Racial Misidentification of American Indians/Alaska Natives in the HIV/AIDS Reporting Systems of Five States and One Urban Health Jurisdiction, U.S., 1984–2002

Jeanne Bertolli, PhD[a]
Lisa M. Lee, PhD[b]
Patrick S. Sullivan, DVM, PhD[c]
AI/AN Race/Ethnicity Data Validation Workgroup[d]

SYNOPSIS

Objectives. We examined racial misidentification of American Indians/Alaska Natives (AI/AN) reported to the human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) Reporting Systems (HARS) of five U.S. states and one county.

Methods. To identify AI/AN records with misidentified race, we linked HARS data from 1984 through 2002 to the Indian Health Service National Patient Information and Reporting System (NPIRS), excluding non-AI/AN dependents, using probabilistic matching with clerical review. We used chi-square tests to examine differences in proportions and logistic regression to examine the associations of racial misidentification with HARS site, degree of AI/AN ancestry, mode of exposure to HIV, and urban or rural location of residence at time of diagnosis.

Results. A total of 1,523 AI/AN individuals was found in both NPIRS and HARS; race was misidentified in HARS for 459 (30%). The percentages of racially misidentified ranged from 3.7% (in Alaska) to 55% (in California). AI/AN people were misidentified as white (70%), Hispanic (16%), black (11%), and Asian/Pacific Islander (2%); for 0.9%, race was unspecified. Logistic regression results (data from all areas, all variables) indicated that urban residence at time of diagnosis, degree of AI/AN ancestry, and mode of exposure to HIV were significantly associated with racial misidentification of AI/AN people reported to HARS.

Conclusions. Our findings add to the evidence that racial misidentification of AI/AN in surveillance data can result in underestimation of AI/AN HIV/AIDS case counts. Racial misidentification must be addressed to ensure that HIV/AIDS surveillance data can be used as the basis for equitable resource allocation decisions, and to inform and mobilize public health action.

[a]Office of Health Disparities, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA (current affiliation: Behavioral and Clinical Surveillance Branch, Division of HIV/AIDS Prevention, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA)

[b]HIV Incidence and Case Surveillance Branch, Division of HIV/AIDS Prevention, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA

[c]Behavioral and Clinical Surveillance Branch, Division of HIV/AIDS Prevention, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA

[d]Please see acknowledgments for a complete listing of Workgroup members.

Address correspondence to: Jeanne Bertolli, PhD, 1600 Clifton Rd. NE, Mailstop E46, Atlanta, GA 30333; tel. 404-639-8500; fax 404-639-8640; e-mail <jbertolli@cdc.gov>.

Historical treaty obligations and the unique government-to-government relationship between the United States and federally recognized American Indian/Alaska Native (AI/AN) tribes entitle tribal members and dependents to receive federally funded health-care services, many of which are now provided through tribally operated health-care facilities.[1] For the purposes of epidemiologic assessment and program design, monitoring, and evaluation, AI/AN people who use Indian Health Service (IHS) funded health-care services differ from AI/AN people who are unconnected with AI/AN tribes or health-care institutions, but who still identify as AI/AN in whole or in part because of their ancestry. Although the public health surveillance systems that provide data for epidemiologic analyses, to direct public health programs, and to monitor and evaluate program progress allow for identification of AI/AN race, they are not designed to make distinctions between AI/AN people based on their connection (or lack of connection) with AI/AN tribes or AI/AN health-care institutions.

Accurate coding of AI/AN race in human immunodeficiency virus (HIV)/acquired immunodeficiency syndrome (AIDS) surveillance systems, along with completeness of case reporting, is essential to ensure that resources for HIV-related services are channeled to the system of AI/AN-serving health-care facilities and public health systems and that AI/AN communities are able to support and focus HIV prevention measures effectively. Even when all HIV/AIDS diagnoses are reported to a public health surveillance system, racial misidentification may reduce the effectiveness of surveillance to provide "information for action" (e.g., through underestimation of the need for HIV-related services) and by erroneously supporting the notion that HIV infection is a problem outside of, rather than within, AI/AN communities.[2]

In 2003, 220 AIDS diagnoses among adult and adolescent AI/AN were reported nationally and 136 HIV diagnoses among adult and adolescent AI/AN were reported from 41 areas with confidential name-based HIV infection reporting; the rate of AIDS was 10.4 per 100,000 AI/AN people.[3] Undercounting of AI/AN cases due to racial misidentification is a documented problem in cancer, injury, renal disease, and sexually transmitted disease (STD) monitoring systems, and with birth and death certificates.[4–14] This undercounting in other public health surveillance systems has prompted concern that racial misidentification might also affect the accuracy of data on HIV/AIDS in AI/AN populations.[15–18]

In the early 1990s, three exploratory studies conducted to assess racial misidentification of reported AI/AN cases of AIDS compared race coding on state surveillance case reports with self-reported AI/AN race,[19,20] or with AI/AN race as determined by eligibility for IHS care.[21] As proof of eligibility, the IHS requires a Certificate of Degree of Indian Blood (CDIB),[22] issued by the Bureau of Indian Affairs (BIA) to those who submit an application and evidence of lineal descent. Based on a very small number of cases in Los Angeles (LA) County, Seattle-King County, and Washington State, and using different methodologies, the findings of these assessments were not generalizable, but they did point to a potential problem of undercounting of AI/AN AIDS cases due to misidentification of race/ethnicity.[19–21]

Two later evaluations, by Kelley et al.[23] and Lee et al.,[24] both comparing race coding on surveillance case reports with self-reported race, were consistent in showing AI/AN as the group with the most disagreement between self-reported race and race reported on a surveillance case report (57% and 65% disagreement, respectively). No further evaluations comparing race/ethnicity data of AIDS surveillance case reports against eligibility for IHS services funded through IHS have been published in the 15 years since the first limited evaluation. We describe the findings of such an evaluation in five U.S. states with the highest cumulative numbers of HIV/AIDS diagnoses among AI/AN reported through 2003, and in a large urban health jurisdiction within one of these states.

