Best Practices for American Indian and Alaska Native Data Collection

Current standard data collection practices by many federal, state, and local entities effectively omit or misclassify American Indian and Alaska Native (AI/AN) populations, both urban and rural. This is particularly concerning in the midst of the COVID-19 pandemic as these current standards of practice are resulting in a gross undercount of the impact COVID-19 has on Native people. Two major problems that are seen in data collection for Native populations include multiple descriptions of Native people found in data sources between federal, state, and local public entities and methodologies for collection, analysis, and presentation of data are inconsistent in available datasets.

To address these incomplete, inaccurate, and unreliable standard data collection and analysis practices, Urban Indian Health Institute (UIHI), a Tribal Epidemiology Center, has created best practices for methods to collect, analyze, and present data on AI/AN populations. The following data collection best practices recommendations are grounded in and stem from Indigenous values and practices.

Urban Indian Health Institute
A Division of the Seattle Indian Health Board

Our mission is to decolonize data,
for indigenous people, by indigenous people.
611 12th Avenue South, Seattle, WA 98144
206–812–3030 | info@uihi.org | www.uihi.org

Recommendations

General Data Collection

Best Practices for Advocacy and Decision Making for Tribal Leadership, Policy Makers, Urban Indian Organizations, and Community Members

Mandate collection of race and ethnicity in health data that utilizes local, state, federal, and territorial funds. Include enforcement mechanisms for non-collection of this data. Provide best practices, training, and technical assistance for mandated agencies.

In data collection, AI/AN should always be defined as AI/AN alone; and, if the AI/AN individual identifies as another race, include the individuals who are AI/AN in any combination with any other race and include those who identify as Latinx/Hispanic. In the event the definition cannot be as inclusive as stated above, the next less inclusive definition should be used, i.e. AI/AN alone.

Data tools used for collection of race and ethnicity should allow for selection of multiple races with the ability to disaggregate the data once collected. Data collection tools that do not allow for disaggregation are not recommended as they will effectively eliminate AI/AN in the data. AI/AN are one of the largest growing multi-racial groups in the United States, and data collection should reflect this diversity.[1]

Collect tribal affiliation. If using electronic data collections tools, this should be an inclusive list of all federal- and state-recognized tribes with a write-in option for First Nations or other tribal affiliation not listed. If using paper, ensure there is a space allocated to write in tribal affiliation.

- UIHI recommends using “tribal affiliation” in contrast to “tribal citizenship,” as only tribes determine and define tribal citizenship. Federal, state, and local entities do not have the ability to verify tribal citizenship. Use of tribal affiliation allows for the collection of what tribe(s) an individual identifies with, without impeding on tribal sovereignty.

- Caution should be taken when releasing data on tribal affiliation publicly. UIHI recommends working with local tribes and Native organizations in the region to obtain recommendations on when and how the data should be shared.

Do not release tribally specific data without a Data Use Agreement from the tribe that grants such a release.

- For example, if a tribe reports the number of COVID-19 infections to the state, the state cannot release that tribe’s data in a way that identifies the tribe without their permission. The release of this data without permission is a direct violation of the tribe’s sovereignty, which grants them authority to govern any release of this data. When a Data Use Agreement is executed between government agencies and tribes regarding use of this data, it protects both the tribe and the government agency and should be standard practice for all data shared between tribes and government agencies.

National Electronic Disease Surveillance System access should be granted to Tribal Epidemiology Centers who were established as Tribal Public Health Authorities under the 2010 Affordable Health Care Act. Tribal Epidemiology Centers have the unique ability to work with tribes and urban Indian populations and are governed by urban and rural Native leadership.

Aggregate data on AI/AN populations. Aggregate data across time that includes a longer time frame for the analysis builds larger samples, which assists in overcoming the challenge of small populations analysis.

- For example, analyze data over three or five years rather than a single year. Another consideration for aggregating data is to combine several adjacent counties into one group, or present data at the state level to reflect demographics and outcomes of AI/AN.

Use weighted sampling for AI/AN populations.

- Weighted sampling is the practice that allows for the population that is being analyzed to accurately reflect how it's proportion in the total population is being represented from which it is being abstracted from. This gives increased strength to small populations.

Limit stratification in analysis to restrict reduction of sample size. Increased breakdowns often reduce sample sizes to very small numbers.

