

Health Policy Brief

November 2020

Limited Access to Health Data on American Indian and Alaska Natives Impedes Population Health Insights

Tara L. Becker, Susan H. Babey, Riti Shimkhada, AJ Scheitler, and Ninez A. Ponce

“The lack of reporting on AIAN populations and their health issues challenges policymakers charged with addressing disparities.”

SUMMARY: American Indian and Alaska Native (AIAN) people are underrepresented and often invisible in public health data and research. AIAN health data capacity is impeded by the quality of information collected, released, and reported on AIANs in population-based surveys. AIANs are either put in a residual “other” category or, typically, depicted as single-race non-Latinx AIANs. Though the figure varies substantially across federal surveys, fewer than two in five AIANs report as single-race non-Latinx AIANs. Drawing population inferences based on single-race AIANs also fails to capture the considerable segment of the AIAN population that comprises individuals of more than one race.

To promote more accurate insights on the health of the AIAN population, we examined eight population datasets that focus on health, health status, health behaviors, and health access. For each survey, we assessed whether it is possible to identify AIAN respondents from (1) readily accessible public-use data files and (2) restricted data files. We also assessed the extent to which data users can distinguish three major subgroups within the AIAN race category from these data files: single-race non-Latinx AIAN, single-race Latinx AIAN, and AIAN of more than one race. In this policy brief, we discuss the implications of the limited accessibility of AIAN information in population datasets and provide some recommendations that may help improve the availability of AIAN health information.

Improving the health of an estimated 5.7 million American Indians and Alaska Natives (AIANs) in the nation¹ is hampered by population data systems that hide or obscure their representation in federal data systems. In most national, population-based survey data, the number of AIAN respondents interviewed is too small to allow the inclusion of race/ethnicity information that reports AIAN as a separate category in the publicly available data files. Doing so presents a disclosure risk, because AIAN respondents could potentially be identified by data users. The public and survey respondents expect privacy, and survey managers ensure this commitment to guard against disclosure breaches, which could lead

to community distrust. Although protecting respondents is of primary importance for survey managers, one result of this limitation is that conducting research that focuses on AIAN populations is often not possible using publicly available data. The lack of reporting on these populations and their health issues challenges policymakers and program developers charged with addressing disparities.

Individuals who identify as AIAN are likely to also self-identify as another race and are more likely to identify with Latinx ethnicity. (In this report, “Latinx” is used to denote those who are Latino/Latina as well as those who are Hispanic or of Spanish-

“In public-use data files, often only single-race non-Latinx AIAN are classified as AIAN, substantially reducing the sample size of the population.”

speaking origin.) The most common race-classification systems in public health and most other population-based data combine those who report more than one race into a single category, which makes it impossible to identify AIANs who report being two or more races. In addition, most public health datasets contain at least one measure that combines Latinx ethnicity with the five major Office of Management and Budget race categories: non-Latinx American Indian or Alaska Native, non-Latinx Native Hawaiian/Pacific Islander, non-Latinx Asian, non-Latinx black or non-Latinx African American, and non-Latinx white. This means that when AIAN respondents can be identified in public-use data files, often only single-race non-Latinx AIAN are classified as AIAN, substantially reducing the sample size of the AIAN population.

This policy brief describes the accessibility of information on AIAN in each of eight population health surveys (see box, this page). For each survey, we assessed whether it is possible to identify AIAN respondents in (1) public-use data files, which are readily

accessible, and (2) restricted-use data files. We also assessed the extent to which each data file allows users to distinguish three important subgroups within the AIAN race category: single-race non-Latinx AIAN, single-race Latinx AIAN, and AIAN of more than one race. Exhibit 1 provides information about the eight surveys evaluated.

