EFFECT OF RACE AND ETHNICITY CLASSIFICATION ON SURVEY ESTIMATES: ANOMALY OF THE WEIGHTED TOTALS OF AMERICAN INDIANS AND ALASKA NATIVES

Sunghee Lee, PhD, Delight E. Satter, MPH, and Ninez A. Ponce, PhD

Abstract: Racial classification is a paramount concern in data collection and analysis for American Indians and Alaska Natives (AI/ANs) and has far-reaching implications in health research. We examine how different racial classifications affect survey weights and consequently change health-related indicators for the AI/AN population in California. Using a very large random population-based sample of AI/ANs, we compared the impact of three weighting strategies on counts and rates of selected health indicators. We found that different weights examined in this study did not change the percentage estimates of health-related variables for AI/ANs, but did influence the population total estimates dramatically. In survey data, different racial classifications and tabulations of AI/ANs could yield discrepancies in weighted estimates for the AI/AN population. Policy makers need to be aware that the choice of racial classification schemes for this racial-political group can generally influence the data they use for decision making.

INTRODUCTION

Racial classification is a paramount concern in data collection and analysis for American Indians and Alaska Natives (AI/ANs) and has far-reaching implications in health research. This article evaluates how different race/ethnicity-based survey weights affect health-related estimates for the AI/AN population in California. We first consider how variations in classifying and tabulating racial and ethnic groups affect the development of survey weights for the AI/AN sample in a population-based survey, propose a new weighting method, and then evaluate how these weights impact the rates and counts of important indicators of health status, health behaviors, utilization, and access to healthcare for the AI/AN population.

Scientific surveys employ probability samples to ensure that a given survey represents the population of interest with known nonzero probabilities (Kish, 1966; Cochran, 1977; Särndal, Swesson, & Wretman, 1992). Each unit in the sample carries its selection probability derived from the sampling procedure. The inverse of the selection probability serves as the "base weight." Simply speaking, the weight is the number of population units that a particular sample represents in a particular survey, and the sum of weights is equal to the population size.

Due to operational glitches arising in survey practice (such as nonresponse), the theoretical statistical representativeness of the sample becomes dampened. In order to compensate, the base weights are rescaled through an adjustment process. There are a number of approaches one might consider for this adjustment (Kalton & Flores-Cervantes, 2003). Typically, adjustment factors are calculated by controlling for some sample characteristics to match the known population distribution, and are applied to the base weights to create "final weights." By applying the final weights in the estimation, survey estimates are expected to approximate the population quantities (Tompkins & Kim, 2007).

Race and Ethnicity in Survey Weighting

Weighting adjustment is a key process in preparing survey data, and race and ethnicity are important common variables controlled for in the weighting adjustment, as survey results are often presented by their subgroups. In most survey practice, small racial and ethnic categories are lumped together in weighting. For example, the National Health Interview Survey (NHIS) uses the categories Hispanic, non-Hispanic Black, and non-Hispanic other race (Botman, Moore, Moriarity, & Parsons, 2000). Because the non-Hispanic other race category includes multiple race/ethnicity groups, it is possible that the individual proportion of these groups may be distorted in the weighting process.

Race/Ethnicity Data Collection and Classification in the United States

Race/ethnicity data play an important role in research. Health research, in particular, analyzes survey data using race/ethnicity as an independent variable, because this information serves as a proxy for unmeasured social factors (Mays, Ponce, Washington, & Cochran, 2003) and an indicator of health disparities and health care access (Winker, 2004), as well as an independent effect from other demographic and socioeconomic characteristics (Swift, 2002).

Race/ethnicity is a construct far more complex than one would expect in demographically diverse societies like the U.S. The change of race/ethnicity measurement in the federal government is well documented (Burhansstipanov & Satter, 2000; Mays et al., 2003). Problems associated with

