New 2020 Census Rules Make It Harder to Navigate Native American Data

AUGUST 2023

Matthew T. Gregg

Federal Reserve Bank of Minneapolis

H. Trostle

Federal Reserve Bank of Minneapolis

Carolyn A. Liebler

University of Minnesota

New 2020 Census Rules Make It Harder to Navigate Native American Data

Matthew T. Gregg*
Senior Economist
Center for Indian Country
Development
Federal Reserve Bank of
Minneapolis

H. Trostle
Senior Policy Analyst
Center for Indian Country
Development
Federal Reserve Bank of
Minneapolis

Carolyn A. Liebler
Associate Professor
Department of Sociology
University of Minnesota

August 2023

Abstract

The U.S. Census Bureau has developed a new set of safeguards for the 2020 Census to protect respondent confidentiality. This paper discusses the implications for Indian Country of two of the new safeguards: differential privacy, a statistical framework that infuses random noise into raw data; and dynamic population thresholds, parameters used to determine whether detailed American Indian tribes and Alaska Native (AIAN) tribes and villages receive demographic data. We show that differential privacy greatly affects the accuracy of published, place-based statistics on smaller American Indian, Alaska Native, and Native Hawaiian (AIANNH) areas. We also find that the minimum population thresholds used to report demographic statistics on detailed, self-identified American Indian and Alaska Native respondents living in AIANNH areas are so high that roughly 80 percent of all tribal groups who previously received a full suite of demographic data would no longer have their age data reported by sex. We conclude with recommendations on how to increase the utility of 2020 Census data on tribes and tribal areas.

I.   Introduction

Tribes have always had to work around data gaps in Indian Country (Lozar 2023). Even with the known undercount in tribal communities, tribal planners have come to rely on the decennial census (NPR 2022). The census provides a critical source for demographic and housing information on the diverse Native American (i.e., American Indian, Alaska Native, and Native Hawaiian) population. For example, Economic Development Director Michael Southard (Choctaw Nation of Oklahoma) uses census data, supplemented with tribally collected data, for everything from making business decisions about where to develop a grocery store or a daycare center to projecting his tribe's future population growth. Without good data, the tribe could make expensive missteps in planning that would be particularly devastating for the regional economy since, as Southard points out, "the Tribe is the only economic engine for many of these [rural Oklahoma] counties."

In a change that dramatically impacts data from the 2020 census about American Indian, Alaska Native, and Native Hawaiian (AIANNH) Areas[1], the Census Bureau has changed its safeguards. Essentially, the Census Bureau, aiming to produce high-quality data while protecting respondent confidentiality, decided that old methods to protect respondent confidentiality are insufficient. There are perceived new threats to exposing the confidential individual data used to produce aggregated published tables.[2] In this paper, we provide an overview of how the Census

---

[1] AIANNH Areas include the following legal and statistical areas: federally recognized American Indian reservations and off-reservation trust lands, state recognized reservations, Native Hawaiian Home Lands, Oklahoma Tribal Statistical Areas, Alaska Native Village Statistical Areas, tribal designated statistical areas, and state designated tribal statistical areas.

[2] While this paper is not focused on the effect of the pandemic, the enumeration of the 2020 Census coincided with the COVID-19 pandemic. This required the Census Bureau to pause its operations in Spring 2020 and extend the self-response period until October 2020 which subsequently led to a delay in the Post-Enumeration Survey, a main tool to evaluate the quality of the decennial data (U.S. Census Bureau n.d.-a, PRB 2022, U.S. Census Bureau n.d.-b). These surveys revealed that the 2020 Census generated larger errors, whether due to overcounting or undercounting, than those in the 2010 Census (Cohn and Passel 2022). With respect to the AIAN population, the error rate on

Bureau's new suite of privacy methods, including "differential privacy" and "dynamic population thresholds," affect the accuracy of 2020 census data about AIANNH Areas and about people who reported affiliation with American Indian tribes[3] and Alaska Native villages[4] (which are collectively referred to by the Census Bureau as "tribal grouping" data).

We find that differential privacy affects the accuracy of the total population (of any race) living in AIANNH Areas, especially among smaller AIANNH Area communities.[5] Differential privacy, in particular, greatly affects the accuracy of the AIAN alone population counts living in AIANNH Areas. This is especially concerning since many AIANNH places contain tribal governments that need information about their population and subpopulations for governance and planning purposes. We also show that the sheer quantity of tribal grouping data, especially when collected at the AIANNH Area level, will substantially decrease due to stricter minimum population thresholds for reporting data. Taken together, the greater inaccuracy of AIAN place-based data caused by differential privacy and the relative scarcity of AIAN people-based data will harm the ability of self-governing tribes to meet the needs of their citizens.

---

reservations was 5.6 percent, which was statistically similar to the 2010 error rate of 4.9 percent (Khubba, Heim, and Hong 2022).

[3] Census respondents who identified themselves as American Indian or Alaska Native were asked to report their enrolled or principal tribe(s) in the write-in line. The detailed AIAN tribal grouping data reflects the entries reported in the write-in line. For example, if enough respondents wrote "White Earth Chippewa" as their principal tribe, then the Census Bureau would publish statistics on this population. The Census Bureau also combines individual American Indian tribes, such as Bois Forte Chippewa, Lake Superior Chippewa, White Earth Chippewa, and so on, into a general category, which the Census Bureau refers to as the "Chippewa tribal grouping alone" population. In this paper, we refer to statistics from the write-in portion of the AIAN race question as either "tribal" or "tribal groupings" data since these data contain both aggregated tribal data and data on individual sub-tribal populations.

