Reviewer's report

Title: Tufts PACE Clinical Predictive Model Registry: Update 1990 through 2015

Version: 0 Date: 2017-07-28

Reviewer: Joie Ensor

Reviewer's report:

The manuscript presents the findings of an update to an existing registry of clinical prediction models (CPMs) which predict individual's risk of CVD outcomes. The authors should be commended for undertaking the large amount of work involved in the development and updating of such a registry, as I am sure this will be a useful resource for many clinicians and researchers in the field.

The aim of the registry (and therefore this update to the registry) is stated as "to aid clinicians and researchers in understanding the state of CPM development across the spectrum of CVD". To this end the authors have achieved their aim, providing a searchable online database of available CPMs; however despite the amazing work done to date, there is further potential for this work in the future.

Below are a number of minor comments for the authors to consider, however some comments are outside of the scope of this article and I have been clear where I do not expect additional work for this manuscript.

Minor comments

1. Page 5 line 41 - states that the manuscript is concerned with 'de novo' CPMs. However within this group of 'new development' studies there were internal and external validations performed (page 8 lines 41-49). Given this it is potentially important to present the statistics on reporting of calibration, split across these three types (development/internal validation/external validation), as it is unlikely that calibration is reported as part of a model development because it will be by definition perfect. Therefore the statistics and trend reported for this may be misleading. For example it may be that calibration was always reported in external validations, appropriately. Could the authors break this down, or at least make this potential bias clear in the manuscript?

2. Following this, the statement in the discussion (page 9 lines 36-41), should potentially be revised to reflect the uncertainty about reporting of calibration across different types of model study.
3. Further to the above, it would have been interesting to have broken down the information further so that the kind of calibration or discrimination measures could have been examined. For example, are the majority of articles reporting only Hosmer-Lemeshow test results for calibration? Particularly in terms of the presentation of the model, how many articles reported the actual underlying model - this is important and essential information to allow future use of the model for validation and updating of the model. A substantial amount of work has been undertaken by the authors to extract the available information already, and I do not expect the authors to revisit the articles to extract this information for this manuscript, but perhaps the usefulness of further information could be considered for future work on the registry.

4. The authors state in their discussion that an important limitation is the potentially missed CPMs. In searching the registry briefly it seems that the registry does have some missed models (e.g HERDOO2 model for VTE recurrence risk prediction - doi: 10.1503/cmaj.080493). This is only an observation based on one small example in which I am aware of available models, but does raise questions about the breadth of the search strategy, and selection process used in the original and update to the registry. I sympathise with the authors regarding the difficulties in identifying CPMs in the literature and I do not expect the authors to revise their search for this manuscript, but this is an important area for future research and potentially broadening the search strategy could be something worth investigating in future work for the registry.

5. Numbers do not add up in the flow diagram presented in Figure 1. The included articles is said to be 244, and the articles from previous search is said to be 503. The total articles in registry is then given as 740. However 244+503=747. Please correct or provide elaboration as to why 7 articles were excluded from the registry.

6. Following the above, the original article appears to find 506 included articles, but in this manuscript 503 articles are carried forward from previous searches. Could the authors explain this discrepancy please?

7. I think it is important that the article is explicit in stating how many additional models were found through the update as this is the focus of this manuscript. I also think it is important to clarify in the text what the search dates were for the update, it is currently unclear. Only the flow diagram given in Figure 1 eludes to this, but in the interest of reproducibility I think the dates could be given in the text.

8. Also I think it may improve the transparency and aid the reading of this manuscript to report the inclusion criteria used for the study. While this is available in the original manuscript, I think it would aid the reading of this manuscript, as it is an important
element of the work. Similarly, while the authors may not wish to include an example of their search strategy I see no reason not to include a list of databases searched.

9. The split of logistic (55%) to Cox (33%) regression models is interesting, given that the majority of models were classified as prognostic (1060). It may have been helpful to have a measure of time recorded, as I assume that many of these models are therefore predicting short-term outcomes in which censoring could be arguably ignored? Although I realise that the aim of the registry is not to define which models are 'good', only what models exist.

**Level of interest**

Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**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