measuring the race of Hispanic/Latino populations in the 1990 U.S. Census, and other issues, led the U.S. Office of Management and Budget (OMB) to release new Statistical Policy Directive No. 15 (U.S. OMB 1995, 1997). The main changes are: (1) the ethnicity question precedes the race question; (2) there are now five race categories: White, Black, American Indian or Alaska Native (AI/AN), Native Hawaiian or Other Pacific Islander (NHPI), and Asian (Asian and Pacific Islander became separate categories, unlike the previous definition); (3) the "some other race" category no longer exists; and (4) people can identify with more than one race group described above. These changes were reflected in the 2000 U.S. Census data collection, with the OMB-approved exception that the Census questionnaire also provided the "some other race" category. Roughly speaking, the 2000 U.S. Census raw data provide the total number of people in the population identifying with Hispanic/Latino ethnicity and each of the following six race variables: NHPI, AI/AN, Asian, Black or African American, White, and some other race. The sum of the race counts exceeds the total population counts, because 2.4% of the population chose more than one race in Census SF-1 Table P7. Given the complexity of the collected data and the need to summarize the information into a single variable with mutually exclusive categories, the U.S. Census Bureau (2002) released a modified race data summary file assigning individuals in the "some other race" category into one of the OMB race categories.

It should be noted that it is common for persons in the U.S. to use race/ethnicity terms without understanding their meanings, origins, or current implications (Cruz-Jansen, 2002). It has long been stated by the federal government that the racial and ethnic classification standards for the U.S. are not based in science, but are responses to expressed needs of politicians (Forbes, 1990; U.S. OMB 1997). Nonetheless, racial categories and counts based on the OMB guidelines are the major organizing tool for public health data (Tashiro, 2002).

The AI/AN Population

The AI/AN population, the only federally recognized political minority in the U.S., is increasing at about 1.8% a year, not including tribes gaining federal recognition. According to the 2000 U.S. Census, 4.1 million (1.5%) adults in the U.S. are AI/AN or AI/AN in combination with one or more other races (Ogunwole, 2002). The number of residents who reported as AI/AN in combination with one or more races increased 110% between the 1990 and 2000 U.S. Censuses, while the number of residents who reported as AI/AN alone increased 26%. More AI/ANs (627,562 people in 2000) live in California than any other state in the U.S. (Ogunwole). There are more federally recognized tribes in California (107) than any state except Alaska, as well as numerous non-federally recognized tribes.

NEW CONTRIBUTION

The fundamental data issue we explore is the potential discrepancy of weighted estimates of health-related variables for the AI/AN population that results from different racial classification and tabulation of AI/ANs in preparing and using survey data. If the weights alter the estimates substantially, the weighting schemes should be subject to reconsideration. Using a very large population-based sample of AI/ANs, this is, to our knowledge, the first article to examine the impact of racial classification on survey weights and the consequences of different weights on AI/AN health data. Our new contribution is to inform health researchers who use population-based survey data when studying different racial and ethnic groups, particularly the AI/AN population, of the importance of understanding how the survey weights are created and for what they can and cannot account.

DATA AND METHODS

This study used data from the 2001 California Health Interview Survey (CHIS), a state-wide random-digit dial telephone survey assessing the health of California's general population. The sample was drawn by geographically stratifying the state, mostly at the county level so that reliable county-level estimates could be obtained. As the largest state health survey in the U.S., CHIS also has the ability to provide accurate data on the AI/AN population due to modifications made at different development stages (including survey design, interviewer training protocol, and sampling). The survey instrument has been through extensive cultural and linguistic review and adaptation (Ponce et al., 2004). The format of race/ethnicity questions follows that of the 2000 U.S. Census, as respondents may report more than one race. These questions were specifically designed to avoid problems faced by other surveys collecting AI/AN data. Persons who answered as AI/AN on any race question were asked to further identify their tribal affiliation(s), whether they are an enrolled member of a tribe, and in which tribe(s) they were enrolled. The data allow us to distinguish individuals from California tribes and non-California tribes, as well as those without tribal affiliations. CHIS 2001 oversampled AI/ANs, yielding the largest sample of AI/AN for a population-based cross-sectional data collection in U.S. history, a total of almost 4,000 AI/AN individuals (Yen & Satter, 2002). The interviewers received an in-person cultural competency training about the unique cultural and linguistic issues that can arise when interviewing AI/ANs. The cultural competency training materials were incorporated into the standard interviewer training protocol for subsequent administrations of CHIS (Satter, Veiga-Ermert, Burhansstipanov, Pena, & Restivo, 2005).

In weighting the CHIS data, the base weights were adjusted through 10 stages (CHIS, 2005). The last stage was a statistical adjustment process called ratio-raking which controls the marginal distributions of the raking variables (Deming & Stephen, 1940). Eleven dimensions of demographic and socio-economic variables and their cross-classifications shown in the Appendix were controlled in this stage.

