**Reviewer’s report**

**Title:** Self-reported data in environmental health studies: Mail vs. Web-based surveys

**Version:** 0 **Date:** 18 Jul 2019

**Reviewer:** Morgan Millar

**Reviewer's report:**

Review of BMRM-D-19-00288

Self-reported data in environmental health studies: Mail vs. Web-based surveys

This is a study to compare mail vs. web survey response outcomes in a cross-sectional study of households in rural areas in Denmark. The study design entailed offering households the option of responding by paper or web. Modes were compared in respect to the impact of reminder letters on response rate increases, demographic makeup of responding individuals, and estimates from the survey of health symptoms and attitudes toward environmental stressors. This is a relevant topic of inquiry, and an interesting question to examine in this particular context. My specific suggestions, questions, and comments are detailed below.

**Abstract**

* Line 6-7: It is not clear exactly what is meant by "statistical aspects"; this phrase is somewhat vague. I also found the statement that the number of comparisons of web and mail survey responses is limited to seem strange, as this seems to be a popular topic of recent research.

**Background**

* Page 3, line 6: what is meant by "practicality"? Is this referring to cost-effectiveness? Or some other advantage of web surveys?
Page 3, line16-17: similar to comment made about abstract, I feel that there is more recent research (from the past decade) that could be incorporated regarding comparisons of web and paper survey modes. I would not classify this topic as one that has not been much explored. For example, a study that may be particularly relevant for comparing results is one from 2010 in a semi-rural area in the U.S.: Using the internet to survey small towns and communities: limitations and possibilities in the early 21st century, by Smyth et al. (published in American Behavioral Scientist). This article and subsequent work that cites it might be useful reference material to incorporate into the background as well as to draw on for comparisons in terms of discussion of this study's results.

Methods

Page 5, Line 9-10: can the authors explain the significance of surveying when application of animal slurry is restricted by law? The relevance of this may not be clear to all readers. Is this a timeframe chosen based on lower levels of environmental contaminants? Is this done to get a low-end estimate of health effects? Or is this relevant for optimal timing for the population to be responding to surveys? Or is there some other reason?

Page 7, Line 11-12: what values of the 5-point health symptoms scale were coded as low vs. increased frequency?

Page 8, lines 4-6: it seems like education level was not an adjustment variable in the multivariable models. This could be one of the most highly relevant demographics to control for when assessing differences in response by web and mail, since educational attainment is associated with Internet use.

Also, is there a reason that models did not adjust for region? Because the regions included were selected based on their different levels of environmental contaminants, this seems important to control for when predicting responses to questions on environmental attitudes and health symptoms.
Results

* Page 8 and Table 1: it would be helpful for readers if the Table and the text provided a more thorough overview of the basic overall response rate results. For example, include overall total response rate/percent responded for the survey rather than just the total number of responses. I would recommend starting the results with a paragraph focused solely on overall response patterns (survey as a whole, by region, and by mode) before going into the results before/after reminder letter.

* Similarly, it would also be helpful to see response rates by region in the Table.

* It would also be helpful to add percentages representing what percent of responses were by mail and what percent were by web—that is, to add percentages to the "Total" columns in the table (making each the percent of total responses received)

* Table 1, footnote b: were these two individuals' responses excluded from analysis? The statement is not clear about if they are included or excluded. I would think excluding them would be most appropriate.

* Page 8-9 (results comparing responses before/after reminder): My impression is that these results need to be reassessed/reconsidered from several angles, and in more detail. The interpretation of these results that is provided—that reminders had a greater effect on the web response than the mail response—is incomplete, and something that cannot be directly determined in a study in which respondents were offered both modes and self-selected mode of response.

* For example, rather than looking at the percent of responses within mode before/after the reminder, it might be more informative to look at percent of responses across modes before and after the reminder.

* Another interpretation of the results could be that more people responded by mail rather than web after the initial mailing because a paper questionnaire was provided, making that option more convenient. Upon receipt of a reminder, the convenience of a provided paper questionnaire was presumably no longer as applicable. At this point, for either mode, respondents would have to expend some effort to access the questionnaire: find the previously delivered paper questionnaire or go online to access the web version. Perhaps a greater proportion of web responses came after the reminder because that was more accessible for those who’d discarded the paper questionnaire.
I could encourage the authors to reexamine these results from various perspectives and consider additional explanations.

* Section 3.2/Table 2: Do the regions differ in terms of demographic profiles? E.g., are average age, education level, etc. similar across these regions?

* Section 3.3, Similar to last comment: It would be helpful if there was more information about the populations being studied. Are there any Census data or some other similar source of information on the demographics of the regions being studied that results could be compared to? For example, what is the average age of residents in these areas?

* Also, is there any way to know what percent of residences in these rural areas specifically have internet access? The discussion section mentions that 5% of the full Danish population does not access the internet daily, but it is highly probable that lack of internet access is lower in rural areas.

* Table 3: could overall characteristics of the full responding sample also be displayed?

Discussion

* Page 12, lines 1-10. The findings are indeed similar to other studies examining demographics and mode choice, but I feel some statements are more speculation than observed results (e.g., web respondents had more computing skills, retired persons are more accepting of posting mail). I'm not sure the delineation between observed trends and speculation are made clear.
Conclusion

* The conclusion suggests that the response differences observed across the survey modes is indeed a measurement difference between modes, and therefore implies hesitancy to combine mail and web survey results. An alternative conclusion, and a more optimistic one, is that the use of two modes might have attracted a more diverse set of respondents with different experiences. Without more information on the entire populations of interest, it isn't possible to fully assess this, but it is also not possible to conclude there is a true mode effect on responses when individuals are self-selecting response mode. The authors seem to dismiss the utility of their results. There could be more elaboration on possible interpretations.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published
**Declaration of competing interests**

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal