Reviewer’s report

Title: Multi-cohort modeling strategies for scalable globally accessible prostate cancer risk tools

Version: 0 Date: 15 Mar 2019

Reviewer: Luis Esteban

Reviewer's report:

Report on "Visualization and internal validation strategies for building multi-cohort prediction models with an application to prostate cancer" by Johanna Tolksdorf et al.

The manuscript proposes a new strategy to validate the suitability of centers in a multi-cohort study. This is a valuable effort to provide a practice tool to help to build predictive models in this type of studies. I think the manuscript it is well written and quite clear. My main concerns lie in how it was focused:

1) The selection strategy analyzes changes in AUC and Hosmer Lemeshov statistics parameters with the presence or absence of a center in the predictive model. This is an incomplete view point of the properties of a predictive model. Calibration and discrimination have been analyzed in a well established study, but no comments appears about the clinical utility of predictive models. This is very important because the applicability of predictive models lies in their clinical utility.

In this type of problem, choosing a probability cut-off point, the number of avoiding biopsies with a minimum loss of high grade cancers is the key to know the clinical utility of the model.

I think the manuscript must be complete with a clinical utility analysis. You can estimate the best cut-off for the model built with the entire cohort, calculate net benefit, and then in a similar fashion of the rest of study analyzes changes in net benefit when you drop a cohort.

2) From what follows from the text, because the form of the models does not appear explicitly, in some models centers have been introduced in the predictive models as random effects, but no reference in the manuscript appears about p-value of this random effect. Please comment.

Minor points:

1) In the text, the expression "underneath" of AUC must be changed by "under", which is the English term.

2) Page 5, line 73, in my opinion the sentence "Creating models on multiple different cohorts is a superior approach" is not appropriate, I think is a more robust way to build predictive models, but not superior, a previous validation of models is necessary to apply it in a new center not included in an analysis, but the matching in characteristics and the calibration is what makes that can be better or worse using a multicenter or a unique center study. In fact, this is what authors comment in the Discussion section, page 13, lines 250-252.

3) Page 9, lines 163, the sentence "as well" is repeated, please delete one.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

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