Reviewer’s report

Title: Unweighted regression models perform better than weighted regression techniques for respondent-driven sampling data: Results from a simulation study

Version: 0 Date: 12 May 2019

Reviewer: Luis Rocha

Reviewer's report:

This is an important study given that (I agree) the use of regression models in RDS data has increased in the medical literature however standards and routines are not well-tested and well-established. The authors consider various parameters and models, covering diverse practical scenarios, that is positive. For the same reason, I think the presentation of the paper should improve to facilitate understanding the rationale and the different analysis.

Some suggestions and comments:

I suggest that authors create for example a diagram to guide readers on the various cases (particularly for section "Data Analysis").

It is a bit confusing which tables and which graphs go to main text or appendix.

In p6, it would be interesting to report the homophily of the original real network for reference.

In p6, row 23-24. By more variation, do authors mean that they control the "variance" of the degree distribution, keeping the mean/median?

In p6, row 34. By population, do authors mean "a fixed set of parameters"?

In p6, row 34. Though "data sets" are correct, nicer to use "samples" since you are generating samples from the same set of parameters.

In p8, row 14. Not sure what you mean by knowing "precisely the degree". One may know the degree distribution with a "1-person" resolution but this does not mean that the degree (or network size) was accurately reported in the survey. Several studies have fine-tuned numbers (in contrast to ranges, that seems to be the point of the authors).

In p12, row 8. I think there is a typo in the sentence "prevalence, as shown in Figure 2. 8 in Appendix B details the"?

In p12, row 51. This is minor but I would rather write "suggest" or "recommend" instead of "prefer".

In p16, row 46. I think here the authors could make a strong conclusion and use "recommend" or "suggest" rather than "feel".

In p16, row 44. What do you mean by "very low prevalence"? 10% or less? 10% is not very low.

In p16, row 42. Not sure what authors mean by saying that prevalence is more likely to be known than homophily in hard-to-reach-populations. Homophily can be estimated using RDS.

In p16, row 17. Not sure I agree with "nice" number. Rounding does happen but collecting "non-nice" numbers is common as well. I am not aware of studies pointing that rounding (to 5, 10, 20 and 100) is the standard. In your Fig3, these numbers are not always the most relevant choice.
In Fig.4, I am not sure why this degree distribution does not resemble much Fig.3? Specially this valley between degrees 30 and 40 that looks unconventional. Could you comment on that? Have you tried any theoretical cases using for example the log-normal distribution with same mean (possibly standard deviation) as the real data?

In Table 3, I am not sure I understand the systematic under-estimation of the prevalence using RDS-II (and it's worse estimation than the naive method), that is as high as 20% lower for 10% homophily. Is this a result of some particularities of your network model?

The word "cluster" has multiple meanings in social networks, it may be related to clustering of traits (homophily) or links (either triangles or network communities). They affect RDS in different ways (see e.g. Rocha et al. (2017) Respondent-driven sampling bias induced by community structure and response rates in social networks. J Royal Statistical Society A). The authors should clarify these differences in the introduction. I understand that the study focuses on homophily but it would be interesting to include a short discussion on how clustering of links may affect the results (but I don't think new analysis should be done since this would be too much).

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.

Yes

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:

Acceptable

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