Author’s response to reviews

Title: Harnessing repeated measurements of predictor variables for clinical risk prediction: A review of existing methods

Authors:

Lucy Bull (lucy.bull@manchester.ac.uk)

Mark Lunt (Mark.Lunt@manchester.ac.uk)

Glen Martin (glen.martin@manchester.ac.uk)

Kimme Hyrich (Kimme.Hyrich@manchester.ac.uk)

Jamie Sergeant (jamie.sergeant@manchester.ac.uk)

Version: 1 Date: 07 Apr 2020

Author’s response to reviews:

Covering letter & Responses to Reviewers' comments

Dear Dr Maarten van Smeden,

Please find enclosed the revised manuscript of our paper entitled, “Harnessing repeated measurements of predictor variables for clinical risk prediction: A review of existing methods” to be considered for publication in BMC Diagnostic and Prognostic Research.

We would like to thank the reviewers and editors for their comments and suggestions as well as for the opportunity to revise the manuscript. We appreciated all of the reviewers’ comments and we have revised the manuscript in response to them. Changes to the manuscript have been tracked for ease of reference. Below is a point-by-point response to each comment with details and locations of the amendments. We hope that these changes strengthen the work sufficiently for the manuscript to be accepted for publication.

Yours sincerely,

Lucy Bull

For ease of reference, the letters R and A have been used to refer to the reviewers' comments and the authors' responses, respectively.

(R = Reviewer, A = Authors)

Reviewer 1:
R: The manuscript "Harnessing repeated measurements of predictor variables for clinical risk prediction: A review of existing methods" is a useful addition to the literature. I also very much liked the precise definition of terms in the Methods section and section 3.2 (including Figure 2).

-----------------------
A: Thank you for your supportive comments. We are pleased that you agree that our manuscript is a useful addition to the literature.

-----------------------
R: Below are some comments that I would like the authors to address:

# Comments
## Background
p.4, line 8/9:
R: The authors only refer to binary or time-to-event outcomes (throughout the manuscript). Why are continuous outcomes not included in this review?

-----------------------
A: The review focussed on the prediction of future binary and time-to-event outcomes as these are the most common outcomes reported for clinical prediction models (CPMs) (e.g. for CPMs developed on electronic health records (i.e. readily-available longitudinal data) between 2009 and 2014 (1), and for all CPMs published in 2008 (2)). Although we do not specifically consider continuous outcomes, predicting the future average continuous outcome for an individual is touched upon in the identified methods for harnessing repeated measurements in the first stage of a two-stage CPM.

The following text has been added to the Background section of the manuscript [Pg. 4]:

“Binary and time-to-event outcomes are of primary interest here as they are the most commonly reported amongst the prediction-modelling literature (6, 7).”

1. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. Journal of the American Medical Informatics Association. 2016;24(1):198-208.
2. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS medicine. 2012;9(5):1-12.

-----------------------
## Methods
p.7 line 11:
R: The authors mention that they extracted the computer software used from the screened articles. Could the authors provide some information about how often the software used was reported at all, and what type of software is being used in the field of CPMs? I realize that, for an applied researcher, searching for a solution to his/her prediction problem, this may not be of primary interest, but I think in the context of a general review of the methodology this type of information could have a place. If there was a lack of reporting of the software used, I think this should also be mentioned on page 27, lines 14-19.

-----------------------
A: Thank you for your suggestion to include more details about the reporting of software use. We believe that the primary interest of an applied researcher would be the packages and functions available in particular statistical software to implement the methods discussed in the review; as such, we focus on presenting this information in Table 4. In general, there was indeed a lack of reporting of the software used, in particular for advanced machine learning algorithms (excluding random forests for example) and widespread/commonly used statistical approaches such as time-dependent survival models. Of note, Table 4 includes all the software reported in the articles we studied, and as such we hope that this addresses the reviewer’s comment about reporting the types of software being used. As regards the frequency and detail of reporting, while we agree with the reviewer that this is an interesting question for the transparency and reporting of CPMs amongst the literature, we feel that adding this information would distract from the messages for applied researchers (our primary goal of this paper).

