List of responses to the comments for the author of: A multinomial generalized linear mixed model for clustered competing risks data

November 15, 2022

Associate Editor

The manuscript has been read by two independent reviewers and myself. It proposes a GLMM approach to handle clustered competing risks with a within-cluster dependence structure.

As you can see below, the manuscript has been positively evaluated, but there are still some issues that must be addressed. In particular, the authors should specify the differences with the existing literature and, especially, with the work by Cederkivst et.al. (2019) that seems to have proposed a very similar model.

Author's response

We are very grateful for the positive evaluation and clarified in the paper the differences between our work and Cederkvist et al. (2019), these clarifications were made in the introduction.

In summary, the differences are: we follow Cederkvist et al. (2019) and work with the Cumulative Incidence Function (CIF) specification proposed by them. However, there, to be able to model and infer the CIF parameters and within-cluster dependence, an elaborated composite likelihood approach was proposed together with an Adaptative Gaussian Quadrature (AGQ) scheme to marginalize the composite likelihood. Extra steps were needed to take to be able to quantify the uncertainty around the parameters. We, on the other side, modeled the same CIF specification via a much simpler and known formulation - a generalized linear mixed model (GLMM) formulation. Besides the simpler formulation (which also provides a full and proper likelihood function) we used a Laplace approximation scheme, simpler, in contrast to the used AGQ. To make this much simpler

framework feasible, we took advantage of state-of-the-art numerical algorithms implementations and computational libraries. Basically, is the safe CIF specification but with different likelihood functions/formulations and parameter estimation routines/schemes.

Reviewer 1

This is an interesting topic, and the authors reports an extended simulation study while proposing an estimation methods for competing risks with clustered data following the proposal by Cedervisk et al (2019). The author should specify with more detail differences, if they are present, between their proposal and the original one not only from the computational side: they declare to propose a new model but it seems the one proposed by Cedervisk et al.

I suggest adding some short motivations could add applicative values, examples on the usage of this class of models that could help readers to figure out in a clearer way the possible applications.

The simulations results underline some situation that can be problematic. This is very interesting because in the original proposal by Cedervisk et al only a real data set and a single simulation study were considered, I suggest stressing this point and add some comparisons with the previous proposal. Especially for problematic situations, have they tried to consider the original computational method proposed by Cedervisk et al? Can they give more insight for simulation cases where the latent effects didn't work well?

Minor

Page 8 row 18, please specify better what is yijt Page 3 row 52 and page 29 row 1 is Vaupel not Valpel

Reviewer 2

This paper proposes a multinomial mixed model (similar to a pattern mixture model) for the cumulative incidence functions of clustered multivariate competing risks data. Inference is based on the (Laplace-approximated) maximum likelihood. The methods are rigorously developed and could be useful in practice. I have a few minor comments to help the authors improve the paper.

1. In simple language, what is the main difference of the proposed approach with Cederkvist et al. (2019)?

2. Recently, Ahn et al. (2022) has proposed a robuts approach to semiparametric regression of multivariate clustered competing risks data. Can you comment on the similarities and differences with the proposed method (a parametric one if I understand it correctly)?

3. The Laplace-approximated maximum likelihood looks similar to the EM algorithm with the latent variables u_i treated as missing data. Can you clarify the difference?

References

Cederkvist, L., Holst, K. K., Andersen, K. K. and Scheike, T. H. (2019). Modeling the cumulative incidence function of multivariate competing risks data allowing for within-cluster dependence of risk and timing, *Biostatistics* **20**(2): 199–217.