[4] For Alaska Native communities, the Census Bureau refers to an Alaska Native tribal grouping as the combination of individual Alaska Native tribes. For example, the "Inupiat tribal grouping alone" category contains such Alaska Native tribes as Native Village of Brevig Mission, Native Village of Deering, and King Island Indian Community, to name a few.

[5] The distortion of differentially private data about smaller populations has been highlighted in other papers. For example, Hauer and Santos-Lozada (2021) show that differential privacy significantly distorts mortality rates and Winkler et al. (2022) show that differential privacy makes net migration rates unusable for half of all U.S. counties.

II.   **Privacy Protection in 2020 Census Data**

The Census Bureau has released four data products from the 2020 census: Apportionment Counts, the PL 94-171 Redistricting File, the Demographic and Housing Characteristics (DHC), and the Demographic Profiles. The procedure called "differential privacy" was applied to all but the Apportionment Counts. The Census Bureau plans to release the detailed DHC-A (which will contain detailed race population data) and DHC-B files (which will contain detailed housing data) in the fall of 2023 and 2024, respectively. These will have different privacy protection strategies and will not include all the information available in prior years.

Census 2020 data products, in particular the DHC File and the Demographic Profiles, about Native communities share similarities to their 2010 counterparts; population and housing tables are again available for each AIANNH Area and will soon be available for detailed American Indian and Alaska Native (AIAN) tribal groupings. As an improvement over the prior policy, the Census Bureau's 2020 race and ethnicity coding procedure captured more of the write-in section of the race question, increasing from 30 to 200 characters (Census 2021). Thus, longer write-in responses, such as individuals who self-identify with multiple tribes, are now captured in the raw data.

However, there is a considerable difference between these 2020 census products and their 2010 counterparts. By design, the Census Bureau's new privacy methods significantly affect the utility of the 2020 census data about AIANHH Areas and the people who report their tribe. As mentioned earlier, a perceived heightened privacy threat was caused by the rapid increase of third-party data that could be used to identify individuals from published census statistics. So, the Census Bureau modernized its approach to protect respondent confidentiality based on the framework of

"differential privacy" (Census 2020).[6] In a nutshell, differential privacy quantifies the amount of private information "leaked out" when a statistic is published. The Census Bureau, through rounds of public consultations, has developed an algorithm that essentially draws a random number (either positive, negative, or zero) from a statistical distribution and adds it to each statistic before publishing. Thus, the accuracy of the published data depends on how much noise is infused to a particular data point. Although a computer program chose the specific random numbers, it was a Census Bureau policy decision to set how much overall noise would be added to the published census statistics and how to allocate that noise to specific statistics.

To illustrate how differential privacy affects the quality of 2020 census data on AIANNH Areas, we use demonstration data provided by the Census Bureau (IPUMS NHGIS 2023). The demonstration data allows us to compare how published statistics from the 2010 census would have looked if differential privacy had been applied to the 2010 data. We focus our analysis on three aspects of 2020 census data about AIANHH Areas: (a) the total population in AIANHH Areas; (b) disaggregated populations in AIANHH Areas (by race, age, and sex); and (c) the 2020 version of the AIAN Summary file. Our purpose is to help users better anticipate how the new privacy safeguards affect the fitness-for-use of the 2020 AIANNH Area data and disaggregated AIAN data.

III.   **The Statistical Accuracy of AIANNH Area Data Decreases with Population Size**

Total population in AIANHH Areas

First, we compare the difference between the reported total population (of all races) for each AIANNH Area to the differentially private version of that statistic, as shown in the demonstration

---

[6] The Census Bureau conducted a "database reconstruction" experiment (Abowd 2021) and concluded that modern computing power and commercial datasets pose additional risks to respondent confidentially that require a new strategy for disclosure protection. Further investigation showed that the alarming results may have been due to random chance (Ruggles and Van Riper 2022). Regardless, the newly developed method of "differential privacy" (Garfinkel 2022) had already been adopted by the Census Bureau.

data. AIANNH Areas have populations ranging from less than 50 people to greater than 10,000 people, and there are many AIANNH Areas with small populations and few AIANNH Areas with very large populations.

We show in Figure 1 how the impacts of differential privacy vary depending on the AIANNH Area population size. Each point represents an AIANHH Area, and its location in Figure 1 reflects the percentage difference between the reported and differentially private total population (of all races) in the AIANNH Area. If a point is at zero, the differentially private algorithm did not affect the accuracy of that AIANNH Area population count. If the difference is positive, the differential privacy algorithm would have reported a higher population than the published count in 2010. If the difference is negative, then differential privacy would have produced a smaller total population count than the published total population count.

Figure 1 reveals two main results. First, the 2010 census AIANNH Area population counts using differential privacy are, *on average*, identical to those published in the 2010 census. The exception is that differential privacy reduces the population counts reported in AIANHH Areas with the smallest populations (between 11 and 50 people). This result implies that if a data user is only interested in pooling all the AIANNH Area data to create summary statistics, then differential privacy will not impact those statistics. Of course, if a data user is interested in estimating relationships, then the additional noise differential privacy creates will attenuate results. In addition, if a user is interested in a specific place, then differential privacy will affect the accuracy of that specific place, especially if the person is using small AIANNH Area data on populations less than 50 people.

