Author’s response to reviews

Title: Dynamic models to predict health outcomes: Current status and methodological challenges

Authors:

David Jenkins (david.jenkins-5@manchester.ac.uk)
Matthew Sperin (matthew.sperrin@manchester.ac.uk)
Glen Martin (glen.martin@manchester.ac.uk)
Niels Peek (niels.peek@manchester.ac.uk)

Version: 1 Date: 15 Oct 2018

Author’s response to reviews:

Response to reviewers’ comments

We would like to thank all reviewers for their careful and thorough reading of this manuscript and for their constructive comments and suggestions, which have improved the quality of this manuscript. Our point-by-point responses to the reviewers’ comments (in bold italics) are given below:

We have also attached our response to reviewers as a word document and included both a track change manuscript and revised manuscript without track changes.

Comments from reviewer 1

In this manuscript, the authors review approaches for refining a prognostic model over time, in order to combat the loss of accuracy that can result from changes at the patient level and system level. I believe that the field of dynamic prediction modeling is sufficiently developed to warrant such a review. The authors clearly present current methods and raise important limitations of current approaches. However, I feel that a major revision of the article is required. Overall, the organization of the results and discussion did not help the reader to distinguish between what was review material and what was new discussion offered by these authors. Better organization and focus of the Results section, along with a more comprehensive Discussion section, would improve the paper. Below are several comments, in no particular order.

R1.1. The authors did not include page numbers in their document.

Author response: Page numbers have now been added to the document
R1.2. Abstract: "calibration drift" is not well defined here.

Author response: Thank you for highlighting this. As a response to this and R2.1 we have edited the text to highlight accurate prediction rather than solely calibration. The updated text now reads (pg. 2):

‘Disease populations, clinical practice, and healthcare systems are constantly evolving. This can result in clinical prediction models quickly becoming outdated and less accurate over time. A potential solution…’

We maintain the use of “calibration drift” in the main manuscript as this is a common term in the literature, ensuring that we define this in the introduction and include an appropriate reference. This is done in the introduction (pg. 4) and reads:

‘Over time, population demographics, prevalence of disease, clinical practice, and the healthcare system as a whole may change, meaning that predictions based on static data can become outdated and hence no longer accurate. This is known as calibration drift9 and is one of the major pitfalls in using CPMs in practice10.’

R1.3. In the first paragraph of the Introduction, the authors appear to be more motivated by cost reduction rather than improved patient care. Both are important, but I believe that patient care comes first. In any case, I think this opening paragraph is a bit long, and could be shortened and more focused. I think the authors should particularly stress the point that with the increasing availability of large-scale data sources, more flexible and responsive approaches (models) are needed.

Author response: Thank you for highlighting this. We agree that patient care comes first and have modified the text to reflect this. We have also reduced the paragraph length. The new paragraph now reads (pg. 3):

‘Healthcare systems have limited resources and their budgets are being reduced1, while there are increasing numbers of people living with one or more long term conditions2,3. This can have a negative effect on health outcomes4, and systems therefore need to be more efficient. One way to improve efficiency is by implementing preventative measures which delay or prevent onset of disease and increase the overall health of the population. Increased data collection in healthcare systems, and availability of large scale data sources provide an opportunity to effectively target care and resources in a data-driven way. This could also be used to guide health policies, assist in healthcare auditing and select appropriate therapies in individual patient management, along with other uses5 to improve the healthcare system as a whole.’

R1.4. Then, in the second paragraph of the Introduction, I believe that clinical prediction models can be used for both diagnosis and prognosis, but the authors here focus on prognostic models. This should be clearer.
Author response: Thank you for this comment. We agree and have modified the text accordingly. The text now reads (pg. 3):

‘Clinical prediction models (CPMs) are used for diagnosis or prediction of future outcomes for individuals\textsuperscript{7,8}, and thus have the potential to be used for decision making and effective targeting of resources…’

We have also added the below sentences to the 3rd paragraph of the results and final paragraph of the discussion, respectively (pgs. 7 and 18).

‘To illustrate the approaches, we will focus on prognosis and consider a regression model…’

‘Also, although we restrict our attention to prognostic models, the findings are generalizable to diagnostic modelling.’

R1.5. "Only models that update in calendar time are considered dynamic here." I'm not sure what the authors mean by this. Are there other time scales over which a model could be updated?

Author response: We thank the reviewer for highlighting this and have reworded this sentence to now read:

‘Both a model and an individual patient’s risk can change over time. Throughout this paper, we focus on scenarios where the model evolves over time; we do not consider an alternative viewpoint where one observes repeated measures for an individual (and therefore model time varying coefficients).’

