Reviewer’s report

Title: Chronic Disease Population Risk Tool (CDPoRT): a study protocol for a prediction model that assesses population-based chronic disease incidence

Version: 0 Date: 18 May 2018

Reviewer: Mark Rutherford

Reviewer's report:

This paper documents a study protocol for a chronic disease prediction model to assess chronic disease incidence. The approaches that will be taken are fairly clearly described and in the large part well justified.

I have a few comments detailed below:

1. I agree that death free of a chronic disease is a competing risk for the first occurrence of any; but is it not also interesting to understand the breakdown of chronic diseases and also if multiple diseases are developed in certain individuals? If so, death from a lethal chronic disease will also be a competing risk and a more advanced modelling approach may be needed.

2. Following on from that, some are much worse outcomes than others and a lot have interlinked risk factors. I'm not 100% convinced that this composite outcome is the most informative outcome of choice? You do refer to multimorbidity as a potential sensitivity analysis - I cannot imagine that the same model will work well for diseases with highly correlated risk factors. Can you further justify the outcome choice and how you choose which are included?

3. Choice of validation approach - is it not better to use all of the data to build the model, and then use a repeated cross-validation technique?

4. For people with multiple records in the survey - will you update their covariate values if they differ from survey to survey?
5. Do you need to categorise BMI and other continuous covariates? Can you not fit continuously and non-linearly, and then report in groups later if needed for reporting purposes?

6. PH checking - you say you need a more interpretable model, and so will sometimes use PH - why? Some of the interpretability will be lost in other ways such as using splines for continuous covariates - transformations, graphical presentations and predictions of the model parameters will be needed then anyway. Best to fit the better model, no?

7. Could a full imputation approach not be incorporated to deal with the missing data?

Level of interest

Please indicate how interesting you found the manuscript:

An article of importance in its field

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.