Second, we find that differential privacy does generate a small number of wildly inaccurate observations. For example, in AIANHH Areas with the smallest populations, several total population

counts would have been reported as 25 percent smaller than their total population count if differential privacy was adopted in 2010, and some AIANNH total population counts would have been about 20 percent greater under differential privacy. Among AIANNH Areas with larger populations, the impact of differential privacy becomes less pronounced; however, some AIANNH Areas between 500 and 1,000 people would have seen an impact in their population count of 8-13 percent (in absolute value) if differential privacy was imposed in 2010. In sum, a data user or policymaker who is interested in using a particular AIANNH Area population to forecast, let's say, future housing needs will struggle to determine if significant relative changes in population from 2010 are due to natural changes to their population or are a by-product of differential privacy.

Disaggregated population in AIANHH Areas

When disaggregating data by race, age, and/or sex, differential privacy creates a lot more trouble for data users interested in a single AIANNH Area. To see this, we use Figure 2 to report the percentage difference between the 2010 differentially private and published counts for the *AIAN Alone* population living in each AIANNHA Area. Like Figure 1, we distinguish between AIANNH Areas of different sizes to show how the effects of differential privacy vary by population size.

Like the analysis of the total population shown in Figure 1, Figure 2 reveals that differential privacy does not affect the *average* AIAN Alone population size across AIANHH Areas. This is especially true for AIANNH Areas with larger populations. However, unlike Figure 1, differential privacy substantially impacts the AIAN Alone population counts in individual AIANHH areas. For example, Figure 2 shows that differential privacy increased the AIAN Alone population size by up to 20 percent in some AIANNH Areas and decreased it by 38 percent in some other AIANHH Areas. Even moderately sized AIANNH Areas are affected by differential privacy when looking at disaggregated population; for example, some AIANNH Areas between 100 and 500 people would

6

have seen their AIAN Alone population count decrease by as much as 29 percent if differential privacy were imposed.

We also found similar results (not shown) when cutting data by sex, age, or sex- by-age. For example, we see the same patterns in differentially private statistics on the total population aged 18 and over, the total female population, and the total female population aged 18 and over, measured at the AIANNH Area level. In sum, sex-, age- and sex-by-age counts by AIANNH Areas are noisier due to differential privacy than total (all races) population counts, especially for AIANNH Areas with smaller populations.

Our analyses indicate that differential privacy will make it harder to understand demographic changes in individual tribal areas.

IV. New Population Thresholds Will Substantially Reduce the Quantity of Tribal-Specific Data[7]

As mentioned earlier, the Census Bureau is planning to publish the detailed DHC-A file (which will contain demographic data for detailed tribal groupings[8]) in the fall of 2023 and the detailed DHC-B file (which will contain housing data for detailed tribal groupings) in the fall of 2024. The 2010 version of the detailed DHC-A and DHC-B files were referred to as the AIAN Summary File and contained 71 population, housing, and group quarter data tables for each 1,567 American Indian and Alaska Native tribal groupings (U.S. Census Bureau 2012). In 2010, the Census Bureau used a population threshold of 100 for every data table in every geographic area to protect the identity of respondents. Thus, for example, if at least 100 respondents from Becker County, Minnesota, wrote in "White Earth Band of the Minnesota Chippewa Tribe," then all 71 data tables (total population, sex-

---

[7] According to the Census Bureau's proof of concept, these thresholds are not final and subject to potential change.
[8] For the Census Bureau's definition of "tribal groupings," see fn3 and fn4.

by-age, and housing counts) for this tribal grouping within Becker County would have been published. If less than 100 people of a particular tribe were enumerated in a specific geographic area, no data on this specific tribe would have been published.

The 2020 version of the AIAN Summary File

For the detailed 2020 DHC-A file, instead of using a 100-person cutoff, the Census Bureau will use dynamic population thresholds.[9] The Census Bureau will only publish the total population count for tribal groupings in many more situations than in 2010. For example, when statistical noise is added to the raw data and the resulting tribal grouping's noise-infused population in a state (or in the U.S.) is less than 500 people, only the noise-infused total population will be published for that state (or the nation). Suppose the detailed tribal grouping's noise-infused population is between 22 and 999 people at a more precise geographic level (e.g., county, tract, place, or AIANNH Area). In that case, the Census Bureau will, again, only publish the total population for that group in that geography.[10] The remaining dynamic thresholds are as follows: If a tribal grouping's population in 2020 is small,[11] their noise-infused total population and four age categories[12] per sex will be published. If a tribal grouping's noise-infused population is moderate,[13] then their noise-infused total population count will be published, as well as nine age categories[14] per sex. And if a tribal grouping's noise-infused population is large,[15] then the Census Bureau will publish the noise-infused

---

[9] The plans for the housing tables that will be available in the 2020 DHC-B file have yet to be published.

[10] If the noised-infused total population is less than 22, then it will not be listed.

[11] "Small" here means between 500-999 people at the national or state level, or between 1000 and 4,999 people at the county, tract, place or AIANNH Area level.

[12] The four age categories are: under the age of 18, 18 to 44, 45 to 64, and 65 and over.

[13] "Moderate" here means between 1000 and 6,999 people at the national or state level, or between 5,000 and 19,999 people at the smaller geographic entity level.

[14] The nine age categories are: under the age of 5, 5 to 17, 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 74 and over.

[15] "Large" here means 7,000 or more at the national or state level, or more than 20,000 people at the detailed geographic entity level.

total population count and 23 age categories per sex.[16] Note that they will not publish more than 23 age categories per sex for tribal groupings of any size.

