Response to Reviewers

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

COMMENTS FROM THE EDITOR:

A big thank you to the author team for addressing the concerns of the reviewers. I would appreciate if the definition of data quality can be added as per suggestion of reviewer 2. Otherwise the manuscript is ready to go.

*Thank you very much. Our response to reviewer 2 is below.*

COMMENTS FROM REVIEWERS:

Reviewer 1:

Many thanks for addressing all points I raised previously. I mainly checked the parts I was not happy about and they now look OK, so I don't have any further objections.

*Thank you very much.*

Reviewer 2:

First of all, I would like to congratulate you on the revised version of your manuscript. The paper is now much clearer and more engaging—it reads very well. Since you have addressed all of my original comments, I will focus solely on the new submission.

*Thank you very much.*

My only remaining suggestion is to clarify in Sections 3.1 and 3.3 how you define data quality—specifically, what criteria or metrics you used to classify data as low or high quality. This is particularly important given that, in Section 3.1, you emphasize the low quality of the external data (lines 420 and 424), followed by a statement in Section 3.*3*, line 466, that "better external data were available," which prompted me to consider how the quality of these datasets was assessed. Providing this clarification would strengthen the manuscript.

*Thank you for this suggestion. We incorporated it as follows.*

*First, in the introduction to Section 3, we added two paragraphs that define the data requirements for an attacker database and state the relevant quality metrics. In particular, we characterize lower-quality and higher-quality attacking data by the extent to which the values of their linking variables match the values that would be reported as of April 1, 2010 by a census respondent. And throughout the manuscript we now refer to “lower-quality” and “higher-quality” attacker databases (COMRCL and CEFatkr, resp.) See lines 416-432 in the resubmission.*

*Then, in Section 3.1, we introduce the COMRCL attacker database and explain that in Section 3.3, we will show that it is a lower-quality attacker database (lines 436-443). We introduce CEFatkr,, and explain why it must be a higher-quality attacker database (lines 452-456).*

*Finally, in Section 3.3, new footnote 10 and lines 487-490 explain what we mean by* "better external data were available."

*We also made minor edits to the supplemental text so that the language was identical to the main text vis-a-vis these revisions.*