:

#	DOMAIN	TOT	W_TOT	W_CE	W_RATE	TOT	W_TOT	W_MATCH	W_RATE	CONS %
1	Am. Indian/AK Native ON reservation	14,438	1,126	1,025	0.910	12,956	985	855	0.868	98
2	OFF reservation Am. Indian/AK Native	4,664	3,701	3,403	0.919	3,833	3,059	2,708	0.885	59
3	Hispanic	100,786	87,934	81,443	0.926	89,586	79,477	69,842	0.879	91
3-A	Hispanic	100,500	87,704	81,257	0.926	89,412	79,328	69,711	0.879	91
4	Non-Hispanic Black	97,767	84,001	76,025	0.905	84,127	72,922	63,717	0.874	95
4-B	Non-Hispanic Black	96,090	82,359	74,489	0.904	82,314	71,120	62,073	0.873	94
5	Native Hawaiian or Pacific Islander	2,890	1,305	1,188	0.910	2,624	1,322	1,134	0.858	75
5-C	Native Hawaiian or Pacific Islander	1,739	848	749	0.883	1,665	949	827	0.871	69
5+A	Native Hawaiian or Pacific Islander	3,176	1,535	1,374	0.895	2,798	1,471	1,265	0.860	69
5+A-C	Native Hawaiian or Pacific Islander	2,025	1,078	935	0.867	1,839	1,098	958	0.872	62
6	Non-Hispanic Asian	29,731	26,067	24,179	0.928	25,153	22,140	20,084	0.907	88
6+C	Non-Hispanic Asian	30,882	26,524	24,618	0.928	26,112	22,513	20,391	0.906	87
7	Non-Hispanic White or "Some Other Race"	462,624	508,766	479,346	0.942	422,298	460,674	430,875	0.935	98
7-D	Non-Hispanic White or "Some Other Race"	458,585	505,137	476,078	0.942	409,936	449,098	420,488	0.936	97
7+B	Non-Hispanic White or "Some Other Race"	464,301	510,408	480,882	0.942	424,111	462,476	432,519	0.935	98
7+B-D	Non-Hispanic White or "Some Other Race"	460,262	506,779	477,614	0.942	411,749	450,900	422,132	0.936	97
D	SOME OTHER RACE	4,039	3,629	3,268	0.901	12,362	11,576	10,387	0.897	28

Table 10(a): Effect of Geography on Hispanic Subgroups (E-Sample)

CE Rates, Hispanic Subgroup by Division

	1	2	3	4	5	6	7	8	9	TOT
CUBAN	0.981	0.853	0.898	0.704	0.938	0.856	0.800	0.760	0.873	0.913
MEXICAN	0.922	0.897	0.914	0.914	0.920	0.889	0.918	0.918	0.940	0.927
PRT RICAN	0.921	0.907	0.943	0.763	0.929	0.883	0.943	0.938	0.908	0.917
OTHER HISP	0.922	0.920	0.934	0.923	0.925	0.900	0.933	0.923	0.947	0.930
HISPANIC	0.922	0.911	0.921	0.908	0.928	0.891	0.921	0.918	0.940	0.926
NOT HISP	0.940	0.932	0.944	0.951	0.930	0.937	0.933	0.933	0.934	0.936
TOTAL	0.939	0.930	0.943	0.950	0.930	0.936	0.930	0.930	0.936	0.935

CE Rates, Hispanic Subgroup by Division, Standardized by Hispanic Status

	1	2	3	4	5	6	7	8	9	TOT
CUBAN	0.059	(0.058)	(0.023)	(0.204)	0.010	(0.035)	(0.121)	(0.158)	(0.067)	(0.013)
MEXICAN	0.000	(0.014)	(0.007)	0.006	(0.008)	(0.002)	(0.003)	0.000	0.000	0.001
PRT RICAN	(0.001)	(0.004)	0.022	(0.145)	0.001	(0.008)	0.022	0.020	(0.032)	(0.009)
OTHER HISP	0.000	0.009	0.013	0.015	(0.003)	0.009	0.012	0.005	0.007	0.004
HISPANIC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOT HISP	0.018	0.021	0.023	0.043	0.002	0.046	0.012	0.015	(0.006)	0.010
TOTAL	0.017	0.019	0.022	0.042	0.002	0.045	0.009	0.012	(0.004)	0.009

