Author’s response to reviews

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

Authors:

Benjamin Wessler (bwessler@tuftsmedicalcenter.org)
Jessica Paulus (jpaulus@tuftsmedicalcenter.org)
Christine Lundquist (clundquist@tuftsmedicalcenter.org)
Muhammad Ajlan (ajlan.muhammad@gmail.com)
Zuhair Natto (z_world@hotmail.com)
William Janes (wjanes@tuftsmedicalcenter.org)
Nitin Jethmalani (Nitin.Jethmalani@tufts.edu)
Gowri Raman (graman@tuftsmedicalcenter.org)
Jennifer Lutz (j旅途z@tuftsmedicalcenter.org)
David Kent (dkent1@tuftsmedicalcenter.org)

Version: 1 Date: 2017-10-12

Author’s response to reviews:

R1. I have found the paper clear, well laid out and well written. Some minor comments:

1. Redundant figures should be avoided (Figure 2 and Figure 4). The same are provided in website.

Thank you for this observation. We have removed the old Figure 4 from the manuscript. We have left Figure 2 in the body of the manuscript since we feel it gives a sense of the scope (and increasing volume of literature) that is contained in this novel resource. If the editors feel that this figure is redundant (and not valuable as part of the body of the manuscript) we are of course happy to work with the editorial staff to remove this figure as well.
2. The following statement "Circulation published 53 (7.1%) and Journal of American College of Cardiology published 45 (6.0%) of the articles included." should be rephrased to make clear to which sample the percentages pertain.

Thank you. This statement refers to the entire registry. This sentence in the Results section has been edited and now reads: “Circulation published 53 (7.1%) and Journal of American College of Cardiology published 45 (6.0%) of the articles included in the Registry.”

3. Tables 2A, 2B, 2C should be merged. Thank you for this clarifying comment.

We agree that this improves the presentation of these data. These data are now presented in a single merged Table 2.

4. It will be interesting to plot the distribution of discriminatory performance of the included stratified according to the "primary index condition".

We agree that this is of interest. We have added these data in a new figure (new Figure 4) to visually show the distribution of discriminatory performance according to index condition.

5. Have the authors captured any items for quality assessment of the selected papers? Such additional information would have been important.

Thank you for this important question. For each CPM included in our Registry we have captured information on common model building quality metrics such as events per variable (EPV), Table 3. Unfortunately, many published CPMs do not report the number of screened variables and thus we are limited to reporting on the number of events per included variable. While reporting of model building techniques may have improved since the recent publication of the TRIPOD
statement, the vast majority of the CPMs in our Registry were developed and published before this statement was released. Given the broad scope of this literature and the historic under-reporting of quality measures, we do not present a formal quality assessment of the published CPMs. Our group is actively exploring some of the parameters that reflect ‘model quality’ though unfortunately assessment at this stage is limited by poor (and variable) CPM reporting. As part of upcoming work, newer models will be evaluated in the context of contemporary reporting standards and CPM quality can be more fully assessed.

Reviewer #2

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?
Thank you for this important observation. As this comment makes clear, calibration is generally reported only for validation exercises. We have clarified this sentence in the Results section of manuscript. This sentence now reads “Of the CPMs included in this Registry, only 425 (39%) report some measure of model calibration as part of a validation exercise (either external or internal) presented at the time of CPM publication.” We present an aggregate statistic (from internal and external validations) because the methods used to create this Registry do not allow for complete assessment of external validations since we have only captured external validations published as part of the de novo CPM publication. Our group is currently working on a grant from the Patient Centered Outcome Research Institute (PCORI) to systematically assess all published external validations of these CPMs through a comprehensive citation search. When this work is completed we will be more fully able to describe the reporting differences between internal and external validations.

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.

Thank you. We agree that this statement should be clearer and we have improved the specificity of this statement. This sentence in the Results now reads “The frequency of reporting CPM discrimination and some measure of calibration (as part of an internal or external validation) as part of the original CPM description increased from 1990 to 2015 (p for trend <0.001 for both).”

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.
Thank you for these important observations. As noted in the manuscript on page 4, in order to be included in the Registry CPMs had to include sufficient information for readers to generate individual predictions. Typically this was in the form of a point score, equation, decision tree, nomogram, or online calculator. We have substantial interest in the calibration and discrimination metrics that are reported across these CPMs and have been collecting more detailed information throughout this update. While we are actively working to more fully characterize these specific metrics, we have added the following sentence to the Results section to more fully describe the measures of calibration that are seen. “Of the CPMs reporting calibration that were published after May 2012, only 93 (56%) report a Hosmer-Lemeshow statistic.”

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.

Thank you for identifying this CPM. We will be adding this CPM to our Registry (and website). We have found that some potentially eligible CPMs are ultimately excluded since they do not permit readers to generate individual predictions based on published results. We have found isolated examples of CPMs that were not identified though our original search. We are continuously updating our Registry based on input from field experts and have included a point of contact on our website so that authors and researchers can reach out to suggest additional CPMs 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.
Thank you for identifying this discrepancy. We have reviewed these primary data and have updated Figure 1. This flow chart now accurately describes that the total number of articles included in the Registry is 747.

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?

Thank you for noting this discrepancy. From the time of the original publication (2015) three articles from the original search were excluded based on further review. These articles either were not focused on CVD (PMID 16308250 and 21818020) or were not an original publication of a de novo CPM (PMID 12123274). The number of articles carried through from the original search is 503.

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.

We agree that it is very important to be clear with the search dates. We have added the specific search date (March 31st 2015) to the text. We have added to the Results section that “374 CPMs were added to the Registry during this update”.

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.
We agree that more details will be helpful to the reader. We have added the following sentence to the Study Search and Selection section: “Briefly, we performed a PubMed search for English-language articles containing newly developed CPMs. A CPM was defined as a model that provides a method to calculate or categorize an individual's risk for a binary outcome.”

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.

We agree that the split between logistic and cox models is interesting, but caution that many more cox models than logistic models were excluded from the registry based on inclusion criteria (since many did not describe the baseline hazard or provide an alternative way to generate individual patient predictions). We agree strongly in a complete analysis of the reported model building methods for the CPMs reported in our Registry. Descriptions of time recorded, how censoring is handled, and also how missing data are handled will offer substantial insights into model building techniques. Future work will focus on evaluating whether certain modeling techniques are correlated with performance measures. Given the volume of data in our Registry we feel this effort is beyond the scope of this Update manuscript. We have added the following sentences to the Results section to more completely describe the follow-up time for these CPMs. “For the top 10 index conditions, CPMs most commonly (51%) predicted mortality over a short timeframe (< 3 months). 44% of these CPMs predicted mortality over a long time frame (> 6 months).”