The dynamic population thresholds will impact different tribal groupings differently. For tribal groupings with less than 100 people, their total noise-infused populations will be published for the first time. However, for tribal groupings with noise-infused populations greater than 100 but less than 500 at the national or state level, the 2020 stricter thresholds will produce significantly fewer data tables for those groups. For example, in 2010, 394 respondents wrote "Prairie Island Indian Community" and, as a result, the Census Bureau published their total population size (394) and 104 age categories per sex for this tribal grouping. Under the new dynamic population thresholds, only one statistic published about people who reported Prairie Island Indian Community affiliation would have been published: the total population.

To illustrate how the dynamic population thresholds will impact the amount of available tribal grouping data in the soon-to-be-released 2020 detailed DHC-A file, we have applied the new 2020 population thresholds to the published 2010 AIAN Summary File. The tribal groupings in the 2010 AIAN Summary File each received a complete set of data, which, in 2010, meant the Census Bureau published their total population and 104 age categories per sex (equivalent to 209 data points). Our goal is to determine— if the dynamic population thresholds were imposed in 2010— what share of 2010 AIAN tribal groupings would have received one total population data point, four additional age tables for each sex (or eight additional data points), nine age tables per sex (or 18 additional data points), or 23 age tables per sex (or 46 additional data points).

---

[16] The 23 data tables can be found on the Census Bureau's population tables for detailed race, ethnicity and AIAN tribes and villages are discussed in U.S. Census Bureau (2023c).

The results of this exercise are shown in Figure 4. The first stacked bar chart shows the share of tribal groupings in 2010 that would have fallen into each data table category[17] using the new national-level population thresholds. This bar shows that roughly 40 percent of all tribal groupings in the 2010 census had less than 500 people in the nation, which means that (if this were in 2020) they would have received only one data point - their total population at the national level.[18] Roughly 16 percent of the 2010 tribal groupings with national-level data would have received four additional age categories, by sex, if these population thresholds were imposed in 2010, and slightly more than 30 percent of all 2010 tribal groupings would have received nine additional age categories of data for each sex. The largest tribal groupings, which would have received 23 additional age data categories per sex, represent only 12 percent of all tribal groupings.

The second stacked bar chart in Figure 4 shows the share of tribal groupings in 2010 that would have fallen into each data table category using the state-level population thresholds (identical to the national-level population thresholds). The state-level data in the 2010 AIAN Summary File contained 6,166 observations.[19] If these new state-level population thresholds were applied in 2010, more than half of the tribal groupings with state-level data would have only received information on their state-level population.

The third and final stacked bar chart in Figure 4 shows the share of the observations in the 2010 AIAN Summary File that would have fallen into each data category using the stricter population thresholds imposed for tribal groupings in AIANNH Areas. Over 80 percent of tribal groupings with

---

[17] Data table categories are: (1.) population only, (2.) total population and four sex-by-age categories, (3.) total population and nine sex-by-age categories, and (4.) total population and 23 sex-by-age categories.

[18] The 2010 AIAN Summary File contained 6,166 rows of state-level data. The average tribal grouping in 2010 received a complete set of data in four states, while, large tribal groupings, such as the "Cherokee alone or in any combination" tribal grouping received a complete set of data in 50 states and in the District of Columbia.

[19] The same tribal grouping can have their population and/or age categories (by sex) reported for multiple sets. For example, in 2010, the "Eastern Band of Cherokees alone" tribal grouping had eight rows of data in the 2010 AIAN Summary File because this tribal grouping's population was greater than 100 in eight different states.

AIANNH Area data in 2010 would have received only data on their total population in a particular

AIANNH Area if the dynamic population threshold had been imposed. For example, the Absentee

Shawnee Tribe of Oklahoma was large enough to have a full suite of sex-by-age data published in

2010 for three different AIANNH Areas: Citizen Potawatomi Nation-Absentee Shawnee Oklahoma

Tribal Statistical Area (OTSA), Kickapoo OTSA, and the Sac and Fox OTSA. Under the new

population thresholds, only three data points (the total population of Absentee Shawnees in each of

the three AIANNH Areas) would be published. Roughly 10 percent of the observations in the 2010

AIAN Summary File would have received data on four additional age groups by sex, while the small

remainder would have received nine or 23 additional age categories for each sex. *In other words,*

tribal leaders should not expect to receive detailed self-reported tribal population data from the 2020

census.

V.   How to Adapt to These Changes

We finish this paper with three recommendations on how to use the 2020 census data products to

increase the utility of the 2020 census data when working with detailed AIAN tribal groupings or

AIANNH Areas.

1. *Aggregate the 2020 DHC data when using census block geographies*. The Census Bureau has

   recommended aggregating U.S. Census block data to larger geographic units in hopes to

   cancel out some of the statistical noise. In other words, the Census Bureau is recommending

   not using individual census block data (Jarmin 2021).[20]

---

[20] The Census Bureau's recommended use of aggregation only applies to census block data. For example,
aggregating block group data within the same census tract will not give you more accurate census tract data.

Of course, if a data user needs AIANNH Area data for a single area, they cannot aggregate; AIANNH Areas do not fit neatly on the census's "geographic spine."[21] The Census Bureau tried to minimize the impact of differential privacy noise on "off-spine"[22] geographies like AIANNH Areas. Still, comparable statistics on off-spine geographies (e.g., the under-18 population for an AIANNH Area with a population of 5,000) will be less accurate than a comparable geographic entity (the under-18 population count for a county with 5,000 people).

