Editors’ and reviewers’ comments in regular type using page, section and line numbers from the version they reviewed; *our responses in italics using page, section and line numbers from both the original and resubmitted text.*

**Comments from the Editors:**

Thank you for this new version of the manuscript, which has much improved and I can see where this is going. As you see in both reviewers' comments, there are still parts where the writing is too lengthy and needs to be improved to allow readers to stay with the paper. Especially the practical implications mentioned by reviewer 2 are of great interest to this audience. Please take a careful look at the suggestions made by both reviewers and in your resubmission indicate what you changed (and why not if not, and how you addressed the point otherwise).

*Thank you for this summary. We have addressed the reviewers’ comments as noted below. In addition to the changes noted below, we found our own errors and corrected them in the following places:*

*Page 4, line 66: The correct number of 2010 census blocks with positive population is 6,207,027. (6,398,202 is the number of blocks with positive housing units or group quarters.)*

*Page 11, lines 339-343: The omitted tables P8 and P10 contained only race data (for all persons and for persons age 18 and over). Corrected here and in the supplemental text page 41, lines 1394-1401.*

*Page 25, line 818: The correct “smallest gain in reidentification precision relative to the proportional guesser” is 32.6% (not 36.8%). Footnote 18 points the reader to the correct cell in supplemental text Table 13.*

*Page 30, lines 920-923: “the reidentification precision rates of rMDFb,t and MDF vary between 21.5% and 21.8%” (not 19.7% and 21.8%). We added that “this is essentially identical to the precision rates for the proportional guesser (20.9% and 20.2%, for COMRCL and CEFatkr, resp.).” Finally, “[n]either the low nor high swap rate experiments deliver precision rates for this vulnerable population below 76.4%” (not 72.2%). Also corrected in the supplemental text at page 55, lines 1856-1864.*

**Comments from Reviewers:**

Reviewer 1: First of all, many thanks to the authors for submitting a completely new paper. Overall, in my view this shortened and streamlined version makes the research much more accessible to the wider audience and leaves out the nitty gritty technical details (not all of them, though, see below) for those interested in the supplementary PDF.

*Thank you. We have addressed your comments as noted below.*

Since almost none of my original comments are relevant to this new version, I'm not going to address them again, and rather review the new submission instead.

*Thank you. That was our intention, although we did address your previous comments, especially in the revised supplementary PDF.*

The abstract now states clearly the overall focus, very nice.

*Thank you.*

Section 1.2 is very clear with the exact research questions, very nice

*Thank you.*

Also lines 132+ are great, I like the critique of "major textbooks" on line 140!

*Thank you (we hope they motivate some revisions to those obsolete recommendations).*

Section 2.5 comes a bit too early, because you haven't showed your algorithm yet (L256: "the algorithm uses" - but what is it?)

*We revised Section 2 as follows. It now contains 8 subsections. Section 2.1 is unchanged. Section 2.2 describes all the inputs for the reconstruction studies (as requested by referee 2). The simple example is relocated to Section 2.3, and redundant definitions have been removed. Section 2.4 (new) describes the basic reconstruction algorithms. Sections 2.5 and 2.6 were edited to improve readability. The extended example was moved to Section 2.7, where it can more effectively build on the rest of the description in Section 2. Section 2.8 was edited for readability.*

I struggled with section 3.6 (you lost me in the first sentence already). Why are you doing this and why is it important? Compared to the previous text, this entire section is much harder to digest - maybe more explicit structure (paragraph headings, etc.) would help the reader?

*We have revised this section (now section 3.7) as follows: the top portion has been rewritten entirely and edits have been made throughout for clarity, and new paragraph headings have been included to provide the reader with a clearer structure and direction for the section.*

*In the same spirit, we shortened Sections 3.2 and 3.3 to focus on the important features of our work, referring the reader to the supplemental text for details we had previously included in the main paper. We left any statistics referenced by the editor or referees in their comments in the main paper.*

Also section 3.8 lacks some high-level intro: what is this again, why should we care, etc. It goes straight to details of the results, but after L724 with "CEF, HDF, rHDF\_b,t, rHDF\_b, MDG, and PRG" I gave up. Somehow I could follow the last part (L756 onwards), but not the technical details.

*We have revised this section (now section 3.9) for clarity by providing a clearer opening and substantially shortening this section to focus principally on the key results for the nonmodal population, referring the reader to the supplemental text for discussion of the corresponding analysis of the less interesting results for the modal populations.*

Section 3.9 is super interesting and quite clear.

