Author’s response to reviews

Title: A Systematic Comparison of Recurrent Event Models for Application to Composite Endpoints

Authors:

Ann-Kathrin Ozga (a.ozga@uke.de)

Geraldine Rauch (geraldine.rauch@charite.de)

Meinhard Kieser (meinhard.kieser@imbi.uni-heidelberg.de)

Version: 1 Date: 13 Nov 2017

Author’s response to reviews:

Dear Editorial Board, dear Reviewers,

Thank you very much for your highly appreciated impact. Your valuable comments and suggestions were all incorporated within the revised manuscript and helped to improve it considerably. All changes made to the original manuscript are highlighted in blue color within the revised manuscript. In addition, please find our point-to-point replies below.

Michael Szarek (Reviewer 1)

This is a relevant topic for clinical trials in many disease settings however the authors should consider modifying the simulation scenarios so that they are more relevant to trials with composite endpoints that occur in a subset of subjects under study. Specifically, In many studies that have a composite endpoint consisting of a fatal event and potentially recurrent non-fatal event:

-- The nonfatal event often consists of several distinct but related events, e.g., in the CV disease setting, MI, stroke, unstable angina, coronary revascularization, etc.-- The fatal event is often a subset of all cause death, e.g., CHD death or CV death. So the deaths that are not part of the composite are a competing risk.
It is true that there is most often more than one single nonfatal event which are likely to be correlated. Moreover, it is also true that there are often several competing causes of death which might not all be disease-related. We completely agree with you that a more complex scenario would include more nonfatal event types and several causes of death, potentially linked by frailty terms. However, we feel that the investigated models are currently not well understood even in the simplistic case of a single nonfatal event and a single cause of death. Therefore, we wanted to keep the simulated scenarios as simple as possible to allow for a basic understanding of the models’ performance. Our focus especially laid on the influence of time-dependent risk changes and on effect changes dependent on the previous event time. If we would mix this already quite complex situation with a situation of competing risks and a higher number of nonfatal event processes, conclusions on the models’ performance are even more difficult as there are then much more factors influencing the final model performance which cannot easily be distinguished. However, we certainly agree that more work is needed after the insights gained from this work to investigate the models in more sophisticated situations. This limitation of our study is now explicitly stated in the Discussion section of our manuscript. Our working group is already investigating in a work in progress the performance of the different models for event processes related by a frailty term. We very much hope that our reply can convince you of the merit of investigating the simple models first, as done within this work. We agree that these considerations must be stated more clearly in the revised manuscript and we therefore added additionally a corresponding paragraph in the Methods section to justify the investigated scenarios.

-- The risk of the recurrent nonfatal event is orders of magnitude higher than the risk of the fatal event.

You are completely correct that the baseline hazard for the nonfatal event is often higher than for the fatal event. This important situation was already captured in our original manuscript by simulation scenario 3f. We revised the text to better highlight the impact of this scenario.

-- The experimental treatment is expected to reduce the risk of the recurrent nonfatal event more than the fatal event

We again agree that the intervention usually has a higher impact on the nonfatal event. This important situation was already captured in our original manuscript by simulation scenarios 1b,
2b, 3b, 4b, and 5b. As before, we revised the text to better highlight the impact of these scenarios.


Thank you for providing these important additional references which we included in our revised manuscript.

Yishu Wei (Reviewer 2)

It is pretty easy to follow and gives good explanations of problem of interest. The topic is very meaningful and is a good guideline in practice.

Thank you very much. We really appreciate your comment.

1. Could BIC being provided in supplementary material to give a rough idea of how different models performs.

This is an excellent idea. We included the models BICs in the supplement material as proposed.
2. Is there any real data application?

We agree that a real data example seems to be appealing. However, the problem with real data is always that the underlying true distributions are not known. As a consequence, the models could only be compared but the performance cannot be judged as the true distributions are not known. We therefore feel that a real data application cannot contribute too much to the general understanding of the models.

3. In simulation result PWP is pretty similar to AG, which is mentioned in P14/Line4-20. However the explanation and result seems still not too convincing, since the difference is small, especially compared with standard deviation. Is it possible to come up with some simulation which will further differentiate these two approaches? Especially since P14/Line22-24 recommend PWP in general.

That is a good point. We added one scenario that shows a greater difference between the AG and the PWP models (scenario 5 f).

4. The bias for all three methods in setting 3a-3e is larger, is there any explanation for this? Since 5000 datasets are simulated in each setting, is it because 200 patients are too small size?

Unfortunately, we are not exactly sure what you mean by “bias” as the presented effects all correspond to “mixed effects” which will naturally deviate from the reported cause-specific effects and this deviation will depend on the functional form of the baseline hazard. Therefore, it is not expected that all scenarios deliver the same effect estimates. In scenarios 3, the baseline hazards are larger as for scenarios 1 and 2 and thus produce more events. Therefore, an existing effect between the treatment and control is less variable. Therefore, in scenarios 3 there is no problem related to sample size. In scenario 3, the effect estimates are more influenced by the event ‘death’. This is because the baseline hazards in scenarios 3 are generally larger, and therefore more terminating events ‘death’ are observed which prevents from experiencing more than one MI event. We very much hope that our explanations helped to clarify this point.