PNAS LETTER

Census officials must constructively engage with independent evaluations

Christopher T. Kenny[a], Cory McCartan[b], Tyler Simko[a], and Kosuke Imai[a,c,1]

Current and former Census Bureau officials Jarmin et al. (1) argue that differential privacy, which underlies the 2020 Census's Disclosure Avoidance System (DAS), satisfies more desirable theoretical criteria than alternatives. They provide detailed criticisms of many published evaluations of the 2020 DAS, including our work. In this letter, we show that their criticisms are unfounded, grossly mischaracterize our research, and ignore critical issues that merit public discussion.

First, there are several points of agreement. We have never disputed that differential privacy has certain known theoretical properties. Like the advocates of differential privacy, including Jarmin et al. and Bun et al. (2), we also believe that any DAS should make statistical inference possible. The key question is how to balance data accuracy and privacy protection (3).

Unlike Jarmin et al., who make theoretical arguments, we have focused on empirical evaluations of the 2020 DAS (3–6). Independent evaluations are essential for improvements of the DAS, as evidenced by previous work that identified issues with earlier implementations. For example, Kenny et al. (4) documented undercounting in racially heterogeneous areas, which was acknowledged by the Bureau and addressed in the final 2020 Census release, contrary to claims by Jarmin et al. (7).

While this represents an encouraging example of transparent interactions between academics and policymakers, it is disappointing to see factual errors and mischaracterizations of our work by Jarmin et al. For example, the authors falsely accuse us in (4) of using "statistics that fundamentally overstate the error" by calculating percent changes, which were never used in any of our analyses. Jarmin et al. also misunderstand us in (3) as attempting to measure disclosure under their preferred "counterfactual" approach, when instead we demonstrated that small-area data increases prior-to-posterior disclosure risk for individuals' racial identification.

Other disagreements are more substantive. Jarmin et al. advance a false dichotomy between "statistical inference" such as BISG and "reconstruction" of microdata, alleging that only the latter constitutes disclosure. But reconstruction is a form of statistical inference. The difference between BISG and the Census reconstruction experiment is that BISG does not incorporate constraints from published table margins, while reconstruction does not integrate auxiliary data using Bayes' rule. Inference combining both approaches would be yet more predictive of individual microdata records.

Whether either of these approaches constitutes disclosure is a separate question. The answer depends on the correspondence between inferred records and confidential microdata, and a particular definition of disclosure. Indeed, the Census' own reconstruction "attack," used to motivate the new DAS, did not measure disclosure the way Jarmin et al. now define it. Depending on the definition of disclosure, however, different inferential approaches may pose heterogeneous privacy risks, such as for minority groups or people with unique combinations of attributes.

In their post-hoc defense of the 2020 DAS, Jarmin et al. have ignored key facts and obscured the diversity of viewpoints on both privacy and data utility. As differential privacy is being considered for broader use in the federal government, debate on these points, and independent researchers' empirical evaluations, remain essential (8).

Author affiliations: [a]Department of Government, Harvard University, Cambridge, MA 02138; [b]Center for Data Science, New York University, New York, NY 10012; and [c]Department of Statistics, Harvard University, Cambridge, MA 02138

Author contributions: C.T.K., C.M., T.S., and K.I. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: imai@harvard.edu.

Published March 5, 2024.

1. R. S. Jarmin *et al.*, An in-depth examination of requirements for disclosure risk assessment. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2220558120 (2023).
2. M. Bun *et al.*, Statistical inference is not a privacy violation (2021). https://differentialprivacy.org/inference-is-not-a-privacy-violation.
3. C. T. Kenny *et al.*, Comment: The essential role of policy evaluation for the 2020 census disclosure avoidance system. *Harv. Data Sci. Rev.*, 10.1162/99608f92.abc2c765 (2022).
4. C. T. Kenny *et al.*, The impact of the US Census disclosure avoidance system on redistricting and voting rights analysis. *Sci. Adv.* **7**, 1–17 (2021).
5. C. McCartan, T. Simko, K. Imai, Making differential privacy work for census data users. *Harv. Data Sci. Rev.*, 10.1162/99608f92.c3c87223 (2023).
6. C. T. Kenny, C. McCartan, S. Kuriwaki, T. Simko, K. Imai, Evaluating bias and noise induced by the U.S. Census Bureau's privacy protection methods. arXiv [Preprint] (2023). https://arxiv.org/pdf/2306.07521.pdf (Accessed 10 February 2024).
7. U. S. Census Bureau, Census bureau sets key parameters to protect privacy in 2020 census results (2021). https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates/2021-06-09.html. Accessed 10 February 2024.
8. J. R. Biden, Executive order on the safe, secure, and trustworthy development and use of artificial intelligence (2023). https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/. Accessed 10 February 2024.