2. *Use the error bands in the detailed DHC-A data*. Unlike the Demographic Profile and the DHC file, the detailed DHC data will publish the margin of error for the total population and sex-by-age categories. We showed that the dynamic population thresholds will greatly impact the quantity of previously released data on tribal groupings. For the select number of tribal groupings whose population will be greater than the population thresholds in 2020, those population counts and sex-by-age data will contain margins of error. In the Census Bureau's detailed DHC-A proof of concept (U.S. Census Bureau, 2023b), the margin of error for detailed tribal groupings at the national or state levels is three, and the margin of error for detailed tribal groupings at the AIANNH Area level is 11. When using state or national-level data, you can be 95 percent confident that the "as-enumerated" count is plus or minus three people away from the published count. Alternatively, when looking at detailed tribal grouping at the sub-state level (like at the AIANNH Area level), you can be 95 percent confident that the "as-enumerated" count is plus or minus 11 people away from the published count. For example, suppose the total population for a detailed tribal grouping, e.g., the Lac

---

[21] The Census Bureau uses the term "geographic spine" to refer to the Census Bureau's standard hierarchy of census geographies that are subsumed by each other. For example, a Census block group is comprised of several Census blocks, a Census Tract is comprised of many Census block groups, a U.S. County is comprised of several Census tracts, a state is comprised of many counties and the U.S.is comprised of 50 states. These geographies are called "on-spine geographies."

[22] "Off-spine" geographies refer to important statistical geographic entities that are not located on the standard hierarch of census geographies. Those include places, minor civil divisions, and, most germane for this paper, AIANNH Areas, to name a few.

Court Oreilles Band of Lake Superior Chippewa, is reported as 1,100 in a particular AIANNH Area. In that case, we can be 95 percent confident that the actual, enumerated count for that AIANNH Area was between 1,089 and 1,111 people.

3. *Focus on the "Invariants."* Some statistics in the 2020 census are not subject to the differential privacy process. The Census Bureau decided to make three variables "invariant" to the new disclosure avoidance system (U.S. Census Bureau, 2020). These are the total population at the state level, the total number of housing units at any geographic level, and the number of group quarter facilities by type at any geographic level. Thus, if a person is interested in accurately counting the number of housing units within a particular AIANNH Area, the published housing count will be identical to the "as-enumerated" housing count, even if this Area has a small population.

VI.   Concluding Remarks

Many tribal communities developed outreach programs around the 2020 census because census data are essential. For instance, the Tulalip Tribes in Washington looked at previous trends of Native undercounts in the 2010 census to develop their public education campaign. They paid for commercials, Facebook ads, messages from the Chair, and even created t-shirts. The goal was to improve the quality of the 2020 census and prevent a Native undercount. Why? Lea Anne Burke, a former tribal planner for the Tulalip Tribes, points out: "Data is vital. It sets the boundaries of your knowledge base, and it helps you understand your capacity."

Without question, the increased measures to protect respondent confidentiality have made the 2020 census data less accurate and accessible. To better protect the confidentiality of census respondents, less detailed tribal data will be available and, when available, will be subject to statistical noise. Traditional users of the decennial census will have to adjust or augment census

data with additional data. For tribes and tribal planners, that may mean collecting their own data, using tribal records, or licensing commercial datasets. These are expensive and time-consuming strategies, though many tribal planners have already adopted them. We hope this paper can help navigate the 2020 census data products in a way that maximizes the usefulness of these crucial datasets and minimizes fruitless efforts to find data that will not be released or work with data that is severely impacted by differential privacy.

References

Abowd., J. 2021. "2010 Declaration of John Abowd, State of Alabama v. United States Department of Commerce." Case No. 3:21-CV-211_RAH-ECM-KCN.

Cohn, D'Vera and Jeffrey S. Passel. 2022. "Key Facts About the Quality of the 2020 Census." *Pew Research Center,* June 8, 2022. https://www.pewresearch.org/short-reads/2022/06/08/key-facts-about-the-quality-of-the-2020-census/.

Garfinkel, Simson, 2022. "Differential Privacy and the 2020 U.S. Census." MIT Case Studies in Social and Ethnical Responsibilities of Computer. https://doi.org/10.21428/2c646de5.7ec6ab93

Hauer, Matthew E., and Santos-Lozada, Alexis R. 2021. "Differential Privacy in the 2020 Census Will Distort COVID-19 Rates." Soxius, 7 https://doi.org/10.1177/2378023121994014

IPUMS, NHGIS. 2023. "Privacy Protected Demonstration Data, Vintage 2023-04-03." Derived from 2010 Privacy-Protected Microdata File (PPMF) for the Redistricting Data and DHC, version 2022-04-03. https://www.nhgis.org/privacy-protected-2010-census-demonstration-data#v20230403.

Jarmin, Ron. 2021. "Redistricting Data: What to Expect and When." *U.S. Census Bureau*, July 28, 2021. https://www.census.gov/newsroom/blogs/director/2021/07/redistricting-data.html.

Khubba, Shadie, Krista Heim, and Jinhee Hong. 2022. *National Census Coverage Estimates for People in the United States by Demographic Characteristics: 2020 Post-Enumeration Survey Estimation Report, PES20-G-01.* Washington, D.C.: U.S. Government Publishing Office, U.S. Census Bureau, March 2022.

Lozar, Casey. 2023. "Indian Country Gains Momentum in Addressing Data Gaps." *Center for Indian Country Development, Federal Reserve Bank of Minneapolis*, March 8, 2023. https://www.minneapolisfed.org/article/2023/indian-country-gains-momentum-in-addressing-data-gaps.

NPR. 2002. "The 2020 census had big undercounts of Black people, Latinos and Native Americans." https://www.npr.org/2022/03/10/1083732104/2020-census-accuracy-undercount-overcount-data-quality.