R1.6. The search strategy and study selection appear to be appropriate.

Author response: Thank you

R1.7. I struggled most with the organization of the Results. The authors indicate that they extracted two general domains: modeling methods; and validation and evaluation. The authors then group the identified papers into three categories: discrete model updating; Bayesian model updating; and varying coefficient modeling. I would suggest that the authors drop the label for 3.1 and 3.2. Then, new subsections should be the three categories (3.1 discrete model updating, 3.2 Bayesian model updating, 3.3 varying coefficient modeling). Under each of these, you should have subsections for your domains (3.X.1 modeling methods and 3.X.2 validation and evaluation). You might also add a 3.X.3 "limitations" or "pros and cons". Presumably, much of this is in the current section 3.3, but I found it difficult to relate these challenges back to the individual methods/categories. Any broader issues that cross-cut the categories should be moved to the Discussion section. This might require some reorganization of the Discussion section, as well.
Author response: Thank you for the above suggestion to help improve our manuscript. We have dropped the labels 3.1 and 3.2 as suggested. We also attempted to restructure the paper as suggested by the reviewer. However, we feel that doing so would make the paper less aligned with the methods section and what we wanted to highlight in the paper.

To clarify, the aim of the paper was to discuss the dynamic modelling field as a whole. Within this we happened to find 3 distinct groups of methods (discrete model updating, Bayesian model updating, and varying coefficient modelling). However, such distinction is only possible for papers that have described model development. The literature around validating a dynamic CPM is much less established, meaning that it was not possible to identify different validation techniques for each of the dynamic models. Therefore, the authors feel that it would be challenging to split the validation section into the three groups. For example, the included papers consider multiple methods but then discuss validation as a general topic. We have added this information into the manuscript on page 13:

‘The literature around validating a dynamic CPM is much less established, meaning that it was not possible to identify different validation techniques for each of the dynamic modelling methods separately.’

Also, the methods section of our paper describes extracting the dynamic modelling methods, validation of dynamic models and challenges/further work (section 2.3). To mirror this, we have formatted the results such that modelling methods is section 3.1, validation is section 3.2 and modelling challenges is section 3.3. Specifically, the validation section (section 3.2) is now a standalone section, rather than being incorporated into “modelling challenges” as previously. We have also removed the subheadings within section 3.3 (3.3.1 – 3.3.3).

Therefore, the revised structure of the results is as follows:

3. Results

3.1 Modelling Methods

3.1.1 Discrete model updating

3.1.2 Continuous model updating

3.2.3 Varying coefficient model

3.2 Model Validation

3.3 Other challenges

We hope this revised structured makes the organisation easier to follow.
R1.8. Related to this, the sentence "...we also extracted any limitations and suggested further work discussed by the authors" is awkward. First, the "authors" might not have realized all the limitations of their method(s), so I'm not sure you should limit yourself to these. Second, I think you mean that you extracted the further work that the "authors" discussed, not that you're simply suggesting further work that was already suggested by the "authors." In particular, the point of a review article is for you to assess progress in this research area and provide a critical assessment. So, while the "authors" might have already done some of this, it's important for you to take a broader view in recognizing limitations and suggesting future research in this area. It's not clear that this has been done.

Author response: Thank you for highlighting this. We have modified this sentence to now read (pg 6.):

’we also extracted any further work discussed by the authors and provide our suggestions for the future work needed in the area’

R1.9. In particular, what needs to be done for these methods to be adopted more broadly in clinical practice? The clinical literature is full of prediction models that never get implemented.

Author response: We thank the reviewer for this comment and agree that this is indeed an important point which needs to be discussed. The topic of why CPMs are not adopted more broadly in clinical practice is multifaceted, and we feel that explicit/detailed discussion of this important topic is out of scope of the current paper. However, we do believe that dynamic models are an important step in the right direction (since they are more likely to remain accurate through time). Further refinements such as improved reporting, better use of existing CPMs (e.g. a focus on external validation rather than de novo development), and incorporating models into hand-held technology (e.g. mobile apps to allow calculation of complex models a patient’s bedside) would be of value. While these are not confined to dynamic prediction models per se, we acknowledge that this is a common problem with prediction models. Therefore, we have added the below text into the discussion section of the manuscript (pg 18):

‘Therefore more practical examples and comparisons of the methods found are warranted for further work. This would help aid the broader adoption of these methods into clinical practice, which is a current issue with CPMs as a whole. While this is not confined to dynamic prediction models, this is a common problem with prediction models and refinements, such as, improved reporting and better use of existing CPMs (e.g. a focus on external validation rather than de novo development) could improve the adoption of CPMs in clinical practice. Also, incorporating models into hand-held technology (e.g. mobile apps to allow calculation of complex models a patient’s bedside) and extending the methods into software with user friendly tutorials would be of value.’

