PNAS LETTER

The counterfactual framework in Jarmin et al. is not a measure of disclosure risk of respondents

Krishnamurty Muralidhar[a], Steven Ruggles[b], Josep Domingo-Ferrer[c], and David Sánchez[c,1]

Jarmin et al. (1) suggest assessing disclosure risk by using a counterfactual method to compare the posterior-to-posterior probability of an inference with and without the target record. They argue that this methodology is superior to the absolute and relative risk assessment methodologies. The counterfactual method, as originally proposed in ref. 2, explicitly rejects the "with and without" target record comparison. Since the counterfactual formulation in ref. 1 uses this inappropriate comparison, their methodology is inextricably linked to differential privacy (DP) but without rigorous formalization, making it impossible to establish a privacy guarantee or prove it satisfies the desiderata.

What is clear is that, unlike the other two methodologies being considered in ref. 1, the counterfactual method does not measure risk to individual respondents. Rather, it assesses whether a protection algorithm satisfies a DP-like requirement. If it does, no one is at risk; if it does not, everyone is at risk. To illustrate this fallacy, consider a population where every respondent is identical and therefore protected against disclosure. The counterfactual methodology in ref. 1 would deem every respondent to be at risk (whereas the preferred option in ref. 2 would not). Having already adopted this measure for the controversial 2020 US Decennial Census (3), Jarmin et al. are attempting to impose a questionable standard by diktat.

The appendix in ref. 1 criticizes four studies based on methodological issues. We now address this criticism.

(1) Ruggles and Van Riper (4) used a simple Monte Carlo simulation to estimate a baseline for evaluating the effectiveness of the Census Bureau's database reconstruction experiment. Jarmin et al. argue that the simulation is invalid because the Census experiment included a previously undocumented rule that "a record in the reconstructed data can be assigned to at most one record in the confidential data." However, it does not make much difference; if the simulation is modified to use each record no more than once, it remains the case that most of the reconstructed individuals have no match in the real population, and most of the matches that do occur would be expected purely by chance.

(2) Muralidhar (5) was trying to show that the reconstruction approach used by the Census was unnecessarily complicated. Criticizing him for using a simpler schema than the one used by the Census is missing the very point of the paper.

(3) Criticizing Francis (6) for not accurately predicting non-modal race/ethnicity is also missing the point. The very purpose of his paper was to show that race/ethnicity for individuals can be predicted accurately with knowledge only of the modal block value.

(4) The criticism of Muralidhar and Domingo-Ferrer (7) rests on the false claim that they assume that suppression methods were used in the Summary File 1 in 2010 tabular data release. Nowhere in their paper do they make that assumption.

In summary, at no point does ref. 1 refute the key conclusion of these studies that in the 2010 Census reconstruction a) most of the matches were random and b) that the reconstruction is primarily due to generalizable inference rather than privacy-violating inference.

ACKNOWLEDGMENTS. Partial support for this work has been received from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871042 ("SoBigData++"), the MCIN/AEI/10.13039/501100011033/FEDER, UE (project PID2021-123637NB-I00, "CURLING"), INCIBE and European Union NextGenerationEU/PRTR (project "HERMES" and INCIBE-URV cybersecurity chair) and the Government of Catalonia (ICREA Acadèmia Prizes to J.D.-F. and D.S. and grant 2021SGR-00115). The third and fourth authors are with the UNESCO Chair in Data Privacy, but the views in this paper are their own and are not necessarily shared by UNESCO.

Author affiliations: [a]Department of Marketing and Supply Chain Management, University of Oklahoma, Price College of Business, Norman, OK 73019; [b]University of Minnesota, Institute for Social Research and Data Innovation, Minneapolis, MN 55455; and [c]Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, United Nations Educational, Scientific and Cultural Organization Chair in Data Privacy, CYBERCAT-Center for Cybersecurity Research of Catalonia, Tarragona, Catalonia E-43007, Spain

Author contributions: K.M. and S.R. designed research; K.M., S.R., J.D.-F., and D.S. performed research; K.M., S.R., J.D.-F., and D.S. analyzed data; and K.M., S.R., J.D.-F., and D.S. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: david.sanchez@urv.cat.

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. D. Kifer *et al.*, Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 census. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2209.03310 (Accessed 30 December 2023).
3. J. M. Abowd, M. B. Hawes, Confidentiality protection in the 2020 U.S. census of population and housing. *Annu. Rev. Stat. Appl.* **10**, 119–144 (2023).
4. S. Ruggles, D. Van Riper, The role of chance in the census bureau database reconstruction experiment. *Popul. Res. Policy Rev.* **41**, 781–788 (2002).
5. K. Muralidhar, "A Re-Examination of the Census Bureau Reconstruction and Reidentification Attack" in *Proceedings of Privacy in Statistical Databases–PSD 2022*, J. Domingo-Ferrer, M. Laurent, Eds. (Springer, 2022), pp. 312–323.
6. P. Francis, A note on the misinterpretation of the US census re-identification attack. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2202.04872 (Accessed 30 December 2023).
7. K. Muralidhar, J. Domingo-Ferrer, Database reconstruction is not so easy and is different from reidentification. *J. Off. Stat.* **39**, 381–398 (2023).