SURVEY ACRONYMS

BRFSS Behavioral Risk Factor Surveillance System

CHIS California Health Interview Survey

MCBS Medicare Current Beneficiary Survey

NHANES National Health and Nutrition Examination Survey

NHIS National Health Interview Survey

NSCH National Survey of Children’s Health

NSDUH National Survey on Drug Use and Health (before 2002: National Household Survey of Drug Abuse)

PATH Population Assessment of Tobacco and Health

Exhibit 1

Description of Eight Population Health Surveys

Survey	Time Period of Survey	Survey Population	Primary Source for...	Longitudinal?
BRFSS	Began 1984 in 15 states; nationwide since 1993	Noninstitutionalized civilian adults in the U.S. and territories	State-level population health information	No
CHIS	Biennial 2001–2009; continuously since 2011	Noninstitutionalized civilian adults, adolescents, and children in California	Population health information for the state of California	No
MCBS	Continuously since 1991	Medicare enrollees	Medicare cost and utilization information	Yes
NHANES	Began 1960s; continuously since 1999	Noninstitutionalized civilian residents of the U.S.	Physical examination data	No
NHIS	Continuously since 1957	Noninstitutionalized civilian residents of the U.S.	National population health information	No*
NSCH	2003, 2007, 2011–2012; annually since 2016	Noninstitutionalized children under age 18 living in the U.S.	National and state-level population health information on children	No
NSDUH	Began 1979; annually since 1990; all 50 states + D.C. since 1999	Noninstitutionalized civilian residents of the U.S. ages 12 and over	Detailed substance use information	No
PATH	Began 2013	Noninstitutionalized civilian residents of the U.S. ages 12 and over	Detailed tobacco use and cessation information	Yes

* Subsample of respondents followed longitudinally through the Medical Expenditures Panel Survey

Accessibility of AIAN Subgroup Information in Public-Use Data Files by Survey

Exhibit 2

Survey	Available in Public-Use Files (PUF)?				Where to Access Restricted Data if Not in PUF?
	All AIAN	Single-race Non-Latinx AIAN	Single-race Latinx AIAN	AIAN More Than One Race	
CHIS	Yes*	Yes*	Yes*	Yes*	CHIS Data Access Center
BRFSS	2001–2012	from 2001	from 2001	2001–2012**	FSRDC, NCHS RDC
NHIS	No	from 1997	from 1997	No**	FSRDC, NCHS RDC
NSDUH†	No	Yes	No	No	FSRDC, NCHS RDC
NSCH	No	No‡	No‡	No	FSRDC, NCHS RDC
MCBS	No	No	No	No	Limited dataset obtained from CMS with data use agreement
NHANES	No	No	No	No	FSRDC, NCHS RDC
PATH	No	No	No	No	ICPSR Virtual Data Enclave

* Available for all age groups in two-year combined public-use files, not in single-year public-use files

** Respondents who reported more than one race were asked for a single main (NHIS) or preferred race (BRFSS), and those who selected AIAN as a main/preferred race can be identified in the PUF data through 2019.

† Public-use data files are available only from 2002.

‡ It is possible to identify single-race AIAN (both non-Latinx and Latinx) who live in the subset of states in which AIANs make up 5% or more of the child population, but not all single-race AIAN.

Note: FSRDC = Federal Statistical Research Data Center
NCHS RDC = National Center for Health Statistics Research Data Center
CMS = Centers for Medicare and Medicaid Services
ICPSR = Inter-university Consortium for Political and Social Research

We selected these surveys based on three primary criteria: (1) how often the survey is used in published research studies, (2) the uniqueness of the information contained in the survey, and (3) the collection of detailed racial/ethnic information that allows the identification of Latinx AIANs and AIANs who report more than one race.

Using these criteria, BRFSS and NHIS were selected based on their extensive use by public health researchers. The remaining six surveys were selected based on the uniqueness of their content or their focus on a specific subpopulation. MCBS is the primary source of information on Americans enrolled in Medicare. The California Health Interview Survey (CHIS) is the largest continuously collected state health survey, and California contains a sizable AIAN population. NSCH is the largest public health survey targeted specifically to children. NHANES is unique for its collection of both survey and biological data. NSDUH contains detailed population health data on drug use, and PATH contains detailed population health data on tobacco

use and cessation efforts. Each of these surveys evaluates collected race information in a way that allows researchers to identify respondents of more than one race, and each collects Latinx ancestry information separately from race. Both of these steps are necessary to allow identification of the total AIAN population and the three AIAN subgroups.

AIAN Information in Population Health Surveys

AIAN Information in Public-Use Data Files

Only two of the eight surveys—CHIS and BRFSS—allow users to identify all four AIAN groups in public-use data files in any time period (Exhibit 2). The remaining datasets present limited or no identifiable information on AIANs in the public-use data files.