There was a significant change in raking process between CHIS 2001 and 2003: Instead of the Census SF-1 data, the California Department of Finance (CA DOF) Population Projections (P-1) for 2003 were used as the external source for control totals. This was done because there were no other sources for intercensal population data for all counties. (For example, the American Community Survey collects data only from selected counties for a given year.) The change in the control source resulted in a change in classification of the raking race/ethnicity variable. The 2001 method controlled the following seven "any mention" (AM, hereafter) race/ethnicity variables separately, as in the 2000 U.S. Census questionnaire: whether or not one is Hispanic/Latino, Black, White, AI/AN, NHPI, Asian, or some other race. At the time, neither the CA DOF projections nor U.S. Census data based on the 1997 OMB Notice "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" were available.

The revised weighting used one race/ethnicity variable following the CA DOF projections pursuant to the 1997 OMB classification. The variable assigned people to one of the following mutually exclusive categories: Latino, non-Latino (NL) White, NL Black, NL Asian, NL AI/AN, NL NHPI, and NL multirace (CA DOF, 2004). In order to keep the data consistent across years, CHIS 2001 data were reweighted using the revised method with the CA DOF 2001 projections.

In addition to these existing weights, this study proposes another set of weights produced by controlling for both the multiple AM race variables and the single CA DOF race variable. Instead of creating weights from the base weights, we started with the existing revised weights and re-adjusted them using both the Census SF-1 data and CA DOF P-1 population projections for different racial and ethnic groups. The proposed weight will be discussed after comparing the original and revised weights described above.

RESULTS

Discrepancy in Weighted Totals for AI/ANs

Overall, the weight revision does not appear to have a major impact on proportions across all racial and ethnic groups in Table 1. The original weighted totals for AM race variables match their control totals from the Census SF-1, and the revised weighted totals for the CA DOF race variable match the control totals from the DOF P-1. The slight decrease in weighted totals compared to the

actual population control is due to the fact that the CHIS target population excludes people in group quarters while control totals include them. When comparing the original and revised weighted totals by race/ethnicity, there seems to be a problem with the AI/AN estimates. Both AM AI/ANs and NL AI/ANs show large discrepancies: The revised weighted total for AM AI/ANs is nearly double the original one (611,468 vs. 1,287,600), and that for NL AI/ANs is nearly quadruple (52,433 vs. 209,385).

Table 1
Race and Ethnicity Distribution in the 2001 California Health Interview Survey

	Unweighted Total		Original Weighted Total		Revised Weighted Total		Control Total[1]	
	Count	%	Count	%	Count	%	Count	%
Census Race/Ethnicity							Census SF-1	
AM Latino[2]	18,872	25.6	10,774,044	32.6	11,380,586	33.6	10,966,556	32.2
AM White	53,359	72.3	20,984,429	63.5	20,489,598	60.4	21,490,973	63.1
AM Black	4,320	5.9	2,370,085	7.2	2,492,683	7.4	2,513,041	7.4
AM Asian	6,113	8.3	4,092,122	12.4	4,232,701	12.5	4,155,685	12.2
AM AI/AN	3,990	5.4	611,468	1.9	1,287,600	3.8	627,562	1.8
AM NHOPI	592	0.8	215,878	0.7	278,552	0.8	221,458	0.7
AM Other	9,806	13.3	6,447,609	19.5	6,256,902	18.5	6,575,625	19.3
DOF Race/Ethnicity							CA DOF P-1	
Latino	18,872	25.6	10,774,044	32.6	11,380,586	33.6	11,082,985	32.6
NL White	42,652	57.8	15,931,533	48.2	15,603,435	46.0	16,047,989	47.1
NL Black	3,272	4.4	1,770,685	5.4	2,141,059	6.3	2,222,816	6.5
NL Asian	4,917	6.7	3,232,261	9.8	3,800,472	11.2	3,746,292	11.0
NL AI/AN[3]	540	0.7	53,325	0.2	210,296	0.6	192,753	0.6
NL NHOPI	237	0.3	62,261	0.2	113,932	0.3	111,200	0.3
NL Multirace	3,327	4.5	1,223,094	3.7	661,355	2.0	639,163	1.9
Total	73,917	100.0	33,050,816	100.0	33,911,135	100.0	34,043,198	100.0

[1] Control totals include the group quarter population for presentation purposes. These numbers are modified in the weighting procedure to exclude those populations, because the CHIS target population excludes them.

[2] AM indicates "any mention." Therefore, the sum of all AM race categories exceeds 100% of the sample.

[3] This includes 3 people who were reclassified because they reported single other race only.

The reasons for the discrepancies are twofold: the unweighted sample distribution and the difference in the race/ethnicity classification methods between the Census SF-1 and the CA DOF P-1. Let us first examine how the original weights were constructed. The sample included 3,990 AM AI/ANs across adults, adolescents, and children (5.4% of the total sample). This proportion was much larger than the proportion of AM AI/ANs (1.9%) in the control totals, Census SF-1. Therefore, these AM AI/ANs were assigned smaller weights than their counterparts in raking. Accordingly, when the original weights were applied, the weighted total of AM AI/ANs matched the population total from the Census SF-1.

In creating the CA DOF race/ethnicity variable with AM variables, 1,288 of the AM AI/ANs were classified as Latinos, and 2,165 as NL multirace individuals; only 537 people were retained as NL AI/ANs. In fact, the left column of Table 2 shows that, among these 3,990 AM AI/ANs, two thirds were also reported as AM White, and slightly less than one third were, in fact, AM Latinos. Because these AM Latinos are a subset of AM AI/ANs, the smaller weights on AM AI/ANs than their counterparts in the original weighting had been carried over for the NL AI/AN group of the CA DOF race classification. For this reason, when the weighted total was calculated for NL AI/ANs using the original weight, it appeared as if there were only 52,433 NL AI/ANs, when the true population total is 192,753.

Table 2
Self-Reported Race Variables (n=3,990)

Census Race/Ethnicity		DOF Race/Ethnicity	
AM AI/AN	3,990	NL AI/AN	537
AM Latino	1,288	Latino	1,288
AM White	2,376	NL White	0
AM Black	372	NL Black	0
AM Asian	75	NL Asian	0
AM NHOPI	32	NL NHOPI	0
AM Other	7	NL Multirace	2,165
		Total	3,990

In addition, the revised raking used the CA DOF race variable instead of AM race variables for control totals. When NL AI/ANs were examined, their proportion in the sample was their population proportion (0.7% vs. 0.6%). However, Table 2 shows that a large proportion of AM AI/ANs were Latinos, whose sample proportion was smaller than the population proportions (see Table 1), indicating that weights for Latinos were larger than for non-Latinos. Therefore, when all AM AI/ANs were combined, their weighted total became larger than it should have been. While these revised weights made the weighted total for NL AI/ANs comparable to the CA DOF control total (210,296 vs. 192,753), they produced far larger estimates for AM AI/ANs than the original

weighted total and the population counts (1,287,600 vs. 611,468/627,562). This discrepancy occurs because AM AI/ANs included not only 537 NL AI/AN respondents but also 1,288 Latinos and 2,165 NL multirace individuals in the CA DOF classification (see Table 2). The discrepancy in the unweighted distribution resides in the dynamics of racial and ethnicity classification and the complexity of combining the multiple AM race/ethnicity variables into one single variable as described above.

Proposed Adjustment

A possible way to reconcile this discrepancy is to control for both the multiple U.S. Census AM race variables and the single CA DOF race variable, unlike the previous two methods that controlled for only one of these two. The methodology of the proposed weighting itself is essentially the same as that of the previous weighting described in CHIS (2005); the only difference is that both race definitions are included in the adjustment. Population control totals presented in the far right column of Table 3 were used in calculating weights so that all weighted race/ethnicity estimates would match their respective population totals well. The results of the proposed weight are shown in Table 3.

