Impacts of Differential Privacy on data about American Indian and Alaska Native People and people on Tribal Lands

Carolyn A. Liebler, Ph.D.
Associate Professor
Department of Sociology &
Minnesota Population Center
University of Minnesota
Data on tribal lands & AIAN people is hard to get! Census data is often the only data source.

Core users include
- Tribal leaders for tribal governance, local governments, state and federal governments, researchers, judges, lawmakers, policy makers, advocacy groups

Some uses of these data
- Funding allocations, policy development, policy evaluation, planning
- Fertility rate denominators, other vital rate denominators
- Water rights and other litigation and judicial decisions
- Enforcement of equal employment, housing, and other anti-discrimination laws
Some Issues with the 2020 Census data

• Collected during a pandemic
• Threat of the “citizenship question” may have impacted response
• New coding system for race and ethnicity may have had a big impact
• Differential Privacy
Differential Privacy (DP)

This is the Census Bureau’s way of protecting the 2020 Census data
• It is a mathematical algorithm applied by a computer program
• It has been used in other contexts but is new for the Census Bureau
• It was used already on the 2020 results you have seen

The Census Bureau is thinking of using DP on other Census data, like the American Community Survey

To learn more about DP, see: https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/differential-privacy.html
Simplified Example of Differential Privacy

• The differential privacy algorithm adds or subtracts a number that’s drawn at random from a distribution of numbers.

• This is done to every number; the results do not always make sense.

How different can the census numbers be while still being useful?

<table>
<thead>
<tr>
<th>County</th>
<th>Real # of AIAN counted</th>
<th>DP adds/sub</th>
<th># AIAN in the public data</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ada County</td>
<td>550</td>
<td>+38</td>
<td>588</td>
<td>7%</td>
</tr>
<tr>
<td>Barhe County</td>
<td>1,234</td>
<td>-148</td>
<td>1,086</td>
<td>-12%</td>
</tr>
</tbody>
</table>

| Dear County    | 99                     | -99         | 0                         | -100%        |
| Etta County    | 345                    | +480        | 825                       | 139%         |
The Census Bureau’s example data show the Impacts of Differential Privacy on 2020 Census data about American Indian and Alaska Native People and people on Tribal Lands.
Abbreviations

**AIAN single-race** = People reported as American Indian and/or Alaska Native in the Census race question and no other races.

**AIAN AOIC** = People reported as American Indian and/or Alaska Native in the Census race question, whether alone or in combination with another race or races.

Note that this is self-described race; tribal membership is not required.
Figure 1: The example data show that differential privacy algorithm causes 2020 Census **AIAN single-race** county population sizes to be mis-reported.

For example: the algorithm allowed one county’s data in this state to appear to **double its AIAN population** and a different county in the state to appear that **all AIAN people were gone!**
Figure 2: The example data show that the differential privacy algorithm causes 2020 Census AIAN AOIC county population sizes to be mis-reported.
Figure 3: The example data show that differential privacy algorithm causes 2020 Census AIAN AOIC adult (age 18+) county population sizes to be mis-reported.
Figure 4: The example data show that differential privacy algorithm causes 2020 Census AIAN AOIC child (ages 0-17) county population sizes to be mis-reported. The data illustrate the impact on county population sizes across the state, with both smaller and larger total AIAN population size changes observed. The largest positive percentage differences range from 100% to 300%, while the largest negative percentage differences range from -200% to -150%, indicating a mis-reporting of county population sizes.
Summary of the impacts of differential privacy on county-level data about AIAN people in the 2020 Census

• Sometimes the 2020 Census data totals are similar to the actual count
• Sometimes, DP makes the population huge! Other times, the population is gone.

We do not currently have any way to officially know from the Census Bureau whether the data in a particular place are wrong.
  • If it looks like a ridiculous number, it probably is!
The Census Bureau’s example data show the Impacts of Differential Privacy on 2020 Census data about American Indian and Alaska Native People and people on Tribal Lands
Tribal lands

Reservations and other land areas defined and used by tribal governments.

History is complex so there are many legal statuses of lands.

• People of any race(s) in these areas
• Borders cross state and county boundaries

Tribal governance requires accurate, geographically-detailed information at the tribal tract level.

Tribal census tracts are commonly used for governance. They nest within tribal lands and have a minimum population of 1,200, a maximum population of 8,000, and an optimum population of 4,000.
In some Tribal Land Areas (even large ones), the DP algorithm makes it seem like there are no AIAN people! Or that there are 4x as many as were actually counted in the 2020 Census.

Figure 5: The example data show that differential privacy algorithm causes 2020 Census AIAN (single race) and AIAN AOIC population sizes in Tribal Land Areas to be mis-reported.

<table>
<thead>
<tr>
<th>Total Population of Tribal Land Area</th>
<th># of Tribal Land Areas of this size</th>
<th>AIAN single race</th>
<th>AIAN AOIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Largest negative % difference</td>
<td>Largest positive % difference</td>
</tr>
<tr>
<td>10 to 199 people</td>
<td>381</td>
<td>-100%</td>
<td>300%</td>
</tr>
<tr>
<td>200 to 999 people</td>
<td>352</td>
<td>-100%</td>
<td>400%</td>
</tr>
<tr>
<td>1,000 to 1,999 people</td>
<td>245</td>
<td>-60%</td>
<td>300%</td>
</tr>
<tr>
<td>2,000 to 2,999 people</td>
<td>310</td>
<td>-43%</td>
<td>35%</td>
</tr>
<tr>
<td>3,000 to 3,999 people</td>
<td>254</td>
<td>-60%</td>
<td>27%</td>
</tr>
<tr>
<td>4,000 or more people</td>
<td>471</td>
<td>-22%</td>
<td>50%</td>
</tr>
</tbody>
</table>
Final thoughts

• Data on Tribal Lands & AIAN people is hard to get! Census data is often the *only* data source.

• The 2020 Census data about AIAN people and Tribal Lands has been deeply damaged by Differential Privacy.

• The **Census Bureau needs to hear from Tribal Leaders** about:
  • Whether Tribal Leaders want DP to be applied to other data like ACS.
  • Impacts of the decisions that have already been made.
Census Bureau Tribal Affairs Team

301-763-6100
OCIA.TAO@census.gov

TRIBAL AFFAIRS TEAM
Dee Alexander, Tribal Affairs Coordinator
B. Kawe Saafi, Tribal Affairs Specialist
Melissa Bruce, Intergovernmental Affairs Specialist
I am grateful to Randall Akee and Yvette Roubideaux for conversations supporting this presentation.

Differential privacy demonstration data were provided by the Census Bureau and distributed by IPUMS NHGIS at [https://www.nhgis.org/privacy-protected-2010-census-demonstration-data](https://www.nhgis.org/privacy-protected-2010-census-demonstration-data)

- Finalized DP settings for 2020 redistricting data, released June 8, 2021, should be cited as: David Van Riper, Tracy Kugler, and Jonathan Schroeder. IPUMS NHGIS Privacy-Protected 2010 Census Demonstration Data, version 20210608 [Database]. Minneapolis, MN: IPUMS. 2020.