Avoid reporting data collected and findings from analysis as 'multi-racial' and 'other' when possible.

Link data sets to correct for racial misclassification. Racial misclassification is when an individual is classified as a different race than they identify with. This often occurs when the data collector makes assumptions based on stereotypical physical appearance instead of asking the individual what their race is. In some instances, racial misclassification occurs when an individual's race is not collected and the data system defaults to "white."

- Data linkages aim to identify two records in two data sets that represent the same person.2 For example, a data linkage between a cancer registry and an IHS patient registration looks for records in the two files that are for the same person. Because the IHS patient registration file includes tribal members only, any individual in the cancer registry who is also in the IHS file is assumed to self-identify as AI/AN. Thus, the record in the cancer registry is corrected to reflect the correct race of the individual who is misclassified as another race.

Oversample the AI/AN population.

- Oversampling is an intentional sampling process, designed to incorporate more (typically low prevalence) members of a certain community (AI/AN population) into your sample and to adjust population distribution of the dataset.

Conduct mixed-methods research (quantitative and qualitative).

- Mixed-methods research includes storytelling, focus groups, and key informant interviews. Often, epidemiologists have findings that are not statistically significant, but that does not mean the data is not important or indicative of change or disparity, especially when working with a small population. In such cases, supplementing qualitative data can support initial results despite quantitative results showing insignificance.

Report limitations of data collection and analysis.

- It is important to list, explain, and discuss limitations so those considerations can be accounted for in evaluating the results and outcomes but also so that future endeavors may seek to address and improve upon these limitations.

Report strength-based and positive outcomes that focus on effective results illustrating the strength and resiliency of Indigenous people.

Surveillance Data Collection

Best Practices for Public Health Jurisdictions

Surveillance data flows from the local level through reports of diseases, conditions, and outbreaks to the state, local (New York City and District of Columbia), or territory and then to the Centers for Disease Control and Prevention (CDC) through the Nationally Notifiable Disease Surveillance System (NNDSS). COVID-19 is a Nationally Notifiable Condition to the Centers for Disease Control and Prevention. NNDSS follows the 1997 Office of Management and Budget (OMB) standards of reporting race in one of five categories, permitting the reporting of more than one race, and race being based on self-identification.

Data Collection

Often, current data collection standards do not provide inclusive categories in data collection tools that properly capture AI/AN. This results in the erasure of the AI/AN populations and limits the ability to understand the health and well-being of the community.

Collection of race and ethnicity. Reporting forms at the local, state, and territorial level must include reporting on race and ethnicity, must include AI/AN as one of the racial categories, and must allow the reporting of multiple races. AI/AN should always be defined as, AI/AN alone, and, if the AI/AN individual identifies as another race, include the individuals who are AI/AN in any combination with any other race, and include those who identify as Latinx/Hispanic. In the event the definition cannot be as inclusive as stated above, the next less inclusive definition should be used, i.e. AI/AN alone.

Collect tribal affiliation. If using electronic data collection tools, this should be an inclusive list of all federal- and state-recognized tribes with a write-in option for First Nations or other tribal affiliations not listed. If using paper, ensure there is a space allocated to write in tribal affiliation.

Public health personnel at the local level (or state or territorial level if case investigation is done at that level) need to receive training on asking people under investigation (PUI) for COVID-19 about their race and ethnicity—to enable PUI to report more than one race if that is how the person self-identifies—and to record these responses correctly on case reporting forms. They should be able to explain in a culturally attuned way to the PUI why this information is being collected and how it will be used. They should be trained to obtain this information on first contact with the PUI and, if not obtained on first contact, to ask when further contact is made. If the PUI is not able to respond because of illness or disability, public health personnel should receive training on how to illicit this information from family, friends, or those who could provide race information in a manner that would be most acceptable to the PUI.

Regular feedback should be provided to public health personnel at the local, state, and territorial level about missing race data along with a plan for quality improvement as problems are noted.

States and territories need to report race/ethnicity information to the CDC. NNDSS provides National Electronic Disease Surveillance System (NEDSS) standards to support the transmission of 1997 OMB standards of race/ethnicity data to the CDC by states, territories, and tribal health departments.