E-Sample Totals, in Thousands, Hispanic Subgroup by Division

	1	2	3	4	5	6	7	8	9	TOT
CUBAN	0.0	0.4	0.1	0.0	2.6	0.0	0.1	0.1	0.2	3.5
MEXICAN	0.1	1.0	4.3	1.1	2.3	0.3	12.9	6.3	22.4	50.7
PRT RICAN	0.9	4.1	0.7	0.0	1.8	0.0	0.2	0.2	0.5	8.4
OTHER HISP	0.9	5.6	1.0	0.3	4.7	0.2	4.0	2.6	6.1	25.4
HISPANIC	2.0	11.0	6.1	1.5	11.4	0.6	17.2	9.1	29.1	88.1
NOT HISP	32.9	90.0	108.0	46.6	121.3	42.8	60.7	37.9	84.6	624.8
TOTAL	34.8	101.1	114.1	48.1	132.7	43.4	78.0	47.0	113.7	712.9

KEY

1	NEW ENGLAND	(CT MA ME NH RI VT)
2	MID-ATLANTIC	(NJ NY PA)
3	EAST NORTH CENTRAL	(IL IN MI OH WI)
4	WEST NORTH CENTRAL	(IA KS MN MO ND NE SD)
5	SOUTH ATLANTIC	(DC DE FL GA MD NC SC VA WV)
6	EAST SOUTH CENTRAL	(AL KY MS TN)
7	WEST SOUTH CENTRAL	(AR LA OK TX)
8	MOUNTAIN	(AZ CO ID MT NM NV UT WY)
9	PACIFIC	(AK CA HI OR WA)

Table 10(b): Effect of Geography on Hispanic Subgroups (P-Sample)

Match Rates, Hispanic Subgroup by Division

	1	2	3	4	5	6	7	8	9	TOT
CUBAN	0.913	0.862	0.915	0.916	0.945	0.867	0.844	0.918	0.912	0.931
MEXICAN	0.839	0.791	0.857	0.901	0.717	0.864	0.883	0.866	0.895	0.875
PRT RICAN	0.892	0.872	0.867	0.943	0.895	0.975	0.922	0.858	0.894	0.882
OTHER HISP	0.873	0.872	0.873	0.895	0.870	0.821	0.859	0.888	0.877	0.872
HISPANIC	0.878	0.865	0.865	0.898	0.865	0.835	0.878	0.874	0.891	0.877
NOT HISP	0.932	0.922	0.941	0.955	0.916	0.919	0.921	0.924	0.923	0.927
TOTAL	0.928	0.914	0.936	0.952	0.910	0.916	0.911	0.913	0.914	0.920

Match Rates, Hispanic Subgroup by Division, Standardized by Hispanic Status

	1	2	3	4	5	6	7	8	9	TOT
CUBAN	0.035	(0.003)	0.050	0.018	0.080	0.032	(0.034)	0.044	0.021	0.054
MEXICAN	(0.039)	(0.074)	(0.008)	0.003	(0.148)	0.029	0.005	(0.008)	0.004	(0.002)
PRT RICAN	0.014	0.007	0.002	0.045	0.030	0.140	0.044	(0.016)	0.003	0.005
OTHER HISP	(0.005)	0.007	0.008	(0.003)	0.005	(0.014)	(0.019)	0.014	(0.014)	(0.005)
HISPANIC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOT HISP	0.054	0.057	0.076	0.057	0.051	0.084	0.043	0.050	0.032	0.050
TOTAL	0.050	0.049	0.071	0.054	0.045	0.081	0.033	0.039	0.023	0.043

P-Sample Totals, in Thousands, Hispanic Subgroup by Division

	1	2	3	4	5	6	7	8	9	TOT
CUBAN	0.0	0.3	0.1	0.0	2.6	0.0	0.1	0.0	0.2	3.4
MEXICAN	0.1	0.9	3.8	0.9	2.0	0.2	11.9	5.7	20.7	46.3
PRT RICAN	0.9	3.6	0.6	0.0	1.8	0.0	0.2	0.1	0.4	7.7
OTHER HISP	1.6	7.2	2.7	1.3	7.4	1.0	3.9	3.2	7.6	35.9
HISPANIC	2.6	12.1	7.2	2.3	13.7	1.3	16.1	9.0	28.9	93.3
NOT HISP	28.7	78.4	95.8	41.3	105.4	36.4	53.1	32.9	75.4	547.3
TOTAL	31.3	90.5	103.0	43.6	119.1	37.6	69.2	41.9	104.3	640.6

KEY

1	NEW ENGLAND	(CT MA ME NH RI VT)
2	MID-ATLANTIC	(NJ NY PA)
3	EAST NORTH CENTRAL	(IL IN MI OH WI)
4	WEST NORTH CENTRAL	(IA KS MN MO ND NE SD)
5	SOUTH ATLANTIC	(DC DE FL GA MD NC SC VA WV)
6	EAST SOUTH CENTRAL	(AL KY MS TN)
7	WEST SOUTH CENTRAL	(AR LA OK TX)
8	MOUNTAIN	(AZ CO ID MT NM NV UT WY)
9	PACIFIC	(AK CA HI OR WA)