We have added the following text to the limitations in the Discussion [pg. 29 - 30]:

“Third, while aggregate details of available software for discussed methods are provided in Table 4, the frequency and the level of detail of software reporting in the identified literature were not assessed.” …
“Despite a general lack of detail in the literature in the reporting of available software, all reported available software has been included in Table 4. Such information will be useful for the implementation of identified methods and can provide an indication of where software may not have been well-reported.”

-----------------------
## Results

p.9 Figure 1:
R: In the "Screening" section, it says that "method[s]" with no application were excluded. Why?
The title of the manuscript states that this is "A review of existing methods". Should that not include purely statistical publications?

A: Thank you for alerting us that this required additional clarification. Although purely statistical papers would be interesting from a methodological point of view, to see which additional methods could be brought to the field, this was not the primary goal of the review. The primary aim of this review was to facilitate the understanding and uptake of existing methodology that has been implemented in real-world data, by systematically identifying them and providing a comprehensive summary of the common themes amongst the literature. We note that many articles that focus on developing statistical methodology in this field would often also include application to real-world clinical data, and would therefore be included in our review.

We have modified the text in the Background section to clarify our intention [pg. 4]:

“To the authors’ knowledge, a broad review of methods used for handling longitudinal data in binary or time-to-event CPMs has not yet been performed.”
“Our objective was to review the literature and provide applied researchers with a comprehensive summary of existing approaches used for harnessing repeated measurements of predictor variables in CPMs.”

-----------------------
R: Why was "patient-reported outcome framework" an exclusion criterion?

A: The lists of reasons within Figure 1 are descriptions of where the publications, which were captured by the search strategy, did not satisfy the inclusion criteria for that particular stage of screening stated in Table 3. Therefore “patient-reported outcome framework” is simply indicating that the database search strategy picked up a paper proposing a patient-reported outcome framework, instead of a clinical prediction model harnessing longitudinal data. Given that we only aimed to review CPMs, this paper was therefore excluded from the review.

R: For completeness, could the authors provide a list of the included articles in an additional supplementary file (for example, a .csv)? This is the data they worked with, after all.

A: Thank you for this suggestion. In light of your comment, we have now added a list of included articles as a supplementary table. We have presented the papers in alphabetical order by first author surname. The supplementary table includes the following information: Authors, Year of publication, Article title, Journal of publication, Volume, Pages, and DOI number.

We have referenced this supplementary table in the text in the Results section under Database search [pg. 8]:

“Supplementary Table 1 lists all articles included in the review.”

p.10 line 17:

R: It is not entirely clear to me what A1 "to better specify the predictor-outcome" relationship" exactly means. I can see how "to better capture the ..." would be an aim (for example, when using a more flexible approach that does impose fewer restrictions and could, therefore, give a more accurate representation of the actual relationship). Is this what the authors mean?

A: As you correctly suggested, repeated observations through time of a predictor allow us to better estimate an association between a predictor and an outcome of interest, especially without imposed restrictions or assumptions. As a simple non-parametric example, within the landmark framework, prediction models are only developed on those who are still at-risk at a particular time-point. Therefore we are estimating the predictor-outcome relationship whilst accounting for the ever-changing at-risk population during follow up.

Thank you for providing an alternative way to communicate this particular aim, and in line with your explanation of how we may need to better “represent” the predictor-outcome relationship within a CPM, we have adopted the phrase “represent” to make this clearer.

The changes in the manuscript (as displayed below) are located in the Abstract [pg. 2], Results [pg.10, 11, and 14] and Discussion [pg.26] sections, respectively.

“Each of these frameworks satisfies at least one of three aims: to better represent the predictor-outcome relationship over time; to infer a covariate value at a pre-specified time, and to account for the effect of covariate change.”
“All of the discovered methods satisfied one or more of these three methodological aims: (A1) to better represent the predictor-outcome relationship, (A2) to infer or predict a covariate value at a pre-specified time, or (A3) to account for the effects of how a predictor changes over time.”

“Methods satisfying A1 tend to utilise repeated observations to represent a time-constant relationship, or better represent a time-varying relationship, between a predictor and the event of interest.”

