Reviewer's report

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

Version: 0 Date: 12 Jul 2019

Reviewer: Reviewer 2

Reviewer's report:

PEER REVIEWER ASSESSMENTS:

OBJECTIVE - Full research articles: is there a clear objective that addresses a testable research question(s) (brief or other article types: is there a clear objective)?
Yes - there is a clear objective

DESIGN - Is the current approach (including controls and analysis protocols) appropriate for the objective?
No - there are major issues

EXECUTION - Are the experiments and analyses performed with technical rigor to allow confidence in the results?
No - there are major issues

STATISTICS - Is the use of statistics in the manuscript appropriate?
No - there are issues with the statistics in the study

INTERPRETATION - Is the current interpretation/discussion of the results reasonable and not overstated?
No - there are minor issues

OVERALL MANUSCRIPT POTENTIAL - Is the current version of this work technically sound? If not, can revisions be made to make the work technically sound?
No - manuscript has some fundamental flaw(s)

PEER REVIEWER COMMENTS:

GENERAL COMMENTS: The generalisability of risk prediction models across different settings is an important and well-studied issue. The authors propose a strategy to develop a risk prediction model with data available from multiple cohorts. However, there is substantial literature on this exact topic which provides similar, and in many ways, more sophisticated approaches to this exact problem. The paper does not refer to this extant literature. There are also some methodological issues with the authors' proposed approach which I detail below.
REQUESTED REVISIONS:

1. There is a large literature on building risk prediction models from data arising across different settings, which this paper does not refer to.

   * https://doi.org/10.1002/sim.5732 presents approaches similar to the first three methods presented in this paper, but give careful thought to how an intercept should be chosen when the prediction model is used.
   * https://doi.org/10.1186/1471-2288-14-3 which is a review of studies like the current paper.
   * https://doi.org/10.1371/journal.pmed.1001886 guidance paper on conducting studies like this one.
   * https://doi.org/10.1002/sim.6080
   * https://doi.org/10.1002/sim.7586
   * https://doi.org/10.1186/s12874-016-0277-1

2. Abstract/Background: 'develop a strategy as recently applied' needs rewording. A strategy is developed, then applied.

3. L94: Was the standardisation performed within sites or across all sites?

4. The description of how missing data is handled should all be in one place (some description on L103, some on L126).

5. The first three methods considered by the authors are usually called one-stage meta analysis while the latter two are called two-stage meta-analysis: e.g. see https://doi.org/10.1371/journal.pone.0060650 and https://doi.org/10.1002/sim.7141

6. L129-134: Here the authors explain how they selected their model. They perform stepwise logistic regression (not specified whether forward or backward) in 'multiple permutations of subsets' (not explained how these subsets were produced), used BIC as a model selection criteria (which is not designed for optimising prediction) and kept terms that appeared in 5% of all model fits. No reference is given for this approach and it doesn't seem intuitively reasonable. In particular, the predictive impact of any one variable in the model would depend on which other variables are also present, and the approach won't capture that. Please use a standard approach for model selection, or give further references and justification for the proposed approach.

7. The Hosmer-Lemeshow test is an extremely limited means of assessing calibration. Best practice is to also consider 1) calibration in the large, 2) calibration slopes, and 3) to visually inspect calibration plots.
8. For the internal validation, the authors use a leave-one-out approach and a somewhat complex permutation approach (in both cases, the unit of observation being the site rather than the patient). The former approach is standard (e.g. as used in Debray et al 2013 Stat Med) while the latter is not. So can the authors better justify inclusion of the latter choice?

9. The confidence interval is also based on the above permutation test at the site-level unit of observation. So it's not entirely clear what this is a confidence interval of? If the confidence interval were constructed using a more standard approach I suspect it would be narrower. In the current study the confidence intervals on the AUC and Hosmer-Lemeshow statistics are so wide that the study provides no information to choose between the different methods.

10. The authors should refer to, and complete, the TRIPOD checklist in developing their risk prediction model: https://www.equator-network.org/reporting-guidelines/tripod-statement/

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.

No

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

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.

This reviewer has been recruited by a partner organization, Research Square. Reviewers with declared or apparent competing interests are not utilized for these reviews. This reviewer has agreed to publication of their comments online under a Creative Commons Attribution License attributed to Research Square and was paid a small honorarium for completing the review within a specified timeframe. Honoraria for reviews such as this are paid regardless of the reviewer recommendation.

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/](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