Ruggles, Steven and Van Riper, David. 2022. "The Role of Chance in the Census Bureau Database Reconstruction Experiment." *Population Research and Policy Review* 41, 781-788. https://doi.org/10.1007/s11113-021-09674-3.

PRB. 2022. "Information and Advice on 2020 Federal Data Quality and Use." Accessed: July 13, 2023. https://www.prb.org/wp-content/uploads/2022/08/2020-federal-data-quality-and-use.pdf.

U.S. Census Bureau. n.d.-a "2020 Census Operational Timeline and Adjustments Due to Covid-19." Last modified: February 3, 2023. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/operational-adjustments.html.

U.S. Census Bureau. n.d.-b "Post-Enumeration Surveys." Last modified: August 17, 2022. https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement/pes.html.

U.S. Census Bureau. 2012. *2010 Census of Population and Housing, American Indian and Alaska Native Summary File: Technical Documentation.* Washington, D.C.: U.S. Government Publishing Office, December 2012. https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/complete-tech-docs/summary-file/aiansf.pdf.

U.S. Census Bureau, 2020. "Invariants Set for 2020 Census Data Products." Last Modified: November 25, 2020. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates/2020-11-25.html000

U.S. Census Bureau. 2023. "2020 Census Race and Ethnicity Data: The Detailed DHC-A Proof of Concept." Last Modified: February 3, 2023. https://www.census.gov/data/academy/webinars/2023/detailed-dhc-a-proof-of-concept.html.

U.S. Census Bureau. 2023b. "2020 Census Race and Ethnicity Data: The Detailed DHC-A Proof of Concept (Video)." Accessed: July 13, 2023. https://www.youtube.com/watch?v=ePSb55MYoII&t=1508s.

U.S. Census Bureau. 2023c. "Population Data Available for 2020 Census Detailed Race, Ethnicity, and American Indian and Alaska Native Tribes and Villages." Last Modified: January 31, 2023. https://www.census.gov/library/fact-sheets/2023/dec/adaptive-design.html.

Winkler, Richelle L., Butler, Jaclyn L., Curtis, Katherine J. and Egan-Robertson, David. 2022. "Differential Privacy and the Accuracy of County-Level Net Migration Estimates." Population Research and Policy Review 41, 417–435. https://doi.org/10.1007/s11113-021-09664-5

Figure 1: The Impact of Differential Privacy on total all-race population counts, by AIANNH Areas
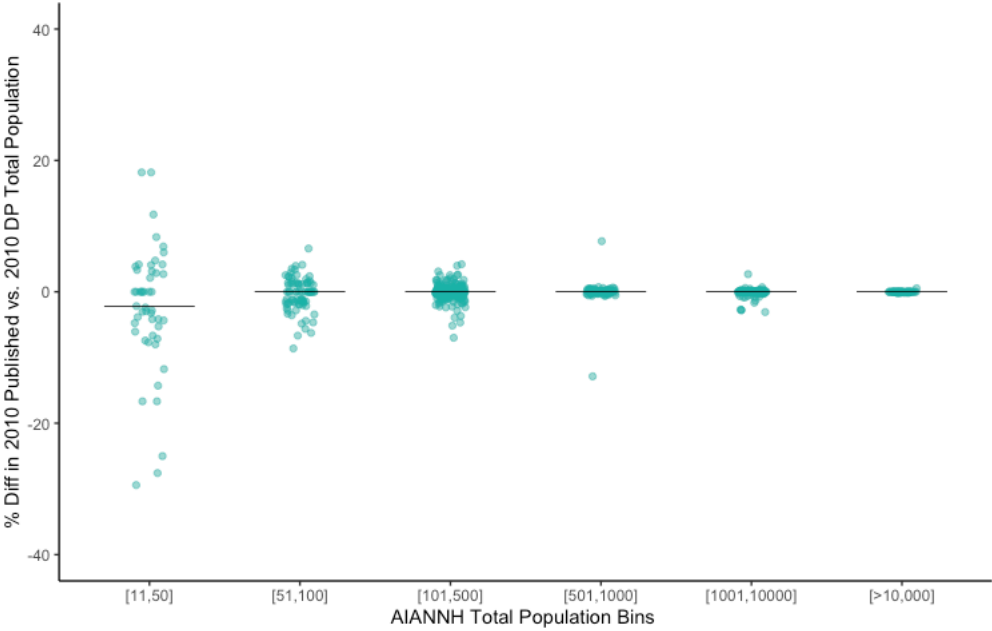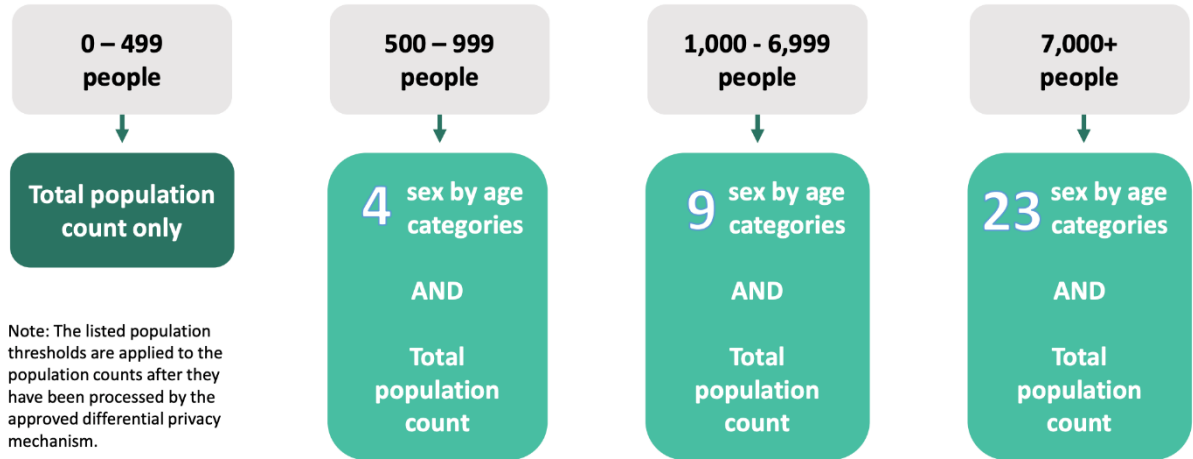