- In CHIS, it is possible to identify all three AIAN race subgroups (single-race non-Latinx AIAN, single-race Latinx AIAN, and AIAN reporting more than one race) for all ages by using the public-use data file that combines two years of data.

“Only two of the eight surveys allow users to identify all four AIAN groups in public-use data files.”

“Restricted-use data files contain more detailed race information and allow researchers to identify each of the three AIAN subgroups.”

- Before 2013, BRFSS included in public-use files detailed information on all races reported by the respondent, as well as a separate indicator for Latinx ancestry. Beginning with the 2013 data, this detailed information is no longer available; single-race AIAN respondents can be identified. However, BRFSS includes a “preferred race” measure, in which respondents who report more than one race choose which race they most identify with. Using this measure, it is possible to identify a subset of AIAN respondents of more than one race who most identify as AIAN.
- Beginning with the 1997 data, NHIS public-use data files have included a race-only measure that identifies single-race AIAN respondents, but combines those who report AIAN and another race with other respondents who report two or more races.¹ When combined with the indicator for Latinx ancestry, NHIS data identify both single-race non-Latinx AIAN and single-race Latinx AIAN. Though it is not possible to identify all respondents who are more than one race, up until 2018, NHIS asked those reporting more than one race to identify which one best represented their race; those who reported AIAN as their main race can thus be identified in the public-use data for each year through 2018.
- In the NSDUH public-use data files, only single-race non-Latinx AIAN respondents can be identified, because all Latinx AIAN are classified as Latinx, while AIANs reporting more than one race are classified with all other respondents who report two or more races.
- In NSCH public-use data files, it is possible to identify single-race AIAN (Latinx and non-Latinx) only within a subset of U.S. states in which at least 5% of the population is AIAN.
- In MCBS, NHANES, and PATH, AIANs cannot be identified in the public-use data files.

Access to Restricted-Use Data Files

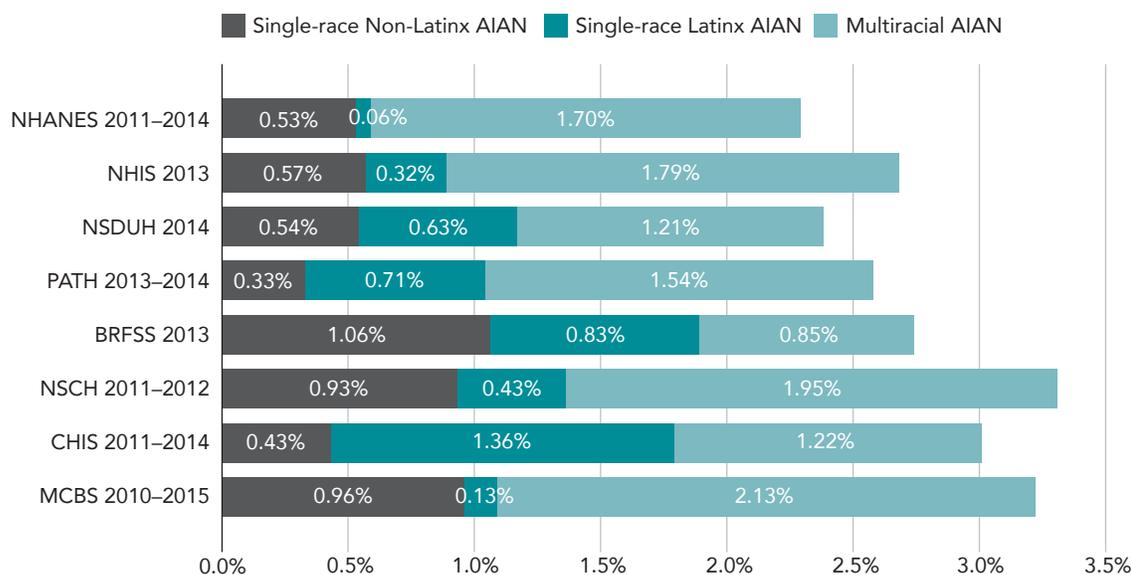
Each of the eight datasets has restricted-use data files that contain more detailed race information and allow researchers to identify each of the three AIAN subgroups (Exhibit 2). Five of these datasets now make their restricted data available through either a federal statistical research data center (FSRDC) or a National Center for Health Statistics research data center (NCHS RDC). Three of these (NHANES, NHIS, and NSCH) have been available this way for several years; access to BRFSS and NSDUH was added more recently.