Table 11: Consistency of Hispanic Subgroups

TABLE OF HISPANIC CODES FOR E-SAMPLE AND P-SAMPLE MATCHED CASES (WEIGHTED)

		P					
		CUBAN	MEXICAN	PRT RICAN	OTHER HISP	NOT HISP	
E	CUBAN	2,177	27	30	134	210	2,579
	MEXICAN	15	31,586	114	2,889	2,521	37,124
	PUERTO RICAN	31	119	4,835	466	615	6,066
	OTHER HISP	198	4,391	592	11,594	1,667	18,442
	NOT HISP	293	1,556	380	11,656	471,573	485,457
		2,715	37,678	5,950	26,739	476,585	549,668

E-SAMPLE HISPANIC CODES ONTO P-SAMPLE HISPANIC CODES (READ ACROSS)

		P					
		CUBAN	MEXICAN	PRT RICAN	OTHER HISP	NOT HISP	
E	CUBAN	84.41%	1.05%	1.16%	5.19%	8.14%	1
	MEXICAN	0.04%	85.08%	0.31%	7.78%	6.79%	1
	PUERTO RICAN	0.51%	1.95%	79.70%	7.69%	10.13%	1
	OTHER HISP	1.08%	23.81%	3.21%	62.87%	9.04%	1
	NOT HISP	0.06%	0.32%	0.08%	2.40%	97.14%	1

P-SAMPLE HISPANIC CODES ONTO E-SAMPLE HISPANIC CODES (READ DOWN)

		P				
		CUBAN	MEXICAN	PRT RICAN	OTHER HISP	NOT HISP
E	CUBAN	80.18%	0.07%	0.50%	0.50%	0.04%
	MEXICAN	0.57%	83.83%	1.92%	10.80%	0.53%
	PUERTO RICAN	1.14%	0.31%	81.26%	1.74%	0.13%
	OTHER HISP	7.31%	11.65%	9.94%	43.36%	0.35%
	NOT HISP	10.79%	4.13%	6.38%	43.59%	98.95%
		1	1	1	1	1

Table 12(a): Common Race/Origin Codes for CUBAN subgroup

ecode	pcode	ecases	pcases	mcases
HW-----	HW-----	3,002	2,243	1,678
HW-----	H-----O	3,002	332	161
H-----O	HW-----	241	2,243	67
HW-----	HW----O	3,002	103	62
HW----O	HW-----	110	2,243	41
H-----O	H-----O	241	332	35
H-B----	H-----O	102	332	24
		3,535	2,780	2,177

Table 12(b): Common Race/Origin Codes for MEXICAN subgroup

ecode	pcode	ecases	pcases	mcases
H-----O	H-----O	22,264	22,564	10,487
HW-----	H-----O	24,725	22,564	8,303
HW-----	HW-----	24,725	11,371	5,984
H-----O	HW-----	22,264	11,371	3,296
HW----O	H-----O	1,910	22,564	763
HW-----	HW----O	24,725	1,547	719
H-----O	HW----O	22,264	1,547	385
HW----O	HW-----	1,910	11,371	328
H----I-	H-----O	606	22,564	205
HW----O	HW----O	1,910	1,547	104
H----I-	HW-----	606	11,371	59
H-B----	H-----O	349	22,564	58
HW---I-	H-----O	169	22,564	46
H----I-	H----IO	606	105	43
HW-----	H----I-	24,725	113	43
H-B----	HW-----	349	11,371	38
HW-----	HW---I-	24,725	120	29
H--A---	H-----O	143	22,564	28
H----IO	H-----O	76	22,564	27
H-B---O	H-B---O	103	75	25
HW---I-	HW-----	169	11,371	24
H-B----	H-B---O	349	75	22
H-----O	HW---I-	22,264	120	22
H-----O	H----I-	22,264	113	22
HW-----	H----IO	24,725	105	21
		50,733	36,267	31,586

Table 12(c): Common Race/Origin Codes for PUERTO RICAN subgroup

ecode	pcode	ecases	pcases	mcases
H-----O	H-----O	3,174	3,265	1,405
HW-----	H-----O	3,969	3,265	1,047
HW-----	HW-----	3,969	2,077	1,043
H-----O	HW-----	3,174	2,077	486
HW-----	HW----O	3,969	255	127
H-B----	H-----O	577	3,265	98
HW----O	H-----O	309	3,265	80
H-B----	H-B----	577	193	77
HW----O	HW-----	309	2,077	65
H-B----	HW-----	577	2,077	40
H-----O	HW----O	3,174	255	35
HWB----	H-----O	82	3,265	25
H-----O	H-B----	3,174	193	24
		8,425	6,070	4,835

