Reviewer’s report

Title: Sample size calculation for estimating key epidemiological parameters using serological data and mathematical modelling

Version: 0 Date: 12 Aug 2018

Reviewer: Andreas Halgreen Eiset

Reviewer's report:

The authors are commended for taking upon themselves the task of exploring mathematical models to estimate the seroprevalence, force of infection, etc. of infectious diseases in relation to assumed age-distribution. The manuscript presents a range of well thought-out figures and tables that improves the understanding.

Unfortunately, I find that the manuscript suffers from a lack of structure and clarity throughout.

Below are some of my major concerns.

In general
1. Both in the text and in legends several abbreviations are not explained, e.g. l. 165 "MISR"
2. Several redundant paragraphs e.g. ll. 228 - 233.
3. Many paragraphs becomes difficult to understand because of confusion in the use of past and present tense (and the unmotivated change in this) as well as unclear references to other parts of the text e.g. ll. 185-187 and l. 188, respectively.
4. Please state at the earliest possible what is meant by "key epidemiological parameters" - this is not universal nor self-evident.
5. The manuscript holds discussion in the results section, results in the methods section, etc. E.g. ll. 221-224
6. Fig. 1: I really like the idea of this and think the manuscript could improve with a much more elaborated version of this

Background
7. The section is unnecessary wordy and would gain from being more concise
8. Many statements are very generalising, e.g. ll. 69-70 that references a new review by the authors in a specific field. It is certainly not the case that in general study designs, even when limiting the scope to the medical field, are not "informed" by mathematical modelling. Another example of excessive generalisation is ll. 90-91.

Methods
9. Please elaborate on how the age specific prevalence obtained via the survey data is used in the models/simulations. As for now the description leaves the impression of indirect
standardisation, if this is indeed the case please state this
10. It is not clear what information (e.g. from the Belgian survey) goes into the models described
    (logistic, MISR, etc.) - how are the regressors chosen?
11. ll. 125-126: The authors state (correctly) that GLMs can be used to estimate sero-prevalence.
    In ll. 155-156 the authors chooses a logistic model. I do not think it can be assumed that every
    reader of the Medical Research Methodology can make the connection - I think the manuscript
    would benefit from a sentence making this clear
12. How does ll. 193-194 relate to the explanations in ll. 155-157?
13. R is cited in ll. 200-201. Has R only been used for the MLE-estimates? If so please reference
    the other statistical software used.

Results
14. In l. 237 and ll. 249-250 there is references to figures in the supplementary materials. These
    figures seems essential (are mentioned in second sentence in the results section!) in answering
    the purpose of the study and should not be "supplementary" but essential i.e. part of the main
    text.
15. l. 253: please elaborate on how "precision" is defined? As 0.5 * 0.95 = 0.475 confidence
    interval? This should be defined the very first time the expression is encountered in the
    manuscript. Also please elaborate on how/on what criterion is it decided if one model/age
    distribution is more precise than another?

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 recommend additional statistical review
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