The original weighted total of AM AI/AN matches the Census SF-1 figure for AM AI/AN very well, but the NL AI/AN total fails to match the CA DOF P-1 total. In a similar fashion, the revised weighted total of NL AI/AN matches the CA DOF P-1 figure, but the AM AI/AN total is very far from the Census SF-1 total. These findings are to be expected, as only one racial/ethnicity classification is controlled in these two weighting schemes. However, the proposed weighted totals of both AM AI/AN and NL AI/AN are closer to the Census SF-1 and the CA DOF P-1 control totals simultaneously, unlike the original and revised weight methods, where the weighted totals matched the control totals of one of the two racial/ethnicity classifications. This is because both classifications are controlled in the proposed weighting. Although the proposed weighting is not perfect, it shows an improvement over the other two weighting methods. It is clear that the proposed weights provide better estimates than the other two weights, as differences examined in NL AI/ANs using the original weights, and in AM AI/ANs using the revised weights, were mitigated. The weighted totals for both AM AI/ANs and NL AI/ANs using the proposed weights were reasonably close to both control totals. This was especially true for AM AI/ANs.

One caveat of the proposed weights is that they distort the distribution of the NL NHPI group. The reason for this distortion may be found in NL NHPIs' small proportion in the population (0.3%) and their sample size (237). The proposed method may be subject to more measurement error as group size decreases.

Table 3
Control Totals

	Original Weight		Revised Weight		Proposed Weight		Control Total	
	Count	%	Count	%	Count	%	Count	%
Census Race/Ethnicity								
AM Latino	10,774,044	32.6	11,380,586	33.6	11,033,116	32.6	10,979,341	32.4
AM White	20,984,429	63.5	20,489,598	60.4	21,419,884	63.2	21,516,027	63.4
AM Black	2,370,085	7.2	2,492,683	7.4	2,508,154	7.4	2,515,971	7.4
AM Asian	4,092,122	12.4	4,232,701	12.5	4,152,713	12.3	4,160,530	12.3
AM AI/AN	611,468	1.9	1,287,600	3.8	620,477	1.8	628,294	1.9
AM NHOPI	215,878	0.7	278,552	0.8	213,900	0.6	221,716	0.7
AM Other	6,447,609	19.5	6,256,902	18.5	6,487,148	19.1	6,583,291	19.4
DOF Race/Ethnicity								
Latino	10,774,044	32.6	11,380,586	33.6	11,033,116	32.6	11,039,991	32.6
NL White	15,931,533	48.2	15,603,435	46.0	16,102,044	47.5	15,985,735	47.1
NL Black	1,770,685	5.4	2,141,059	6.3	2,153,849	6.4	2,214,193	6.5
NL Asian	3,232,261	9.8	3,800,472	11.2	3,671,415	10.8	3,731,759	11.0
NL AI/AN	53,325	0.2	210,296	0.6	131,661	0.4	192,005	0.6
NL NHOPI	62,261	0.2	113,932	0.3	50,425	0.1	110,769	0.3
NL Multirace	1,223,094	3.7	661,355	2.0	752,993	2.2	636,684	1.9
Total	33,050,816	100.0	33,911,135	100.0	33,895,502	100.0	33,911,135	100.0

Impact of Different Weights for the AI/AN Population

Percentage and total estimates of selected general health variables were calculated using the three weights described previously. Because research on AI/AN health may use different definitions of AI/AN, this study examines the estimates for AM AI/ANs and NL AI/ANs with the purpose of showing the importance of classifying the target study population by race/ethnicity.

Across the column in Table 4, the percentage estimates did not appear to differ substantially by weighting schemes. Asthma prevalence rates for AM AI/AN adults appeared to have the largest differences among the three types of weights, with estimates of 19.5%, 15.2%, and 17.8%. As calculation of the 95% confidence intervals follows p + (se(p)*1.96) where p is the estimated proportion, and se(p) is its standard error, one may easily calculate confidence intervals using information from Table 4. For example, when examining 95% confidence intervals of the three estimates above, they all overlapped; i.e., 95% CI of asthma prevalence for AI/ANs using original weight = 19.5% + (1.2%*1.96). However, it should be noted that determining the significance of

differences using confidence intervals is a convenient yet deficient approach (Schenker & Gentleman, 2001). The weighted totals, on the other hand, differed considerably by weighting schemes. For example, the number of currently insured NL AI/AN adults could be projected anywhere from 31,297 with original weights to 130,136 with revised weights, while the proposed weight produced an estimated total of 82,689 insured NL AI/ANs. This result is not surprising because the population totals for NL AI/AN using these three weights diverged substantially in value (53,325; 210,296; and 131,661). In addition to the characteristics in Table 4, 20 other variables were also examined (results not shown). The findings for these variables were consistent with Table 4: The percentage estimates did not differ by weights, but the weighted totals did.