METHODS

As described, different methodologies have been used to assess racial misidentification of AI/AN race in public health data. Because the purpose of this analysis was to explore the implications of AI/AN racial misidentification for AI/AN-serving health-care institutions funded through IHS, we chose a methodology that would allow assessment of racial misidentification among AI/AN people served by these institutions.

Our analysis involved linking the HIV/AIDS Reporting System (HARS), which is used for population-based AIDS case surveillance in all 50 U.S. states and for HIV infection surveillance in states that require HIV infection reporting, with the IHS National Patient Information and Reporting System (NPIRS).

HARS includes case reports of AI/AN people diagnosed with HIV or AIDS, whether or not the HIV/AIDS diagnosis occurred in an IHS-funded health-care institution. NPIRS comprises records of 2.47 million AI/AN patients who received health-care services from IHS-

funded facilities nationwide between 1984 and 2002, including individuals known to be deceased and those who discontinued use of IHS-funded services.

The states with the largest cumulative numbers of AI/AN HIV/AIDS diagnoses reported to HARS through 2003 were (ranked highest to lowest): California, Oklahoma, Arizona, Washington, and Alaska. The combined AI/AN population of these five states represents 38% of the AI/AN population of the U.S. (AI/AN alone or in combination with other races).[25] In all of these states, name-based reporting of AIDS diagnoses has been in place since 1984 or earlier. Arizona, Oklahoma, and Alaska have had confidential name-based HIV reporting since 1987, 1988, and 1999, respectively. Washington implemented name-to-coded patient identifier reporting of asymptomatic HIV infection in 1999; California implemented HIV reporting by coded patient identifier in 2002.

HARS databases from each of these five states, including all cases of all races/ethnicities reported between 1984 and 2002, were linked individually to NPIRS. To assess racial misidentification of AI/AN within a large urban setting, HARS data from LA County were evaluated separately. (LA was chosen because it has the largest AI/AN population and largest cumulative number of AI/AN AIDS cases among cities in the states included in this analysis.) For this analysis, LA County HARS data were excluded from the California HARS database before the California data were linked to NPIRS. Probabilistic matching was performed using Integrity (Ascential Software, Westborough, MA).[26] Because California did not implement HIV reporting until July 2002, only AIDS data were available from California and LA County; both HIV and AIDS data were available from all the other states.

Prior to linkage, HARS and NPIRS databases were each unduplicated; additional records were created for individuals in each HARS database for whom alias names had been reported (one record per alias); and non-AI/AN dependents were removed from the NPIRS database. Inclusion in NPIRS requires documentation of AI/AN ancestry or dependent (spousal, filial) relationship with a person with documented AI/AN ancestry. After removal of non-AI/AN dependents, the remaining NPIRS records were assumed to be of individuals with certification of AI/AN ancestry. Record linkage took place at the offices of participating health departments, on a password-protected stand-alone computer that was not on an agency network or in any other manner linked to another computer.

Each site developed its own matching process, based on the availability and completeness of data elements usable to identify matching records. Three to seven "passes" through the data were made to identify individuals appearing in both the HARS and NPIRS databases. Each pass was defined by a unique combination of matching parameters. Records for which a match was found in a given pass were excluded from subsequent passes.

The first step for each pass was to create blocks of HARS and NPIRS record pairs matching exactly on each of a set of data elements. Blocks were defined more broadly with each successive pass by including fewer data elements on which pairs of records must match. Depending on the pass, the set of data elements used to create blocks included various combinations of the following: last name, first name, gender, date/year/month of birth, soundex (a coded surname index based on the way a surname sounds), and social security number. Data elements were available in all states except as noted: in Washington, only names of people with symptomatic HIV infection were available; social security numbers were unavailable in three states and inconsistently available in the other two states.

The second step for each pass was to identify, among the blocks of records, those with exact or possible matches on the remaining data elements (those from the list previously mentioned not used to create the blocks for that pass, and including one additional element: middle name). Finally, weights were calculated that reflected the probability of a true match, given the degree of variability in data elements used to identify matching records (the software automatically adjusted weights to account for common variations of first names, such as "Thomas" and "Tom"). To determine whether a pair of records matched, the software calculated a weight for each comparison, according to the error probabilities associated with each field.[27]

Two probabilities are specified in advance for each field: the m probability (i.e., the probability that a field agrees given that the record pair is a true match) and the u probability (i.e., the probability that the field agrees, given that the pair is, in fact, unmatched). These error probabilities were estimated from experience with the data; for some very important fields, such as name and social security number, the m probability was set high to force these fields to have a high penalty for disagreeing. For each matching field, the software computed a weight based on these two probabilities, and used the weights from different fields being compared to obtain a composite weight.[27] The composite weights assigned to the record pairs created a distribution of scores.

Within the distribution, cutoff values were defined such that any record pair receiving a weight equal to or greater than the upper cutoff was considered a match,

and any record pair receiving a weight equal to or less than the lower cutoff was considered a non-match. Any record pairs with weights that fell between these two cutoff values were manually reviewed. The cutoffs were determined for data from each jurisdiction by the principal investigator, based on the distribution of composite weights and corresponding differences in critical data fields. The cutoffs reflected the level of confidence that a record pair with a given composite weight was a true match.

The third step involved review of possible matches identified in step 2 for each pass and a decision to accept or reject each possible match based on the degree and nature of differences in the data used to identify matching records. Considerations included the possibility that a difference in a name or date could be due to a typographical error. Each case was reviewed by two or more people and a decision was made by consensus. Exact correspondence of the social security numbers of a HARS record and a NPIRS record provided strong evidence favoring acceptance of a pair of records as a match when there were slight differences in other variables. At the conclusion of the record linkage process, all temporary files that included patient identifiers were permanently deleted.