*Thank you.*

Section 3.10 is a very important conclusion, but it feels buried now in the very technical "chapter" 3. Maybe move it to 5?

*Done. See page 33 lines 982-990,*

Minor suggestions or questions:

L56: "household-level record swapping" is the underlying method for the HDF which is the "gold truth" for further experiments - it would be great to explain exactly what that is (and show an example). Regarding this, I was wondering how much errors are introduced by the swapping. I'm assuming your linkage attacks with the commercial databases are not "swapped", so reconstructing "swapped" data must inherently underestimate the attacker accuracy (due to the inherent noise in swapped HDF). Do you think that, if you were to reconstruct the original clean data, your attack might in fact be much stronger?

*We agree with the sentiment of this comment, and we say explicitly (page 6, lines 178-185) that we did not attempt to undo the swapping, which would have improved the accuracy of the reconstructed data. Future research might attempt this, and we would expect the results to be stronger. Appendix C of the supplemental text goes provides substantial detail about the Census Bureau’s swapping algorithms and variations that we researched.*

L232: CEF ?! Overall, I like the reduction of abbreviations, but maybe there's more that can be un-abbreviated. Abbreviations make it really hard to digest.

*The acronym CEF is now defined in section 1.1 where the CEF and HDF are introduced.*

Overall, very nice and important work; recommend accepting it (with minor edits, see above)

*Thank you.*

Reviewer 2: Dear authors,

Thank you for the opportunity to review your manuscript entitled "A Simulated Reconstruction and Reidentification Attack on the 2010 U.S. Census." The paper presents a valuable contribution, and I appreciate the effort and depth of work that has gone into it. I have carefully reviewed the manuscript and compiled detailed comments to help improve its clarity, structure, and overall readability. Please find my comments below:

*Thank you for your review and comments.*

1. Paragraph Structure: Some paragraphs are quite long and would benefit from being broken into smaller, more digestible parts. For example but not limited to: Page 4 lines 92-114; Pages 6-7 lines 219 - 240; Page 16 lines 529-543; Pages 21-22 lines 718 - 741; Page 22 lines 742 - 765.

*Thank you, we agree that many paragraphs were too long. We have attempted to correct these throughout the manuscript.*

2. Table 1: Please add column titles. The first column is currently unclear — what does it represent?

*Done, here and in the supplemental text. See page 7, lines between 209 and 210.*

3. I suggest including a dedicated section introducing all data sources used in the study. This would help avoid repetition and improve flow. For instance, the information in Page 7 lines 203 and 204 and Page 6 line 184 and Page 6 line 173 and Page 7 line 205 are repetitions.

*Thank you for this suggestion. We did not do this for the following reasons. In reviews of earlier versions of this paper, possibly including yours, reviewers expressed frustration at not being able to distinguish between the reconstruction study and the reidentification study. For this reason, we limited the discussion of data sources in Section 2 to the ones used only in the reconstruction and its evaluation: Summary File 1, as summarized in Table 1, and the Hundred-percent Detail File, as described in the introduction and elaborated in Section 2.4. We describe the additional sources used for reidentification in Section 3.1 (new).*

*We do make frequent references to Table 1 in the supplemental text, which is exactly what you asked for.*

*That said, we also looked for redundancy in our descriptions—the ones you noted and others throughout the manuscript—and edited them out of the revised text.*

4. Page 7 lines 213 - 218 should refer to Table 1 using "Panel A" and "Panel B" terminology. Also consider include "b" and "t" to reflect the notation.

*We did this in the text. We believe that labeling the panels in the table using A, B, etc. is HDSR style. If that’s not the case, we can label them B and T.*

5. Geographic context: On Page 7 line 221, "At the tract level and above", please, provide a geographic description, as this may not be general knowledge for all readers.

*Done, and references provided. See pp. 8-9, lines 252-258.*

6. On Page 10 line 302, should "solvar" be capitalized?

*No, this is a math equation and solvar is the variable name.*

7. Reference needed: On Page 12 lines 352 - 356, please include a reference indicating where this information can be found.

