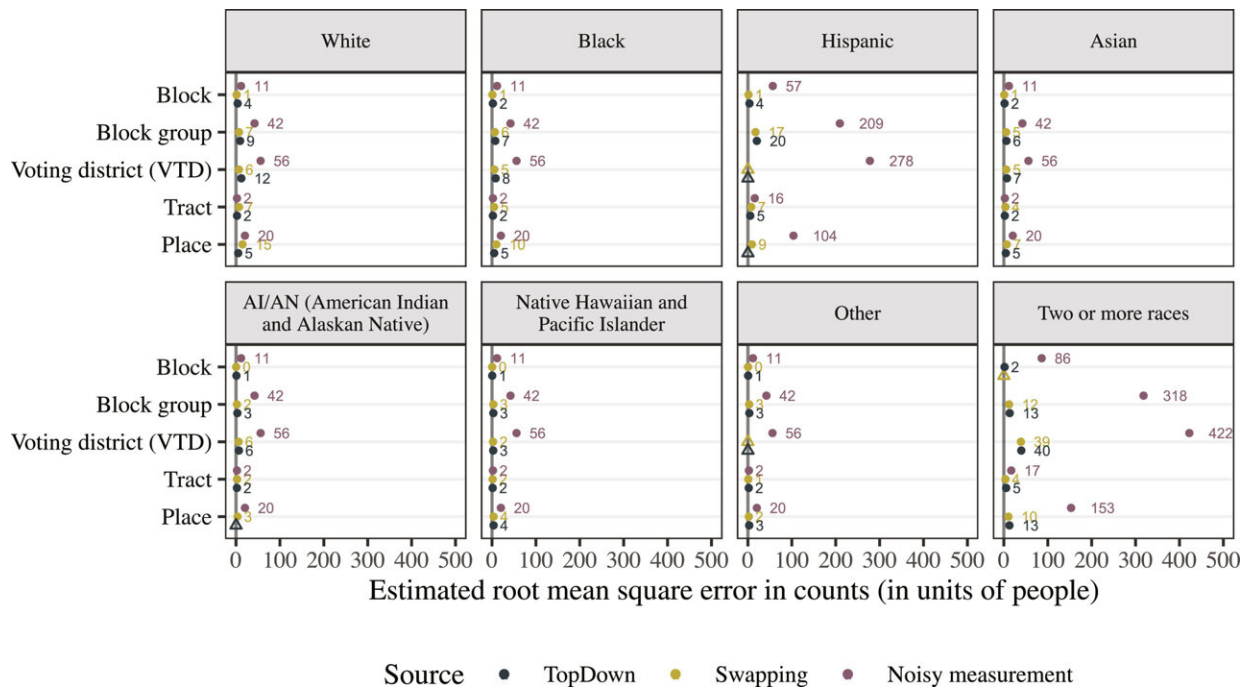


Study of new method used to preserve privacy with US census data suggests accuracy has suffered

May 6 2024, by Bob Yirka



Estimated root mean square error (RMSE) for population counts of a race/ethnicity group, at each geographic level. The RMSE quantifies the average magnitude of error for a given geography for a particular geographic unit. Triangles for RMSE indicate that the estimated mean square error was negative and hence was set to zero. Credit: *Science Advances* (2024). DOI: 10.1126/sciadv.adl2524

A small team of political scientists, statisticians and data scientists from Harvard University, New York University, and Yale University, has found that by switching to a new method to better protect privacy, the U.S. Census Department has introduced factors that reduce accuracy in some cases.

In their [paper](#) published in the journal *Science Advances*, the group describes how they analyzed a file provided by Census officials to measure [accuracy](#) in publicly available census data and their results.

Prior to the 2020 U.S. census, officials with the U.S. Census Bureau worried about the privacy of the people who provide answers to the census, opted to change the method by which they ensured [data security](#).

The old method was called, "swapping." It involved swapping data from people living in one block of a city with people in another block, thereby preventing people from being identified based on their data. The new method is called "differential privacy" and it involves adding what the Bureau describes as "noise" to each piece of data that is collected.

In this new effort, the research team could find no instance of an outside entity conducting research to determine if the new method did indeed provide more privacy or if the processed data was more or less accurate than had been the case when swapping was used. So, they began one of their own.

The study began with the research team asking the Census Bureau to give them access to what is called the noisy measurement file (NMF)—the one used for the 2020 census. The Bureau denied the request, which led the team to sue them. Eventually, the [lawsuit](#) was dropped when the Bureau agreed to give the team the NMF associated with the much smaller 2010 census—one that was carried out as a way to test the new method, and involved both swapping and differentiating.

The researchers then analyzed that file as a way to study the impact on accuracy of changing to the new system. In so doing, they found that overall, the two systems provided roughly equal accuracy on a broad scale. But they also found evidence of a reduction in accuracy at the block level of a type that could adversely impact minorities and multiracial populations.

More information: Christopher T. Kenny et al, Evaluating bias and noise induced by the U.S. Census Bureau's privacy protection methods, *Science Advances* (2024). [DOI: 10.1126/sciadv.adl2524](https://doi.org/10.1126/sciadv.adl2524)

© 2024 Science X Network

Citation: Study of new method used to preserve privacy with US census data suggests accuracy has suffered (2024, May 6) retrieved 18 July 2024 from <https://phys.org/news/2024-05-method-privacy-census-accuracy.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.