Table 4
CHIS 2001 Estimates of Health-Related Variables for the AI/AN Population Using Different Weights

	Original Weight			Revised Weight			Proposed Weight		
	Weighted Total	(%)	SE (%)	Weighted Total	(%)	SE (%)	Weighted Total	(%)	SE (%)
General health: Fair, Poor									
AM AI/AN	89,569	21.4	1.1	187,395	22.9	1.4	88,428	22.8	1.5
NL AI/AN	8,952	24.5	2.6	37,918	24.9	2.9	23,402	24.3	2.9
Arthritis									
AM AI/AN	112,468	26.9	1.1	198,327	24.2	1.3	106,974	27.6	1.6
NL AI/AN	10,982	30.1	2.7	46,862	30.7	3.1	28,873	30.0	3.1
Asthma									
AM AI/AN	81,522	19.5	1.2	124,393	15.2	1.0	69,153	17.8	1.4
NL AI/AN	7,708	21.2	2.6	34,728	22.8	3.0	21,243	22.1	3.0
Diabetes									
AM AI/AN	30,828	7.4	0.6	63,869	7.8	0.8	30,172	7.8	0.8
NL AI/AN	3,700	10.2	1.7	14,543	9.6	1.8	9,104	9.5	1.8
Hypertension									
AM AI/AN	105,411	25.2	1.1	183,000	22.4	1.2	92,480	23.9	1.3
NL AI/AN	10,476	28.8	2.8	44,632	29.4	3.0	28,115	29.3	3.1
Consume fruits and vegetables at least five times a day									
AM AI/AN	203,484	49.9	1.4	391,252	49.0	1.6	180,199	47.5	1.7
NL AI/AN	16,063	44.9	3.3	65,589	43.8	3.5	40,727	43.0	3.6
Overweight									
AM AI/AN	253,120	61.8	1.4	502,496	62.9	1.6	238,857	63.1	1.6
NL AI/AN	22,893	63.7	3.4	98,926	66.1	3.3	60,720	64.1	3.7

continued on next page

Table 4, Continued
CHIS 2001 Estimates of Health-Related Variables for the AI/AN Population Using Different Weights

	Original Weight			Revised Weight			Proposed Weight		
	Weighted Total	(%)	SE (%)	Weighted Total	(%)	SE (%)	Weighted Total	(%)	SE (%)
Food security among Federal poverty level 200% or below									
AM AI/AN	99,283	64.5	2.0	238,812	65.9	2.4	101,166	65.6	2.3
NL AI/AN	10,440	67.3	4.2	41,364	67.7	4.6	27,781	69.7	4.7
Drank any alcoholic beverage in past month									
AM AI/AN	239,749	57.3	1.3	472,431	57.7	1.6	225,065	58.0	1.6
NL AI/AN	19,517	53.5	3.2	80,046	52.5	3.5	49,570	51.4	3.7
Binge drinking among those who had alcoholic beverage in past month									
AM AI/AN	84,378	20.2	1.2	182,400	22.4	1.4	87,638	22.7	1.8
NL AI/AN	7,218	20.1	2.6	29,749	19.9	2.7	18,197	19.2	2.7
Current smoker									
AM AI/AN	113,153	27.0	1.2	215,877	26.3	1.4	109,279	28.2	1.7
NL AI/AN	12,030	33.0	2.9	49,619	32.5	3.1	30,352	31.5	3.1
Currently insured									
AM AI/AN	346,079	82.7	1.1	651,971	79.6	1.4	316,456	81.6	1.3
NL AI/AN	31,297	85.7	2.0	130,136	85.3	2.2	82,689	85.8	2.1
Insured all past 12 months									
AM AI/AN	274,626	73.4	1.3	521,209	70.4	1.6	247,301	71.2	1.8
NL AI/AN	23,949	76.6	2.9	104,211	77.7	2.9	66,016	78.3	2.9
Covered by Indian Health Service									
AM AI/AN	16,645	4.3	0.5	42,355	5.9	0.7	22,933	6.6	0.7
NL AI/AN	7,553	22.4	2.5	27,401	19.0	2.6	16,761	19.0	2.5
Delays or not getting medical test/treatment in past 12 months									
AM AI/AN	44,143	10.6	0.8	74,377	9.1	0.8	36,181	9.3	0.9
NL AI/AN	3,410	9.3	1.6	15,001	9.8	1.8	9,176	9.5	1.8
Discriminated against in receiving health care in past 12 months									
AM AI/AN	36,727	8.9	0.7	65,256	8.1	0.8	36,457	9.5	1.3
NL AI/AN	3,274	9.1	1.8	14,899	9.9	2.1	9,113	9.5	2.1