R1.10. Table 2: Grouping headers for "discrete model updating" and "Bayesian model updating" would be helpful.
Author response: Thank you for this suggestion to improve table 2.

We have now updated the table to include this information (see attached in the response to reviewer document or pg 25 in the manuscript).

Comments from reviewer 2

This is an interesting manuscript on an important topic. I like the separation in 3 categories of updating approaches.

The main point suggested is 'validation'. Another point of attention is the tuning of a forgetting factor / decay? The latter seems very reasonable for a continuous in time approach; other methods are more relevant for discrete data collections? Please sharpen your view here a bit.

Author response: We thank the reviewer for this feedback to improve our manuscript. If data are arriving in discrete time blocks (e.g. yearly) then we agree that we can probably add little with continuous updating. However, the aspiration of a learning health system is that data will arrive in streams (e.g. every 24 hours) for models to be updated, and hence the continuous updating methods we discuss become more important.

R2.1. Calibration drift: The focus in the Abstract seems to be on calibration drift; the more general aspect is 'accurate predictions', which I would suggest can be attributed to temporal or geographic differences in baseline risk and / or predictor effects. The latter can be summarized in a calibration slope, which may also reflect some overfitting, more than a true external validity issue. Calibration drift suggests a focus on baseline risk differences. This is indeed a frequent problem with external validation, but not necessarily so.

Author response: We thank the reviewer for this excellent point and have edited the abstract to focus on the general aspect of accurate prediction. The abstract introduction has now been edited to read (pg 2.):

‘Disease populations, clinical practice, and healthcare systems are constantly evolving. This can result in clinical prediction models quickly becoming outdated and less accurate over time. A potential solution is to develop 'dynamic' prediction models capable of retaining accuracy by evolving over time in response to observed changes. Our aim was to review the literature in this area to understand the current state-of-the-art in dynamic prediction modelling and identify unresolved methodological challenges’

Throughout the paper we focus on temporal change over time and some of the findings are also applicable to geographic updating and so we have added in the following text to our discussion (pg 19.):

‘Although the focus of this review was in methods accounting for temporal differences over time, some of the methods and issues raised would apply to geographic or contextual updating,
for example, where a model is to be used in a different population to which it was originally
developed. We also restrict our attention to prognostic models but the findings are generalizable
to diagnostic modelling.’

R2.2. In line with these more general issues with accurate predictions we recently proposed a test
may be with an extra predictor / biomarker: https://www.ncbi.nlm.nih.gov/pubmed/27678479

Author response: Thank you for highlighting this information which we can use to improve our
manuscript. We have included this reference and added the below text (pg 18.):

‘A close test procedure has previously been used34 to select which discrete updating method
should be used when updating your model. However, this has only been used for transportability
to a new population, opposed to updating regularly over time. Exploring this method to address
calibration drift, as well as, extending the method to include Bayesian updating and decide
when/if updating should occur would be extremely useful and increase the utility of the
approach.’

R2.3. The term 'dynamic modeling' might not be used in the current context, although I see that
many papers did so before. See e.g. Putter's book: dynamic prediction, which takes as a time
axis: follow-up per patient. Here calendar time may be considered, and 'updating' may be a far
easier term, perhaps less sexy, but more appropriate? Updating applies not only to calendar time,
but also to geography, which is advantageous.

Author response: Thank you for this suggestion and we agree that retaining accuracy across
geography is very important. However, our main aim was to consider temporal change over time
(within any given population) and the methods which can be used to address this, rather than
transportability of models and retaining accuracy in this setting. Also, dynamic means that the
updating itself is integral to the model, whereas 'updating' suggests that we change/renew the
model. Therefore, we have retained the term dynamic modelling for these reasons.

R2.4. Some references may have been missed; please consider searching with the term '
updating'. If the claim is 'extends the findings by conducting an up to date literature search’,
then the search needs to be comprehensive.

Author response: Thank you for this suggestion. Our search included the terms’ model updating’
and ‘dynamic model* and updat*’. Although our review was not a complete systematic review,
we believe our search to be comprehensive. We feel that the term updating alone would have
resulted in a vast increase in papers which were clearly out of scope and that the current search
sufficiently covers the area of interest. See R2.4a for more information.

