Racial Misclassification of American Indians in Oklahoma State Surveillance Data for Sexually Transmitted Diseases

Douglas A. Thoroughman,1,2 Deborah Frederickson,3 H. Dan Cameron,4 Laura K. Shelby,2,5 and James E. Cheek2

The burden of sexually transmitted diseases (STDs) is high in American Indian/Alaska Native (AI/AN) populations. In addition, race is often misclassified in surveillance data. This study examined potential racial misclassification of American Indians in STD surveillance data in Oklahoma. Oklahoma State STD surveillance data for 1995 were matched with the Oklahoma State Indian Health Service Patient Registry to determine the number of AI/AN women who had one of three STDs but were not listed in Oklahoma surveillance data as AI/AN. Accounting for racial misclassification increased the rate of chlamydia for AI/AN women in Oklahoma by 32% (342/100,000 vs. 452/100,000) in the overall population. For gonorrhea, the rate increased by 57% (94/100,000 vs. 148/100,000) and for syphilis by 27% (15/100,000 vs. 19/100,000). Misclassified AI/AN women most often were classified as "White," and the likelihood of misclassification increased with a lower percentage of AI/AN ancestry. These findings indicate that STD rates may be underestimated for AI/AN populations nationwide. Racial misclassification in state surveillance data causes inaccuracies in characterizing the burden of infectious diseases in minorities. Am J Epidemiol 2002;155:1137–41.

Sexually transmitted diseases (STDs) are occurring in epidemic proportions, with approximately 15 million new STD cases diagnosed in the United States each year (1). Although the incidence of some STDs has declined to historic lows (e.g., syphilis), other STDs, such as gonorrhea and chlamydia, have experienced a resurgence or have remained at high levels, especially in younger age groups (2). The burden of these diseases has been particularly high among some US subpopulations, including American Indian/Alaska Native (AI/AN). In 1997, AI/AN persons had higher rates of chlamydia than the general US population: 485.0/100,000 versus 196.8/100,000; although rates among the AI/AN population were slightly lower than those in the general population for gonorrhea (99.7/100,000 vs. 123.6/100,000) and syphilis (2.0/100,000 vs. 3.2/100,000), AI/AN persons had the second highest rates when compared with all other racial groups (3).

Unfortunately, racial misclassification in STD surveillance data may contribute to a substantial underestimate of the number of reported STD cases in AI/AN populations. Rates for several public health outcomes have been underestimated in AI/AN populations because of racial misclassification on medical records, in documents, and in data registering systems. Computer linkage of large data sets has illuminated the extent of this problem in AI/AN populations. As early as the 1960s, linkage of birth and death certificates made evident that infant mortality was underestimated for AI/AN populations (4). This problem has consistently been shown in studies conducted since then (5–9).

Studies linking the Indian Health Service (IHS) Patient Registry to other public health data sources have revealed similar racial misclassification of AI/AN persons. On death certificates, race was misclassified for 12.8 percent of American Indians in Washington State between 1985 and 1990 (10). Linking the Puget Sound Surveillance, Epidemiology, and End Results (SEER) database for 1974–1989 to the IHS Patient Registry showed that only 60 percent of AI/AN patients with invasive cancer and 69 percent of patients with in situ cervical cancer were listed as AI/AN in this database (11). Annual injury rates for AI/AN patients listed in the Oregon Injury Registry increased by 68 percent when adjusted for racial misclassification (12), and rates of end-stage renal disease increased from 267.5/million to 311.6/million when Northwest Renal Network data were linked to the IHS Patient Registry for 1988–1990 (13). A recent report of racial misclassification on death certificates based on US national vital statistics was consistent
with these findings for American Indians and Hispanics (14). Percentage of Indian ancestry, often called “blood quantum,” has previously been associated with racial misclassification (10, 11). No previous studies of racial misclassification in routinely reported infectious disease monitoring systems have been found in the current public health literature.

The purpose of this study was to assess the extent of racial misclassification of American Indians in STD surveillance data. Oklahoma was chosen as the site for this examination because of the widespread dispersal of the AI/AN population throughout the state. In addition, the AI/AN population in Oklahoma (=280,000) constitutes approximately 8 percent of the state’s population, the highest number in any US state. It was hypothesized that the rates of chlamydia, gonorrhea, and syphilis would increase substantially when we accounted for racial misclassification.