CONCLUSION

Race/ethnicity, one of the most important weighting variables in survey data, can be classified in many different ways, and choice of classification impacts the public health statistics for AI/ANs. Different weights examined in this study did not change the percentage estimates of health-related variables for AI/ANs but did influence the weighted totals. Although one type of

race/ethnicity variable is controlled, if others are not, it is possible that the estimated population totals for uncontrolled variables could diverge from the true population counts.

It is reasonable to assume that a similar pattern may emerge for other small racial groups, such as NHPI. Weights created for these groups might not be as stable as weights for other groups because of their small proportions in the population, small sample sizes, the complexity in measuring race/ethnicity and the dynamics in its classification, and the availability of the data for weighting control totals. This instability was shown in the proposed weights—as the precision for AI/ANs improved, there was a negative effect on NL NHPI. For small racial and ethnic groups, a reasonable strategy might be to take the percentage estimates from the survey and multiply them by their known population totals from external sources such as the U.S. Census or official intercensal population statistics to estimate weighted counts.

It has been shown that classification and tabulation rules can affect both counts and predictors of health status, risks, and health needs of some populations by race/ethnicity (Mays et al., 2003). In addition, variants in classification and tabulation can potentially affect the rarest population groups in weighting survey data. In California, AI/ANs are greatly affected: California is home to the largest population of AI/ANs in the U.S., but overall AI/ANs are one of the smallest populations in this diverse state. As survey data are widely used for policy planning purposes, policy makers need to be aware that the choice of racial tabulation for weighting variables affects the data they use for decision making. Ideally, the race/ethnicity variables used in analyses will be consistent with variables controlled in weighting, and account for vulnerable and small populations. AI/ANs as a racial minority—and the only U.S. federally recognized political minority—are underrepresented in public health data collection systems. Imprecise estimates caused by inconsistency between the race/ethnicity variables being analyzed and the variables controlled in the weighting could cause policy makers to overlook the health needs of this racial-political group and result in serious resource misallocations in public health.

Sunghee Lee, PhD
UCLA Center for Health Policy Research,
UCLA Department of Biostatistics
10960 Wilshire Blvd., Suite 1550
Los Angeles, CA 90025
Phone: 310/794-2399

REFERENCES

Botman, S.L., Moore, T.F., Moriarity, C.L., & Parsons, V.L. (2000). Design and estimation for the National Health Interview Survey, 1995-2004. *Vital Health Statistics, 2*(130), 1-31.

Burhansstipanov, L. & Satter, D.E. (2000). Office of Management and Budget racial categories and implications for American Indians and Alaska Natives. *American Journal of Public Health, 90*(11), 1720-1723.

California Department of Finance. (2004). *Race/ethnic population with age and sex detail, 2000-2050*. Sacramento, CA.

California Health Interview Survey (2005). *CHIS 2001 methodology series: Revised CHIS 2001 weights*. Los Angeles, CA: UCLA Center for Health Policy Research.

Cochran, W. (1977). *Sampling techniques*. New York: John Wiley & Sons.

Cruz-Jansen, M. (2002). Lives on the crossfire: The struggle of multiethnic and multiracial Latinos for identity in a dichotomous and racialized world. *Race, Gender & Class, 9*(2), 47-62.

Deming ,W.E. & Stephen, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics, 11*, 427-444.

Forbes, J. (1990). Undercounting Native Americans: The 1980 Census and the manipulation of

racial identity in the United States. *Wicazo Sa Review, 6*(1), 2-6.

Kalton, G. & Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics, 19*(2), 81-97.

Kish, L. (1966). *Survey sampling*. New York: John Wiley & Sons.

Mays, V.M., Ponce, N.A., Washington, D.L., & Cochran, S.D. (2003). Classification of race and ethnicity: Implications for public health. *Annual Review of Public Health, 24*, 83-110.