NNDSS access should be granted to Tribal Epidemiology Centers who were established as Tribal Public Health Authorities under the 2010 Affordable Health Care Act. Tribal Epidemiology Centers have the unique ability to work with tribes and urban Indian populations and are governed by urban and rural Native leadership.

Data Analysis

AI/AN are frequently not analyzed or are placed in a category with other smaller racial groups and analyzed as “other”. The declaration of a small population or not statistically significant stems from the practices of incorrectly identifying and defining AI/AN and misclassifying them as “other” races or ethnicities. Without further breakdown, disaggregation, and enumeration of the different racial combinations that Native people self-identify with, AI/AN are erased, omitted, and/or suppressed from reports.

Numerator: Include people who are AI/AN alone and, if multi-race, include people who are AI/AN in any combination with other races. Include all ethnicities.

Denominator: Counts of AI/AN alone or in any combination can be obtained from data.census.gov using the American Community Survey. Data are available down to the county level in the United States. Data are available by gender and age down to the state/territory level. Additional data on AI/AN alone or in any combination with other races may be obtained from state, territorial, and tribal population forecasting organizations.

Counts: If the numbers of AI/AN with COVID-19 are too small to protect privacy, consider aggregating the data of several adjacent counties or presenting data at the state level. Take into consideration how surveillance data for other conditions with small numbers is presented and discussed, and aggregate data across time that includes a longer time frame for the analysis. For example, analyze data over three or five years rather than a single year.

Defined settings: If a significant proportion of cases are due to outbreaks in people who live in defined settings (long-term care facilities, jails and prisons, homeless shelters, etc.), consideration should be given to analyzing these cases separately from cases assumed to have been exposed elsewhere in the community. This serves two purposes: 1) it can highlight which defined settings pose a specific morbidity risk to AI/AN and 2) if a defined setting makes up a substantial proportion of deaths in a county or state, including those deaths in the analysis of the community can hide the true mortality burden. As an example of the second case, if there were many deaths in long-term care facilities serving primarily older, non-Hispanic White people, including these in the population under analysis may mask that mortality outside of those facilities is disproportionately among AI/AN.

COVID-19 Mortality Data in Native Populations

Best Practices for Medical Examiners, Coroners, and Funeral Homes

Unlike surveillance data where the best practice is to ask, if possible, the PUI for COVID-19 which racial and ethnic groups they self-identify as, this is not possible with the loved one who has passed away. In most instances where someone dies, the funeral home is responsible for working with the decedent's next of kin or informant to fill out the demographic portion of the death certificate, including how the decedent would have described their race and/or ethnicity. In some instances, the medical examiner or coroner may be responsible for obtaining this information from the decedent's next of kin or informant. Numerous studies have demonstrated that a high proportion of AI/AN people who die are misclassified as white on their death certificates.[3,4]

Data Collection

Due to subjective observation by the funeral home, medical examiner, or coroner, Native people are born AI/AN and die classified as white. In addition, if the next of kin is asked, funeral homes, medical examiners, and coroners may not properly offer multi-race options.

Funeral homes, medical examiners, and coroners must ask the next of kin or informant how the decedent would have described their race/ethnicity and tribal affiliation and should abstain from placing information on the death certificate based on subjective observation.

AI/AN should always be defined as AI/AN alone, and, if the AI/AN individual identifies as another race, include the individuals who are AI/AN in combination with other race, and include those who identify as Latinx/Hispanic. In the event the definition cannot be as inclusive as stated above, the next less inclusive definition should be used, i.e. AI/AN alone.

Regular feedback should be provided to funeral homes, medical examiners, and coroners about unknown or refused race responses on the death certificates along with a plan for quality improvement when a problem is identified.

Provisions in state and territorial law should be made for the next of kin or informant to amend the death certificate after it is filed if the race/ethnicity information is incorrect or is unknown at the time the death certificate is filed.

Data Analysis

Since 2018, all state and territorial death certificates comply with 1997 Office of Management and Budget (OMB) standards of reporting race in one of five categories, permitting the reporting of more than one race, and race being based on self-identification.