To estimate the effect of correcting racial misidentification among AI/AN people accessing IHS-funded health institutions on the overall AI/AN-specific AIDS rates, we adjusted the rates of AI/AN people diagnosed with AIDS at the end of 2001 for each site, based on the reassignment of non-AI/AN AIDS cases with matching records in NPIRS to the AI/AN race category. AI/AN-specific AIDS surveillance rates are estimated by dividing unduplicated AI/AN case counts from HARS by the AI/AN population as determined by the most recent census or post-censal estimate. Assignment of race in the HARS data used for the analysis was according to the four categories specified in the 1977 federal Office of Management and Budget (OMB) standards (i.e., each person was assigned a single racial category).[28]

Vintage 2001 post-censal estimates of the AI/AN population were used for the denominators of the site-specific rates. Because the 2000 census allowed respondents to select one or more race categories when responding to a query on their racial identity, to match the racial categories in the HARS data, the racial categories of the rate denominators had to be estimated by bridging the vintage 2001 post-censal estimates to the single AI/AN race category allowed under the 1977 OMB standards. The bridging methodology was developed by the National Center for Health Statistics, Centers for Disease Control and Prevention.[29]

We adjusted the AI/AN-specific AIDS rate estimate from AIDS surveillance data for each health jurisdiction participating in the analysis by multiplying the jurisdiction's rate estimate by the following correction factor:

$$\frac{\text{Number of AI/AN diagnosed with AIDS in 2001 (based on record linkage results)}}{\text{Number of AI/AN diagnosed with AIDS in 2001 (based on AIDS surveillance)}}$$

(In accordance with the purpose of this analysis, this adjustment only corrects racial misidentification of individuals eligible for and served by IHS-funded health facilities.)

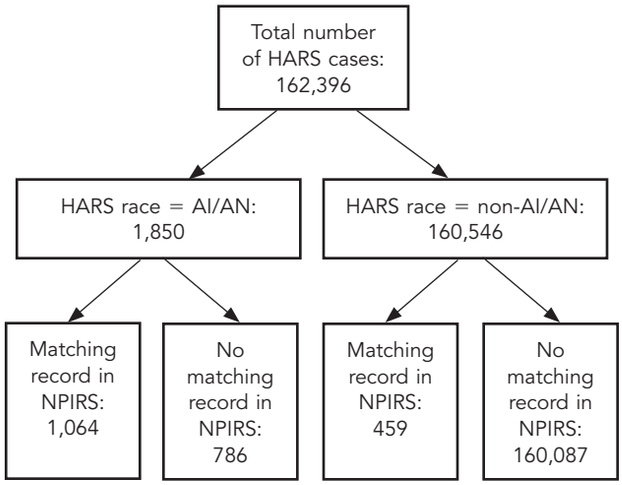
All analyses were conducted using SAS statistical software, version 8.02.[30] For each site, we used chi-square tests to examine differences in proportions of people who were racially misidentified, by gender, age at the end of 2002, degree of AI/AN ancestry (from NPIRS), date of HIV/AIDS diagnosis, and urban/rural location of residence at time of HIV/AIDS diagnosis (i.e., metropolitan statistical area [MSA] or non-MSA as defined by the OMB),[31] type of facility (public or private) where the HIV/AIDS diagnosis was made, and mode of exposure to HIV. We combined the data from all sites and developed a logistic regression model through a backward stepwise procedure (including site in the model to control for differences between jurisdictions) that examined the independent associations of racial misidentification with gender, age, degree of AI/AN ancestry, mode of exposure to HIV, date of diagnosis of HIV/AIDS, and location of residence at time of HIV/AIDS diagnosis. The criterion for removal of variables from the model was $p=0.05$.

RESULTS

The number of HARS records linked with the 2.47 million records in NPIRS varied by site from 762 to 81,079. Figure 1 summarizes the results of the record linkage. Among the total of 162,396 HARS records from all sites combined, 1,523 with a matching record in NPIRS were identified. Of these 1,523 cases, 1,064 (70%) had been assigned AI/AN race in HARS, and 459 (30%) had been assigned a non-AI/AN race descriptor (i.e., were misidentified).

The 1,064 AI/AN HIV/AIDS cases in HARS with a corresponding record in NPIRS accounted for 57% of the total of 1,850 cases with AI/AN race in HARS. For the remaining 786 cases (43%) with AI/AN race in HARS, we found no corresponding record in NPIRS to verify the accuracy of race coding. (These 786 cases included 15 in Alaska, 152 in Arizona, 97 in LA County,

Figure 1. Combined results from linkage of HARS records of five U.S. states and one urban health jurisdiction with records of the IHS NPIRS, 1984–2002



HARS = Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome Reporting System

IHS = Indian Health Service

NPIRS = National Patient Information and Registration System

AI/AN = American Indian/Alaska Native

330 in California [excluding LA County], 54 in Oklahoma, and 138 in Washington.) Among the 160,546 cases with non-AI/AN race in HARS, 459 (0.3% of all non-AI/AN cases in HARS) were determined to be AI/AN and the remaining 160,087 (99.7%) were presumed to be non-AI/AN because no corresponding record was found in NPIRS.

Table 1 presents the results of the record linkage by site. The percentages of AI/AN HIV/AIDS cases that were racially misidentified in HARS ranged from 3.7% (in Alaska) to 55.0% (in California). Correction of the documented race of racially misidentified AI/AN people increases the proportion of AI/AN among cumulative HIV/AIDS cases from 22.7% to 23.4% in Alaska, from 3.0% to 3.4% in Arizona, from 0.3% to 0.4% in LA County, from 0.5% to 0.7% in California (excluding LA County), from 6.2% to 8.8% in Oklahoma, and from 1.7% to 2.0% in Washington.

Table 2 shows the increases in the rates of AI/AN people diagnosed with AIDS in 2001, per 100,000 population, for each site, after inclusion of racially misidentified AI/AN cases. The rate increased by 43.3% in LA County (with the addition of four cases to the numerator), 40.8% in Oklahoma (with the addition of seven cases), 16.3% in California (with the addition of three cases), 8.7% in Washington (with the addition of one case), and 8.1% in Arizona (with the addition of two cases); the rate in Alaska did not change.

