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: 1 Date: 28 Nov 2019

Author’s response to reviews:

Dear Reviewers,

Thank you 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 have submitted a protocol for a systematic review of recurrent events for chronic diseases. With the caveat that I have no idea whether or not similar reviews in this area exist, this sounds like an important topic.

Thank you very much for your positive feedback.

I have some queries about the goal and design of the review, some of which could possibly be clarified through some changes to terminology used in the protocol. I hope the authors feel comfortable in any response to say if I have misinterpreted or am otherwise just mistaken - a great opportunity for me to learn from the discussion.
It took me a while to work out exactly what the content of the review would be. I think this is partially a matter of terminology. The title speaks of this being a review of prognostic models. I don't really know the answer, but I wonder if we should make a distinction between particular models (for a particular application: particular disease, setting, outcome, covariates etc) and methods for constructing such models?

What was confusing was that the title made me think I was going to read a review about the former, rather than the latter. But then the authors stated their aims to be things like to "evaluate existing methodology" (line 68) and to "provide insight into the development of new methods" (line 28, and abstract). I took the latter to mean that they will use the results to inform their own method development (for example, by establishing the state of the art). However, when we look at the methods of the review (inclusion criteria etc) it does indeed seem that the authors intend to perform a review of particular applications, rather than of methodologies per se.

The objectives of the review should be much clearer now. To clarify, the primary aim is to identify what methods are being used in practice for prognostic models for recurrent event data as they are not well known or applied in research based on our knowledge and experience. Although a secondary aim would be to evaluate the performance of the models if that data was available, this is not our main objective and we do not anticipate there to be many results which allow us to do this. I can also confirm we do not intend to use the results to develop our own method, as instead we want to use the results to inform researchers what methods are available and which are commonly used in practice.

I think this prompts the question of whether the decision to search for particular applications, rather than for methods and stats papers, will allow the authors to address their goal of identifying available methods? Presumably some of the relevant work will be methodological rather than applied, and I wonder if the search strategy and inclusion criteria employed by the authors will catch these papers (or indeed, if they are turned up by the search, would they be excluded if they didn't include a genuine model development on real data? - and would that be the right thing to do?)

I think my main concern would be that the authors might look at which methods have, as a matter of fact, been used, identify apparent gaps in the methodology on this basis, and to start doing methods work on that basis. Meanwhile, the same work may have been done (perhaps, tested and rejected) already, in the stats journals. I worry that if the authors do not pick up this literature as part of their review, they might waste time duplicating existing work.
Indeed, the authors say (line 14): "whilst various prognostic models for recurrent events have been applied in research, most are underutilised." Does this mean that there are many methods, but they haven't been applied in clinical research? Or does it mean that there are many models (developed for particular applications, particular settings, diseases, outcomes, predictors etc) that aren't used in clinical practice. If the latter, a natural question is, how do we know this? If the former (methods are not used to develop particular applications), then will these methods appear in a review of applications? I am raising these points so that the authors can consider whether or not their proposed approach will give them what they need.

Having run the search strategy during pilot work, we can confirm that we find hits covering both applications and statistical methodology. Therefore we are confident that our proposed approach will lead to relevant hits which will inform our work as to frequently used statistical models for analysing recurrent event data.

Some other features of the protocol strike me as perhaps unnecessary for the authors' goals.

- The authors plan to extract performance measures for the various models (c-statistics or whatever). So we will end up with a review containing information on the performance of different models for different disease areas? Is this useful? If we want to conduct a review for clinicians and patients working in a particular clinical area, then shouldn't we have a dedicated review for each condition and population?

The aim of extracting the performance measures is not to produce a pooled estimate, but to inform performance in line with the underlying event rate being modelled. We agree that a review of each clinical area would be needed if our focus was clinicians and patients working in a particular clinical area.

