Reviewer’s report

Title: Development and validation of a prediction model to estimate the risk of liver cirrhosis in primary care patients with abnormal liver blood test results: protocol for an electronic health record study in Clinical Practice Research Datalink

Version: 0 Date: 06 Feb 2019

Reviewer: Laura Bonnett

Reviewer’s report:

This is a well-thought-out protocol which promises to lead to a well-developed and suitably validated prediction model. I have only minor comments, the most pressing of which is the need to add a number of references to justify (the appropriate) statistical decisions. I have specified these below. The authors should be commended on their excellent knowledge of appropriate statistical techniques.

1. Line 160 - Are there any other prediction models that are similar to the one you plan to develop? We should avoid publishing new models just for the sake of it, but this is perfectly valid if existing models are methodologically poor, or do not exist.

2. Line 206 - I'm pleased to see that you have referenced the Riley paper about using big data sets for prediction modelling. It is always a difficult choice whether to opt for Internal-External Cross Validation (IECV) or to leave some data out of the development data set. I am happy with your choice but would appreciate some mention of IECV within your protocol even if to say that it is an alternative option which has not been considered in your case, and justify why not.

3. Table 1 - add references where possible to justify cut-offs.

4. Line 257 - How do you define "up-to-standard"?

5. Lines 270-274 - Please consider revising this sentence as I find it difficult to follow.

6. Sensitivity analyses - clarify that full model development methods will be undertaken for each sensitivity analysis i.e. confirm that bootstrap resampling will be performed to account for model optimism.

7. Line 351 - I believe that the word "events" is missing after 251.

8. Line 372 - On line 365 you say that events may not be missing completely at random. However, on line 365 you say because data are missing at random. Please resolve this conflict.


10. Line 420 - Define S_0(t) and explain how it will be estimated.
11. Lines 425 - 433 - Add references for all methods i.e. Cox's centiles, Royston's D, R2D and Harrell's c.


13. Line 457-459 - IECV might be appropriate here?


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.