Figure 2 shows the distribution of misidentified AI/AN people according to incorrectly assigned racial/ethnic category for all sites combined, including four people for whom race was listed as unknown. Overall, AI/AN people were most likely to be misidentified as white (70% of the AI/AN misidentified cases), and this was also true at each site. But, as shown in Figure 3, there was some variation across sites in the distribution of other race categories to which AI/AN cases were

Table 1. Results of linking HARS records, from five U.S. states and one urban health jurisdiction, with records of the IHS NPIRS, 1984–2002

Site	*Total number of HARS records*[a]	*Total number of HARS/NPIRS matches*	*Number of HARS/NPIRS matches with AI/AN race in HARS*	*Number of HARS/NPIRS matches with non-AI/AN race in HARS*	*Percent of HARS/NPIRS matches racially misidentified in HARS*
Alaska	762	164	158	6	3.7
Arizona	16,582	420	346	74	17.6
California[b]	81,079	231	104	127	55.0
Los Angeles County	44,013	68	31	37	54.4
Oklahoma	6,252	500	332	168	33.6
Washington	13,708	140	93	47	33.6
Total	162,396	1,523	1,064	459	30.1

[a]On date of record linkage

[b]California data exclude records from Los Angeles County.

AI/AN = American Indian/Alaska Native

HARS = Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome Reporting System

IHS = Indian Health Service

NPIRS = National Patient Information and Registration System

Table 2. Rates of AI/AN people diagnosed with AIDS in 2001, per 100,000 population, before and after correction for racial misidentification, by site—from linkage of HARS records, from five U.S. states and one urban health jurisdiction, with records of the IHS NPIRS, 1984–2002

	Before correction		*After correction*	
Site	*Number of cases*	*AI/AN rate*[a]	*Number of cases*	*AI/AN rate*[a]
Alaska	5	6.6	5	6.6
Arizona	27	14.7	29	15.9
California[b]	20	12.9	23	15.0
Los Angeles County	7	25.2	11	36.1
Oklahoma	17	7.6	24	10.7
Washington	12	15.0	13	16.3

[a]Numerators of rates were estimated from HARS, with adjustment for reporting delay; denominators were estimated from the 2000 U.S. Census by "bridging" from the multirace categories to the race taxonomy used prior to the 2000 Census. (National Center for Health Statistics [US]. Documentation for bridged-race vintage 2001 [July 1, 2000–July 1, 2001] postcensal population estimates for calculating vital rates. Available from: URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datasets/nvss/bridgepop/DocumentationBridgedPostCenV2001.doc [cited 2005 May 1].)

[b]California data exclude records from Los Angeles County.

AI/AN = American Indian/Alaska Native

AIDS = acquired immunodeficiency syndrome

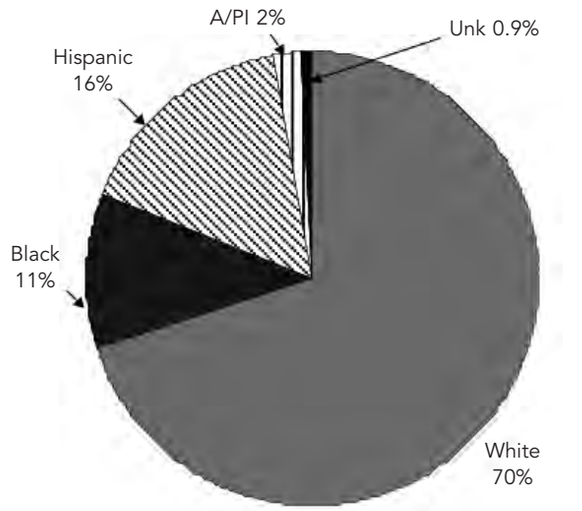
HIV = human immunodeficiency virus

HARS = HIV/AIDS Reporting System

IHS = Indian Health Service

NPIRS = National Patient Information and Registration System

Figure 2. Distribution of misidentified AI/AN people (*n*=459) reported to HARS by incorrectly assigned racial/ethnic classification—combined data for five U.S. states and one urban health jurisdiction, 1984–2002—results from linkage with the IHS NPIRS



AI/AN = American Indian/Alaska Native

HARS = Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome Reporting System

IHS = Indian Health Service

NPIRS = National Patient Information and Registration System

A/PI = Asian/Pacific Islander

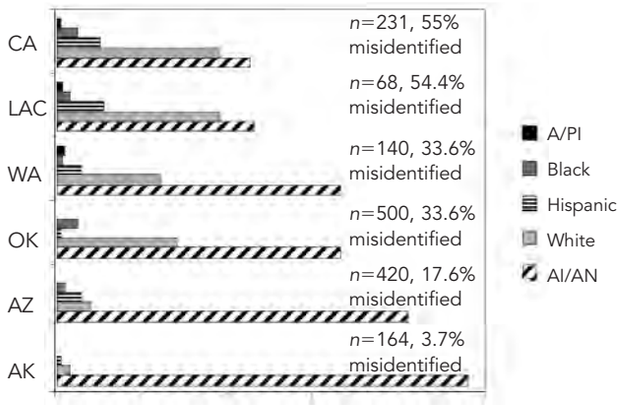
Unk = unknown

misidentified. In Alaska, Arizona, LA County, California (excluding LA County), and Washington, the second most common race/ethnic category to which AI/AN were misidentified was Hispanic, whereas in Oklahoma the second most common category was black.

Among 481 people recorded in NPIRS as having full AI/AN ancestry (100% AI/AN ancestry of both parents) and who had records in both the NPIRS and HARS databases, 83% were correctly identified as AI/AN in the HARS database, 10% were misidentified as white, 3% each as black or Hispanic, and 1% as Asian/Pacific Islander. People with 50% or more AI/AN ancestry were as likely as people with full AI/AN ancestry to be racially misidentified, but the proportion misidentified increased as degree of AI/AN ancestry dropped below 50%: 29% of people with at least 25% but less than 50% AI/AN ancestry were misidentified, as were 47% of those with less than 25% AI/AN ancestry. For 127 (8%) of 1,523 people whose records existed in both NPIRS and HARS, there was no information in NPIRS on degree of AI/AN ancestry.