MATERIALS AND METHODS

To determine the total number of AI/AN persons in Oklahoma diagnosed with chlamydia, gonorrhea, or syphilis, two data sets were matched: the Oklahoma State registry of STD patients and the IHS Patient Registry. The Oklahoma STD Patient Registry lists all persons who have tested positive for one of several STDs in the state of Oklahoma. Oklahoma conducts a two-stage surveillance system for STDs. All laboratories (state, tribal, and private) that perform STD testing for facilities in Oklahoma are required to submit to the state health department any positive test results for all patients tested. Providers are asked to supply a state report form on any patient tested for STDs. Laboratory reports that arrive at the State Health Department but have no corresponding provider report are followed up with a letter to the provider and phone calls until the report is submitted; thus, nearly 100 percent case ascertainment and collection of pertinent variables for STD patients seen in health care facilities are achieved. All STD patient information is entered into a central database and is catalogued in yearly data sets. Race is listed on the provider report form and could be self-identified or based on a judgment by clinic personnel. However, the way in which a person’s race is determined is not recorded in the STD surveillance data.

The Oklahoma STD database used for this study consisted of three separate data sets, which included all persons diagnosed during calendar year 1995 with chlamydia, gonorrhea, and syphilis, respectively, regardless of gender. Each record of these data sets contained approximately 20 variables (e.g., name, date of birth, race, sex, age, visit location, and provider). The chlamydia surveillance data set recorded 4,829 persons diagnosed with chlamydia, the gonorrhea data set included 4,605 cases, and the syphilis data set had 603 cases.

The IHS Patient Registry lists all patients who have accessed IHS health care in Oklahoma in any form in the past. To be eligible to receive care at an IHS facility, a person must be an enrolled member of a federally recognized AI/AN tribe or the spouse of an AI/AN receiving obstetric care (but spouses are not considered American Indian and are listed as non-American Indian in the data). This data set is historical and may include multiple entries for a given person if he or she attended more than one IHS or tribal health facility. However, IHS personnel eliminated overt duplication of entries before the data were used in this study. The IHS Patient Registry was current through September 1996 and contained 492,804 records.

The two data sets were matched by using a staged process of automated matching and subsequent automated and hand elimination of inappropriate matches. Computer matching was accomplished by using Automatch software (version 3.0; originally manufactured by MatchWare Technologies, Inc., Burtonsville, Maryland (1995)), which uses a probabilistic model in its matching procedure. For this study, six matching runs were executed between each of the three STD data sets and the IHS Patient Registry. A 90 percent probability level for accurate matches was chosen for matching on all variables. Only last name, first name, middle initial, full date of birth, year of birth, month of birth, day of birth, sex, and age were available variables for matching purposes.

After the Automatch procedure was completed, appropriate and inappropriate matches were identified. Matches were considered definite or inappropriate on the basis of several characteristics (e.g., name similar but birth date differing for month and day vs. name similar and birth date matching exactly). Matches that were not eliminated or considered definite at this stage were examined by hand and were included as definite if differences were clearly the result of simple data entry or recording errors. Other questionable matches were examined by a blinded panel of three researchers who primarily considered uniqueness of name; those on which all three researchers agreed were true matches were included as definite.

All rates reported in this paper are for women only. Rates were age adjusted by the direct method using the 1990 US Census estimates for women to make rates comparable for the different diseases. Denominators used to calculate rates were based on 1995 census estimates of the female AI/AN population in Oklahoma or the IHS female Oklahoma user population, depending on the numerator derivation. This paper will note when these different denominator estimates were used. The sensitivity of identifying AI/AN women in the Oklahoma STD surveillance data was calculated for each STD by using data for women who had ever received health care from the IHS as the “gold standard.”

RESULTS

Chlamydia

Oklahoma State had a record of 4,250 women who had been diagnosed with chlamydia at least once during calendar year 1995. Of these women, 562 (13.2 percent) had been recorded as American Indian in the Oklahoma data, resulting in a rate of 342/100,000 when the 1995 IHS estimate of the age distribution of the AI/AN female population in Oklahoma was used as the denominator for the crude rate in each age group and rates were standardized to the 1990 US Census female population. (In this paper, 1995 IHS estimates refer to those based on yearly extrapolations from the
After data were matched to the IHS Patient Registry, 742 women (an additional 180 matches) were identified as AI/AN in the Oklahoma data, corresponding to a rate of 452/100,000, an increase of 32 percent in the reported chlamydia rate in the female AI/AN population of Oklahoma. These numbers represent all women identified as AI/AN in Oklahoma STD surveillance data and those identified as AI/AN through matching with the IHS Patient Registry.

