

Debate over new census privacy measures overlooks larger issues with data error in Title I funding

25 August 2022



Credit: Pixabay/CC0 Public Domain

Controversy surrounds the U.S. Census Bureau's new measures to preserve privacy, but a new study examines how existing data error can pose an even larger problem for evidence-based policies. The cornerstone of the Census Bureau's updated privacy measures, differential privacy, requires injecting statistical uncertainty, or noise, when sharing sensitive data. Scholars, politicians, and activists have raised concerns about the effect of this noise on crucial uses of census data. Yet most analyses of trade-offs around differential privacy overlook deeper uncertainties in census data. In a new study, researchers examined how education policies that use census data misallocate funds as a result of statistical uncertainty.

The study found that misallocations due to noise injected for privacy can be small or negligible, compared to misallocations due to existing sources of data error such as misreporting or non-response. But the study also finds that simple policy reforms could help funding formulas address unequal distribution of uncertainty from data error and smooth the way for new privacy protections, offering an avenue for compromise between

targeted policy, equity, and better privacy protections.

The study, conducted by researchers at Carnegie Mellon University (CMU) and published in *Science*, focuses on Title I of the Elementary and Secondary Education Act, which provides financial assistance to school districts with high numbers of children from [low-income families](#) to help ensure that all children meet state education standards. Federal funds are allocated through formulas based primarily on Census estimates of poverty and the cost of education in every state. In 2021, the U.S. government appropriated more than \$16.5 billion in Title I funds to more than 13,000 [school districts](#) and other local education agencies.

In this study, researchers used an exact simulation of the Title I allocation process to compare the policy impacts of noise injected for privacy to the impacts of existing statistical uncertainty. Specifically, they compared the impacts of quantified data error and of a possible differentially private noise injection mechanism. For example, of the \$11.7 billion in 2021 Title I funds this study examined, \$1.06 billion were allocated away from some districts in an average run of the simulation due to data error alone. This figure increased by just \$50 million when the researchers injected noise to provide relatively strong privacy protection.

"We paid special attention to the way Title I implicitly concentrates the negative impacts of statistical uncertainty on marginalized groups," explains Ryan Steed, a Ph.D. student at CMU's Heinz College, who led the study. "Weakening privacy protection does little to help these groups, and for them, participating in a Census survey can be especially risky."

The results show that misallocations due to

statistical uncertainty particularly disadvantage marginalized groups (e.g., Black and Asian students; districts with large populations of Hispanic students). Whether a demographic group lost funding depended on whether its members tended to live in high- or low-poverty districts, including those in denser, usually urban districts.

"However, we also identified policy reforms that could reduce the disparate impacts of both data error and privacy mechanisms," notes Steven Wu, assistant professor at CMU's School of Computer Science. "For example, using multi-year averages, rather than estimates from a single year, decreased both overall misallocation and disparities in outcomes."

Among the study's limitations, the authors point out their study does not account for systematic undercounts and many other unquantified forms of statistical [uncertainty](#) that affect poverty estimates, including previous measures to protect privacy such as data swapping.

"Our results suggest that the impacts of differential [privacy](#) relative to other sources of error in [census data](#) could be minimal," notes Alessandro Acquisti, professor of information technology and [public policy](#) at CMU's Heinz College, who coauthored the study. "Simply acknowledging the effects of data [error](#) could improve future policy design for both funding formulas and avoiding disclosure."

More information: Ryan Steed et al, Policy impacts of statistical uncertainty and privacy, *Science* (2022). [DOI: 10.1126/science.abq4481](https://doi.org/10.1126/science.abq4481)

Provided by Carnegie Mellon University

APA citation: Debate over new census privacy measures overlooks larger issues with data error in Title I funding (2022, August 25) retrieved 28 November 2022 from <https://techxplore.com/news/2022-08-debate-census-privacy-overlooks-larger.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.