Results from the logistic regression model combining data from all sites are shown in Table 3. After the backward stepwise elimination procedure, the variables remaining in the logistic regression model were site (jurisdiction contributing the data), degree of AI/AN ancestry, mode of exposure to HIV, risk category, and urban residence (in an MSA) at diagnosis. The results indicated that urban location of residence at diagnosis, degree of AI/AN ancestry, and mode of exposure were significantly associated with racial misidentification of AI/AN cases of HIV/AIDS reported to HARS. Age, gender, and years since HIV/AIDS diagnosis were excluded from the final model because they were not significantly associated with racial misidentification. People whose HIV infection or AIDS was diagnosed in an urban area had almost twice the odds of being racially misidentified as people diagnosed in a rural area. The odds of racial misidentification increased with decreasing degree of AI/AN ancestry. Whereas AI/AN people with more than one-half AI/AN ancestry were not significantly more likely to be racially misidenti-

Figure 3. Percentage distribution of AI/AN people reported to HARS by assigned race/ethnicity and site, five U.S. states and one urban health jurisdiction, 1984–2002—results from linkage with the IHS NPIRS



A/PI = Asian/Pacific Islander

AI/AN = American Indian/Alaska Native

HARS = Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome Reporting System

IHS = Indian Health Service

NPIRS = National Patient Information and Registration System

CA = California

LAC = Los Angeles County

WA = Washington

OK = Oklahoma

AZ = Arizona

AK = Alaska

Table 3. Factors independently associated with racial misidentification of AI/AN HIV/AIDS cases reported to HARS in five U.S. states and one urban health jurisdiction, 1984–2002, from multivariate logistic regression analysis[a]

Characteristic	*AOR*	*95% CI*
Degree of AI/AN ancestry		
100%	—	—
50%–99%	1.2	0.7, 1.9
25%–49%	2.0	1.2, 3.2
<25%	4.6	2.9, 7.4
Unknown	4.3	2.7, 6.9
Risk category		
MSM	—	—
IDU	0.9	0.6, 1.3
MSM/IDU	0.7	0.5, 1.1
Heterosexual contact	0.6	0.3, 1.1
Unspecified[b]	2.1	1.3, 3.3
MSA[c]		
Non-MSA	—	—
MSA	1.9	1.3, 2.7

[a]Results are from a logistic regression model, which included all the covariates listed in the first column, as well as a covariate representing the health jurisidiction contributing the data, which was included in the model to control for differences between the jurisdictions.

[b]The association of unspecified risk with racial misidentification may reflect that incompleteness of data used to assign risk category tends to coincide with incompleteness of information on race/ethnicity.

[c]MSA refers to the U.S. Office of Management and Budget classification (Office of Management and Budget [US]. Update of statistical area definitions and guidance on their uses; OMB Bulletin No. 05-02. November 2004 [cited 2006 Dec 27]. Available from: URL: http://www.whitehouse.gov/omb/bulletins/fy05/b05-02.html).

AI/AN = American Indian/Alaska Native

HIV = human immunodeficiency virus

AIDS = acquired immunodeficiency syndrome

HARS = HIV/AIDS Reporting System

MSM = men who have sex with men

IDU = injection drug user

MSA = metropolitan statistical area

CI = confidence interval

AOR = adjusted odds ratio

fied than people with full AI/AN ancestry, those with one-half to one-quarter AI/AN ancestry had twice the odds of being misidentified, and those with less than one-quarter AI/AN ancestry had 4.6 times the odds of being racially misidentified. Unrecorded degree of AI/AN ancestry and unknown mode of exposure to HIV infection were also significantly associated with racial misidentification.

DISCUSSION

We describe racial misidentification of AI/AN cases of HIV/AIDS reported to surveillance systems in five states and one urban health jurisdiction. This effort addresses concerns that such racial misidentification contributes to undercounting of AI/AN HIV/AIDS cases, potentially leading to under-allocation of resources and lack of support for services and programs tailored to AI/AN needs.[15–18] In this evaluation, we have considered inclusion in NPIRS to be verification of the AI/AN race of HIV/AIDS cases reported to HARS. While NPIRS is likely one of the most comprehensive national databases of people with AI/AN ancestry, it has a number of limitations. Because NPIRS likely does not include all those who identify culturally as AI/AN, it is limited as a standard for identifying people in the HARS database who are culturally AI/AN. This limitation has important implications for using the results of our data linkage project, e.g., our results represent a conservative estimate of racial misidentification of

reported AI/AN HIV/AIDS for use in adjusting estimates of prevention and care needs.

Nonetheless, this conservative estimate indicates substantial racial misidentification of AI/AN in some sites. In four of the six participating sites, racial misidentification exceeded 30%; in LA County and California (excluding LA County), misidentification exceeded 50%. Our finding that race was misidentified for 33% of AI/AN HIV/AIDS cases in Oklahoma is similar to the results of an earlier evaluation of racial misidentification of AI/AN syphilis and chlamydia cases reported in Oklahoma. This evaluation involved an analogous method of linking a surveillance database (in this case, a sexually transmitted disease [STD] surveillance database) with IHS patient registration data.[8] In the STD evaluation, 27% and 36% of AI/AN syphilis and chlamydia cases were misidentified, respectively; the proportion of AI/AN gonorrhea cases identified as non-AI/AN was even higher (56%).