“Similarly to the TDCM framework, the primary methodological aim of Generalised Estimating Equations (GEE) is to utilise repeated observations from the same individual to better represent the association between the predictor variables and the event of interest.”

“This methodological review has identified three ways in which available methods can utilise longitudinal information to enhance the performance of CPMs: (A1) to better represent the predictor-outcome relationship; (A2) to infer a covariate value at a pre-specified time, and (A3) to account for the effects of predictor change over time.”

-----------

p.11 line 1:
R: I think the definition of A2 should not use the "true value". I don't think it is possible to get the true value in most settings. I suggest using, for instance, to infer or predict (or estimate?) the value of a covariate at a pre-specified time.

-----------------------

A: Thank you for pointing this out and suggesting to remove the term “true value”. The inferred value is and always will be an estimate. This has been rephrased (as displayed below) in the Results [pg. 11] and Discussion [pg. 26] sections of the manuscript, respectively.

“...:(A1) to better represent the predictor-outcome relationship; (A2) to infer a covariate value at a pre-specified time, and (A3) to account for the effects of predictor change over time.”

-----------------------

p.12 line 1:
R: The statement that methods F4 to F7 require repeated measurements at the time of prediction is not entirely clear to me. (1) does this mean that a value measured at exactly that timepoint has to be available? Or does it mean that we can only do prediction for subjects for whom some (at least one?) of the repeated measurements is already available?

-----------------------

A: We acknowledge that “require” may not be the correct word here, and we have rephrased this to “being able to harness” repeated measurements at the time of prediction. We simply wanted to highlight that some identified approaches, despite harnessing subject-level repeated measurements on a population for model development, only require cross-sectional patient information at the time of prediction. Other methods and frameworks are able to harness repeated measurements if they are available at the time of prediction.

Adjusted text in manuscript [pg.12]:

“All identified methods require subject-level longitudinal information on a study population for CPM development. The TSM, JM, and TC frameworks (as well as some ML algorithms) can
also harness a subject’s repeated measurements at the time of prediction. While, as stand-alone frameworks, the TDCM, GEE and LA frameworks only require a subject’s most recent observations (i.e. a maximum of one measurement for each predictor) at the time of prediction.”

R: (2) I’m not sure if this is correct (at least in my interpretation of what is a requirement):
With a two-stage approach or JM, for instance, the value at the exact time of prediction could be estimated from the model imposed on repeated measurements that are available at other time points for the same patient.
If a patient has no values at all at the time of prediction, technically, it should still be possible to impute a value based on the distribution estimated by the fitted model conditional on other observed covariates. (Maybe not in all estimation frameworks, but the authors make a general statement here.)
I expect that in most cases in practice such an imputed value will not be very meaningful/reliable, and maybe the available software is not capable of handling this. However, since there is a way to handle cases with no information, in theory, I would not consider it a requirement of the method.
From my personal Bayesian missing data point of view, allowing for cases without any longitudinal information would even be a natural thing.

A: Thank you for this viewpoint. We recognise that some longitudinal methods still work with only a single measurement of any predictor, which is why we have relaxed the terminology and exchanged the word “require” to “can harness”. With regards to imputing an estimated predictor value where we have no predictor values for a subject at the time of prediction, we thank you for highlighting this opportunity with advanced Bayesian methods. As this particular opportunity was only discussed when Multiple Mixed-effects Models were employed in a two-stage model amongst the landmark framework, it has limited attention within the review.

To clarify that the review only reports methods from the articles that the systematic literature search picked up on, and to not eliminate possibilities of adopting additional or extended approaches from the statistical literature, the following statement has been added to the manuscript [pg. 11]:

“It is important to highlight that the content of this methodological review only covers methods reported in the literature identified via the database search, and that other valid approaches may exist but have not yet been applied in this field of clinical risk prediction.”

p.12 and following:
R: I think it would be helpful for readers if the authors would state within the first few sentences of each of the subsections about the different modeling frameworks explicitly if the approach can handle both types of outcomes or only survival/binary.