The FSRDCs are located throughout the country. We accessed the FSRDC located at UCLA. NCHS RDCs open to nonfederal researchers are located in two locations in the greater Washington, D.C., metro area. Direct access using the NCHS RDC is advertised as taking from 8 to 12 weeks. However, acquiring access to an FSRDC can take more than three months, due to NCHS proposal review and federal background checks needed for acquiring Special Sworn Status through the U.S. Census Bureau. Researchers can incur substantial fees and user costs (the costs for our project were \$4,500 for access to four NCHS datasets, and \$15,000 for FSRDC access). The NCHS fees are based on the number of datasets and years requested. The FSRDC fees support the operations of the local FSRDCs, and the cost varies depending on the staffing needs for the project. Once approved, a researcher can conduct analyses in the secure site of the FSRDC, then submit the output to NCHS. After disclosure review by NCHS and the U.S. Census, NCHS emails the output to the researcher.

Three of the restricted datasets we used were accessed through other means. Researchers who would like to use the CHIS restricted data can do so only by submitting a research application to the CHIS DAC and paying

Weighted Percentage of AIANs by AIAN Subgroup and Survey

Exhibit 3



Note: Analyses used weights provided by administrators to account for the design of each survey.

a setup fee of \$500 and a user fee that on average ranges from \$1,000 to \$3,000. Although they are not granted direct access to the CHIS data, researchers can submit analysis requests or statistical analysis programs to a CHIS statistician, who will then produce the requested output. Researchers who want to use the MCBS limited dataset (restricted data) can request the data from CMS. Once CMS has approved the request, researchers enter into a data use agreement (DUA) with CMS. In addition, effective this past June, researchers incur a fee of \$600 per year of MCBS data requested. Once fees are paid, the restricted MCBS data are provided to the researcher. Access to restricted-use data for PATH can be obtained through the ICPSR Virtual Data Enclave.

Implications of Inaccessibility of AIAN Information in Health Datasets

Most often, when researchers can identify any AIAN respondents in the public-use data, the only AIAN subgroup that can be identified is single-race non-Latinx. If most AIANs fell into this category, this would have

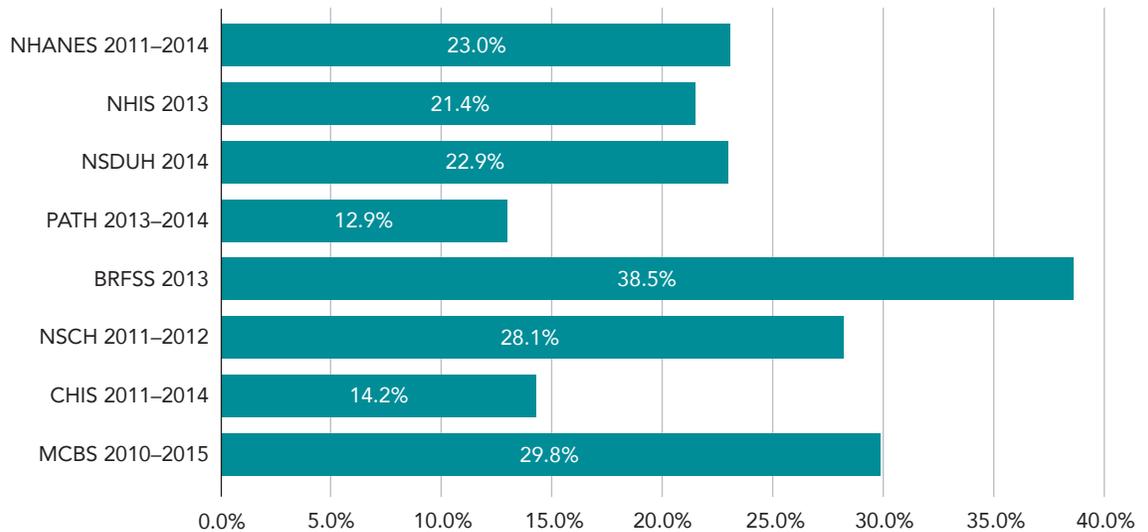
less impact on our understanding of health-related outcomes for AIAN populations. However, this is not the case. Using restricted-access files for all datasets except CHIS and BRFSS, Exhibit 3 shows the percentage of the population that falls into each of the AIAN subgroups within each survey.