Information recorded in medical records by hospital or clinical staff is the primary source of demographic information included in HIV/AIDS surveillance case reports. Health data on race/ethnicity are problematic for a number of reasons.[32–34] Data on race/ethnicity may be derived from an individual's self-report of his/her race or from a provider's perception of a client's race/ethnicity based on observation of physical characteristics, names, or other factors.[33] Furthermore, clients and their health-care providers may not understand the taxonomy of race/ethnicity categories used for surveillance purposes,[20,34] and race codes on forms may vary from place to place. In addition, a person with less than 100% AI/AN ancestry may identify with different racial groups in different settings.[7] Some AI/AN individuals with HIV/AIDS may be reluctant to identify themselves as AI/AN in a health-care setting because of stigma associated with HIV or risk behaviors, or to avoid unwanted referral to IHS.[23]

In our evaluation, we were unable to ascertain whether the source of miscoded race on case reports was the patient or the provider. However, our conservative method of identifying miscoding of race guaranteed that all those whose race we found documented incorrectly were listed in NPIRS, indicating that they had self-identified as AI/AN at least twice, to obtain a CDIB and to register for IHS services; 17% of people of full AI/AN ancestry listed in both the NPIRS and the HARS databases were misidentified in HARS as non-AI/AN, suggesting that true misidentification of those with exclusively AI/AN ancestry occurs relatively frequently.

It is possible that some non-AI/AN cases were misidentified in HARS as AI/AN. The finding of 786 cases identified as AI/AN in HARS with no matching record in NPIRS might be construed as support for this notion. However, the inability to locate matching records in NPIRS does not necessarily confirm these cases as non-AI/AN; it may instead be due to the limitations of the record linkage process, of using NPIRS, or both. In our record linkage process, we considered social security numbers as the ultimate proof of a match between HARS and NPIRS records, but social security numbers were not available in three of the participating sites, and inconsistently available in the other two. It is also possible that we may have found more matches between the records of the two databases if we had been able to use additional variables to identify matches. The use of NPIRS also has limitations, mainly that AI/AN people who receive health care from public or private providers rather than from IHS are not included in NPIRS, and that those who receive services from tribal facilities are included only if the tribal facility sends patient registration data to IHS. Had we linked HARS to tribal membership rolls or to a list of people to whom the BIA has issued a CDIB, we might have been able to verify the AI/AN race of more of the cases identified as AI/AN in HARS. Although we did not have this information, there is some evidence that racial misidentification of non-AI/AN people as AI/AN occurs less frequently than misidentification of AI/AN as non-AI/AN.[24]

Our results describe wide variation across sites in the percentages of reported AI/AN HIV/AIDS cases misidentified by race, from 3.7% (six of 164) in Alaska to 55.0% (127 of 231) in California. Geographic variation has been described in other evaluations of racial misidentification of AI/AN reported to various public health surveillance systems. This geographic variation is postulated to be directly related to the AI/AN proportion of the population (i.e., the larger the presence of AI/AN, the greater the likelihood of correct racial identification), and to the proportion of AI/AN people receiving care from AI/AN-serving health facilities (i.e., those who are diagnosed in AI/AN-serving facilities are more likely to be reported to surveillance systems as AI/AN).[8] Results of this evaluation support this correlation. Alaska—where AI/AN people make up more than 19% of the population, and where IHS and tribal health facilities' location in all rural regional hubs and urban centers gives generally easier access, on average, than for AI/AN residing in the lower 48 states—had the smallest proportion of racially misidentified AI/AN HIV/AIDS cases.

Harwell and colleagues[12] found that AI/AN people who live near reservations are more likely to be identified correctly as AI/AN on death certificates than

those who reside in urban areas. Results of another investigation indicated that variations in coding of AI/AN race on death certificates were related to whether the coding was done by tribal officials or by non-Indian funeral directors.[35] In California, where comparatively few AI/AN live on tribal lands and where AI/AN-serving health facilities are sparse,[36] one might expect greater racial misidentification of AI/AN, as our evaluation found.

We also report statistically significant associations of racial misidentification with percentages of AI/AN ancestry of one-half or less, as well as with urban location of residence at time of diagnosis. These results are concordant with those of other investigations[4,6,8–10,12] and corroborate the expectation that AI/AN people who may not conform to common notions of AI/AN physical features, and who are not otherwise identifiable by their residence on or near a reservation or their receipt of health care from an AI/AN-serving provider, tend to be racially misidentified.

Probabilistic matching allowed more flexibility in identifying matching records in HARS and NPIRS than an exact matching routine would have. However, among the pairs of records that the computer identified as possible (not exact) matches, the ultimate decision for acceptance or rejection as a match was made according to a site-specific algorithm considering the degree and type of difference in values of the matching variables, i.e., criteria that are not 100% error-proof.

The rates and changes in rates we present should be interpreted with caution. Ambiguity of group membership, and changes in group identity over time pose problems not only for identifying people as belonging to a particular group for case counts, but also complicate estimation of disease rates, because of the difficulty of determining compatible population estimates.[34] For example, demographic projections by the Bureau of the Census have underestimated the AI/AN population by as much as 35% in the past three decades, due to both ambiguity about AI/AN group membership and shifting criteria for identity.[34,37] The use of multirace categories, starting with the 2000 U.S. Census, further complicates estimation of population size for denominators of rates.[2,12,38]

IMPLICATIONS

The extent of racial misidentification of AI/AN people reported with HIV/AIDS documented in this report has implications for addressing prevention and service needs, and for the visibility of HIV/AIDS among AI/AN. Due to the direct link between case counts and funding for services, undercounting of AI/AN cases may contribute to underfunding of AI/AN-targeted services, and the development or perpetuation of health disparities.[39] It is possible that this undercounting also contributes to denial of HIV/AIDS as a problem in AI/AN communities.[18] Accurate AI/AN HIV/AIDS case counts depend in part on the application of an AI/AN group definition. Adjustment of national case counts to account for racial misidentification is complicated by the variation in degrees of misidentification across geographic areas, and by the lack of a clear concept of AI/AN race that is consistently understood and applied and that has direct relevance to the uses of surveillance data.[2,35]