When we considered only those women who had accessed IHS health care services (the IHS user population, which excludes non-tribally registered AI/AN persons), the number of women identified as AI/AN in the Oklahoma State STD data was 409 (9.6 percent of 4,250). Matching with the IHS Patient Registry increased this number to 589. The corresponding rates of chlamydia (using the 1995 IHS user population estimate as a denominator for the crude rate in each age group and standardizing rates to the 1990 US Census female population) were 234/100,000 and 336/100,000, respectively, which represents an increase of 44 percent between the Oklahoma-based rate and the matched rate. Thus, the sensitivity of identifying AI/AN women (based on use of IHS health care services) was 69 percent (409/589).

Gonorrhea

A total of 2,500 Oklahoma women were diagnosed as having gonorrhea at least once in 1995; of these, 153 (6.1 percent) were listed as AI/AN. The corresponding rate of gonorrhea diagnosis for AI/AN women was 94/100,000 when the 1995 IHS estimate of the age distribution of the AI/AN female population in Oklahoma was used as the denominator for the crude rate in each age group and rates were standardized to the 1990 US Census female population. When data were matched with the IHS Patient Registry, an additional 85 women were included, resulting in a total of 238 AI/AN gonorrhea diagnoses and increasing the rate by 57 percent to 148/100,000.

When we limited case patients to IHS users only, 108 (4.3 percent of 2,500) AI/AN IHS users were identified as AI/AN in the Oklahoma syphilis data, resulting in a rate of 62/100,000 when the 1995 IHS user population estimate was used as a denominator for the crude rate in each age group and rates were standardized to the 1990 US Census female population. Matching increased the number of case patients to 22 and the rate to 151/100,000, a rate increase of 36 percent. Regarding syphilis surveillance, 73 percent (16/22) of women were identified correctly as AI/AN in the Oklahoma STD data on the basis of their use of IHS health care services.

Syphilis

Oklahoma reported 306 female syphilis patients in 1995. Of these patients, 23 (7.5 percent) were recorded as AI/AN, resulting in a rate of 15/100,000 when the 1995 IHS estimate of the age distribution of the AI/AN female population in Oklahoma was used as the denominator for the crude rate in each age group and rates were standardized to the 1990 US Census female population. An additional six case patients were identified through matching with the IHS Patient Registry, increasing the total to 29 and the rate to 19/100,000, a 27 percent increase.

Regarding the IHS user population, only 16 (5.2 percent of 306) IHS users were identified as AI/AN in the Oklahoma syphilis data, resulting in a rate of 11/100,000 when the 1995 IHS user population estimate was used as a denominator for the crude rate in each age group and rates were standardized to the 1990 US Census female population. Matching increased the number of case patients to 22 and the rate to 15/100,000, a rate increase of 36 percent. Regarding syphilis surveillance, 73 percent (16/22) of women were identified correctly as AI/AN in the Oklahoma STD data on the basis of their use of IHS health care services.

Misclassification

Among those STD patients identified as IHS users but listed as a race other than AI/AN in the Oklahoma data (i.e., those who were misclassified), most were classified as “White.” For all three STDs, “White” was the primary misclassification; “Black” was the second most prominent misclassification; and “Hispanic,” “Asian,” and “Other” or “Unknown” were represented only slightly (figure 1).

Likelihood of misclassification

For each STD, we examined the relation between racial misclassification in the Oklahoma surveillance data and the percentage of AI/AN heritage, or blood quantum (i.e., <25
percent, 25–49 percent, 50–99 percent, and 100 percent AI/AN ancestry). The lower the percentage of Indian ancestry, the greater the likelihood of racial misclassification in the Oklahoma data (figure 2). The chi-square test for trend was significant for chlamydia ($\chi^2 = 38.2, p < 0.001$) and gonorrhea ($\chi^2 = 8.4, p = 0.004$) but did not approach significance for syphilis; results for syphilis were unstable because of the limited number of syphilis cases among AI/AN women ($n = 22$).

**DISCUSSION**

This study indicates that racial misclassification of AI/AN populations in 1995 Oklahoma State STD surveillance data resulted in large underestimates of the burden of chlamydia, gonorrhea, and syphilis in the female AI/AN population of this state. When misclassification occurred, race was most likely to be misclassified as “White” or “Black.” Misclassification of AI/AN persons was more likely to occur inversely to the percentage of AI/AN ancestry of the case patient.

