WHOSE 2010 CENSUS RESPONSES CAN BE RECONSTRUCTED WITH CERTAINTY?

Aloni Cohen and JN Matthews
University of Chicago

The US Census Bureau is required by law to protect the confidentiality of individual respondents. In the lead-up to the 2020 Census, the Bureau came to view the disclosure avoidance system used in 2010 as vulnerable to *database reconstruction attacks*, whereby the microdata underlying a set of published tables "is recovered merely by finding a set of microdata that is consistent with the published statistical tabulations" [4].

There has been debate about the true extent of the disclosure risks for the 2010 Census, from the tables in Summary File 1 (SF1).[1] Census Bureau claimed reconstructed microdata perfectly matched 46% of the confidential microdata [1]. Ruggles and Van Riper argue that the 46% rate is uninterpretable absent "a null model of random guessing" [6]. Muralidhar argues that, moreover, the system of equations underlying database reconstruction for the 2010 Census can be so under-constrained that any individual solution is meaningless [5]. Abowd et al. respond, showing that reconstruction beats random guessing, and that the reconstruction system of equations is often fully determined [2]. Namely, 70% of blocks (with 31% of the US population) have a single valid solution. Even so, the extent of the disclosures from reconstructing 2010 SF1 is still unclear. First, Abowed et al. only consider solution variability at the block level [2]. This underestimates the disclosure risk, as individual microdata may be fully determined even if the containing block is not. Second, they reconstructs *agebins*—a 38-bin schema reported in various Census tabulations—not single year *ages* [2]. This overestimates the disclosure risk, as uncertainty about 1-year age may remain.

In this work, we ask: What fraction of the (swapped) 2010 Census responses are fully determined by the 2010 tables, and how does it vary by demographics? Like [2], we are primarily concerned with what can be inferred with certainty. But we focus on individuals rather than blocks, and 1-year ages rather than agebins. Prior works sample feasible reconstructions, sometimes also optimizing a function of the result. We instead symbolically analyze the full space of feasible reconstructions at the tract level. This allows us to identify *frozen variables*: those whose value is fully determined. Each variable corresponds to the multiplicity of a *persona*—a combination of age×sex×race×ethnicity×block. E.g., one persona is ⟨16-year-old, female, Black-alone, Non-Hispanic, census block 371559602021013⟩. If the corresponding variable is frozen to 2, there are exactly 2 such microdata records in every possible reconstruction. Summing the values of all frozen variables gives the population whose microdata responses are fully determined.

Figure 1 shows results on the state of North Carolina in the 2010 SF1. Observe that the total frozen population in the sate is 17.5%. But this total obscures significant disparities among different groups. Overall, smaller minority populations are frozen at much higher rates. For example, over 80% of NHPI responses in NC are fully determined by the 2010 SF1 and individuals living in smaller census blocks have a greater risk of being frozen.

We also consider inferences that can be drawn directly from the tables without the complexity of full reconstruction. Specifically the P12 tables report *agebin* at the *block* level, and *age* at the *tract* level. When there is only one block in a tract with population in a given agebin we can easily infer the 1-year age at the block level. With this simple observation, we are able to recover age×sex×race×block for 1.3% of the US population, which is not released in any census publications. Table 1 shows the rate of direct 1-year inference broken down by race for the country and three states. Once more we see a large disparity in the vulnerability by race, and the smallest racial groups facing the greatest disclosure risk.

Above we restricted ourselves to conclusions that can be drawn about the microdata *with certainty*. We also extend the work of Dick et al [3], who explore the possibility of making high-confidence *membership inferences*—learning which personas have non-zero multiplicity. Solving a linear optimization problem to generate many (approximate) reconstructions, they rank order the resulting personas by frequency. We improve their methodology by changing both the reconstruction engine and the ranking. With these techniques we can improve on the accuracy and confidence of

[1] The SF1 tables aggregate microdata that has already been subjected to *swapping*. Perfectly reconstructing the data underlying the SF1 would only yield the swapped microdata. Our goal is to reconstruct as much of this swapped microdata as possible. For the remainder of this abstract, we mostly elide the swapped/unswapped distinction.

