Author’s response to reviews

Title: Consent to Data Linkage in a Large Online Epidemiological Survey of 18-23 Year Old Australian Women in 2012-13

Authors:

Deirdre Mclaughlin (deirdre.mclaughlin@unimelb.edu.au)
Anna Graves (anna.graves@newcastle.edu.au)
Janni Leung (j.leung1@uq.edu.au)
Jennifer Powers (jenny.powers@newcastle.edu.au)

Version: 1 Date: 11 Feb 2019

Reviewer 1: The paper is well written. Opt-in consent and record linkage is an important topic.

A few caveats nevertheless: the authors claim that no prior study examined opt-in linkage requests in web surveys. This seems inconsistent with the literature cited in the conclusion and my reading of the literature. Joe Sakshaug, Public Opinion Quarterly, Vol. 78, No. 1, Spring 2014, pp. 166-176 is one example.

Response: We did not say this. However we would like to thank the reviewer for the reference which we have added to the introduction. ‘In a web survey of employment a small increase in consent rates was found when the time-saving benefit of linkage was mentioned.’
One weakness of the study is the absence of a probability-based survey. While the methods section refers to a different paper to learn more details, it should be noted in the context of this study why the sampling scheme does not negatively affect the conclusions drawn.

Response: To the extent that it is possible we have compared the sample with the 2011 Australian Census. ‘Comparison with the 2011 Australian Census showed that women in the sample were broadly representative of women of the same age nationally (Census 49.0% versus 52.6% aged 18-20; Census 74.5% versus 75.0% living in major cities excluding missing data) although a lower proportion of women only had school qualifications (Census 60.1% versus 51.4%).’ We have revised the conclusions and explained why the sampling scheme does not negatively affect the conclusions drawn.

Also the stronger effect of a gift compared to cash-incentive is surprising, given from what we find elsewhere in the survey literature. It would be interesting to know why the authors think this might be the case.

Response: We have clarified that ‘The leggings were very fashionable and highly desirable at the time of the survey.’

Recruitment strategies and incentives vary quite a bit across the data analyzed here. I would recommend parsing this out to avoid any confounders if possible.

Response: Recruitment and incentive variables have been included in the revised analyses.

Also a table or graph explaining the study design would help. Currently it is easy to get lost in the comments on earlier studies, earlier recruiting etc.

Response: We have rewritten the section on study design including recruitment and incentives. To further clarify we have deleted the comments on earlier studies and recruiting.
The model selection strategy is not entirely convincing. Since the authors do not have hypothesis on what effects the outcome (consent), and if those relationships are linear functions, I would suggest re-running the analysis and relying on non-parametric tree based models.

Response: We thank the reviewer for this suggestion and have re-run the analysis using Classification and regression trees (CART), a non-parametric model. The methods and results have been adjusted accordingly.

The result statement (69% consent) is a strong statement since appears as if the authors talk about a population estimate. Given the recruitment strategy, this seems not justifiable (neither do we know the probabilities of selection or the non-response mechanism and thus can't adjust for either).

Response: We have reworded the result statement to indicate that we are referring to the sample estimate and not the population estimate. We have also made some comparisons with 2011 Australian Census and showed that women in the sample were broadly representative of women of the same age nationally.

Reviewer 2: This study is timely. The following may improve the readability of the manuscript.

Methods: sample size

Response: We have clarified that the sample consists of 25,541 women who completed the online survey.

Abstract: consent to which administrative dataset? Demographics? SES? MRN numbers? You may want to use terms such as PHI or potential PHI which may infer level of privacy.

Response: We have clarified that they are health-related administrative datasets.
Background: It would be beneficial for the audience if you could clarify 1) what literature concretely reports about consent linkages in the online-based study like yours, 2) what literature concretely reports regarding consent linkages in the face-to-face studies. You somewhat mentioned but there is room to improve in clear communication. Both are completely different and irrelevant in some degree.

Response: We have revised the background to clarify literature that reported on the online-based or face-to-face consent linkages studies.

103: women- define age category
Response: We have clarified that we are evaluating differences between young women who did or did not provide consent to data linkage.

114: insert "length of questions and range of sample size (url)" instead of "details of the study can be found" phrase. It would look clearer for the audience with the same length of the sentence.
Response: We have rewritten the section on study design including recruitment and incentives.

135: specify age range
Response: The age range is already specified in the study design section of the methods. ‘In 2010, the ALSWH was provided with funding by the Australian Government Department of Health to recruit a new cohort of 18-23 year old women.’

198: report likelihood with numbers.
Response: Data have been re-analysed and the results rewritten accordingly.