Despite differences across surveys in the size of the overall AIAN population, we can see a clear pattern. In six of the eight surveys, AIANs reporting more than one race make up the largest proportion of the overall AIAN population. Reflecting the high percentage of Latinx residents in California, single-race Latinx AIAN respondents in CHIS comprise the largest AIAN subgroup. In BRFSS, single-race non-Latinx respondents comprise the highest percentage of the overall AIAN population. Further, as seen in most population surveys, the population of AIANs reporting more than one race is substantially larger than that of single-race non-Latinx AIANs.

“AIANs reporting more than one race make up the largest proportion of the overall AIAN population.”

Exhibit 4

Weighted Percentage of AIANs Who Are Single-Race Non-Latinx AIAN by Survey



Note: Analyses used weights provided by administrators to account for the design of each survey.

“The practice of focusing on single-race non-Latinx AIAN respondents significantly reduces the size of the AIAN population.”

The practice of focusing on single-race non-Latinx AIAN respondents significantly reduces the size of the AIAN population. This can be seen in Exhibit 4, which shows the percentage of the overall AIAN population in each survey representing single-race non-Latinx AIAN. Though the number varies substantially across surveys, fewer than two in five AIANs report as single-race non-Latinx AIAN. In Wave 1 of PATH, which was collected in 2013–2014, only 12.9% of AIANs reported as single-race non-Latinx AIAN. In comparison, in 2013 BRFSS, just under two-fifths (38.5%) of all AIANs reported as single-race non-Latinx AIAN. *Although single-race non-Latinx AIANs comprise less than 40% of the total AIAN population nationally—and in some surveys, less than 20%—this is the AIAN subgroup most often available to researchers. As a result, this is the AIAN subgroup most often used when making inferences about the health of the AIAN population.*

Conclusions and Recommendations

Few population health datasets include measures that allow researchers to identify any AIAN populations in their public-use data. Population-level public health research on AIANs remains scarce, in part because of the difficulties researchers face in accessing

information about this population. When these datasets are the primary source of information on a particular topic (such as PATH’s wealth of information on tobacco use) or are of a specific type (such as the NHANES biological data collection), researchers’ ability to use the data to assess these health measures for AIAN populations is nearly foreclosed.

The convention of tabulation and racial/ethnic reporting in publicly available data is intended to (1) report single-race non-Latinx and (2) assign all respondents of more than one race to one multiracial category. As shown in this brief, this convention obscures insights about the population that has substantial subpopulations who are Latinx and are of more than one race.

Restricted data generally provide the disaggregation needed for the AIAN population, but access to restricted data is often difficult, expensive, and time-consuming. This discourages the use of such data, thus significantly limiting access to information about the health-related disparities faced by the AIAN population and rendering their needs invisible in public health policy.

There may be significant health-related differences across the subgroups we examined; the presence of such differences would suggest that not accounting for the experience of all three of these subgroups of the AIAN population leads to an incomplete understanding of the health-related challenges faced by AIANs. Unless there is better measurement of the barriers to accessing data on the AIAN population, and a lowering of these barriers, these subgroups will continue to be underrepresented in public health research and public health programs.