Ogunwole, S.U. (2002). *The American Indian and Alaska Native population: 2000. Census 2000 brief*. Retrieved February 1, 2009 from http://www.census.gov/prod/2002pubs/c2kbr01-15.pdf

Ponce, N.A., Lavarreda, S.A., Yen, W., Brown, R.E., DiSogra, C., & Satter, D.E. (2004). The California Health Interview Survey 2001: Translation of a major survey for California's multiethnic population. *Public Health Reports, 119*, 388-395.

Särndal, C-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Satter, D.E., Veiga-Ermert, A., Burhansstipanov, L., Pena, L., & Restivo, T. (2005). Communicating respectfully with American Indian and Alaska Natives: Lessons from the California Health Interview Survey. *Journal of Cancer Education, 20*(1), 49-51.

Schenker, N., & Gentleman, J.F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician, 55*(3), 182-186.

Swift, E.K. (2002). *Guidance for the national healthcare disparities report. Committee on guidance for designing a national healthcare disparities report*. Washington, DC: Institute of Medicine.

Tashiro, C. (2002). Considering the significance of ancestry through the prism of mixed-race identity. *Advances in Nursing Science, 25*(2),1-21.

Tompkins, L. & Kim, J. (2007). *An application of alternative weighting matrix collapsing approaches for improving sample estimates*. Paper presented at the annual Joint Statistical Meeting, Salt Lake City, UT.

U.S. Census Bureau. (2002). *Census 2000 modified race data* [MR(31)-CO.txt]. Retrieved February 1, 2009 from http://www.census.gov/popest/archives/files/MRSF-01-US1.pdf

U.S. Office of Management and Budget. (1995). *Standards for the classification of federal data on race and ethnicity. Federal Register Notice*. Retrieved February 1, 2009 from http://www.whitehouse.gov/omb/fedreg/race-ethnicity.html

U.S. Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity. Federal Register Notice*. Retrieved February 1, 2009 from http://www.whitehouse.gov/omb/fedreg/ombdir15.html

Winker, M.A. (2004). Measuring race and ethnicity. Why and how? *Journal of the American Medical Association, 292*(13), 1612-1614.

Yen, W. & Satter, D.E. (2002). *Unexpected large sample yield of American Indian and Alaska Natives in the 2001 California Health Interview Survey*. Paper presented at the annual meeting of the Academy for Health Services Research and Health Policy, Washington, D.C.

Appendix
Original and Revised CHIS 2001 Ratio-Raking Dimensions

Control Geography	Control Variable: Original Ratio-Raking	Control Variable: Revised Ratio-Raking
Stratum	1. Age x Sex 2. Age	1. LA SPA, Alameda county 2. Age x Sex 3. Age 4. Race2 x Age
Collapsed Stratum	3. Any mention Latino x Age 4. Any mention African American x Age 5. Any mention White x Age	
Region		5. DOF Race
State	6. Any mention Other race x Age 7. Any mention NHOPI [1] x Age 8. Any mention AI/AN [2] x Age 9. Any mention Asian x Age 10. Age x Sex 11. Nontelephone adjustment [4] combining AFDC (Aid to Families with Dependent Children) participation, # of child, # of adults, and Race1	6. DOF Race x Age 7. Age x Sex 8. Asian group [3] x Age 9. Education 10. # of adult 11. Nontelephone adjustment combining household tenure, # of adults and education level

Source: CHIS Technical Report – Revised California Health Interview Survey 2001 Weights available at http://www.chis.ucla.edu/pdf/reweight_technical_chis01.pdf

[1] Native Hawaiian Other Pacific Islander

[2] American Indian/Alaska Native

[3] Non-Latino Chinese, Non-Latino Korean, Non-Latino Filipino, Non-Latino Vietnamese, Other, or Non-Asian

[4] Nontelephone adjustment dimension for 2001 revised ratio-raking is the same as its counterpart in 2001 original ratio-raking

Note: Race 1: See Appendix in CHIS Technical Report – Revised California Health Interview Survey 2001 Weights

Race 2: Latino, Non-Latino White, Non-Latino African American, Non-Latino Asian, Non-Latino American Indian and Alaska Native, Non-Latino Native Hawaiian/other Pacific Islander, Non-Latino multiple race