Record Linkage of California Case-based Sexually Transmitted Disease Surveillance Data and Indian Health Service Race Classification Data

Purpose

National statistics about sexually transmitted diseases (STDs) show disparities between American Indians and Alaska Natives (AIAN) and the rest of the U.S. population. In California, AIAN are often reported as having a lower STD burden than AIAN from other geographic regions of the United States.[1] However, race misclassification can contribute to the underestimation of AIAN STD cases. Specifically, previous research studies have found that national AIAN health surveillance data can have high race misclassification rates when there is a geographic co-ethnic concentration of other races with AIAN and a large number of AIAN living off reservation and/or in urban areas.[2,3,4] Given that California has a higher proportion of other races to AIAN and that a large number of California AIAN access health care services at off-reservation, non-Indian Health Service (IHS) providers that do not require evidence of tribal lineage to receive health care, it is anticipated that many California AIAN are incorrectly classified as non-AIAN.

The purpose of this study was to assess California AIAN race misclassification by linking case-based STD surveillance data from the California Department of Public Health (CDPH) with IHS National Patient Information Reporting System (NPIRS) data. This study also sought to describe characteristics of misclassified versus correctly classified California AIAN to enhance the interpretability of existing AIAN case-based STD surveillance data.

Linkage Procedure

Patient identifiers (e.g., first and last name, date of birth, and sex) were used to match 2007-2012 CDPH case-based chlamydia, gonorrhea, and primary and secondary (P&S) syphilis surveillance data with NPIRS data from AIAN who accessed a California Indian health clinic at least once between 1984 and 2013. AIAN included in the IHS NPIRS registry were California residents with proof of Indian descent or enrollment in a federally recognized tribe as well as a member of the Indian community served by a local tribal health facility. Registrants also included AIAN from non-federally recognized tribes who were descendents of AIAN living in California on June 1, 1982; who were descendents of the California judgment rolls; or who held trust land in California.

---

[1] Centers for Disease Control and Prevention and Indian Health Service. Indian Health Surveillance Report – Sexually Transmitted Diseases 2011, Atlanta, GA. http://www.cdc.gov/std/stats/ihs/default.htm. Accessed November 17, 2015.

[2] Bertolli J, Lee LM, Sullivan PS. 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. *Public Health Rep.* 2007;122:382-92.

[3] Jim MA, Arias E, Seneca, DS, Hoopes MJ, Jim CC, Johnson NJ, Wiggins DC. Racial Misclassification of American Indians and Alaska Natives by Indian Health Service Contract Health Service Delivery Area. *Am J Public Health.* 2014;104(S3):S295-302.

[4] Thoroughman DA, Fredersickson D, Cameron HD, Shelby LK, Cheek JE. Racial misclassification of American Indians in Oklahoma State surveillance data for sexually transmitted diseases. *Am J Epidemiology.* 2002;155:1137-41.

Epidemiologists at IHS worked with representatives from the California Tribal Epidemiology Center (CTEC) and the CDPH STD Control Branch to obtain institutional review board approvals and secure and link 2007-2012 CDPH case-based STD surveillance data with 1984-2013 IHS NPIRS race classification data. After the data linkage, IHS epidemiologists provided a de-identified, delimited data file to CTEC staff for subsequent data analysis. The de-identified, delimited linked data file included CDPH data, IHS NPIRS data, or *both* CDPH and IHS NPIRS data from 7935 AIAN.

Research Questions and Findings

How many AIAN STD cases were initially identified in the 2007-2012 CDPH case-based surveillance data?

A total of 3374 AIAN STD cases were identified in the 2007-2012 CDPH case-based STD surveillance data. These cases are considered pre-linkage AIAN STD cases, meaning that they have not been adjusted for race misclassification. See Table 1.

Table 1. Pre-Linkage CDPH AIAN STD Cases

Chlamydia Cases	Gonorrhea Cases	P&S Syphilis Cases	All STD Cases
2807	520	47	3374

Based on the linkage with 1984-2013 IHS NPIRS data, what were the overall race classification and misclassification rates of AIAN in the 2007-2012 CDPH case-based STD surveillance data?

- The de-identified, delimited linked data file used for the linkage included CDPH data, IHS NPIRS data, or *both* CDPH and IHS NPIRS data from **7935 AIAN.**

- **1207 of 7935 AIAN** (15.2%) were classified as AIAN in both CDPH and IHS NPIRS data, confirming AIAN race classification in both data sources.

- **4561 of 7935 AIAN** (57.5%) were classified as non-AIAN in the CDPH data and AIAN in the IHS NPIRS data, *meaning that 57.5% of AIAN were misclassified in the CDPH data.*

- **2167 of 7935 AIAN** (27.3%) were classified as AIAN by CDPH but did not have matching IHS NPIRS records to confirm AIAN race classification status.

Based on the linkage with 1984-2013 IHS NPIRS data, what were the race misclassification rates of AIAN in the 2007-2012 CDPH case-based STD surveillance data by type of STD?

CDPH AIAN race misclassification rates ranged from approximately 45% to 58%, depending on type of STD. **Misclassification rates were 58.1% for chlamydia**, **54.8% for gonorrhea, and 44.7% for P&S syphilis**. See Table 2.