Numerator: Include decedents who are AI/AN alone and, if multi-race, include decedents who are AI/AN in any combination of another race. Include all ethnicities. For guidance on how to use provisional death data to count deaths from COVID-19 see: https://www.cdc.gov/nchs/data/nvss/coronavirus/Alert-3-Final-COVID-19-Guidance-and-Provisional-Death-Counts.pdf

Denominator: Counts of AI/AN alone or in any combination can be obtained from data.census.gov using the American Community Survey. Data are available down to the county level in the United States. Data are available by gender and age down to the state/territory level. Additional data on AI/AN alone or in any combination may be obtained from state, territorial, and tribal population forecasting organizations.

Small counts: Understand state, territorial, and tribal laws about the confidentiality of mortality data. In some areas, mortality data from death certificates is public data with no assumption of privacy. In other areas, there are laws regarding the release of data. If privacy is a concern and numbers of AI/AN with COVID-19 are too small to protect privacy, consider aggregating the data of several adjacent counties or presenting data at the state level. Take into consideration how mortality data for other conditions with small numbers is presented and discussed.

Defined settings: If a significant proportion of deaths are in people who live in defined settings (long-term care facilities, jails and prisons, homeless shelters, etc.), consideration should be given to analyzing these deaths separately from cases assumed to have died elsewhere in the community. This serves two purposes: 1) it can highlight which defined settings pose a specific mortality risk to AI/AN and 2) if a defined setting makes up a substantial proportion of deaths in a county or state, including those deaths in the analysis of the community can hide the true mortality burden. As an example of the second case, if there were many deaths in long-term care facilities serving primarily older, non-Hispanic White people, including these in the population under analysis may mask that mortality outside of those facilities is disproportionately among AI/AN.

It may be possible to check for racial misclassification using other data sources. Link data sets to correct for racial misclassification. Data linkages aim to identify two records in two data sets that represent the same person.[2] For example, a data linkage between a cancer registry and an IHS patient registration looks for records in the two files that are for the same person. Because the IHS patient registration file includes tribal members only, any individual in the cancer registry who is also in the IHS file is assumed to self-identify as AI/AN. Thus, the record in the cancer registry is corrected to reflect the correct race of the individual who is misclassified as another race.

The Urban Indian Health Institute recommendations are grounded in the principles of Indigenous data sovereignty

In order to conduct Indigenous epidemiology, we must honor and uphold tribes' inherent right to govern their peoples, lands, resources, and data. We use these practices and elements to assess and evaluate AI/AN populations to provide accurate and meaningful data that is relevant and reflects the unique cultures, traditions, and health needs of urban and rural Native communities. To address the impact of COVID-19 in Native communities, we must use a model of Indigenous health equity,[5] which demands collaboration with Native communities and tribal leadership for meaningful data collection and analysis.

When undertaking any efforts toward improving data collection among AI/AN people, come to Indigenous people because we have the answers, not because you think we have the most problems. The answers to preventing the spread of COVID-19 in AI/AN communities are carried in our stories, our practices of honoring elders, and generational practice of ensuring a great future for the next generations.

Data for Indigenous people, by Indigenous people.

Urban Indian Health Institute is available for technical assistance requests regarding these recommendations.

Phone: (206) 812–3030
Email: info@uihi.org
Visit: uihi.org/request-technical-assistance

References

1 Norris, T., Vines, P. L., & Hoeffel, E. M. (n.d.). American Indian and Alaska Native Population: 2010.

2 Johnson, J. C., Soliman, A. S., Tadgerson, D., Copeland, G. E., Seefeld, D. A., Pingatore, N. L., … Roubidoux, M. A. (2009, June). Tribal linkage and race data quality for American Indians in a state cancer registry. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4274940/

3 Arias E, Schauman WS, Eschbach K, Sorlie PD, Backlund E. The validity of race and Hispanic origin reporting on death certificates in the United States. National Center for Health Statistics. Vital Health Stat 2(148). 2008.

4 Jim, M. A., Arias, E., Seneca, D. S., Hoopes, M. J., Jim, C. C., Johnson, N. J., & Wiggins, C. L. (2014). Racial Misclassification of American Indians and Alaska Natives by Indian Health Service Contract Health Service Delivery Area. American Journal of Public Health, 104(S3). doi: 10.2105/ajph.2014.301933

5 Echo-Hawk, A. (n.d.). Indigenous Health Equity. Retrieved from https://www.uihi.org/resources/indigenous-health-equity/