As Satter[40] points out, treating disparate tribes as one group obscures important cultural distinctions that are relevant to health and social service delivery. These distinctions may be especially important for HIV prevention education, because gender and family roles, attitudes toward sexual orientation, and health-care beliefs vary across groups.[41] Burhansstipanov and Satter[2] and others[5,7–9,14,20] propose some recommendations for addressing racial misidentification of AI/AN in public health surveillance data, including training health-care providers to document their patients' self-reported race using a standard nomenclature, and regular matching of surveillance databases with tribal membership rolls. The latter solution requires formalized collaboration between state and county health departments and tribal governments. Recent efforts by the Council of State and Territorial Epidemiologists and the National Alliance of State and Territorial AIDS Directors to engage tribal health authorities are a step in this direction.[16,42]

CONCLUSION

Our findings add to the evidence that racial misidentification of AI/AN in HIV/AIDS surveillance data contributes to underestimation of the numbers of AI/AN people diagnosed with HIV/AIDS. A meaningful and practical concept of AI/AN group and subgroup membership, as well as formalized collaboration between regional health departments and tribal governments are needed for two reasons: to address racial misidentification and to ensure that HIV/AIDS surveillance data can be used not only as the basis for equitable resource allocation decisions, but also to inform and mobilize AI/AN communities and tailor prevention programs.

The opinions expressed are those of the authors and do not necessarily reflect the viewpoint of the Indian Health Service.

The Workgroup includes: Jennifer K. Baham, MPH, Office of AIDS, Epidemiologic Studies Section, California Department of Health Services, Sacramento, CA; Keith Bletzer, PhD, MPH,

Office of HIV/AIDS, Arizona Department of Health Services, Phoenix, AZ; Dan Cameron, PhD, Planning & Partnership Development, Oklahoma City Area Indian Health Service, Oklahoma City, OK; Penelope Cordes, PhD, HIV/STD Program, Alaska Department of Health and Social Services, Anchorage, AK; Maria Courogen, MPH, Infectious Disease and Reproductive Health Assessment Unit, Washington State Department of Health, Olympia, WA; Rick DeStephens, Office of HIV/AIDS, Arizona Department of Health Services, Phoenix, AZ; Douglas M. Frye, MD, MPH, HIV Epidemiology Program, Los Angeles County Department of Health Services, Los Angeles, CA; Virginia Y. Hu, MPH, HIV Epidemiology Program, Los Angeles County Department of Health Services, Los Angeles, CA; Deborah Frederickson Klinghoffer, PhD, HIV/STD Service, Oklahoma State Department of Health, Oklahoma City, OK; Elizabeth Lowery, MPH, Office of Health Disparities, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA; Anna Meddaugh-Baskapan, Infectious Disease and Reproductive Health Assessment Unit, Washington State Department of Health, Olympia, WA; Emily Puuka, Northwest Portland Area Indian Health Board, Portland, OR; Maya Tholandi, MPH, Office of AIDS, Epidemiologic Studies Section, California Department of Health Services, Sacramento, CA; Mark Turner, MPH, HIV/STD Service, Oklahoma State Department of Health, Oklahoma City, OK; Mi Suk Yu, MSPH, HIV Epidemiology Program, Los Angeles County Department of Health Services, Los Angeles, CA.

