Rejoinder: We Can Improve the Usability of the Census Noisy Measurements File

Cory McCartan[1] Tyler Simko[2] Kosuke Imai[2,3]

[1]Center for Data Science, New York University, New York, New York, United States of America,
[2]Department of Government, Harvard University, Cambridge, Massachusetts, United States of America,
[3]Department of Statistics, Harvard University, Cambridge, Massachusetts, United States of America

We thank John Abowd for providing comments on our article (McCartan et al., 2023a). We appreciate the opportunity to directly engage with an architect of the 2020 Census Disclosure Avoidance System (DAS) who played a central role in its development and implementation.

In our article (McCartan et al., 2023a), we offered specific recommendations to the U.S. Census Bureau on how the Noisy Measurements File (NMF)—a key output of the new privacy protection system developed for the 2020 Census—can be improved. In his commentary, Abowd (2024) provides valuable insight into the decision-making process during the design and implementation of the new DAS. He suggests that future research should prioritize improvements to the final data products instead of intermediate DAS components like the NMF. Below, we clarify a point of agreement between Abowd and us, and reiterate two main points of our original article that were not directly discussed by Abowd in his commentary.

First, Abowd and we agree that the NMF can be made easier to use through changes, including additional processing, more efficient formatting, and improved documentation. As Abowd notes, the NMF was not originally designed for public use, documentation was not prioritized, and working with the NMF in its current form requires a working understanding of a complex DAS code base used by the Census Bureau. These complexities only make it less likely that researchers will use the NMF to "inform the design of future decennial census data products in a meaningful way" (Abowd, 2024). The NMF is no longer a purely experimental product, but instead a core part of a new privacy protection system. We encourage the bureau to constructively engage with public feedback like ours to improve future NMF releases, as we are nearly halfway to the 2030 Census.

Second, we reiterate that there are tangible, immediate, and relatively straightforward steps the Census Bureau, possibly through collaboration with academic researchers, can take to reprocess the NMF to facilitate its broader use. Specifically, we describe a procedure in our original article ([recommendation no. 3](#)) that would drastically reduce the number of statistics from the trillions in the published NMF to the billions in the published data. This sort of preprocessed release would largely remove the current requirement that NMF users be familiar with the large and complex DAS codebase. Further, possibly through collaboration with academic researchers, the bureau can release premade assignment files that enable users to link NMF geographies with existing census geographies users are already familiar with, as we describe in our original article. Recommendations like these are relatively easy to implement and yet drastically improve the usability of NMF.

Third, improving access to new DAS products will facilitate research on ways to understand the implications of the DAS and improve its design for future censuses. There are two ways in which the improved access to the NMF can help inform the future design of the DAS. First, as demonstrated in a separate paper (Kenny et al., 2024), the NMF can be used to evaluate the potential bias and noise of various disclosure avoidance systems, including those of the 2010 and 2020 Censuses. Second, the improved access to the NMF

can help to inform the design of future iterations of the NMF and DAS. As Abowd notes, the complex structure of the current NMF was largely induced by the DAS constraints placed by bureau senior executives. The complications further increased once it was determined the NMF must be publicly released, as it was not originally designed to be. These policy growing pains need not be repeated. Additional evaluation from both inside and outside of the bureau can improve the release of future products, but only the bureau can ensure that the published NMF is usable by the wide range of stakeholders impacted by its decisions. Independent evaluation by various stakeholders can help the bureau and census data users understand how newer releases like the DHC-A may serve as a model for the census moving forward.

The NMF is an essential part of the new disclosure avoidance system adopted by the U.S. Census Bureau. As Abowd (2024) himself acknowledges, however, the current release of NMF is not suited for public use due to its complexity and lack of proper documentation. Fortunately, as shown in our original article, the bureau can take relatively straightforward steps to address these issues (McCartan et al., 2023b). As suggested by Abowd (2024), further improvements to the NMF could also be made through collaborations between academics and the bureau. These efforts could build on successful existing collaborations, such as the multiple rounds of feedback and improvement during the development of the new DAS (e.g., Kenny et al., 2021a, 2021b).

Disclosure Statement

Cory McCartan, Tyler Simko, and Kosuke Imai have no financial or non-financial disclosures to share for this article.

References

Abowd, J. (2024). Noisy measurements are important, the design of census products is much more important. *Harvard Data Science Review, 6*(2). https://doi.org/10.1162/99608f92.79d4660d

Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E., Simko, T., & Imai, K. (2021a). *The impact of the U.S. Census Disclosure Avoidance System on redistricting and voting rights analysis*. ArXiv. https://doi.org/10.48550/arXiv.2105.14197

Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E., Simko, T., & Imai, K. (2021b). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census. *Science Advances, 7*(7). https://doi.org/10.1126/sciadv.abk3283

Kenny, C. T., McCartan, C., Kuriwaki, S., Simko, T., & Imai, K. (2024). Evaluating bias and noise induced by the U.S. Census Bureau's privacy protection methods. *Science Advances, 10*(18). https://doi.org/10.1126/sciadv.adl2524

McCartan, C., Simko, T., & Imai, K. (2023a). Making differential privacy work for census data users. *Harvard Data Science Review, 5*(4). https://doi.org/10.1162/99608f92.c3c87223

McCartan, C., Simko, T., & Imai, K. (2023b). Researchers need better access to US Census data. *Science, 380*(6648), 902–903. https://doi.org/10.1126/science.adi7004