It is important to note that rates may be impacted by possible underreporting of race in the denominator. Studies have shown serious underreporting of minority races in population counts and population estimates (14). This underreporting in the denominator could negate, to some degree, the quantitative impact of underreporting in the numerator when rates are calculated. However, even when racial misclassification was accounted for in the denominator, mortality rates for American Indians were still underestimated by 21 percent in the study cited (14). Thus, the impact of denominator misclassification would not be expected to eradicate the effects of misclassification in the numerator, especially in light of the extreme rate increases seen in our study. In addition, the methods relied on in this study do not account for racial misclassification of AI/AN persons not in the IHS patient system, which could further increase the rate disparities we found.

The effect of recategorizing race because of AI/AN misclassification would, by definition, lower the estimated rates of disease for the other races for which AI/AN persons were originally classified. This adjustment would not be expected to change the rates of disease in these other racial categories substantially, because the actual number of cases among AI/AN persons is small (180 for chlamydia, 85 for gonorrhea, and six for syphilis) with respect to the population of cases of other races.

The results of this study must be generalized with caution. The AI/AN population of Oklahoma has historically been intermixed with non-AI/AN groups and not separated on reservations, as is true with some other AI/AN populations. Less racial misclassification would be expected in states or areas in which the AI/AN population is more grouped and less likely to attend non-IHS health care facilities for STD services. Those who attend IHS facilities would automatically be classified as AI/AN, so misclassification cannot occur easily in these settings. However, in areas with a similar lack of boundaries for AI/AN groups (e.g., California, parts of the northeastern United States, and urban areas), racial misclassification may occur in proportions similar to those observed in these data. In 1990, over half (59 percent) of American Indians and Alaska Natives lived in urban areas (15) in which conditions were similar to those described here.

The implication of these findings is that traditional estimates of the burden of STDs in the AI/AN population probably underestimate the true burden. Although racial misclassification may be limited in reservation areas, many regions have similar mixing of AI/AN and non-AI/AN persons, increasing the likelihood of misclassification. If misclassification is occurring at the levels found in this study, the resources available for STD control, already limited for AI/AN populations, may be further lacking and in need of rapid increases to stem the tide of this epidemic. A racial disparity exists in the burden of STDs in the United States, and racial misclassification illuminates the depth of this disparity. The long-term results of chlamydia exposure can lead to several deleterious public health outcomes, all of which can be prevented by early diagnosis and treatment. The availability of resources to facilitate early diagnosis and treatment depends on accurate estimates of the burden of disease in the community.

The implications for state-collected data on reportable diseases are clear for AI/AN populations: if racial information is not collected in a situation where it can be confirmed (i.e., in an IHS facility), there is a fairly high probability of misclassification for AI/AN patients. For STDs, state-reported rates are underestimated, which could be true for many or most other states and reportable infectious diseases. Compared with other states, Oklahoma has an effective STD surveillance system; thus, similar racial misclassification issues would be expected to arise in other states were their surveillance systems investigated similarly. Race has always been a difficult demographic variable to accurately collect in surveillance data; unfortunately, racial misclassification is reflected in national estimates of STDs summarized and reported by the Centers for Disease Control and Prevention. Without a standard method of adjustment or more accurate data collection techniques, these discrepancies will continue.
This study is limited in that tribal rolls were not accessed to more accurately characterize the extent of racial misclassification in the Oklahoma data. Such an assessment was not possible given the number of tribes in Oklahoma. However, if tribal rolls had been accessed, the estimate of misclassification would have only increased because more persons identified as a race other than AI/AN could have been found to be AI/AN. Additionally, Social Security numbers were not available in the Oklahoma State data, making accurate matching more difficult. In either case, the results of this study should stand as probable minimum estimates of racial misclassification of AI/AN female STD patients in Oklahoma State.

In conclusion, racial misclassification of AI/AN persons in 1995 Oklahoma State STD surveillance data resulted in large underestimates of the burden of chlamydia, gonorrhea, and syphilis in the female AI/AN population of this state. American Indian women were most likely to be misclassified as “White” or “Black,” and misclassification was more likely to occur for those with a lower percentage of AI/AN ancestry. This is the first known study to examine racial misclassification of infectious disease surveillance data for American Indians. On the basis of this and several other studies, caution should be used in interpreting all disease rates (infectious and chronic) for American Indians; rates have consistently been underestimated because of racial misclassification.

REFERENCES