A: Thank you for this suggestion. We agree that this information would be valuable for the review’s readership.
This information has therefore been added to the manuscript for the TDCM [pg. 13], GEE [pg. 14], LA [pg. 15], TC [pg. 22-23] and ML [pg. 23] frameworks, respectively. Outcome types had already been confirmed for the TSM and JM frameworks.

“Therefore TDCM falls under the first methodological aim (A1), and as the timing of each event is required, can only handle time-to-event outcomes.”

“However, unlike the TDCM framework, GEE models account for within and between individual correlation, can directly harness repeated events per individual (28, 29), and can model either binary or survival outcomes.”

“The flexibility in choice of model used to develop the CPM at each landmark time point allows for both binary and survival outcomes to be modelled under this framework.”

“Mixed-effect (ME) models have also been employed to classify longitudinal trajectories for binary events or categories, these methods have been grouped under the trajectory classification framework for this review (52, 76, 77).”

“To extend this approach to predict time-to-event outcomes, covariate trajectories may be classified into categories that can then be used as a predictor within a survival model. This extension can be performed under the two-stage modelling or joint-modelling framework, the latter approach is referred to as the joint latent class model in Additional File 1.”

“The majority of reported machine learning algorithms were employed to classify data for binary outcomes, with very limited attention on time-to-event outcomes (81-83)”

-----------------------

p.13 lines 4-6:
R: How does "those still at risk" go together with "those who have ...experienced an event at that specific time", since the latter are no longer under risk?

Please re-phrase.

-----------------------

A: Thank you for highlighting that the clarity of this sentence could be improved. “Those still at risk” refers to those subjects who have not had an event before time t.

We have rephrased this sentence to state that the method compares the most recently observed covariates of those who were still at risk “just before” each event time, for those who do and do not experience an event at that specific time.

The manuscript text now states [p.13]:

“Conceptually, this approach compares the most recent covariate values for those still at risk just before each event time for those who have and have not experienced an event at that specific time.”

-----------------------

p.13 lines 11/12:
R: The authors write that applying a baseline CPM to follow-up data would lead to overestimated risk predictions. If the risk (of having an event) is overestimated, this means that the survival probability is underestimated.

The given reference, however, states the opposite:

"...time-fixed models may tend to overestimate survival if they are applied to follow-up data."

-----------------------
A: Thank you for pointing this out; this has now been corrected within the manuscript. To highlight the key difference for applied researchers between risk predictions and survival probabilities, the phrase has been changed to [pg. 13]:

“Applying baseline CPMs to patient data collected during follow-up would lead to under-estimated risk predictions, and over-estimated survival predictions.”

-----------------------
p.19 lines 18 and following
R: It might be worth mentioning that the term "joint model" generally refers to any models that are estimated jointly. The term "joint model" is often used as an abbreviation of "joint models for longitudinal and survival data", and probably most joint models in the prediction context are of that type. However, there are also other joint models, such as, for example, a joint model for a binary outcome and repeatedly measured covariates.
In my experience, when first encountering a new area of methodology, it can be difficult to distinguish whether a term is the name of one particular type of model or refers to a group of models. It might be useful for readers if this context is provided more clearly.
-----------------------
A: Thank you for this suggestion, which we agree will be helpful for readers. The following statement has been added to the beginning of joint-modelling framework sub-section [p. 20]:

“The joint-modelling (JM) framework addresses the limitations of the TSM framework by simultaneously estimating the longitudinal sub-model and the survival or binary outcome sub-model (13, 51, 52). The term “joint model” more broadly can refer to any statistical models estimated jointly, but here the literature focussed on jointly modelling a longitudinal model and a survival or binary outcome model.”

-----------------------
p.20 lines 14 and following
R: I find this section a bit confusing. It is my understanding that the issue is that, for a new patient, the random effects, which are used in the linear predictor of the survival sub-model, are unknown. To deal with this, random effects can be sampled from the posterior distribution that depends on the new patient's longitudinal measurements up until the time of prediction, the fact that the patient is still alive at this time, and the parameter estimates obtained from the fitted joined model.
Using Monte Carlo simulation allows us to "integrate out" the unknown random effects and thereby take into account the added uncertainty in the estimate of the survival probability for the new patient.
Is this what the authors mean? Maybe they can find a way to re-phrase.
-----------------------
A: This is exactly what we mean, and we can see how the original wording may come across as confusing.