*In HDSR style, the public replication archive, which includes all the information in this paragraph, is cited in the section “Data availability,” which appears at the end of the paper just before the references. We made note of this in the text at the paragraph you cited. See page 13, lines 404-405 and the “Source:” line in Table 3.*

8. On Page 14 lines 453, "the algorithms in this paper", I suggest adding the reference to the supplemental text.

*Done. See page 16, line 492.*

9. On Page 16 line 517, exclude "say" for a more formal tone.

*Done. See page 18, lines 555-556.*

10. In Section 3.4, specify which is Algorithm 1 and which is Algorithm 2 for clarity.

*Done. See page 18, lines 566-568.*

11. Page 17 line 563. Why is same italicized? Keep italics only for variable names or where strictly necessary.

*Done. We also removed all other uses of italics that were not definitions or variable names in the main and supplemental text with one exception: page 4, line 85. We italicized “prior to the 2010 Census” because we wish to stress that the Census Bureau* did *understand in the relevant time frame that the microdata in the HDF were insufficiently protected by swapping to be released without additional SDL.*

12. Page 18 line 592. The difference between {race, ethnicity} and "race x ethnicity" is unclear. Please specify.

*On first use of race x ethnicity (now in section 2.6), we have a footnote that explains the notation. See page 9, lines 271-274 and footnote 6. We also re-read the main and supplemental text to ensure that we used these consistently throughout.*

13. Page 18 line 625. The single sentence could be more effectively presented. Consider using a colon to introduce the following list.

*Done. See Section 3.7.2 on page 21, lines 680-683.*

14. On Page 19 line 626 the statement " the feature set {race, ethnicity} can also be inferred using the name of the target individual" is very strong. Please, provide a reference to support this claim.

*Done. See page 20, lines 684-686.*

15. Page 19 footnote 13. Please include the number of ties mentioned in the supplemental text.

*The main text and the supplemental text have identical discussions of the definition of the modal guesser and the relevant footnote. In neither place did we report the number of ties. We have added additional statistics, including the ones you requested, to this footnote (now 15) and the supplemental text that clarify what we did. See page 21, lines 705-708 and footnote 15.*

16. Page 19 footnote 13. Specify what is meant by "block-group-level".

*Done. Block group is defined in footnote 15 in Section 3.7.3 on page 21.*

17. Table 5: Clarify whether the values in Table 5 correspond directly to Equation 3.2 (confirmed) and Equation 3.3 (precision).

*The table displays the numerators only for putative and confirmed reidentifications. We clarified this in the main and supplemental text. See page 22, lines 744-745.*

18. On page 22, are there any bi- or multi- modal results? If so, how were they treated?

*The answer to this question is now in footnote 15 on page 21 (originally footnote 13) with additional explanatory detail. There are ties in the definition of the mode, and they were broken by using the mode at the next higher level of the geographic hierarchy. When this didn’t work, the national mode was used. There are no ties in the national mode (non-Hispanic White alone).*

19. Page 22 line 736 Capitalize "Equation".

*Thank you for catching this. HDSR style is Equation NN (capital E, no parentheses). We have corrected all equation references in the main and supplemental text.*

20. Highlighting key insight: Page 25 lines 804 - 806: this appears to be a crucial finding. Consider emphasizing it more.

Done. The paragraph now reads:

*Statistical baselines are either almost always wrong (for the modal guesser) or wrong for 4 of 5 persons (for the proportional guesser) in this vulnerable population. These results establish that the published SF1 tables result in clear confidentiality breaches. The high statistical precision of the reconstructed microdata is only possible because the target person's data were used in the SF1 tabulations. Had the target person not participated in the 2010 Census, these inferences would be impossible. Therefore, the results in Table 6 definitively establish that SF1 permits very high accuracy reidentifications of population-unique individuals with characteristics that differ from their neighbors. These are not statistical inferences, because they disappear if that person does not participate in the census. They are confidentiality breaches.*

*See page 26, lines 843-851.*

21. Page 26 lines 825 - 827: These are strong statements and would benefit from additional elaboration.

*It is elaborated in the cited reference (Jarmin et al.). We added the exact page number. See page 27, line 863. We also added a clarifying citation in the new paragraphs of Section 3.7 (page 18, lines 600-609) where we also give an example of this error (page 19, lines 618-625) and specifically identify it as the problem with Kenny et al. in footnote 12 at page 19, line 625.*

22. Page 31 line 915: Please add a reference to Section 2 of the 1965 Voting Rights Act.

*Done, at first reference to the law, on page 27, lines 865-866.*

23. Page 31 line 934: Standardize the reference style to match that used in the introduction.

*Done. See page 33, line 980.*

24. Page 35 line 1108 and Page 36 line 1141: Clarify why the month of the publication is included.

*The month of the publication was included automatically by the HDSR style in bibliographies. We removed the “month” from the .bib file. It no longer appears in the references.*

Overall, this paper has strong potential, and I hope these suggestions help improve its clarity and impact. I look forward to seeing the revised version.

*Thank you.*