Author’s response to reviews

Title: Protocol for a Systematic Review of Prognostic Models for Recurrent Events in Chronic Conditions

Authors:

Victoria Watson (victoria.watson@liverpool.ac.uk)
Catrin Smith (cat1@liverpool.ac.uk)
Laura Bonnett (ljbcmshe@liverpool.ac.uk)

Version: 2 Date: 07 Jan 2020

Author’s response to reviews:

Dear Reviewers,

Thank you again for your comments on our protocol. We have now revised our work in line with your suggestions and hope that our changes meet with your expectations. Our point by point responses follow.

Kind regards,

Victoria

Reviewer 1

The authors haven't signposted the changes made to the manuscript in the response to reviewers, and I can't see a tracked change version (apologies if I missed it) so it is quite difficult to see exactly what changes have been made.

Thank you very much for your feedback. Tracked changes have now been provided in the recent submission so it is clear exactly what has been edited. Apologies that this was not provided in the first resubmission.
I'm not sure that all points have been acknowledged in the response. For example, I previously asked whether it was useful to critically appraise the studies, but the response to this point does not address this query at all:

"Thank you for raising these concerns. We have modified the text to better outline our aims. However, rest assured that our review is not aimed at clinicians or patients, but instead those who will be developing the prognostic model to analyses recurrent events. Hopefully the results can be used to provide guidance regarding which methods are out there and which may be better suited to them given their data from a statistical point of view as opposed to from a clinical perspective. However, we may find that certain clinical areas used a specific type of modelling technique dependant on how the model assumptions meet the disease area they are modelling. Whilst our review is aimed at those who will be building the model to analyse the data, hopefully the review will also inform the importance of prognostic models in recurrent event data to clinicians and patients."

Whilst we admit that critically appraising journal articles from such diverse clinical areas may be futile, we feel that it is an important part of this review. Indeed, as mentioned in the manuscript, although the primary aim is not to assess the quality of included studies there is a need to summarise the quality, as low-quality studies may not be using the most appropriate statistical method for example. To further clarify this point we have added an additional sentence: “Similarly, any conclusions drawn from the included study regarding the results and performance measures of the identified recurrent event method should be interpreted with caution for lower quality studies.”

Similarly, the question re: why the authors are interested in performance measures for various particular applications is not clearly answered. In relation to this point, some of the text suggests that effect measures and performance measures will allow comparison between different approaches, although presumably two different approaches would have to be applied to the same population to render such a comparison meaningful? (e.g. lines 181 to 197). I don't know how realistic that is (by this I mean what I say - I really don't know), and differences between the studies might render such comparisons meaningless. Do we care that a hazard ratio is different in two models, perhaps applied to different populations, adjusted for different covariables etc?
Thank you for your comments. To clarify, we do not intend to compare effect estimates such as hazard ratios across different studies, but instead for included studies which use more than one method for recurrent event data and report multiple hazard ratios. Based on our pilot search, we found studies often applied multiple recurrent event methods to the same dataset as opposed to just one. Therefore, extracting data regarding the magnitude and direction of the effect such as the hazard ratios and confidence intervals allows us to compare between recurrent event methods which are applied to the same dataset.

The results from this could be useful in determining if a certain method over or under-estimated the effect or has a wider degree of uncertainty for example when compared to other methods used in the study. This is particularly useful when comparing these findings to other studies which also use multiple recurrent event methodology in the same paper, as we may find that a certain method tends to commonly over or under estimate the magnitude of the effect for example thus giving a biased result. Similarly, if performance measures are reported we wish to extract these also as we may find that a certain method tends to have poor predictive performance more so than others for example.

The text in the protocol has now been edited accordingly to represent this and clarifying our reasoning for extracting these measures.

So it remains reasonably unclear to me that the methodology matches a clear aim with this revision. There is scope for the review to end up with a variety of applications, each with their performance in a particular population summarised and with the quality of the methods appraised. I say these things in part because some of this stuff (e.g. quality assessment of studies) will represent a huge burden on the first author, without clear gain. That said, the authors are amongst the people who will be the users of the review, so if they think that this protocol suits their needs then, in some sense, it probably doesn't matter if I am not fully convinced; they are members of the target audience for the review! I think I might advise the authors to reflect on what the eventual paper will look like, and how some of the information they collect will (or won't) be useful for drawing conclusions.

I think I'm suggesting that the authors should try to be their own critics here, because they are not only committing themselves to conduct the review outlined in this protocol, but also presumably to use it as the basis for future work.

Jack Wilkinson
Thank you for raising these concerns. The reasoning why we wish to assess the quality of included studies is similar to why we are extracting the estimates and degrees of uncertainly and performance measures. When assessing recurrent event methods which have been applied to the same dataset in an included study, we would also like to assess the quality of the study to ensure any conclusions drawn from the paper regarding predictive performance of the identified methods are not unreliable or biased which may be the case for low quality studies. If we do find the conducted study was of a lower quality, we know to interpret the findings with caution when evaluating the applied method for recurrent event data in prognostic models.

Thank you again for this feedback, hopefully we have better clarified and justified our reasons for exacting data such as performance measured and quality assessment.