Table 2. Race Classification and Misclassification Rates by Type of STD

Classification Status	Chlamydia Cases (n=6700)	Gonorrhea Cases (n=1150)	P&S Syphilis Cases (n=85)	All STD Cases (n=7935)
Classified as AIAN in both CDPH and IHS NPIRS data	15.4% (n = 1032)	14.7% (n=169)	7.1% (n=6)	15.2% (n=1207)
Misclassified as non-AIAN in CDPH data	58.1% (n=3893)	54.8% (n=630)	44.7% (n=38)	57.5% (n=4561)
Classified as AIAN in CDPH data but no IHS NPIRS record to confirm classification	26.5% (n=1775)	30.5% (n=351)	48.2% (n=41)	27.3% (n=2167)

What race(s) were most frequently reported for AIAN who were misclassified as non-AIAN in the 2007-2012 CDPH case-based STD surveillance data?

In general, AIAN who were misclassified as non-AIAN in the CDPH data were **most frequently misclassified as Unknown, Hispanic, or White.** However, AIAN who were misclassified as non-AIAN in the CDPH data and who had P&S syphilis were most frequently misclassified as White, Hispanic, or Black. See Table 3.

Table 3. Race Misclassification Rates by Type of STD

Race Misclassified as in CDPH Data	Chlamydia Cases (n=3893)	Gonorrhea Cases (n=630)	P&S Syphilis Cases (n=38)	All STD Cases (n=4561)
Asian	1.4% (n=55)	1.3% (n=8)	2.6% (n=1)	1.4% (n=64)
Black	5.4% (n=212)	7.5% (n=47)	13.2% (n=5)	5.8% (n=264)
Hispanic	25.4% (n=988)	26.0% (n=164)	34.2% (n=13)	25.5% (n=1165)
Multi	0.3% (n=13)	0.2% (n=1)	5.3% (n=2)	0.4% (n=16)
Other	2.2% (n=85)	2.2% (n=14)	0.0% (n=0)	2.2% (n=99)
Unknown	40.5% (n=1577)	40.0% (n=252)	7.9% (n=3)	40.2% (n=1832)
White	24.7% (n=963)	22.9% (n=144)	36.8% (n=14)	24.6% (n=1121)

What is the calculated adjustment factor to correct for AIAN race misclassification in 2007-2012 CDPH case-based STD surveillance data?

The overall calculated adjustment factor for AIAN race misclassification in the 2007-2012 CDPH case-based STD surveillance data is **2.35**. The STD-specific adjustment factors are **2.39 for chlamydia**, **2.21 for gonorrhea, and 1.81 for P&S syphilis**.[5]

Using the calculated adjustment factor to correct for AIAN race misclassification, how many additional AIAN STD cases were identified in the 2007-2012 CDPH case-based STD surveillance data?

Applying the adjustment factors for race misclassification to the number of pre-linkage CDPH AIAN STD cases listed in Table 1 led to the identification of **4555 additional STD cases than were originally identified** (135.0% increase from 3374 to 7929 cases), including 3902 additional chlamydia cases (139.0% increase from 2807 to 6709 cases), 629 additional gonorrhea cases (121.0% increase from 520 to 1149 cases), and 38 additional P&S cases (80.9% increase from 47 to 85 cases).[6] See Table 4.

Table 4. Adjusted CDPH AIAN STD Cases

Chlamydia Cases	Gonorrhea Cases	P&S Syphilis Cases	All STD Cases
6709	1149	85	7929

What characteristics were associated with AIAN race misclassification in the 2007-2012 CDPH case-based STD surveillance data?

Results of chi-square analyses indicated that AIAN residing in rural areas were misclassified at significantly lower rates than AIAN residing in urban areas (*chi-square value* = 300.25, *p-value* < .001).[7] Neither gender nor age at time of STD diagnosis was significantly associated with AIAN race misclassification in the 2007-2012 CDPH case-based STD surveillance data.

Conclusion

This study found that race misclassification contributes to the underestimation of AIAN STD cases in the 2007-2012 CDPH case-based STD surveillance data. Specifically, results of this record linkage indicate that AIAN race misclassification rates are nearly 60% in the 2007-2012 CDPH case-based STD surveillance data, and AIAN residing in

---

[5] Calculated adjustment factors "are the ratio of the actual number of AIAN deaths revealed on the linked IHS NPIRS data between 1986-1988 to the number of AIAN deaths reported on state death certificates for that time period. This methodology assumes that the rate of under-reporting AIAN deaths has remained relatively constant before and after that 3-year period." *Adjusting for the Miscoding of of Indian Race on State Death Certificates*. Hendler A, et al., Division of Program Statistics (Office of Planning, Evaluation, and Legislation), Indian Health Service; 1996:1-36.
[6] Increase in percentage was calculated with this equation: (Table 4 adjusted AIAN STD cases – Table 1 pre-linkage AIAN STD cases)/(Table 1 pre-linkage AIAN STD cases). Example: (7937 – 3374)/(3374) = 135.2%.
[7] Office of Management and Budget. 2010 Standards for Delineating Metropolitan and Micropolitan Statistical Areas. *Federal Register*. 2010,75(123):37246-52.

urban areas are more likely to be misclassified as non-AIAN than AIAN residing in rural areas.

Findings from this study can be used to inform research, practice, and policy. First, public health researchers and practitioners should consider using adjusted CDPH case-based chlamydia, gonorrhea, and P&S syphilis rates to more accurately portray the STD burden of the California AIAN population. Second, CDPH administrators should develop and enforce best practices for obtaining accurate race information from patients without increasing the overall burden of patient registration or screening. For example, CDPH administrators could make structural changes to electronic health records so that race is routinely collected and reported as part of electronic case reporting. Finally, to meet the increased public health need revealed by the underestimation of AIAN STD cases in the CDPH case-based STD surveillance data, national and state policymakers should allocate additional funding for STD health disparity awareness campaigns, sexual health education initiatives, and STD prevention and treatment services for the California AIAN population.

Acknowledgements