Notes: The green dots represent the percent difference between 2010 reported total population (of all races) and the 2010 differentially private total population (of all races). The black lines represent the median difference between the differentially private 2010 total population and the 2010 published total population for each population bin. We bin each AIANNH Area by the total population count. Because very small statistics are highly brittle to noise injections, we do not include AIANNH Areas with a 2010 published AIAN alone count of 10 people or fewer. The resulting AIANNH Area population bins are [11,50], [51,100], [101,500], [501, 1000], [1001,10000], [>10,000].

Figure 2: The Impact of Differential Privacy on AIAN Alone population counts, by AIANNH Areas
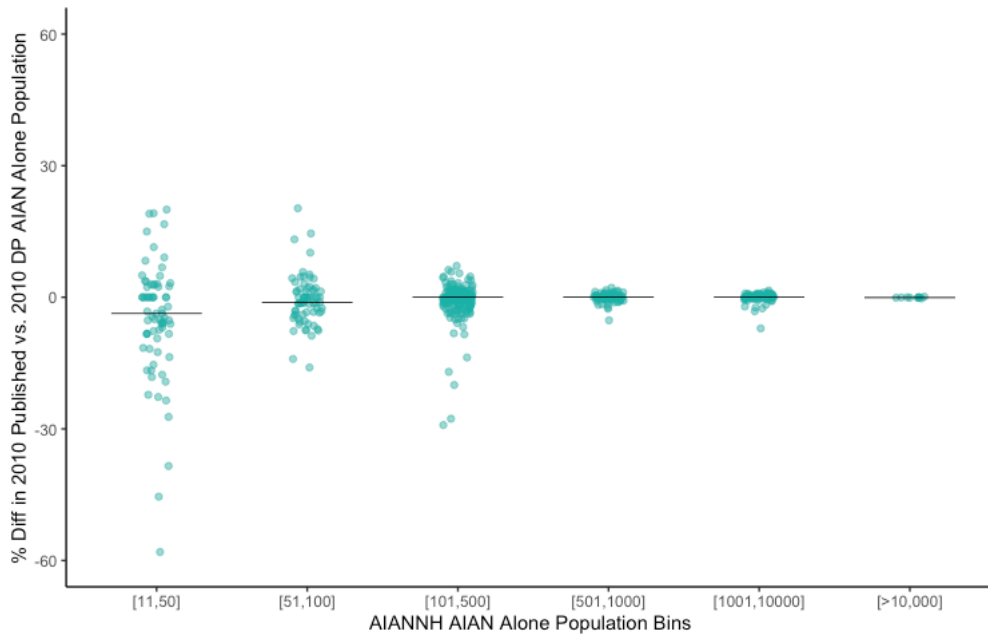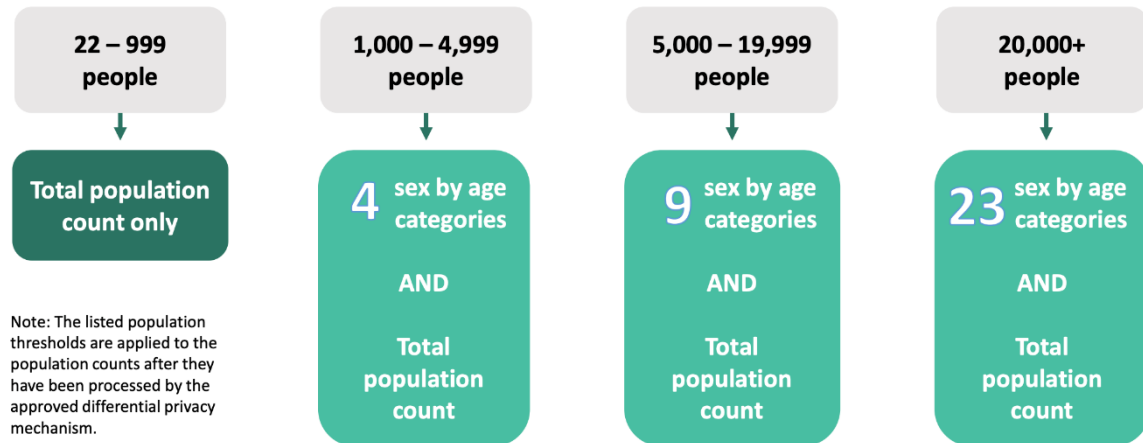

Notes: The green dots represent the percent difference between the 2010 reported AIAN Alone population count and the 2010 differentially private AIAN alone population count. The black lines represent the median difference between the differentially private 2010 AIAN Alone population count and 2010 published AIAN Alone population count for each population bin. We bin each AIANNH Area by the AIAN alone count. Because very small statistics are highly brittle to noise injections, we do not include AIANNH Areas with a 2010 published AIAN alone count of 10 people or fewer. The resulting AIANNH Area population bins are [11,50], [51,100], [101,500], [501, 1000], [1001,10000], [>10,000].