REFERENCES

1. Shelton BL. Issue brief: legal and historical roots of health care for American Indians and Alaska Natives in the United States. Menlo Park, California: Henry J. Kaiser Family Foundation; 2004 [cited 2005 Apr 25]. Available from: URL: http://www.kff.org/minorityhealth/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=31330
2. Burhansstipanov L, Satter DE. Office of Management and Budget racial categories and implications for American Indians and Alaska Natives. Am J Public Health 2000;90:1720-3.
3. Centers for Disease Control and Prevention (US). Cases of HIV infection and AIDS in the United States, 2003. HIV/AIDS Surveillance Report 2004;15:14.
4. Frost F, Taylor V, Fries E. Racial misclassification of Native Americans in a surveillance, epidemiology, and end results cancer registry. J Natl Cancer Inst 1992;84:957-62.
5. Sugarman JR, Holliday M, Ross A, Castorina J, Hui Y. Improving American Indian cancer data in the Washington State cancer registry using linkages with the Indian Health Service and tribal records. Cancer 1996;78(Suppl 7):1564-8.
6. Sugarman JR, Soderberg R, Gordon JE, Rivara FP. Racial misclassification of American Indians: its effect on injury rates in Oregon, 1989 through 1990. Am J Public Health 1993;83:681-4.
7. Sugarman JR, Lawson L. The effect of racial misclassification on estimates of end-stage renal disease among American Indians and Alaska Natives in the Pacific Northwest, 1988 through 1990. Am J Kidney Dis 1993;4:383-6.
8. Thoroughman DA, Frederickson D, Cameron HD, Shelby LK, Cheek JE. Racial misclassification of American Indians in Oklahoma State surveillance data for sexually transmitted diseases. Am J Epidemiol 2002;155:1137-41.
9. Epstein M, Moreno R, Bacchetti P. The underreporting of deaths of American Indian children in California, 1979 through 1993. Am J Public Health 1997;87:1363-6.
10. Frost F, Tollestrup K, Ross A, Sabotta E, Kimball E. Correctness of racial coding of American Indians and Alaska Natives on the Washington State death certificate. Am J Prev Med 1994;10:290-4.
11. Frost F, Shy KK. Racial differences between linked birth and infant death records in Washington State. Am J Public Health 1980;70:974-6.
12. Harwell TS, Hansen D, Moore KR, Jeanotte D, Gohdes D, Helgerson SD. Accuracy of race coding on American Indian death certificates, Montana 1996–1998. Public Health Rep 2002;117:44-9.
13. Sorlie PD, Rogot E, Johnson NJ. Validity of demographic characteristics on the death certificate. Epidemiology 1992;3:181-4.
14. Stehr-Green P, Bettles J, Robertson LD. Effect of racial/ethnic misclassification of American Indians and Alaskan Natives on Washington State death certificates, 1989–1997. Am J Public Health 2002;92:443-4.
15. The National Congress of American Indians. Improving the National Infectious Disease Surveillance System for Native America. Resolution 99-068, 1999 [cited 2006 Dec 27]. Available from: URL: http://www.ncai.org/index.php?id=105&print=1&type=123&selectpro_resid=14
16. Greabell L, Jorstad C. Native Americans and HIV/AIDS: key issues and recommendations for health departments. Washington: The National Alliance of State and Territorial AIDS Directors, 2004 [cited 2006 Dec 27]. Available from: URL: http://www.nastad.org/Docs/Public/Publication/2006110_2004121NativeAmericansandHVADSKeyssuesandRecommendationsf.pdf
17. University of Oklahoma Health Sciences Center. Native Americans and HIV/AIDS. Oklahoma City, OK; March 2000 [cited 2005 Feb 15]. Available from: URL: http://ccan.ouhsc.edu/AIDS.pdf
18. Rowell R, Bouey P. What are American Indian/Alaska Natives' HIV prevention needs? San Francisco: University of California at San Francisco, Center for AIDS Prevention Studies, AIDS Research Institute [cited 2006 Dec 27]. Available from: URL: http://www.caps.ucsf.edu/pubs/FS/nativeamerican.php
19. Hurlich MG, Hopkins SG, Sakuma J, Conway GA. Racial ascertainment of AI/AN persons with AIDS—Seattle/King County, WA 1980–1989. IHS Primary Care Provider 1992;17:73-4.
20. Lieb LE, Conway GA, Hedderman M, Yao J, Kerndt PR. Racial misclassification of American Indians with AIDS in Los Angeles County. J Acquir Immune Defic Syndr 1992;5:1137-41.
21. Smyser M, Helgerson SD, Hess M. Racial misclassification among AI/AN reported with class IV HIV infection in Washington State. IHS Primary Care Provider 1992;17:74-5.
22. Indian Health Care. Washington: US Congress, Office of Technology Assessment; OTA-H-290, US Government Printing Office; April 1986.
23. Kelly JJ, Chu SY, Diaz T, Leary LS, Buehler JW. Race/ethnicity misclassification of persons reported with AIDS. The AIDS Mortality Project Group and The Supplement to HIV/AIDS Surveillance Project Group. Ethn Health 1996;1:87-94.
24. Lee LM, Lehman JS, Bindman AB, Fleming PL. Validation of race/ethnicity and transmission mode in the US HIV/AIDS reporting system. Am J Public Health 2003;93:914-7.
25. Census Bureau (US). Census 2000 brief: the American Indian and Alaska Native population: 2000. 2002 [cited 2005 Feb 16]. Available from: URL: http://www.census.gov/prod/2002pubs/c2kbr01-15.pdf
26. Ascential Software Corporation: The integrity data re-engineering environment: version 4.0. Westborough (MA): Ascential Software Corporation; 2002.
27. Defining SuperMATCH Pre-built Procedures. In: The integrity data re-engineering environment: user guide. Westborough (MA): Ascential Software Corporation; 2000. p. 8-6, 8-7. Also available from: URL: http://www.documentation.com/images/sample_projects/vality.pdf [cited 2006 Dec 27].
28. Office of Management and Budget (US). Revisions to the standards for the classification of federal data on race and ethnicity. Federal Register 62FR58781-58790, October 30, 1997 [cited 2006 Sep 22]. Available from: URL: www.whitehouse.gov/omb/fedreg/1997standards.html
29. National Center for Health Statistics (US). Documentation for bridged-race vintage 2001 (July 1, 2000–July 1, 2001) postcensal population estimates for calculating vital rates [cited 2005 May 1]. Available from: URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datasets/nvss/bridgepop/DocumentationBridgedPosCenV2001.doc
30. SAS Institute, Inc. SAS: Version 8.02 for Windows. Cary (NC): SAS Institute, Inc.; 2003, 2005.
31. Office of Management and Budget (US). Update of statistical area definitions and guidance on their uses; OMB Bulletin No. 05-02. November 2004 [cited 2006 Dec 27]. Available from: URL: http://www.whitehouse.gov/omb/bulletins/fy05/b05-02.html

32. Hahn RA. The state of federal health statistics on racial and ethnic groups. JAMA 1992;267:268-71.
33. Hahn RA. Why race is differentially classified on U.S. birth and infant death certificates: an examination of two hypotheses. Epidemiology 1999;10:108-11.
34. Hahn RA, Stroup DF. Race and ethnicity in public health surveillance: criteria for the scientific use of social categories. Public Health Rep 1994;109:7-15.
35. Hahn RA, Wetterhall SF, Gay GA, Harshbarger DS, Burnett CA, Parrish RG, et al. The recording of demographic information on death certificates: a national survey of funeral directors. Public Health Rep 2002;117:37-43.
36. Kunitz SJ. The history and politics of US health care policy for American Indians and Alaskan Natives. Am J Public Health 1996; 86:1464-73.
37. Passel JS. The growing American Indian population, 1960-1990: beyond demography. In: Sandefur GD, Rindfull RR, Cohen B, editors. Changing numbers, changing needs: American Indian demography and public health. Washington: National Academy Press; 1996. p. 79-102.
38. Sondik EJ, Lucas JW, Madans JH, Smith SS. Race/ethnicity and the 2000 census: implications for public health. Am J Public Health 2000;90:1709-13.
39. Thomas SB. The color line: race matters in the elimination of health disparities. Am J Public Health 2001;91:1046-8.
40. Satter DE. Culturally competent HIV/AIDS prevention for American Indians and Alaska Natives. In: Cultural competence for providing technical assistance, evaluation and training for HIV prevention programs. CRP, Inc.; Washington: CRP, Inc.; 1999.
41. Weaver HN. Through indigenous eyes: Native Americans and the HIV epidemic. Health Soc Work 1999;24:27-34.
42. Landen M. Increasing involvement of tribal epidemiologists in CSTE. Presented at the 2004 CSTE Annual Conference; 2004 Jun 5–9; Boise, ID. p. 24. Also available from: URL: http://www.cste.org/conference/2004FullAgenda.pdf [cited 2006 Dec 27].