Survey leaders and researchers should consider the following to improve measurement and availability of information about the health of the AIAN population:

- **Oversampling** methods can be used to increase the number of AIAN respondents to mitigate disclosure risk. This is expensive, but it is still an obvious and important strategy for improving federal health data on the AIAN population.
- **Release pooled multiyear public-use files** that include more detailed race/ethnicity data. Pooling multiyear data reduces the disclosure risk that may be associated with providing more detailed race data. For example, CHIS provides two-year public-use data files that include sufficient racial/ethnic detail to identify all of the AIAN subgroups.
- **Include information on the overall AIAN population.** When sample sizes prohibit the inclusion of any of the AIAN subgroups, datasets should include information on the overall AIAN population. It is preferable to include the overall AIAN category along with the single-race AIAN category to expand tabulation options for the AIAN population. In cases where disclosure risk prevents the release of measures identifying any AIAN subgroup, such as single-race non-Latinx AIAN, survey administrators should consider publicly releasing at a minimum an overall AIAN indicator that combines single-race AIAN and AIAN of more than

one race into one category. This indicator would allow continued access to data on AIAN populations when disclosure concerns prevent the analysis of AIAN subgroups. The inclusion of at least the overall AIAN population category would allow researchers interested in the AIAN population to use the dataset.

- **Note when conclusions about AIAN health are based solely on single-race non-Latinx AIANs.** When analyzing data in which only single-race non-Latinx AIANs are identifiable as AIAN, researchers should make it clear that results pertain only to this population, not to AIANs more generally. Much of what we know about the health of the AIAN population is based on single-race non-Latinx AIANs. However, this subgroup comprises less than about 40% of the total AIAN population in BRFSS and only about 13% in PATH.
- **Reduce cost/time for accessing restricted data.** If it is not possible for AIAN information to be included in public-use datasets for some surveys, the accessibility of that survey's restricted data becomes extremely important for obtaining information about the AIAN population. This is especially critical for surveys that are the primary source of information on a health topic, such as NSDUH and PATH.

Author Information

The authors of this policy brief hold the following positions at the UCLA Center for Health Policy Research: Tara L. Becker, senior public administration analyst; Susan H. Babey, PhD, senior research scientist; Riti Shimkhada, MPH, PhD, senior research scientist; AJ Scheitler, EdD, director of stakeholder relations; Ninez A. Ponce, MPP, PhD, professor at the UCLA Fielding School of Public Health and director of the UCLA Center for Health Policy Research.

Funder Information

This policy brief was made possible through funding from the Office of the Assistant Secretary of Planning and Research (ASPE), U.S. Department of Health and Human Services.



The California Health Interview Survey (CHIS) covers a wide array of health-related topics, including health insurance coverage, health status and behaviors, and access to health care. It is based on interviews conducted continuously throughout the year with respondents from more than 20,000 California households.

CHIS is a collaboration among the UCLA Center for Health Policy Research, California Department of Public Health, California Department of Health Care Services, and the Public Health Institute. For more information about CHIS, please visit chis.ucla.edu.

10960 Wilshire Blvd., Suite 1550
Los Angeles, California 90024



The UCLA Center
for Health Policy Research
is part of the
UCLA Fielding School of Public Health.

UCLA
FIELDING
SCHOOL OF
PUBLIC HEALTH

The analyses, interpretations, conclusions,
and views expressed in this policy brief are
those of the authors and do not necessarily
represent the UCLA Center for Health Policy
Research, the Regents of the University
of California, or collaborating
organizations or funders.

PB2020-8

Copyright © 2020 by the Regents of the
University of California. All Rights Reserved.

Editor-in-Chief: Ninez A. Ponce, PhD

Phone: 310-794-0909
Fax: 310-794-2686
Email: chpr@ucla.edu
healthpolicy.ucla.edu

Acknowledgments

For their thoughtful review and feedback, the authors wish to thank Delight Satter, MPH (Confederated Tribes of Grand Ronde), senior health scientist, Office of Tribal Affairs and Strategic Alliances, Center for State, Tribal, Local, and Territorial Support at the Centers for Disease Control & Prevention (CDC); Jennifer Parker, PhD, director, Division of Research and Methodology at the National Center for Health Statistics; and Randall Akee, PhD, associate professor, UCLA American Indian Studies and Luskin School of Public Affairs.

Suggested Citation

Becker TL, Babey SH, Shimkhada R, Scheitler AJ, Ponce NA. 2020. *Limited Access to Health Data on American Indian and Alaska Natives Impedes Population Health Insights*. Los Angeles, Calif.: UCLA Center for Health Policy Research.

Endnote

- 1 U.S. Census Bureau. American Community Survey, 2018 1-Year Estimates, Table B02010. Generated by Ninez Ponce using data.census.gov (2 August 2020).