Figure 3: 2020 Census Minimum Dynamic Population Thresholds

0 – 499 people	500 – 999 people	1,000 - 6,999 people	7,000+ people
Total population count only	**4** sex by age categories AND Total population count	**9** sex by age categories AND Total population count	**23** sex by age categories AND Total population count

Note: The listed population thresholds are applied to the population counts after they have been processed by the approved differential privacy mechanism.

Panel A: Minimum Thresholds for detailed tribal groups at national and state levels

22 – 999 people	1,000 – 4,999 people	5,000 – 19,999 people	20,000+ people
Total population count only	**4** sex by age categories AND Total population count	**9** sex by age categories AND Total population count	**23** sex by age categories AND Total population count

Note: The listed population thresholds are applied to the population counts after they have been processed by the approved differential privacy mechanism.

Panel B: Minimum Thresholds for detailed tribal groups in AIANNH Areas

Source: These figures are taken from the Census Bureau's detailed DHC-A proof of concept presentation on January 31, 2023. For more information, see U.S. Census Bureau (2023). Four sex-by-age categories means there will be four age categories published for each sex (for a total of 8 data points). Nine sex-by-age categories means there will be nine age categories published for each sex (for a total of 18 data points). Last, 23 sex-by-age categories means there will be 23 age categories published for each sex (for a total of 46 data points).

Figure 4: How New Rules Affect the Quantity of Tribal Grouping Data
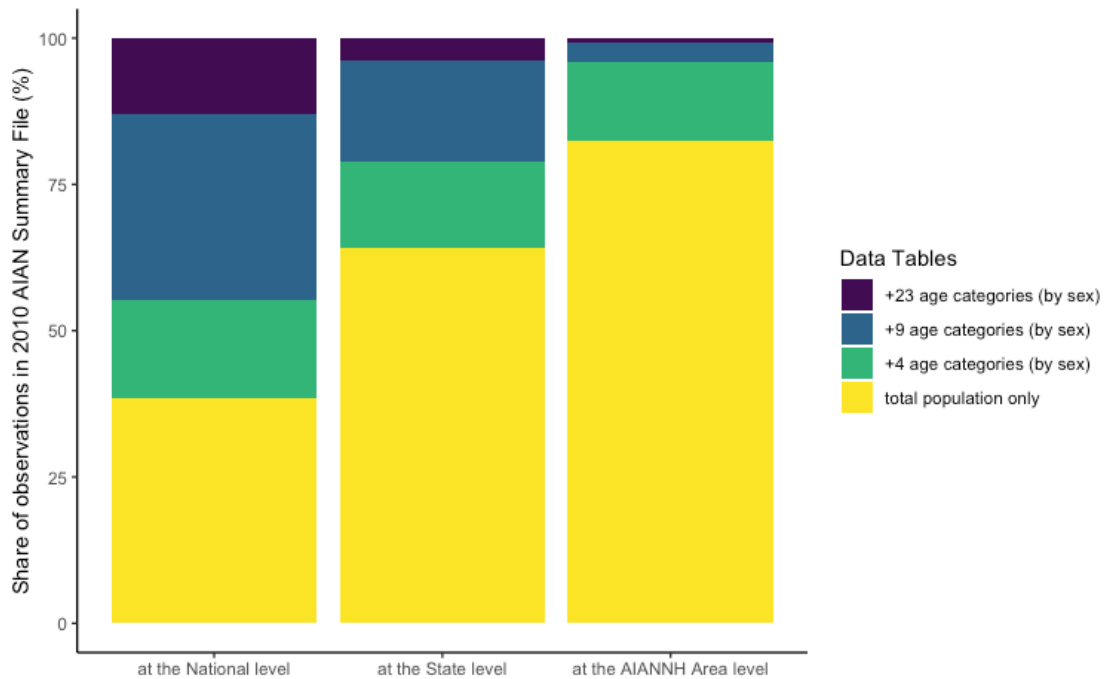
Notes: The y-axis reports the share of the 2010 American Indian and Alaska Native Summary
File that would have been placed in each category (total population only, four additional age
categories (by sex), nine additional age categories (by sex), and 23 additional age categories (by
sex)) if the Census Bureau had applied the census 2020 adaptive design. The first and second
stacked bar chart shows how the detailed tribal groupings would have been grouped based on
the national or state population thresholds. Those population thresholds are as follows: if a tribe
or tribal grouping had less than 499 people at the state or national level, the Census Bureau
would have published only the total population for that group. If a detailed population group
had between 500 and 999 living at the national or state level, four age categories would have
been published for each sex. If a tribe or tribal grouping had between 1,000 and 6,999 people
living at the national or state level, the Census Bureau would have published nine age
categories for each sex. And if a tribe or tribal grouping had 7,000 or more people living at the
national or state level, the Census Bureau will publish 23 age categories per sex. The second
stacked bar chart shows the shares of 2010 tribal groups in each category based on their 2010
population count in AIANNH Areas. The population thresholds at the AIANNH Area level are
as follows: if a tribe or tribal grouping had less than 999 people living in an AIANNH Area, the
Census Bureau would have published only the total population for that group. If a tribe or tribal
grouping had between 1,000 and 4,999 living in an AIANNH Area, four age categories would
have been published per sex. If a detailed population group had between 5,000 and 19,999
people living in an AIANNH Area, the Census Bureau would have published nine age
categories for each sex. And if a tribe or tribal grouping had 20,000 or more people living in an
AIANNH Area, the Census Bureau will publish 23 age categories per sex.