Author response: Thank you for highlighting this. Van Houwelingen’s paper was captured within our search criteria and was originally excluded on the basis of the paper being an applied paper after reading the abstract. As you have mentioned, your work follows Vans thinking and so this was clearly an important and novel manuscript for the progression of the field. We have therefore included this in the review and have double checked the abstracts of the papers which were excluded on the basis of applied work. We acknowledge this to be a limitation of the work and discuss that the exclusion criteria could result in some important papers being excluded on page 18 in the manuscript:

‘Our search focussed on the methodological papers and it was not possible to go through all of the applied work. This may have resulted in some methods, or adaptations of existing methods, not being captured within our search. Nevertheless we believe that we have identified the main methodological approaches to dynamic model development, updating, and validation.’

In response to the inclusion of this paper, figure 1 and table 1 have now been updated in the manuscript. We have also added the references to this paper in the appropriate results section, for example, see individual slope updating description on page 8.

The following text has also been added to the validation section of the results on page 13:

‘Van Houweingen16 conducts a split sample validation on the original CPM and uses this to determine if an update is needed as the new data is collected, however, validation of the updating model was not undertaken.’

If I look at my own work (simply search in PubMed with 'updating' I find some studies that may be relevant (among 26: https://www.ncbi.nlm.nih.gov/pubmed/?term=steyerberg+updating ) that consider temporal impact and model adaptations:

https://www.ncbi.nlm.nih.gov/pubmed/17162015: more recent patients + difference setting (screen vs clinic) Also, an example is the development of a diagnostic model for Lynch syndrome, where the testing procedure was modernized after some point in time: https://www.ncbi.nlm.nih.gov/pubmed/17003395


Cardiology: https://www.ncbi.nlm.nih.gov/pubmed/21367834
Author response: Thank you for highlighting this. We have realised that our inclusion criteria were not completely clear in the manuscript as we only include methodology papers and exclude papers that solely present applications of established methods. We have therefore edited section 2.2 (selection of studies) to now read:

‘The eligible criteria for inclusion were original methodological peer reviewed journal articles which considered: 1) dynamic prediction models (DPMs); 2) model updating methods that could be performed in real time; or 3) model coefficients as functions of time. Exclusion criteria …Applied research, without any methodological work, was excluded because our interest was around the current state-of-the-art and methodology in the area.’

We acknowledge that applications of methods are also important and do discuss that to our knowledge the varying coefficient approach has not been applied to this problem. We also discuss this as a limitation in the discussion section of the manuscript, which reads ‘Our search focussed on the methodological papers and it was not possible to go through all of the applied work. This may have resulted in some methods, or adaptations of existing methods, not being captured within our search’.

We have also added some additional references in the paper to provide some example on some of the applied work which include two of your suggestions above. These are on page 18 and the two papers chosen were:


R2.5. The formulation 'stream across real time' is ugly, and seems to refer to computer science?

Author response: Thank you for highlighting this. We have edited this text to now read:

‘We define dynamic models (DMs) as those which acknowledge the real time of each point, are designed to evolve over time and address the problem of calibration drift. The model could, in principle, change after a single new observation, which could be a structural change or a coefficient change.’

R2.6. The term $\beta_0(t)$ may confuse some as relating to baseline hazard in survival?

Author response: To our knowledge $h_0$ is typically used for baseline hazard and $\beta$ or $\alpha$ are used for the coefficients of a prediction model. As we use alpha elsewhere we have retained the use of $\beta_0(t)$ for the intercept term.
R2.7. Validation: the description as "This is typically done for static CPMs by split sample, bootstrapping or external sample validation" is a bit awkward; cross-validation (e.g. 10-fold) or bootstrap validation are to be recommended; a single, random split sample validation should never be done; dominated by cross-validation / bootstrap validation.

Author response: Thank you for raising this point. We agree and have edited the text to now read

‘We therefore need to formally validate our models. For static CPMs, cross-validation and bootstrap validation are the recommended methods over split sample or external sample validation12, but validation is more complex when it comes to DPMs...’

R2.8. Step change: this could easily be modeled with a specific factor, see e.g. for the introduction of cardioplegy: https://www.ncbi.nlm.nih.gov/pubmed/9243151

Author response: We thank the reviewer for this suggestion and have added this reference (29) and have modified the text to read:

‘An example of a step change in clinical practice is the introduction of less invasive coronary surgery10. This change in surgery, along with a change in the case-mix of the population undergoing cardiac surgery resulted in the EUROSCORE CPM28 largely over predicting patient risk7. One way to model these changes in a CPM would be to include a time factor29 but it has yet to be discussed in the literature how well dynamic models react to these changes and which models provide the most accurate predictions and should be used in these circumstances. This assumes that a step change is anticipated for a known reason. However, in practice it is not always anticipated or known. Therefore, it would also be advantageous to account for, and model, unexpected step changes.’