Table 12(d): Common Race/Origin Codes for OTHER HISPANIC subgroup

ecode	pcode	ecases	pcases	mcases
H-----O	H-----O	5,609	8,916	1,279
HW-----	H-----O	6,329	8,916	924
HW-----	HW-----	6,329	5,252	888
H-----O	HW-----	5,609	5,252	347
HW----O	H-----O	839	8,916	196
HW-----	HW----O	6,329	905	181
H-----O	HW----O	5,609	905	95
HW----O	HW-----	839	5,252	82
H-B----	H-----O	339	8,916	51
H----I-	H-----O	250	8,916	33
HW----O	HW----O	839	905	29
H-B----	HW-----	339	5,252	22
		13,904	15,975	4,411

Table 13: Correct Enumeration and Match Rates for Hispanic Subgroups

How Domains Are Arranged Now:

#	DOMAIN	TOTAL	W_TOTAL	W_CE	W_RATE	TOT	W_TOT	W_MATCH	W_RATE
1	Am. Indian/AK Native ON reservation	14,438	1,126	1,025	0.910	12,956	985	855	0.868
2	OFF reservation Am. Indian/AK Native	4,664	3,701	3,403	0.919	3,833	3,059	2,708	0.885
3	**Hispanic**	**100,786**	**87,934**	**81,444**	**0.926**	**89,586**	**79,476**	**69,841**	**0.879**
4	Non-Hispanic Black	97,767	84,001	76,026	0.905	84,127	72,922	63,717	0.874
5	Native Hawaiian or Pacific Islander	2,890	1,305	1,188	0.910	2,624	1,322	1,134	0.858
6	Non-Hispanic Asian	29,731	26,067	24,179	0.928	25,153	22,140	20,084	0.907
7	Non-Hispanic White or "Some Other Race"	462,624	508,766	479,345	0.942	422,298	460,673	430,875	0.935
		712,900	**712,900**	**666,610**	**0.935**	**640,577**	**640,577**	**589,214**	**0.920**

Breaking Out Hispanic Subpopulations:

#	DOMAIN	TOTAL	W_TOTAL	W_CE	W_RATE	TOT	W_TOT	W_MATCH	W_RATE
3A	CUBAN	2,963	3,535	3,229	0.913	2,805	3,365	3,134	0.931
3B	MEXICAN	59,174	50,683	46,964	0.927	54,032	46,269	40,505	0.875
3C	PUERTO RICAN	9,476	8,399	7,701	0.917	8,143	7,648	6,749	0.882
3D	OTHER HISPANIC	29,173	25,317	23,550	0.930	22,609	20,262	17,894	0.883
3E	DETAILED HISPANIC ETHNICITY UNKNOWN					1,997	1,932	1,559	0.807
		100,786	**87,934**	**81,444**	**0.926**	**89,586**	**79,476**	**69,841**	**0.879**

Table 14: Standard Errors for Hispanic Subgroups

CE ESTIMATES/INTERVAL

	Est (ln)	SE (ln)	EST	INTERVAL	
CUBAN	2.355	0.102	0.913	0.896	0.928
MEXICAN	2.536	0.041	0.927	0.921	0.932
PUERTO RICAN	2.402	0.067	0.917	0.906	0.926
OTHER HISPANIC	2.590	0.042	0.930	0.925	0.935
NOT HISPANIC	2.688	0.018	0.936	0.934	0.938

CE ESTIMATES/INTERVAL (NO RACE/ORIGIN IMPUTES)

	Est (ln)	SE (ln)	EST	INTERVAL	
CUBAN	2.458	0.102	0.921	0.905	0.935
MEXICAN	2.542	0.044	0.927	0.921	0.933
PUERTO RICAN	2.463	0.075	0.922	0.910	0.932
OTHER HISPANIC	2.650	0.045	0.934	0.928	0.939
NOT HISPANIC	2.773	0.036	0.941	0.937	0.945

MATCH ESTIMATES/INTERVAL

	Est (ln)	SE (ln)	EST	INTERVAL	
CUBAN	2.604	0.135	0.931	0.912	0.946
MEXICAN	1.950	0.050	0.875	0.864	0.886
PUERTO RICAN	2.013	0.070	0.882	0.867	0.896
OTHER HISPANIC	1.920	0.044	0.872	0.862	0.881
NOT HISPANIC	2.544	0.043	0.927	0.921	0.933

MATCH ESTIMATES/INTERVAL (NO RACE/ORIGIN IMPUTES)

	Est (ln)	SE (ln)	EST	INTERVAL	
CUBAN	2.602	0.135	0.931	0.912	0.946
MEXICAN	1.951	0.052	0.876	0.864	0.886
PUERTO RICAN	2.022	0.072	0.883	0.868	0.897
OTHER HISPANIC	2.041	0.050	0.885	0.875	0.895
NOT HISPANIC	2.551	0.044	0.928	0.922	0.933