We have now clarified this as follows [p.21]:

“One challenge of using random effects in CPMs is estimating the risk of a future event for a new subject, as their random effects are unknown. To resolve this, random effects can be sampled from their posterior predictive distribution, which depends on the population-level
distribution of random effects from the fitted joint model, the new subject’s covariate values until the time of prediction, and the subject still being at risk at the time of prediction (72, 73).

A more popular choice is to employ the Monte Carlo simulation approach as it takes into account the uncertainty around the survival or event probability estimate (57, 72, 73). Monte Carlo simulation is, conceptually, a procedure that repeatedly samples parameter estimates and random effects based on their estimated posterior distributions from the fitted joint model (72, 73). A new individual’s random effects can be simulated from their posterior predictive distribution, as stated above (72, 73). Repeatedly sampling from the posterior distributions allows for an empirical distribution around the estimated survival or event probability (73). Monte Carlo simulation has been employed independently of the model estimation process and is reported to be computationally efficient in contrast to the joint model specification (74, 75).’’

R: It is not entirely clear to me how this approach works. I can follow the description of the model in lines 6-10. In the Bayesian framework, we could then obtain the posterior distribution for the unknown binary outcome for a new patient, given his/her repeatedly measured covariate information. Something similar should, of course, be possible using likelihood theory. I do not understand, however, what the authors mean in lines 10-12. Could the authors please clarify?

A: Thank you for pointing out that this area would benefit from further clarification. The original statement:

“In practice, when the outcome is unknown, the subject can be assumed to be an event subject and non-event subject for prediction and then these predictions can be compared.”

…has now been changed to the following [p.22]:

“In practice, when the outcome is unknown, separate distributions of the longitudinal predictor values can be estimated based on the event and non-event ME model parameters, as well as the new subject’s observed longitudinal values (78-80). Both of these distributions can then be used to produce a discrimination score, which can later be used to classify the subject, or to produce a posterior probability that the subject will experience the event (78-80).”

We hope that this helps the readers’ understanding of the approach, and how its primary goal is to classify trajectories into binary or categorical groups, by defining the distribution of repeated values conditioning on the subject being an event or non-event subject, and using these distributions either to define a discrimination score, a likelihood ratio statistic or a posterior probability.

R: Moreover, is this latent class approach related to the joint latent class model detailed in the supplementary materials? It might be useful to include a mention of this connection (or the difference).

A: Thank you for highlighting this link. We agree that mentioning it would be a valuable addition to the review. The link between the joint latent class model and the trajectory
classification has now been included, alongside the framework’s ability to harness time-to-event outcomes.

This TC section now states the following [pg. 22-23]:

“To extend this approach to predict time-to-event outcomes, covariate trajectories may be classified into categories that can then be used as a predictor within a survival model. This extension can be performed under the two-stage modelling or joint-modelling framework, the latter approach is referred to as the joint latent class model in Additional File 1.”

Reviewer 2:

R: I have reviewed the manuscript titled “Harnessing repeated measurements of predictor variables for clinical risk prediction: A review of existing methods”. This manuscript is well planned and presented by authors. However, there are following points that required clarification from authors;

Thank you for your supportive comments, and for the opportunity to provide clarification on the points that you raise.

R: As title of this review suggests that it is a review of existing method. So being a reader you expect all key features (such as available methods, sample size, validation methods etc.) of clinical prediction models. However, this article has only focused on the available frameworks. Other aspects either overlooked or left for future research. Being a reader, I may be over expecting to provide each and every information in just one article. However, a brief description for each key feature is expected with some good citations so readers can get some further useful information.

A: Thank you for your comment. Our aim was to provide a summary of existing approaches used to harness repeated measurements of predictor variables for clinical risk prediction, primarily for applied researchers. We did not describe all aspects from the 217 papers in the text as this would represent a vast amount of information and would not directly address our aim. Rather than assess the quality of each paper, we wanted to inform the readers of the common themes amongst the CPM research field, and clearly summarise the similarities and differences between discussed approaches. We foresee this review being a starting point for an applied researcher wishing to make the most of any available longitudinal data they may have access to for CPM development, or available for risk estimation in practice.