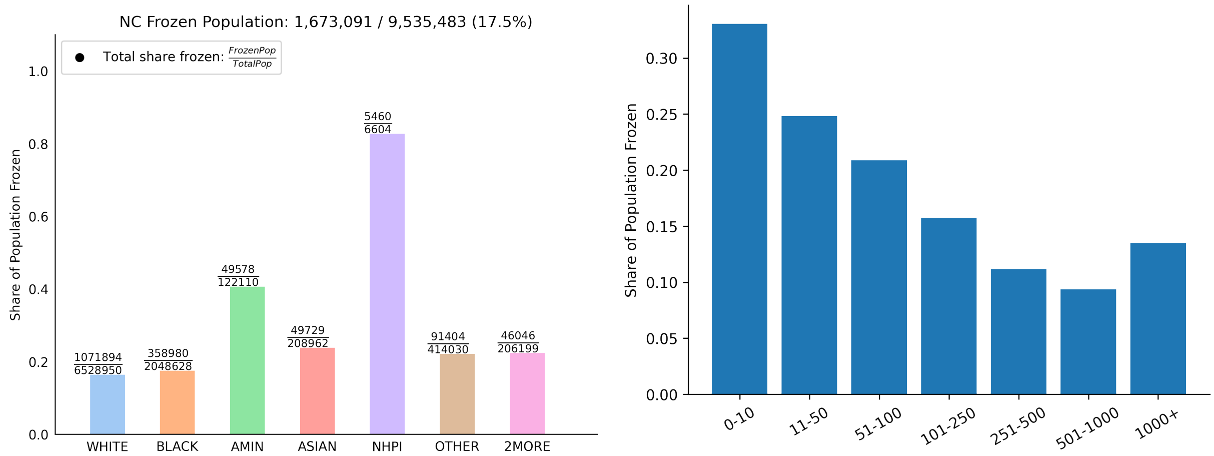

Figure 1. Variation in share of frozen population in North Carolina across race (left) and block size (right). The frozen population is computed by summing the values of frozen variables corresponding to all personas in that category.

	2010 population by race							
	Total (1,000s)	%White	%Black	%AIAN	%Asian	%NHPI	%Other	%2MORE
US	308,144	72.5	12.5	1.0	4.8	0.2	6.2	2.9
California	37,254	57.6	6.2	1.0	13.0	0.4	17.0	4.9
Kentucky	4,339	87.8	7.8	0.2	1.1	0.1	1.3	1.7
Wyoming	564	90.7	0.8	2.4	0.8	0.1	3.0	2.2

	Rate of direct 1-year age inference by race							
	%Total	%White	%Black	%AIAN	%Asian	%NHPI	%Other	%2MORE
US	1.324	0.089	2.092	22.924	4.901	39.65	3.53	8.855
California	1.368	0.171	4.92	27.237	1.267	43.87	0.97	4.118
Kentucky	1.368	0.056	4.332	69.555	20.634	72.211	18.07	18.115
Wyoming	1.334	0.003	26.98	12.32	32.106	74.941	7.912	12.046

Table 1. Results shown for the nation and the three states with largest (CA), median (KY), and smallest (WY) 2010 populations. Percentages in the upper table are fractions of the total population of the region. Percentages in the lower table are fractions of that group's population.

membership inferences as well as address the harder problem of guessing multiplicities of the personas, which Dick et al do not attempt.

References

[1] John Abowd. Apr. 2019. URL: `https://twitter.com/john_abowd/status/1114942713810640896`.

[2] John M Abowd et al. *The 2010 Census Confidentiality Protections Failed, Here's How and Why*. Tech. rep. National Bureau of Economic Research, 2023.

[3] Travis Dick et al. "Confidence-ranked reconstruction of census microdata from published statistics". In: *Proceedings of the National Academy of Sciences* 120.8 (2023), e2218605120.

[4] Simson Garfinkel, John M. Abowd, and Christian Martindale. "Understanding Database Reconstruction Attacks on Public Data: These attacks on statistical databases are no longer a theoretical danger." In: *Queue* 16.5 (Oct. 2018), pp. 28–53. DOI: `10.1145/3291276.3295691`.

[5] Krishnamurty Muralidhar. "A re-examination of the Census Bureau reconstruction and reidentification attack". In: *International Conference on Privacy in Statistical Databases*. Springer. 2022, pp. 312–323.

[6] Steven Ruggles and David Van Riper. "The role of chance in the Census Bureau database reconstruction experiment". In: *Population Research and Policy Review* (2021), pp. 1–8.