- Otherwise, if this really is a review of methodology, then isn't this a bit like, say, a review of methods for analysing crossover trials including the estimates of all the different treatment effects from the different studies (of different treatments in different conditions etc…)

- Similarly, why are we critically appraising the studies? Again, this would make sense if we were doing a Cochrane-style prognostic review of models for a particular health condition - the goal there would be to inform patients and clinicians about the quality of the evidence. Does this have any place in a methods review? Is it a good use of the authors' time?
Thinking more about this point - hopefully the authors will not present this review as offering information for clinician and patient stakeholders. This would require clinical expertise. Moreover, restricting findings to studies that looked only at recurrent events (and not, for example, first event only) would result in a review containing an unusual subset of the prediction studies in that disease area, so overall conclusions are unlikely to be reliable in that regard.

To be clear, I *don't think* that the authors do have these intentions, but some of the methods they describe would be more appropriate if they did - I'm raising these points so that the authors may reflect on whether or not their design is likely to address their objectives.

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 analyse 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.

I think this protocol would really benefit from the authors being a bit clearer about what their objectives are (review of methods or particular models) and reflecting on whether they have selected the most appropriate methods to meet that objective. At present, a clear distinction is not made, and I am left somewhat confused about what the authors hope to achieve (and so, whether or not they are likely to achieve it). If the authors wish to identify the cutting edge in methods in this area, I would encourage them to move away from Cochrane-style methods for prognostic reviews and to consider conducting a methodological review. If, on the other hand, the authors want to review current practice (and to then comment on the state of play using their expertise of what could or should be done), then I think that is a good objective too. Again however, I wouldn't really see the value of quality assessment of studies (unless the authors wish to critique study design in this field? But not what they have described as their aims) or performance assessment of models for that goal. Finally, if the authors want to provide information for clinicians and patients (which is where it would make sense to look at study quality and model performance) then I would suggest that this probably isn't a great idea. Bespoke reviews would be needed for the individual chronic conditions, drawing on topic experts. I'm 99% sure that this isn't the authors intentions, as stated previously, but the current description leaves some room for the reader to doubt.
Thank you for this feedback – hopefully you will agree that we have made our objectives much clearer in the modified protocol, and better justified our reasoning for extracting performance measures and undertaking quality assessments of included studies.

Reviewer 2

Reviewer #2: This protocol summarises the authors proposed methodology for a review of existing methods for prognostic modelling of recurrent events. Best practice for systematic reviews is followed, with appropriate checklists, quality appraisal tools and guidelines to be used. The search strategy seems suitable, though I am not an expert. I wonder if the authors might consult an information specialist regarding their search strategy, and if they have considered the use of prognosis filters (e.g. Geersing). Further, are there other terms that may have been used to describe recurrent event studies, for example the authors refer to chronic conditions throughout (though I realise this may explode the number of hits).

Many thanks for your comments, they were very useful and the protocol has been adjusted accordingly. We can confirm that we consulted with a specialist in systematic reviews when building the search strategy based at the University of Liverpool library who assisted us to ensure we captured as many as possible hits for modelling recurrent events in prognostic models. We considered the Ingui and Hayes filters but instead opted for the Ensor search strategy (https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-2-91) as that led to more hits which we felt would ensure we missed fewer relevant studies.

My only other query is the value in the secondary objective. The current wording suggests that models identified will be evaluated in terms of their performance and how accurately the model predicts recurrence of disease. This seems relevant perhaps for specific conditions where a model may be used, but does not allow for a comparison of recurrent event models with standard survival methods, which is perhaps of more interest methodologically. This may be a misunderstanding due to wording. Quantitative synthesis is not considered, but I believe this is justified given the potential heterogeneity in results (models from various conditions for example).

A section has now been added (lines 186-190) to clarify predictive performance will only be extracted for recurrent event models and if studies do also report the time until only the first event analysis alongside multiple event analysis the effect size and degree of uncertainty could be evaluated to compare between the two.