Note that we have provided a full list of all of the articles that were included in the review in Supplementary Table 1.

We have referenced this supplementary table in the text in the Results section under Database search [pg. 8]:

“Supplementary Table 1 lists all articles included in the review.”
We have provided relevant references when specific examples or variations of each method have been introduced in the review, and we hope that these citations are useful for the readers to obtain further information. More details about additional methods considered in the review and their corresponding references have also been provided in Supplementary File 1.

-----------------------

R: Authors have also considered machine learning (ML) models as one of the classification. However, no specific search terms were used to find ML models related to longitudinal data. For example, random forest for survival (FR-S), CForest, and Rangers etc. So, in my opinion, ML longitudinal models terminology should have been used in key words to find appropriate articles. Following references may be useful for time-to-event data;
Korepanova, N., et al. (2019). "Survival forests under test: Impact of the proportional hazards assumption on prognostic and predictive forests for amyotrophic lateral sclerosis survival.” Stat Methods Med Res: 0962280219862586. https://doi.org/10.1177/0962280219862586
Wang, P., et al. (2019). "Machine Learning for Survival Analysis: A Survey.” ACM Comput. Surv. 51(6): 1-36. https://doi.org/10.1145/3214306

-----------------------

A: The search strategy was not designed to discover either statistical or machine learning methods, since we find this to be a false dichotomy. Rather, we aimed to identify relevant analytical methods that have been used; therefore, while we did not specifically use “machine learning” terminology in our search, we have equally avoided “statistical” terminology. The search strategy was designed to find prediction models (of any analytical form) that have been developed for clinical purposes, and then the methods were extracted from the identified papers. We hoped not to bias it one way or another, but to report any methods that were identified for event prediction. In the majority of cases, these naturally fall into model-based methods. In order to capture machine learning literature, we included the term “algorithm” as well as any other terminology known to refer to a predictive tool of some sort.

Thank you for the suggestion about survival random forests. We have explicitly stated where extensions to survival outcomes are possible.

Unfortunately, both suggested papers on survival random forests were published after the database search was completed, which is why they are not described in the review manuscript.

The following statement has been added to the manuscript so readers are aware that other valid approaches may be available that the database search has not picked up on [pg. 11]:

“It is important to highlight that the content of this methodological review only covers methods reported in the identified literature via the database search, and that other valid approaches may exist but have not yet been applied in this field of clinical risk prediction.”

-----------------------

R: Inclusion Criteria Set B, point 2 says at least one predictor is used for abstracts and full-texts screening. However, at least two predictors should be considered to form a multivariable risk prediction model.

-----------------------

A: We apologise if this was unclear. Each CPM did indeed contain at least two predictors, yet it was only a requirement to have repeated measurements of at least one predictor. Therefore the
clinical prediction model could contain multiple time-fixed predictors and a single longitudinal predictor, in order for it to be a multivariable CPM. We hope this has clarified the inclusion criteria.

-----------------------

R: As model validation is an important part of any clinical risk prediction model. Therefore, authors should also consider one more table showing common methods for model validation (i.e. discrimination and calibration) for each framework. Readers of this article would like to know this info in addition to the available frameworks.

-----------------------

A: Thank you for this suggestion. We agree that this information would be valuable for the reader. However, considering details about each of the validation methods for each of these frameworks would represent a whole body of work in itself, and worthy of its own review. As the use of longitudinal data within clinical prediction models is currently scarce, most papers are simply assessing the predictive accuracy of longitudinal CPMs in the same way they would be for cross-sectional CPMs. Here, one simply assesses the model’s predictive performance when making predictions for patients at different time points during follow-up. We fully agree that tailored model validation techniques would be more appropriate for each framework, but this may require further methodological research.

To acknowledge this point, we have included the following statement regarding the performance assessment of longitudinal CPMs across the literature within the Discussion [pg. 29]:

“Model validation techniques used remain similar to those for cross-sectional CPMs where applicable, and the quantification of predictive improvement is often performed using differences in C-index, which lacks clinical interpretability (108).”

-----------------------